Proposal for a common simple and widely applicable model for right whale population assessments

D.S. Butterworth and A. Ross-Gillespie

Summary

A simple model is presented whose aim is to be applicable across a number of right whale populations, in particular so as to provide results that can be compared across these populations. For this reason, the model is designed to require only very limited data, specifically a time series of comparable annual calf counts without too many missing values. Right whales are assumed to calve at either three- or five-year intervals, with the associated proportions changing over time. Similarly, the value of the parameter (X) reflecting the product of the proportion of births that are female and the first-year survival rate may change over time. An initial application to calf count data for the South African right whale population suggests that such data do not contain sufficient information for annual variations in both the X parameter and in the annual proportion of calving intervals that are three years to be estimated. Fixing X and estimating annual changes in the proportion of three-year calving intervals only appears to provide the best performing approach.

Introduction

The basic idea underlying this document is to develop an as-simple-as-possible model that can be applied to all the various Southern Hemisphere right whale populations to assess demographics and allow for comparison across the populations. The proposed model is outlined in this document, and results are provided for a preliminary application to the South African right whale data to assess feasibility. The full details of the model proposed can be found in the Appendix, but a broad outline of key features is provided here:

- Two key population components are estimated for each year: the total number of adult (past the age at first parturition) females (N_y^t) and the total number of calving females (N_y^c).
- The model assumes that each female will reproduce after either a three-year or a five-year interval.
- The model is fit to annual calf counts (which are assumed not to miss any animals).
- The model estimates:
    - the starting adult female population size for N_{y0-n}^t,
    - the annual proportion of calving females that will enter a three-year calving cycle,
    - a (potentially time-varying) parameter that accounts for juvenile mortality as well as the proportion of calves that are female, and
    - the initial (assumed to be steady) population growth rate.

Results

In order to commence the population's dynamics, the model is started a fair number of n years before the year (y_0) for which the first calf count is available. Several assumptions have been made for this initial analysis, such as:

(a) a constant growth population growth rate (R),

(b) a constant juvenile mortality rate (X_0) and

(c) that a constant proportion (P_0) of calving females will enter a three-year calving cycle during these n years.

This leads to an inter-dependence of three key model parameters *R,* X_0 and P_0. In the interest of simplicity, and as a first attempt for fitting the model, two approaches have been taken for the results presented in this document.

1. The annual proportion of calving females that will enter a three-year calving cycle (p_y) is assumed to be constant with time. P_0 is fixed at a range of values, *R* is estimated freely, and X_0 is determined by the values of P_0 and *R* (see the Appendix for further details).
2. The juvenile mortality (plus proportion of calves that are female) parameter (X_y) is assumed to be constant with time. X_0 is fixed at a range of values, *R* is estimated freely, and P_0 is determined by the values of X_0 and *R*.

Results are presented for five runs – runs 1a and b as per approach (1) above for two different values of P_0, and runs 2a-c as per option approach (2) above for three different values of X_0.

Table 1 lists key parameter values and negative log-likelihood components for the five runs. Figure 1 plots the estimated population trajectories, as well as the trajectories for X_y and p_y.

Discussion

Some key discussion points are provided in bullet point form below.

