INTEGRATED POPULATION MODEL SELECTION

RACHEL MCCREA

ROYAL STATISTICAL SOCIETY ANNUAL CONFERENCE

EXETER, SEPTEMBER 2015

Collaborators

- Takis Besbeas
- Byron Morgan

OUTLINE

INTEGRATED POPULATION MODEL

$$L_G(\Theta|d_1, d_2, \ldots, d_K) = L_1(\theta_1|d_1) \times L_2(\theta_2|d_2) \times \ldots \times L_K(\theta_K|d_K)$$

where

$$\Theta = \theta_1 \cup \theta_2 \cup \ldots \cup \theta_K$$

INTEGRATED POPULATION MODEL

- Advantages
  - unites capture-recapture and time-series methodology;
  - provides a simultaneous description of all the data;
  - increased precision of parameter estimates ;
  - coherent estimation of parameters otherwise not estimable from individual component analyses alone (new methods to diagnose estimable parameters are given in Cole and McCrea, 2015).

- Wide number of applications
  - eg. lapwings, grey herons, cormorants, greater snow geese, soay sheep;
  - A review of applications is given in Schaub and Abadi (2011).

INTEGRATED POPULATION MODEL

This talk focuses on combining **census** (or abundance indices) and **demographic data** such as ring-recovery data.

- state-space models can be used to describe census data, and if Gaussian assumptions are made the Kalman filter can be used to form the likelihood function (Besbeas et al, 2002; Brooks et al, 2004);
- large range of models for capture-recapture data and they are generally of product-multinomial form (McCrea and Morgan, 2014).

PARAMETERS OF THE INTEGRATED MODEL

- annual survival probabilities, $\{\phi_a\}$, which vary with age up to age $A > 1$, and then remain constant with increasing age;
- productivity ρ;
- recovery probability, λ;
- measurement error variance σ^2;

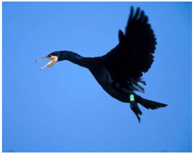
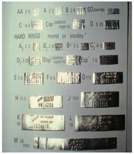
STATE-SPACE MODEL

State-space models are based on two equations, the **transition** equation (1) and the **observation** equation (2).

$$\begin{pmatrix} N_{1,t} \\ N_{2,t} \\ \vdots \\ N_{A-1,t} \\ N_{A+,t} \end{pmatrix} = \begin{pmatrix} 0 & \rho\phi_1 & \cdots & \rho\phi_1 & \rho\phi_1 \\ \phi_2 & 0 & 0 & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \phi_{A+} & \phi_{A+} \end{pmatrix} \begin{pmatrix} N_{1,t-1} \\ N_{1,t-1} \\ \vdots \\ N_{A-1,t-1} \\ N_{A+,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{1,t} \\ \vdots \\ \epsilon_{a-1,t} \\ \epsilon_{A,t} \end{pmatrix} \quad (1)$$

$$y_t = \begin{pmatrix} 0 & 1 & \cdots & 1 \end{pmatrix} \times \begin{pmatrix} N_{1,t} & N_{2,t} & \cdots & N_{A+,t} \end{pmatrix} + \eta_t \quad (2)$$

Ring-recovery data

- Cohorts of individuals are marked and released back into the population;
- When individuals die, they may be recovered dead or their rings/marks may be recovered;
- Data can be summarised by the statistics:
  - R_i: number of marked individuals released at occasion t_i;
  - d_{ij}: number of individuals released at occasion t_i and recovered dead in the time interval (t_{j-1}, t_j).

