SC/68A/ASI/11

# Estimation of population trajectories from fitting population models to individual identification data: some issues

Justin Cooke

INTERNATIONAL
WHALING COMMISSION

# Estimation of population trajectories from fitting population models to individual identification data: some issues

Justin Cooke

Centre for Ecosystem Management Studies
Höllenbergstr. 7, 79312 Emmendingen, Germany *jgc@cems.de*

Note: this was a working paper for the ASI Pre-Meeting, May 8-9 2019, but was promoted to an SC document so that it did not disappear.

SUMMARY

A population model is always required for interpreting mark-recapture data, although for some standard methods it may be a minimal model, such as a closed population model. The requirements for the population model are partly dictated by the nature of the data. For example, if the data are collected on a calving ground, it is necessary to model how often the whales visit the calving ground, which in turn requires that calving intervals be modelled. Experience shows that there is usually considerably heterogeneity in sighting probability related to factors such as age, sex, and reproductive status, and that such heterogeneity can change over time, such that the best model for sampling probability may contain many interaction terms. The popular practice of treating capture probabilities in different years as equal, whenever this yields a lower AIC, is discouraged, because it amounts to treating sample size as an index of relative abundance, without regard to sampling effort. This is inconsistent with the way the Committee normally considers abundance data.

Population models can be bulk models or individual-based, or hybrids between these. Individual-based models are in practice fitted using Bayesian methods and the result expressed as a posterior sample of population trajectories from which all quantities of interest can be computed. Prior distributions need to be chosen carefully so that posteriors for quantities of interest are valid and normalizable, and not unduly influenced by the priors, especially for small datasets. A method is proposed for defining implicit priors for sampling model parameters that ensures that the posteriors for biological quantities of interest, including population size, are independent of these. While a scale-invariant prior can be used for population size, the question of the appropriate priors for other biological parameters has not yet been definitively answered.

Verification of methods can involve substantial work. Testing methods by applying them to a limited suite of small test data sets is proposed as a simpler and quicker way to detect major problems.

## 1. BACKGROUND AND INTRODUCTION

### 1.1 Population models

In principle all estimations of animal populations using capture-recapture involve fitting a population model, albeit often a very simple one. For example, the two-sample Peterson estimator involves fitting a closed population model. In a closed population there are a fixed number of individuals and no births or deaths. This assumption is only acceptable when the time period is short enough for

births and deaths to be negligible. Open population models include births, deaths, and sometimes migration and other factors, and often sex and age structure.

The primary result of interest from fitting a population to capture-recapture data is an estimate of the trajectory of population size and composition over time, and a measure of the uncertainty in this. Other quantities of interest, such as survival rates, are in principle defined by the trajectory, even if they also enter the model as parameters in their own right. The population trajectory may be restricted to the period of data, or may include projections forward into the future, or back projections into the past.

For more complex models, it can be helpful to start each population trajectory well before the period of data, in order to reduce the sensitivity of the results to assumptions about the initial population composition.

### 1.2 Bulk population models

Population models can be of bulk type or individual-based. A very simple unstructured deterministic bulk open population model may be defined by just three parameters: an annual survival rate, an annual reproductive rate, and an initial population size. The equation of the model is:

$$N_{t+1} = SN_t + RN_t \qquad (t = 0, 1, ..., T) \tag{1}$$

where $R$, $S$ and $N_0$ are the parameters.

The parameters $R$, $S$ and $N_0$ may be quantities of interest in themselves. The final population size, $N_T$ is a quantity of interest that is not itself a parameter of the model but is a function of the parameters.

If we only have bulk abundance data, without individual recognition, $S$ and $R$ are not be separately estimable, but only the combination $R + S$. With individual abundance data, they are separately estimable.

Population models are generally extensions of the basic model (1) to include factors such as age and sex structure, the reproductive cycle for species with a multi-year breeding cycle, perhaps some spatial structure, and variation in reproductive and/or survival rates.

With a bulk model, the population trajectory, and hence any other quantities of interest, are determined by the parameters. This applies also to stochastic bulk models, if the parameters are taken to include the random effects.

