Integrated Population Modelling

Byron J.T. Morgan

University of Kent

Collaboration and grant support

- Joint research, with
  - Takis Besbeas,
  - Steve Brooks,
  - Ted Catchpole,
  - Stephen Freeman,
  - Ruth King,
  - José Lahoz-Monfort,
  - Jean-Dominique Lebreton,
  - Eleni Matechou
  - Rachel McCrea,
  - Giacomo Tavecchia
- Research grant support from Max Planck, CEH, BBSRC, NERC and EPSRC

Outline

- Introduction

- Capture-recapture

- Modelling ecological time series

- Kalman filter

- Integrated Population Modelling; not Integral Projection Modelling

- Illustrative results

- Extensions

- New research and discussion

- References

Workshop Bayesian integrated population modelling (IPM) using BUGS and JAGS

- Michael Schaub, Marc Kéry and David Koons
- 25 - 29 July 2016
- Utah State University, Logan,
- Integrated population models (IPMs) represent the powerful combination, in a single Leslie-type of model, of different data sources that are informative about the dynamics of an animal population
- Typical IPMs combine one or more time-series of counts with another data set that is directly informative about survival probabilities, such as ring-recovery or capture-recapture.
- However, many other sources of demographic information may also be envisioned.
- Currently, for non-statisticians the only practical manner to develop and fit an IPM is using BUGS software (WinBUGS, OpenBUGS, JAGS).

Ringing wild birds

- Birds are given unique marks
- They are then observed later, dead or alive
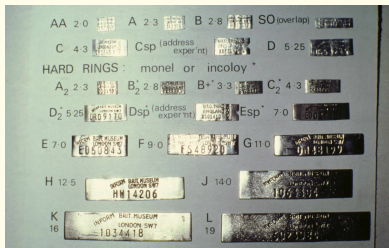
- Rings take different forms, and birds may be given more than one: colour for resighting and metal for dead recovery

Examples for identification of wild animals

Cormorant ring-recovery data

| Number | Number Recovered in Year (-1983) | | | | | | | | | | Never |
Released	1	2	3	4	5	6	7	8	9	10	Recovered
30	1	1	1	1	0	0	0	1	0	0	25
147		10	1	1	0	3	1	2	0	1	128
128			3	1	0	1	3	1	1	0	118
199				6	3	0	3	0	4	1	182
291					7	4	5	4	3	0	268
201						5	1	1	4	0	190
179							3	0	2	0	174
242								3	5	0	234
173									2	0	171
45										2	43

Cormorant ring-recovery data for breeding individuals released from 1983-1991.

Product multinomial likelihood

Making use of the assumption of independence of individuals between cohorts, the data can be modelled by a product of multinomials, and the log-likelihood is given by

$$\log(L_r) = \text{constant} + \sum_{i=1}^{T} \sum_{j=i+1}^{T} d_{i,j} \log(p_{i,j}) + \sum_{i=1}^{T} u_i \log(q_i) \tag{1}$$

- In a T-year study, $d_{i,j}$ individuals are reported dead at time t_j, from a cohort of R_i individuals ringed at time t_i, and the number of animals that are not recovered from the ith cohort are denoted by u_i, so that $u_i = R_i - \sum_j d_{i,j}$.
- The probability corresponding to the $d_{i,j}$ is denoted by $p_{i,j}$, and we write $q_i = 1 - \sum_j p_{i,j}, i = 1, \ldots, T$.
- We let ϕ_i denote the annual survival probability of individuals aged $i = 1, \ldots, a-1, a^+$, corresponding to a age groups, and λ denote the probability that an individual which dies is reported dead.
- The $p_{i,j}$ are modelled in terms of the annual survival probabilities. For example, for $j \leq a^+, p_{i,j} = \phi_1 \ldots \phi_{j-i}(1 - \phi_{j-i+1})\lambda$. Probabilities may be time dependent in general.