Ring-recovery model

- For individuals marked as young, with age-dependent survival up to age A, and $T(> A)$ recovery occasions, the probabilities corresponding to the observed data are:

	2	3	$\cdots$	T
R_1	$(1-\phi_1)\lambda$	$\phi_1(1-\phi_2)\lambda$	$\cdots$	$\prod_{a=1}^{A-1}\phi_a\phi_A^{T-A}(1-\phi_A)\lambda$
R_2		$(1-\phi_1)\lambda$	$\cdots$	$\prod_{a=1}^{A-1}\phi_a\phi_A^{T-A-1}(1-\phi_A)\lambda$
$\vdots$				$\vdots$
R_{T-1}				$(1-\phi_1)\lambda$

Selecting the age-structure for an IPM

- Integrated population model selection with regard to age-structure in survival (or other parameters) involves modelling across state-space model dimensions;
- An alternative approach to modelling age-variation in IPM is to form a maximal SSM, with specific age-structures obtained as special cases of this model.

TWO MODELLING APPROACHES

A two age-class state-space model could be defined by either of the following transition equations:

$$\begin{pmatrix} N_{1,t} \\ N_{A,t} \end{pmatrix} = \begin{pmatrix} 0 & \rho\phi_1 \\ \phi_A & \phi_A \end{pmatrix} \begin{pmatrix} N_{1,t-1} \\ N_{A,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{A,t} \end{pmatrix} \tag{3}$$

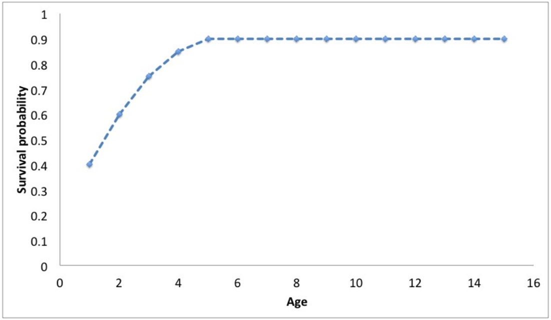
$$\begin{pmatrix} N_{1,t} \\ N_{2,t} \\ N_{A,t} \end{pmatrix} = \begin{pmatrix} 0 & \rho\phi_1 & \rho\phi_1 \\ \phi_A & 0 & 0 \\ 0 & \phi_A & \phi_A \end{pmatrix} \begin{pmatrix} N_{1,t-1} \\ N_{2,t-1} \\ N_{A,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{A,t} \end{pmatrix} \tag{4}$$

Model selection strategies

- Demographic data alone (eg. Besbeas et al, 2002);
- Demographic data alone used as starting point, then performing some model selection on integrated model (eg. McCrea et al, 2010);
- Integrated data.

Currently AIC is used for model selection of IPM.

AGE-DEPENDENT SURVIVAL PROBABILITIES



- Step-wise approach using likelihood-ratio tests (or score tests - eg. McCrea and Morgan, 2011)
  - Step 1: $\phi(1, 2+)$ vs $\phi(1, 2, 3+)$;
  - Step 2: $\phi(1, 2, 3+)$ vs $\phi(1, 2, 3, 4+)$ etc.

SIMULATION STRUCTURE

- Simulation based on grey heron data, commonly used in the IPM literature;
- Simulated 20 years of ring-recovery data and 71 years of census data;
- Range of parameter values have been investigated, representative results presented here.

Results

Table: 150 replications; number of times model is selected using the different approaches

		Number of adult age classes						
		1	2	**3**	4	5	6	7
min AIC	RR	0	40	**74**	16	4	9	7
	Int	0	0	**62**	38	19	18	13
parsimony AIC	RR	0	56	**85**	3	2	3	1
	Int	0	30	**91**	11	5	8	5
LRT	RR	0	63	**84**	3	0	0	0
	Int	0	36	**105**	9	0	0	0

- Minimum AIC performs better for demographic data alone than integrated data;

Results

Table: 150 replications; number of times model is selected using the different approaches

		Number of adult age classes						
		1	2	**3**	4	5	6	7
min AIC	RR	0	40	**74**	16	4	9	7
	Int	0	0	**62**	38	19	18	13
parsimony AIC	RR	0	56	**85**	3	2	3	1
	Int	0	30	**91**	11	5	8	5
LRT	RR	0	63	**84**	3	0	0	0
	Int	0	36	**105**	9	0	0	0