Stochastic models generally have more parameters than deterministic models. For example, if variability in reproduction is added to the above model, then there would be a separate parameter to estimate for each annual reproduction. The annual rates would not normally be estimated as free parameters, but a prior distribution would be assumed, such as normally distributed (after appropriate transformation) with an unknown variance to be estimated.

In principle, a model used to fit capture-recapture data is an individual-based model, at least for the known individuals. However, some popular approaches to fitting capture-recapture data, both in simple and more complex cases, are of a hybrid nature, whereby an implicitly stochastic model for the known individuals is embedded in a deterministic bulk model for the whole population, with reproduction depending on the bulk population. In this sense, bulk models can be fitted to capture-

recapture data, although they Are not strictly bulk models. This distinction is rarely discussed explicitly.

### 1.3 Individual-based population models

An individual-based version of the model (1) would involve $N_0$ individuals alive at time $t = 0$. Thereafter, each individual has an annual probability $R$ of reproducing, and an annual probability $S$ of surviving. This formulation assumes that reproduction does not depend on survival; a slightly different version would have that dependency.

With an individual-based model, the parameters alone do not determine the population trajectory exactly: it also depends on individual events, such as births and deaths. With very small populations, particularly those in danger of extinction, the individual events can be important.

### 1.4 Sampling models

In order to fit a population model to data, including capture-recapture data, a model for the sampling process is required. An example of a simple sampling model is that each sample is a simple random sample of the population alive at that time, where a "sample" may consist, for example, of all individuals sampled in a given field season, treated as a sample without replacement.

More complex sampling models may take account of the relationship between sampling probability and factors such as sex, age and reproductive status, in addition to individual variation in sampling availability, and perhaps spatial factors.

The parameters of the sampling model are nuisance parameters, in the sense that quantities of interest are not functions of the nuisance parameters. However, the likelihood of the data, and hence the estimates of quantities of interest, do depend on the nuisance parameters. This means that we are, in principle, open to approaches that do not require explicit estimates of nuisance parameters, provided that the role of the nuisance parameters is correctly taken into account in inferences about the quantities of interest.

The parameters of the biological model can be called substantive parameters by contrast. The term "parameters of interest" can be ambiguous, because it is not necessarily the parameter values themselves that are of interest, but the quantities which depend on them.

### 1.5 Fitting the models and portrayal of uncertainty

The sampling model defines the likelihood of any given population trajectory. For bulk models, the likelihood is a function of the parameters only. For individual-based models, each individual trajectory has a different likelihood.

In Bayesian analyses, the uncertainty in the population trajectory is straightforwardly reflected by drawing a sample of the posterior distribution of population trajectories, provided that this posterior distribution has been generated in a valid way.

This applies to both bulk and individual-based models, except that for individual-based models a two-step process is often used. Each population trajectory in the posterior sample is generated by selecting a random set of parameter values, and then a random individual trajectory is generated

using these parameter values. Fully individual-based models are in practice fitted only using Bayesian methods.

With non-Bayesian analyses, there are different ways to express the uncertainty. If the log-likelihood can be adequately approximated by quadratic function, then a variance-covariance matrix of the parameters can be computed. The variance of any quantity of interest can be approximated using the delta method: this requires that the derivatives of each quantity of interest with respect to each parameter are computed. If these approximations are not satisfactory, then the likelihood profile for any quantity of interest can be computed, but this must be done for each quantity of interest separately.

To use the variance-based approach, the parameterization of the model should be chosen to yield a log-likelihood that is approximately quadratic. This is good practice even in the Bayesian case, to facilitate efficient sampling of the posterior.

## 2. SOME MODELLING ISSUES

### 2.1 Choice of sampling models

It is a feature of capture-recapture data that biological effects are often confounded with sampling effects. An inappropriate sampling model is quite likely to result in erroneous inferences with respect to population parameters. Therefore, an appropriately wide range of sampling models should be explored.

The sampling model should capture the major sources of heterogeneity in the availability for capture. This can in turn drive the need to incorporate such factors into the biological model.

For example, if mothers with calves are more likely to be recorded than other whales, then it will be necessary to model the distribution of gaps between successive calvings, because the likelihood of the capture history of an individual will depend on its calving history.