- Approach (2) (a time-invariant proportion of births that are female and juvenile survival rate, X_y) seems in general to be able to provide better fits to the data. Furthermore approach (1) (constant proportion p_y of three-year calving cycles) appears to require a strong temporal trend in the X_y parameter trajectory in order to fit the data, which may be beyond the range of biological plausibility. For these reasons, it would seem that approach (2) is preferable to approach (1), having greater "flexibility".
- A key feature in the South African right whale data is the noticeable drop in calf counts in very recent years. Approach (2) tries to address that by substantially reducing the p_y proportions in the most recent years considered in the model, i.e. it explains the reduction in calf counts by assuming that a large majority of adult females have entered five- rather than three-year cycles. This supposition can be validated only given future data, as it implies an expected imminent increase in number of calves over the next few years.
- The impact of changing the adult survival rate assumed for these analyses ($S = 0.97$) needs to be explored. This value is somewhat lower than the value of 0.99 that has been estimated in the application of a more complex population model (using more detailed data) for the South African right whale population (Brandão *et al* 2018). The reason for the choice for this analysis was that the model exhibited slightly more stable behavior for this somewhat lower survival rate, given the interdependence of X_0, P_0, *R* and *S* for the initial year configuration of the model (see equation A4 of the Appendix).
- Note that the scale of the abundance estimates output by the model is determined by the assumption that the calf counts do not miss any animals (and that there are no mortalities resulting from non-natural causes). If the proportion missed remains about the same over time, abundance estimates would simply need to be scaled upwards by the inverse of that proportion; however, if there was a temporal trend in the proportion missed, the impact on results could be more complex.
- Many more variations of these five runs could be explored (such as estimating both X_y and p_y and fixing *R* – though preliminary attempts at this suggest that the model has difficulties in distinguishing variations in X_y from those in p_y from the limited data – annual calf counts - available), but this document is primarily intended to outline the proposed model and to provide some preliminary results. Overall the fit of the model to the data (particularly for approach (2)) is not unreasonable, and the estimates of overall population size appear to be fairly consistent across the five variants, suggesting that the model has some potential to be used as a common widely applicable model for the right whale populations. However, further exploration and development, as well as trial applications to other right whale populations, should first be pursued.

- For applications to other right whale populations, the following information would need to be provided:
  - a few values (considered to be plausible) for the non-juvenile survival rate S;
  - a value for the age at first parturition t_m; and
  - a time series of annual calf counts (desirably complete, though the approach can accommodate missing values for a few of the years).

Reference

Brandão, A., Vermeulen, E., Ross-Gillespie, A., Findlay, K. and Butterworth, D.S. 2018. Updated application of a photo-identification based assessment model to southern right whales in South African waters, focussing on inferences to be drawn from a series of appreciably lower counts of calving females over 2015 to 2017. International Whaling Commission: SC/67b/SH22: 18pp.

Table 1: Summary of results for the five runs presented in this document. The first two runs (1a and b), fix the value of P_0 (the value of p_y prior to y_0, where y_0 is the first year for which data are available), estimate *R* and allow X_y to vary after y_0 (the value of X_y prior to y_0 (X_0) is determined by P_0 and *R*, see the Appendix for more details). The next set of three runs (2a-c) fix the value of X_0, estimate *R* and allow P_y to vary after y_0. Analogous to the first two runs, the value of P_0 is determined by the values of X_0 and *R.* In the table below, the symbols are defined as follows:

P_0 is the value of p_y prior to y_0 (the first year for which there are data), i.e. the proportion of calving females entering a three-year (rather than five year) cycle prior to y_0,
μ_p is the mean of p_y after y_0,
X_0 is the value of X_y prior to y_0, i.e. the value of the variable taking juvenile survival rate and proportion of calves that are female into account prior to y_0,
μ_X is the mean of X_y after y_0,
$\sigma_X$and σ_p are the variance parameters for the fluctuations about the means for $X_y$and p_y post year y_0 (see the Appendix for more details),
R is the constant growth rate assumed for the initial period prior to y_0,
$\Delta lnL\ (total)$ gives the difference in total negative log-likelihood points between run 1a and the rest,
$\Delta lnL\ (data)$ gives the difference in negative log-likelihood points for the data component between run 1a and the rest, and
$\Delta lnL\ (penalties)$ gives the different in total negative log-likelihood points for the combined penalties between run 1a and the rest.