- Minimum AIC often selects models with larger number of age classes; AIC known to have problems for state-space models alone - see Bengtsson and Cavanaugh (2006)

Results

Table: 150 replications; number of times model is selected using the different approaches

		Number of adult age classes						
		1	2	**3**	4	5	6	7
min AIC	RR	0	40	**74**	16	4	9	7
	Int	0	0	**62**	38	19	18	13
parsimony AIC	RR	0	56	**85**	3	2	3	1
	Int	0	30	**91**	11	5	8	5
LRT	RR	0	63	**84**	3	0	0	0
	Int	0	36	**105**	9	0	0	0

- Introducing parsimony argument into AIC reduces the number of models with very large number of age classes being selected;

Results

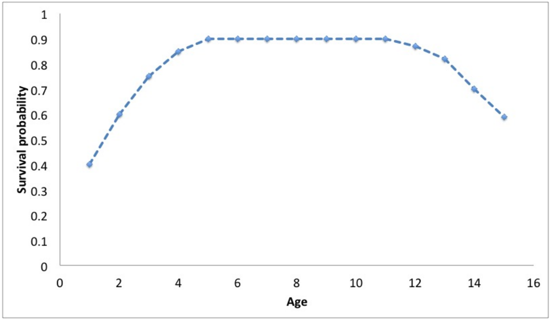
Table: 150 replications; number of times model is selected using the different approaches

		Number of adult age classes						
		1	2	**3**	4	5	6	7
min AIC	RR	0	40	**74**	16	4	9	7
	Int	0	0	**62**	38	19	18	13
parsimony AIC	RR	0	56	**85**	3	2	3	1
	Int	0	30	**91**	11	5	8	5
LRT	RR	0	63	**84**	3	0	0	0
	Int	0	36	**105**	9	0	0	0

- LRT approach performs the best for integrated models.

SENESCENCE

Senescence is a decrease in reproductive output and/or survival with increasing age.



Current model selection approaches for diagnosing senescence are ad hoc - eg. Gaillard et al, (2004).