The modelling of the calving cycle may bring further benefits, such as inferences concerning changes in reproductive rates.

The sampling model should typically be as or more complex than the biological model. If the biological model is structured by age, sex and reproductive status, then sampling probabilities should also be allowed to depend on these factors, at least in principle, even if such dependencies are then pared down during the model selection process.

In the author's experience, in cases where capture probabilities depend on age, sex and reproductive status, the relative capture probabilities of the different population components also vary over time, and between different groups of data collectors (e.g. dedicated vs opportunistic, professional vs amateur, boat-based vs airborne vs shore-based). Such effects may be invisible in small datasets, but become apparent when the dataset is sufficiently large

When such interactions terms are included, the sampling model can contain many parameters, many of which are appropriately modelled as random effects, but note the warning in the following section.

### 2.2 Estimation of capture probabilities

It has become common practice, encouraged by some popular software programmes such as POPAN, to treat capture probabilities in different years as equal, when the estimation of separate capture probabilities is not supported by a lower AIC score. This approach amounts to accepting that sample size alone is a valid index of relative abundance, in other words that survey effort is constant.

This assumption is not specifically related to the method of capture-recapture. However, outside of the capture-recapture framework, the Scientific Committee does not accept the number of animals encountered each year as a valid index of abundance, or that survey effort is constant. On the contrary, the Committee has spent many years steadily refining methods to quantify sighting effort, mainly in the context of distance sampling.

Against this background, it would be irrational for the SC to accept the constant effort assumption, merely because there has been some individual identification, while rejecting it out of hand in all other contexts. Inconsistencies of this kind can arise from compartmentalized thinking, where capture-recapture methods are viewed as completely separate category of estimate from effort-based survey methods, without consideration of the common issues.

If the effective search effort associated with capture-recapture sampling can be quantified, there can be a case for relating capture probability to survey effort, but for most actual capture-recapture datasets, effort is hard to quantify because of spatially unrepresentative sampling. Absent quantification of effort, capture probabilities should be estimated as free parameters, at least for each sampling year (or year/area combination in models with spatial subdivision).

The probabilities may not always be very well estimated, but this fact will be suitably reflected in the variances of the resulting abundance or trend estimates, provided that they are computed correctly.

A nearly equivalent approach is to eliminate the sample-specific capture probabilities as parameters, by making inference conditional on the observed sample sizes. The conditional likelihood approach is used in simple cases, but is computationally harder to implement for complex sampling models.

### 2.3 Prior distributions for nuisance parameters, including capture probabilities

When a Bayesian approach is used, the question arises as to a suitable prior for the capture probabilities. Options that have been used in the literature include uniform on (0,1) or uniform on the log-odds scale. A problem with the latter option is that it is the likelihood is non-zero that the capture probability in the final sample of a dataset is exactly 1. The posterior distribution of the log-odds ratio has infinite area in the right-hand tail, such that the mathematical posterior distribution of the final capture probability is concentrated at 1. The posterior of the log-odds ratio has the archetypal rat's tail distribution, where a seemingly reasonable-looking distribution has a long narrow tail with infinite area. The software used to sample the distribution may fail to detect this tail, and yield a plausible-seeming posterior distribution for quantities of interest. Clearly, it is not satisfactory to use a mathematically degenerate model and rely on incorrect sampling of the posterior to obtain "reasonable" results.

Other choices of prior for the capture probabilities, such as uniform on (0,1) yield normalizable posteriors, but such seemingly innocuous choices can, for small datasets, unduly impact the posterior distribution of quantities of interest.

One solution to the prior distribution problem for nuisance parameters is the method of orthogonal parameterization. Suppose the parameters of the model are $\{ \psi, \phi \}$ where $\psi$ is a vector of

biological parameters and $\phi$ is a vector of nuisance parameters of the sampling model. If we are *a priori* agnostic regarding the values of the nuisance parameters, then we are equally agnostic as to the values of any 1-1 transformation of the vector of nuisance parameters. The model can be reparameterized in terms of { $\psi$, $\lambda$ } where $\lambda$ is a 1-1 function of $\phi$ (that may also depend on $\psi$) without materially changing the situation. Suppose we find a reparameterization such that the likelihood, given the data *y*, of the reparameterized model factorizes into separate components, where each depends on $\psi$ only or on $\lambda$ only:

$$f(y;\psi,\lambda) = g(y;\psi)\,h(y;\lambda) \tag{2}$$

In this case, the normalized posterior for $\psi$ depends only on the factor g, and does not depend on the choice of prior for $\lambda$. We call this an orthogonal parameterization.