Parameter estimates for a time-dependent model

Parameter	MLE	SE	Parameter	MLE	SE
ϕ_1	0.86	0.139	ϕ_6	0.93	0.042
ϕ_2	0.71	0.134	ϕ_7	0.92	0.047
ϕ_3	0.90	0.063	ϕ_8	0.95	0.031
ϕ_4	0.90	0.059	ϕ_9	0.92	0.049
ϕ_5	0.93	0.040	ϕ_{10}	0.98	0.012
			λ	0.22	0.096

- Maximum-likelihood estimates from a model with time-dependent survival probabilities and constant recovery probability fitted to the cormorant ring-recovery data.
- SE denotes standard error, obtained from inverting the Hessian at the maximum-likelihood estimates. Here t_1 represents 1983.
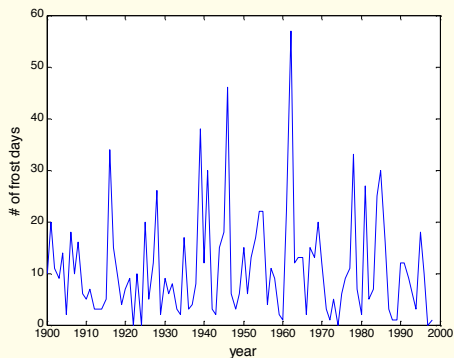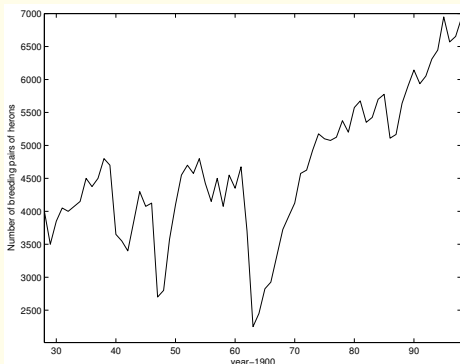
Grey heron: *Ardea cinerea*

The effect of cold winters
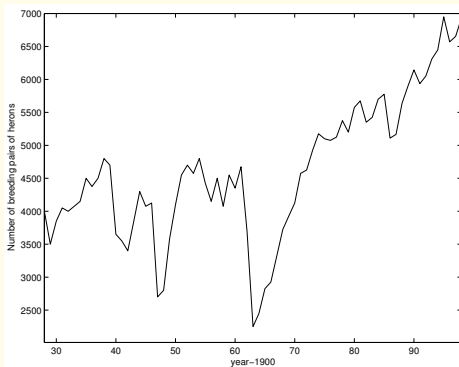
Grey heron

The effect of cold winters, measured by frost days

Transition equation

$$
\begin{pmatrix} N_{1,t} \\ N_{2,t} \\ \vdots \\ N_{a-1,t} \\ N_{a^+,t} \end{pmatrix} = \left(\begin{array}{ccccc|c} 0 & p\phi_1 & \dots & p\phi_1 & p\phi_1 & p\phi_1 \\ \hline \phi_2 & 0 & \dots & 0 & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & \dots & \phi_{a-1} & 0 & 0 \\ 0 & 0 & \dots & 0 & \phi_{a^+} & \phi_{a^+} \end{array} \right) \begin{pmatrix} N_{1,t-1} \\ N_{2,t-1} \\ \vdots \\ N_{a-1,t-1} \\ N_{a^+,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{a-1,t} \\ \epsilon_{a^+,t} \end{pmatrix} \quad (2)
$$

- $N_{r,t}$, $1 \leq r \leq a - 1$ and $N_{a^+,t}$ denote, numbers of female birds aged r-years and greater than $(a - 1)$-years at time t
- the $\{\epsilon_{i,t}\}$ denote process errors.

Modelling process errors

- The process variances are not free parameters, to be estimated.
- They have a particular structure resulting from the modelling.
- Survival may be binomial and recruitment Poisson.
- Thus corresponding to Equation 2, taking $a = 4$ for herons, we have

$$\text{Var}(\epsilon_{1,t}) = p\phi_1(N_{2,t-1} + N_{3,t-1} + N_{4^+,t-1})$$
$$\text{Var}(\epsilon_{2,t}) = \phi_2(1 - \phi_2)N_{1,t-1}$$
$$\text{Var}(\epsilon_{3,t}) = \phi_3(1 - \phi_3)N_{2,t-1}$$
$$\text{Var}(\epsilon_{4^+,t}) = \phi_{4^+}(1 - \phi_{4^+})(N_{3,t-1} + N_{4^+,t-1}). \tag{3}$$

- We make normal approximations for the Poisson and binomial.