Run	P_0	μ_p	X_0	μ_X	σ_X	σ_P	R	$\Delta lnL\ (total)$	$\Delta lnL\ (data)$	$\Delta lnL\ (penalties)$
1a	1.00	1.00	0.33	0.28	0.5	0.5	1.054	0.0	0.0	0.00
1b	0.60	0.60	0.38	0.34	0.5	0.5	1.050	15.9	17.4	-1.50
2a	0.57	0.58	0.40	0.40	0.5	0.5	1.052	-66.1	-67.3	1.21
2b	0.66	0.69	0.35	0.35	0.5	0.5	1.046	-73.8	-78.1	4.36
2c	0.71	0.77	0.30	0.30	0.5	0.5	1.038	-53.3	-59.6	6.35

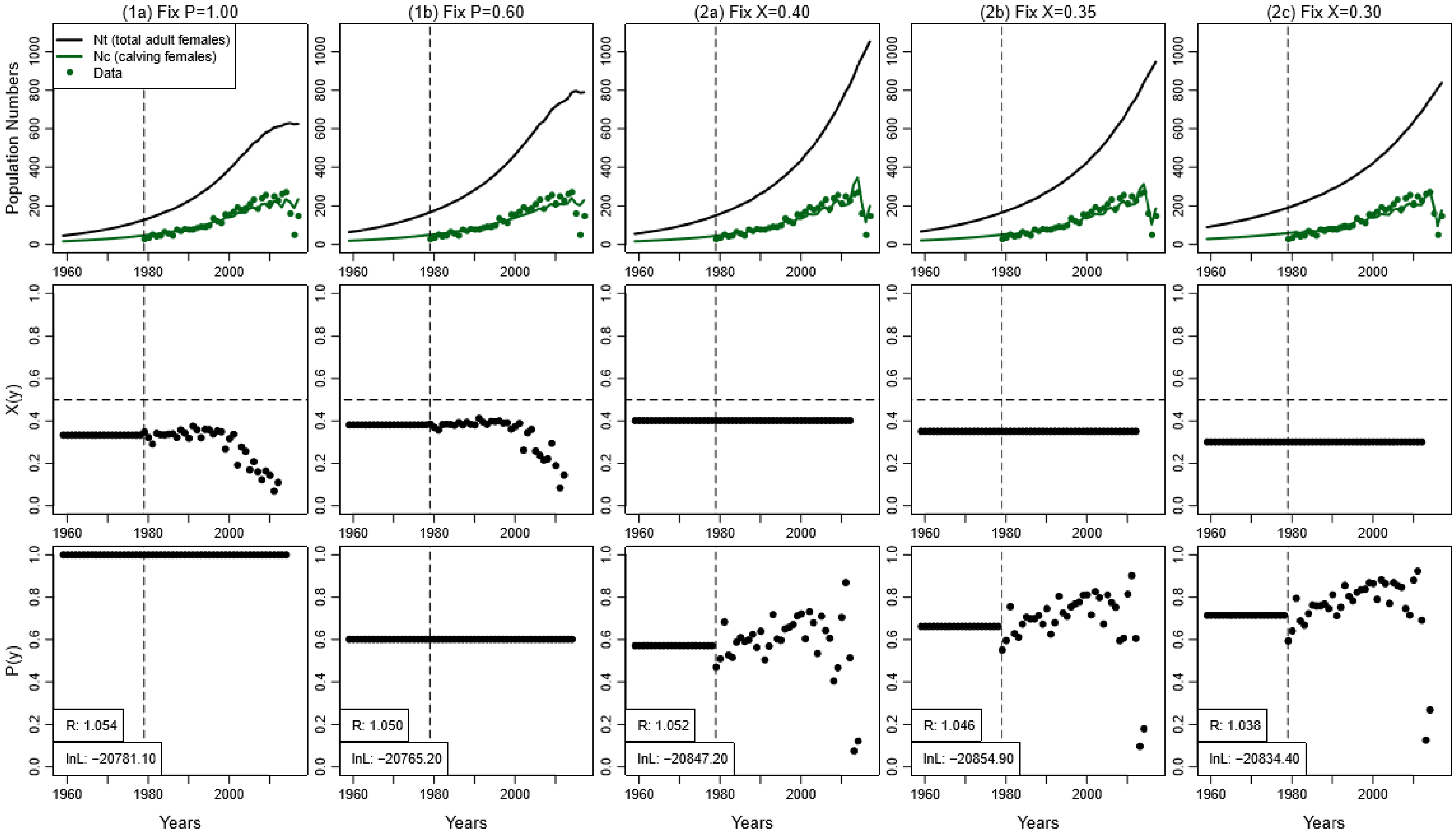

Figure 1: Some graphical output for the five runs presented in this document. The top row shows the population trajectories (in numbers) of the total female population past the age at first parturition (N_y^t), and for the number of females calving each year (N_y^c). The data to which the model is fit (the counts of number of calves per year) are shown by the closed circles. The vertical dashed lines mark the year $y_0 = 1979$, the first year for which data are available. The second row shows the estimates of X_y (combination of juvenile survival rate and proportion of calves that are females) and the bottom row the estimates of p_y (the proportion of calving females each year entering into a three-year calving cycle). The estimated values of the growth rate R and the total negative log-likelihoods are shown in the bottom left corners.

Appendix

Methodology for the proposed right whale common model

The total female population in year *y*+1 is given by:

$$N_{y+1}^t = N_y^t S \; + \; N_{y-t_m+1}^c S^{t_m} X_{y-t_m+1} \tag{A1}$$

where

N_y^t	is the total female population past the age at first parturition in year *y*,
N_y^c	is the number of females calving in year *y*,
S	is the non-juvenile survival rate,
t_m	is the age at first parturition, and
X_y	is an additional (possibly time-varying) parameter to take juvenile (first year) survival into account. X_y needs to be less than 0.50 to account (at least) for the proportion of calves that are female.

$N_{y-t_m+1}^c S^{t_m} X_{y-t_m+1}$ is thus the number of female calves that were born $y - t_m + 1$ years ago and have now reached the age at first parturition.