Assuming that an orthogonal parameterization exists, we do not need to determine it explicitly. We note:

$$\sup_{\varphi} f(y;\psi,\varphi) = \sup_{\lambda} f^{*}(y;\psi,\lambda) = g(y,\psi)\ \sup_{\lambda} h(y,\lambda) \ = \ g(y,\psi)\,k(y) \tag{3}$$

where the final factor *k*(*y*) is a constant factor depending only on *y*, which can be discarded.

Hence we obtain the marginal likelihood for $\psi$ by maximizing the likelihood over $\varphi$. The maximization over $\varphi$ has to be nested within the sampling of the posterior distribution of $\psi$, hence it could be called the Nested Profile Likelihood (NPL) method.

The NPL method depends on the existence of an orthogonal parameterization, even though the latter need not be computed explicitly. One case where such a parameterization exists is when the sampling model is linear on the log-odds scale. For parameter vectors $\psi$ for the biological model and $\varphi$ for the sampling model, and data *y*, the log-likelihood of the full model has the form:

$$L(\psi,\varphi \mid y) \ = \ y^{\mathrm{T}} X\varphi \ + \ h(y,\psi) \ + \ a(\psi,\varphi)$$

where *X* is a constant, known matrix that defines the sampling model structure and *h* and *a* are arbitrary functions. The second derivatives $\partial^2 L \big/ (\partial\varphi)^2$ and $\partial^2 L \big/ (\partial\varphi\ \partial\psi)$ do not depend on the data vector *y*. We seek a parameterization { $\psi$, $\lambda$ } that is orthogonal, such that $\partial^2 L \big/ (\partial\psi\,\partial\lambda) \equiv 0$ . Noting:

$$\frac{\partial^2 L}{\partial\psi\,\partial\lambda} = \frac{\partial^2 L}{\partial\psi\,\partial\varphi}\frac{\partial\varphi}{\partial\lambda} \ - \ \left(\frac{\partial\varphi}{\partial\lambda}\frac{\partial\lambda}{\partial\psi}\right)^{\mathrm{T}} \frac{\partial^2 L}{\partial\varphi\,\partial\varphi}\frac{\partial\varphi}{\partial\lambda}$$

and using the invertibility of the square matrix $\partial\lambda \big/ \partial\varphi$ (because λ is a 1-1 function of $\varphi$), the required parameterization is a solution of the differential equation:

$$\frac{\partial\lambda}{\partial\psi} \ - \ \frac{\partial\lambda}{\partial\varphi}\left(\frac{\partial^2 L}{\partial\varphi\,\partial\varphi}\right)^{-1} \frac{\partial^2 L}{\partial\varphi\,\partial\psi} \ = \ 0$$

We would normally be willing to choose a sampling model with this general structure. The NPL method then obviates the need to choose priors for the parameters of the sampling model.

It is easily shown that in certain special cases, such as when there is just one capture probability to estimate per annual sample, and sampling is with replacement, the NPL method is equivalent to conditioning on the annual sample sizes.

### 2.4 Prior distributions for biological parameters

For the population size in a reference year, a scale-invariant prior, such as uniform on the log scale, is appropriate. If the bulk population model is itself linear on the log-scale, as it will usually be if it does not incorporate density-dependence, then the choice of reference year is immaterial. If the population model is not log-linear, then one approach is to choose the scale-invariant prior for the population size in the year of most interest, on the basis that the prior should be maximally uninformative about this particular population size.

For a survival rate, a prior that is uniform on the log-odds scale is tempting but invalid. If mortality is inferred from the disappearance of animals rather than the recovery of carcasses, then the likelihood of 100% survival is non-zero, even if very small. The posterior distribution of the log-odds survival parameter is then not normalizable, and the posterior, if correctly sampled, would be concentrated at 100% survival. Algorithms that sample the posterior may, as discussed above, fail to detect the "rat's tail" and yield seemingly-plausible posterior distributions of parameters of interest.