Observation equation

- Grey herons are not thought to skip breeding in general, and we equate counting nests to counting breeding females. Breeding starts at age 1.
- The observed counts, $\{y_t\}$, are then given by the corresponding observation equation

$$y_t = (0, 1, \ldots, 1)(N_{1,t}, N_{2,t}, \ldots, N_{a^+,t})' + \eta_t, \tag{4}$$

- The $\{\eta_t\}$ are given a normal distribution (others are possible).
- We have a state-space model.
- We are then able to use the Kalman filter to approximate the likelihood for census data.

Recursions

In general, we write

$$\mathbf{N}_t = \mathbf{\Lambda}_t \mathbf{N}_{t-1} + \boldsymbol{\eta}_t$$
$$\mathbf{Y}_t = \mathbf{Z}_t \mathbf{N}_t + \boldsymbol{\epsilon}_t.$$

We assume $\boldsymbol{\epsilon}_t \sim N(0, \mathbf{Q}_t)$ and $\boldsymbol{\eta}_t \sim N(0, \mathbf{H}_t)$

- The Kalman filter is a recursive procedure to construct the likelihood function for appropriate state-space models.
- We write $\mathbf{a}_t = \mathbb{E}(\mathbf{N}_t)$ and $\mathbf{P}_t = \mathrm{Var}(\mathbf{N}_t)$.
- For starting values $\mathbf{a}_1$ and $\mathbf{P}_1$, the Kalman filter recursions are defined by

$$
\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t \\
\mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \\
\mathbf{G}_t &= \mathbf{\Lambda}_t \mathbf{P}_t \mathbf{Z}_t' \mathbf{F}_t^{-1} \\
\mathbf{a}_{t+1} &= \mathbf{\Lambda}_t \mathbf{a}_t + \mathbf{G}_t \mathbf{v}_t \\
\mathbf{L}_t &= \mathbf{\Lambda}_t - \mathbf{G}_t \mathbf{Z}_t \\
\mathbf{P}_{t+1} &= \mathbf{\Lambda}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{Q}_t.
\end{aligned}
\tag{5}
$$

Kalman filter likelihood

Following the Kalman filter we construct a likelihood function for the unknown parameters, θ, conditional on the observed data $\mathbf{y}_1, \cdots, \mathbf{y}_T$.

- Assuming that the initial state vector $\alpha_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$, the log-likelihood function is defined by

$$\log \mathrm{L}_c(\theta|\mathbf{y}) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right).$$

- The values of $\mathbf{v}_t$ and $\mathbf{F}_t$ are automatically calculated through the implementation of the Kalman filter on the last slide.

- This is the prediction-error decomposition likelihood due to the interpretation of $\mathbf{v}_t$ and $\mathbf{F}_t$.

Bringing it all together

- Models for the two data sets share common survival probabilities, and integrated population modelling exploits this feature.
- The two likelihoods are formed for the component data sets and the product is maximised to produce maximum-likelihood estimates for the complete set of model parameters.
- Important assumption of independence.

$$\log(L_j) = \log(L_r) + \log(L_c), \tag{6}$$

where L_j denotes the joint likelihood, L_r denotes the likelihood for the ring-recovery data, given in Equation 1, and L_c denotes the likelihood for the census data.

Further features of IPM

- The model for the census data includes a productivity parameter, *p*,
- although there is not a data set providing direct information on productivity, as a consequence of integrated population modelling we can estimate this parameter, along with an estimate of its standard error.
- For Grey herons we obtain the estimate of productivity, $\hat{p} = 0.96(0.07)$.

- Similar approaches are to be found in fisheries stock assessment models, where there may be differential weighting of the components in the joint log likelihood.

Formation of capture-recapture likelihood

- It is often the case that the capture-recapture likelihood is obtained from using a stand-alone computer package.

- It is difficult then to combine with other likelihoods to conduct a joint analysis.

- We can resolve this issue by approximating the capture-recapture likelihood near the maximum.