Step up augmented with step down

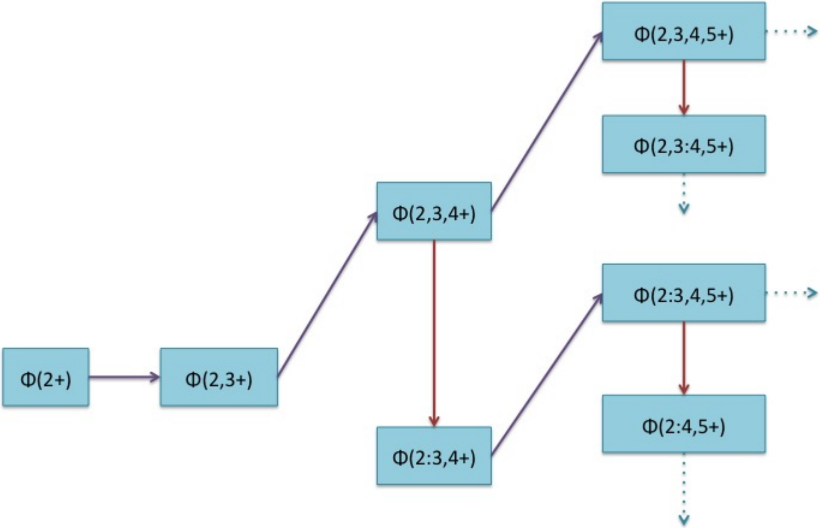


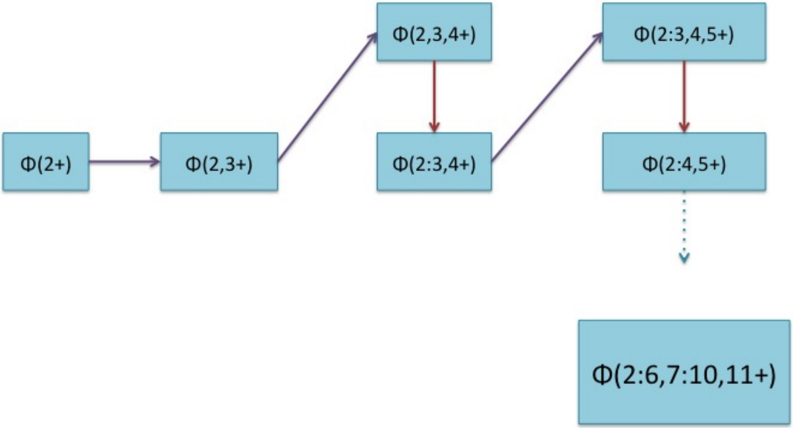
Figure: Ibex, *Capra ibex*

Data description

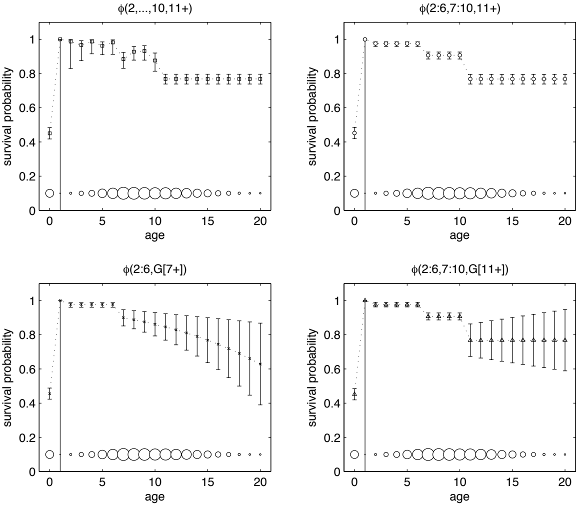
The data arise from a large monitoring study at the Gran Paradiso National Park in Italy.

- Multi-variate census data consisting of counts of adult males and adult females, yearlings and kids have been collected since 1956;
- Adult mark-recapture-recovery data from 1985.
- Kid survival and productivity data from 2000.

Step up augmented with step down



Alternative models



Conclusions

- Optimal model selection for integrated population models has not previously been addressed;
- Contrary to intuition adding information through integrated models can result in deterioration in model selection when AIC is used;
- Simple corrections of AIC are impossible to find as they are model dependent and AIC variants such as AICc, BIC etc are not easily defined in IPM;
- Step-wise approach using LRTs has been found to work well and is robust to irregular survival probability patterns, such as those in a population exhibiting senescence.

REFERENCES

Besbeas, P., McCrea, R. S. and Morgan, B. J. T. (2015) Integrated Population Model Selection in Ecology.

Bengtsson, T. and Cavanaugh, J. E. (2006) An improved Akaike information criterion for state-space model selection. *Computational Statistics & Data Analysis*, **50**, 2635–2654.

Besbeas, P., Freeman, S. N., Morgan, B. J. T. and Catchpole, E. A. (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58**, 540–547.

Brooks, S. P., King, R. and Morgan, B. J. T. (2004) A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, **27.1**, 515-529.

Cole, D. J. and McCrea, R. S. (2015) Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal*. Revision submitted.

Gaillard, J. M., Viallefont, A., Loison, A. and Festa-Bianchet, M. (2004) Assessing senescence patterns in populations of large mammals. *Animal Biodiversity and Conservation*, **27.1**, 47–58.

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of capture-recapture data*. Chapman and Hall/CRC Press. Boca Raton.

McCrea, R. S., Morgan, B. J. T., Gimenez, O., Besbeas, P., Bregnballe, T. and Lebreton, J.-D. (2010) Multi-site integrated population modelling. *Journal of Biological, Agricultural and Environmental Statistics*, **15**, 539-561.

McCrea, R. S. and Morgan, B. J. T. (2011) Multi-site mark-recapture model selection using score tests. *Biometrics*, **67**, 234–241.

Schaub, M. and Abadi, F. (2011) Integrated population models a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology* **152**, 227–237.