To calculate the number of calving females in year *y*, and assumption needs to be made regarding calving interval. For this proposed model, it is assumed that each female will reproduce either after a three-year or a five-year interval. Then:

$$N_y^c = N_{y-3}^c p_{y-3} S^3 + N_{y-5}^c \left(1 - p_{y-5}\right) S^5 + N_{y-tm}^c X_{y-t_m} S^{t_m} \tag{A2}$$

where p_y is the proportion calving each year which will take 3 years until they calve again. Therefore, in equation (A2) above:

$N_{y-3}^c p_{y-3} S^3$	is the number of females that calved three years ago which (a) took a three-year calving interval to reproduce again and (b) survived the three years since last calving,
$N_{y-5}^c \left(1 - p_{y-5}\right) S^5$	is the number of females that calved five years ago and which (a) didn't take a three-year calving interval (which by assumption implies they took a five-year interval) and (b) survived the five years since last calving, and
$N_{y-tm}^c X_{y-t_m} S^{t_m}$	is the number of females reaching age at first parturition in year *y* (i.e. the assumption is made that all females at age of first parturition will produce a calf).

Initial situation (before year y_0)

Start the model some *n* years before the actual first year of interest, y_0, and assume the following for those *n* years.

1. The total number of adult females in year $(y_0 - n)$ is an estimable parameter.
2. Each year a constant proportion ρ of the total population is calving, i.e. $N_y^c = \rho N_y^t$.
3. The proportion of females in three-year calving cycles is constant, i.e. $p_{y+1} = p_y = P_0$.
4. The juvenile mortality and female ratio variable X_y is constant, X_0.
5. The population is growing at a steady rate *R* so that $N_{y+1}^t = R N_y^t$ and $N_{y+1}^c = R N_y^c$.

With these assumptions, equation (A2) becomes:

$$N_y^c = \frac{N_y^c}{R^3} P_0 S^3 + \frac{N_y^c}{R^5} (1 - P_0) S^5 + \frac{N_y^c}{R^{t_m}} X_0 S^{t_m} \tag{A3}$$

Therefore X_0 can be calculated as:

$$X_0 = \left(1 - \frac{S^3}{R^3}P_0 - \frac{S^5}{R^5}(1 - P_0)\right) / \left(\frac{S^{t_m}}{R^{t_m}}\right) \tag{A4}$$