Multivariate normal approximation to capture-recapture-recovery likelihoods

- Following the use of a computer package, we can approximate the likelihood at the maximum by means of a multivariate normal distribution.
- The multivariate normal approximation is defined by

$$2 \log L_r(\boldsymbol{\theta}) = \text{constant} - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

where $\hat{\boldsymbol{\theta}}$ are the maximum-likelihood estimates and $\hat{\boldsymbol{\Sigma}}$ is the estimated variance-covariance matrix.

- We then use the approximation in Equation 6 to form the combined likelihood:

$$\log(L_j) = \log(L_r) + \log(L_c),$$

Examination of the MVN approximation: multistate model
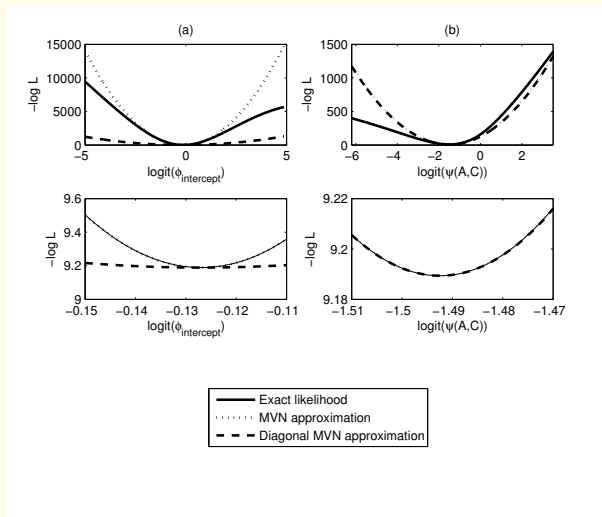
Bayesian approach: no need for Kalman filter etc. DAG
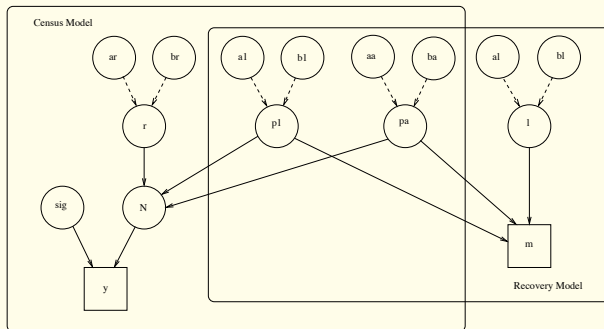
See Chapter 11 of Kéry and Schaub (2012), *Bayesian population analysis using WinBUGS: a hierarchical perspective.*

Advantages of IPMs

- Integrate the results from different surveys into a single analysis.
- Lead to better descriptions and understanding of population dynamics.
- The approach is quite general, eg., count data may be described using N-mixture models and stopover models. Capture-recapture models may be of different types.
- Productivity models are often, though not always, quite simple.
- Allow estimation of unknown parameters, sometimes with complex structure. For example, productivity, migration
- Allow the coherent estimation of standard errors
- Increased precision, which is useful for conservation applications.
- Extend readily to more complex situations, involving
  - several species (Lahoz-Monfort *in preparation*)
  - different animal types, eg., sex, age, breeding
  - different locations

Current applications of IPM: batch marking; stop-over model; dynamic model

Work with Emily Dennis, Marc Kéry, Takis Besbeas, Laura Cowan, Eleni Matechou....
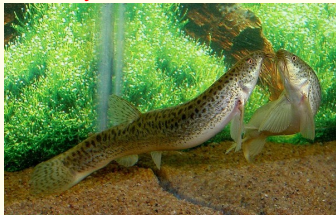
Northern lapwing: *Vanellus vanellus*