A uniform prior for the survival probability on (0,1) avoids this problem but may negatively bias survival rate distributions when the true survival rate is close to 1. A Jeffrey's prior for the survival rate is intermediate between the above extremes and may be reasonable in practice.

It is probably fair to say that a comprehensive, definitive and practical approach to choosing priors for biological parameters has not yet been developed.

## 3. VALIDATION OF MODELS

### 3.1 Components of validation

Aspects where validation is required relate not just to modelling per se but to the full process.

1. Pre-processing of data, such as selection of photo-id observations of sufficient quality in ways that avoid selection bias (that can be caused, for example, by the fact that it is usually easier to conform a match than a non-match, or that a match may become obvious only retrospectively, after further photos of that individual have been obtained).

2. Choice of biological model: is it appropriate for the population in question?

3. Choice of sampling model: is it appropriate for the given dataset for that population, taking into account, inter alia, likely migration habits of that population ?

4. Fitting of the model, mathematical aspects: definition of the likelihood function, choice of parameterization and priors.

5. Fitting of the population model, computational aspects: this includes
   - data handling
   - implementation of the biological model

- implementation of the likelihood function
- fitting algorithm, including sampling of the posterior
- interpretation of outputs: do these correspond to the quantities of interest ?
- software package: is it valid and was it correctly used

In the past the Scientific Committee has considered that models using non-standard software should be subject to software validation. However, in practice this has rarely been accomplished because this is a very labour-intensive tasks and beyond available resources for more than maximally one or two cases per year.

"Standard" software, even if it is bug-free, may nevertheless yield erroneous results for unsuitable choices of model or with datasets that do not meet the assumptions of the method. Interpretation of the output may also be an issue: for example a parameter labelled "survival rate" may in practice reflect a composite of actual survival and age-specific availability, and the user needs to know not to take the estimate at face value.

**3.2 Test data sets**

A more practical approach to validation may be to build up a suite of test data sets to which methods should be applied. This approach would simultaneously test several aspects of model validity, beyond software validity in the narrow sense, and is potentially more efficient in labour terms.

Test data sets could be based on real data (subject to agreement of data owners) and/or on idealised simple datasets designed to highlight common issues.

At least one test dataset should be a "clean" dataset that represents data that would be expected when the assumptions of a reasonably simple model are met. Methods fitted to such a dataset should normally reproduce the true parameters used to construct such a dataset.

Additional data sets should represent more difficult cases typical of whale data. Several standard methods were found to choke on the example of Table 1. There are three sampling occasions and the data consist of capture histories across the three occasions.

Table 1. Example test data set for capture-recapture

| Capture histories | | | Frequency |
|---|---|---|---|
| 1 | 0 | 0 | 85 |
| 0 | 1 | 0 | 85 |
| 0 | 0 | 1 | 85 |
| 1 | 1 | 0 | 5 |
| 1 | 0 | 1 | 5 |
| 0 | 1 | 1 | 5 |
| 1 | 1 | 1 | 5 |

The special property of these test data is that more whales than expected are caught three times, relative to those caught once or twice. This is not unusual for whale data. One would expect a method that ignores such heterogeneity to underestimate the population size somewhat, but this was not always the case for this data set.

### 3.3 Sensitivity to assumptions

All models involve assumptions about a population that do not hold exactly, such as the constancy of survival rates. With sufficiently long data series, these assumptions can be checked, but often this is not possible. The question is not; are the assumptions valid ? but is the main result of interest sensitive to these assumptions. The sensitivity of the results to reasonable departures from the assumptions could be tested , such as by assuming reasonable levels of variability in these assumptions. It would be somewhat laborious to require this in every case, because it leads to proliferation in the quantity of results that need to be documented and presented, even if the computation of additional results does not itself involve much extra work.

Possible ways forward include:

(i) determine in a generic sense which kinds of assumptions can be critical, and require only the sensitivity of the results only to the likely most critical assumptions to be tested.

(ii) develop a simplified and standardized approach to the presentation of sensitivity results such that relevant information can be presented compactly with minimal explanation.