Furthermore, under the assumption that $N_y^c = \rho N_y^t$, equation (A1) can be re-written as:

$$RN_y^t = N_y^t S + \rho N_y^t S^{t_m} X_0/(R^{t_m - 1}) \tag{A5}$$

From this,

$$\rho = R^{t_m - 1}(R - S)/(S^{t_m} X_0) \tag{A6}$$

Model set-up post y_0

The calculations above provide the values for N_y^t, N_y^c, p_y and X_y for the *n* years prior to y_0. From y_0 onwards, equations (A1) and (A2) are used to calculate the population dynamics. The parameters X_y and p_y are estimated as a mean value with annual residuals that are assumed to be normally distributed with a mean of zero and a standard deviation of σ_X and σ_p – more details can be seen in the table below.

Model parameters

The table below lists key model parameters along with further details.

Parameter		Fixed/Estimable
lnN_{y0-n}^t	Total female population size in start year, *n* years before the first year for which data are available ($y_0 = 1979$) for SA right whales. Estimated in log space.	Estimable. For the results in this paper *n* is set at 20 years.
P_0	The constant value of p_y assumed for the initial set-up, *n* years before 1979.	Fixed on input
S	Non-juvenile survival rate	Fixed (at 0.97[1])
t_m	Age at first parturition	Fixed (at 5 years[2])
$X_y = \frac{0.5}{1 + e^{-X_y^*}}$	Parameter to take additional juvenile mortality into account as well as the proportion of calves that are female. The estimable parameter is X_y^*, estimated in logit space so that X_y lies between 0 and 0.5, as the female proportion is assumed not to exceed 0.5 .	Estimable
$X_y^* = \mu_X + \epsilon_y^X$	The model estimates a mean for X_y^* and residuals ϵ_y^X, where $\epsilon_y^X \sim N(0, \sigma_X^2)$, with σ_X fixed on input.	Estimable mean and fixed standard deviation
$p_y = \frac{1}{1 + e^{-p_y^*}}$	The proportion of adult females calving each year that will take three years until they calve again. Similar to *X*, the estimable parameter is p_y^*, estimated in logit space so that p_y lies between 0 and 1.	Estimable
$p_y^* = \mu_p + \epsilon_y^p$	The model estimates a mean for p_y^* and residuals ϵ_y^p, where $\epsilon_y^p \sim N(0, \sigma_p^2)$, with σ_p fixed on input.	Estimable mean and fixed standard deviation

Data and likelihood

The model is fit to number of calves seen each year assuming a Poisson distribution:

$$-lnL = -N_y^{c,obs} lnN_y^c + N_y^c \tag{A7}$$

[1] This value is somewhat lower than the 0.99 estimated in Brandão *et al.* (2018), and was chosen to provide greater stability for these initial explorations.

[2] See Brandão *et al.* (2018).

where

$N_y^{c,obs}$	is the number of calves (male and female) observed in year *y*, and
N_y^c	is the number of females calving in year *y.*

In addition, a penalty is added to the negative log-likleihood for each of the X_y and p_y parameters so that the estimated residuals correspond roughly to a normal distribution with their mean "forced" to be zero.

$$pen_X = w_X\left(\sum_y \epsilon_y^X\right)^2 + \sum_y (\epsilon_y^X)^2/(2\sigma_X^2) \tag{A8}$$

and similarly

$$pen_p = w_p\left(\sum_y \epsilon_y^p\right)^2 + \sum_y (\epsilon_y^p)^2/(2\sigma_p^2) \tag{A9}$$

Lastly, penalties are added to the negative log-likelihood to force some continuity in X_y and p_y when transitioning from the initial setup (before y_0) to the post-y_0 model dynamics.

$$pen_{X,cont} = \left(X_0 - \frac{1}{10}\sum\nolimits_{y_0}^{y_0+9} X_y\right)^2 /(2(0.01)^2) \tag{A10}$$

$$pen_{p,cont} = \left(P_0 - \frac{1}{10}\sum\nolimits_{y_0}^{y_0+9} p_y\right)^2 /(2(0.01)^2) \tag{A11}$$