Lapwing decline: modelling a decline in productivity
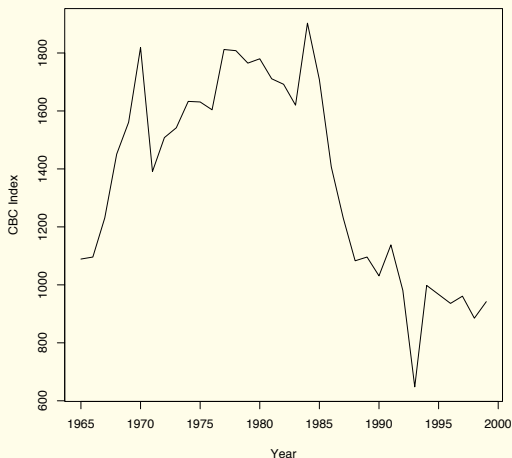
- Decline in population size is either due to a fall in productivity
- or a fall in survival
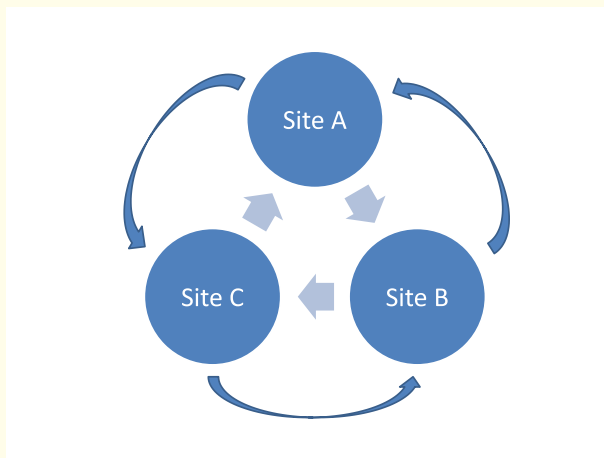- the former is thought to be the case.

Grey heron: *Ardea cinerea*

A threshold model: An illustration of adding complexity to modelling productivity.

- We can model productivity in terms of thresholds.
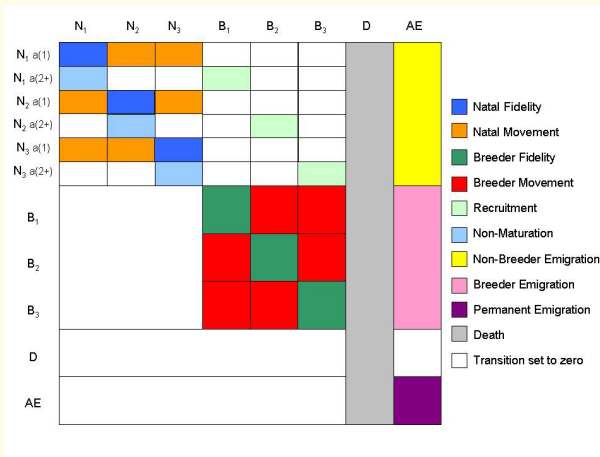- Eg: instead of $\log \rho_t = \kappa_0 + \kappa_1 y_t$, we might set

$$\log \rho_t = \begin{cases} \nu_0 + \nu_1 & \text{if } y_t < \tau, \\ \nu_0 & \text{if } y_t \geq \tau. \end{cases}$$
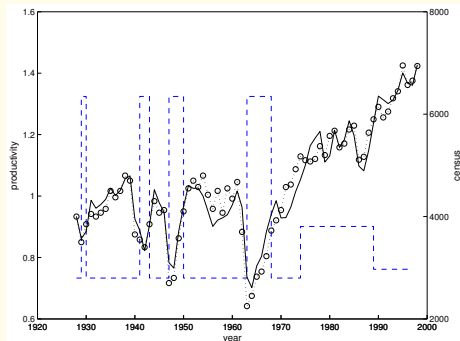
Movement between sites: multi-site model

## Transition between different states: multi-state model

## Issues remaining

Several issues remain for investigation

1. Is variance estimation correct?
2. How do we check for goodness of fit?
3. How do we choose between models?
4. How important is the requirement of independent data from different surveys?
5. Parameter redundancy: what parameters can we estimate?
6. Might IPM be inappropriate? Eg: combination of citizen-science data with transect data. See Pagel *et al Methods in Ecology and Evolution (2014)*.
7. How might one check for consistency between different data sets? See Popescu *et al Ecology and Evolution, (2014)*

1. Variance estimation

- Profile log-likelihoods when there is over dispersion, indexed by θ
- Observation variance may be estimated on a boundary to the parameter space
- Shown here for models of increasing complexity for heron data
- We consider ways of dealing with that.
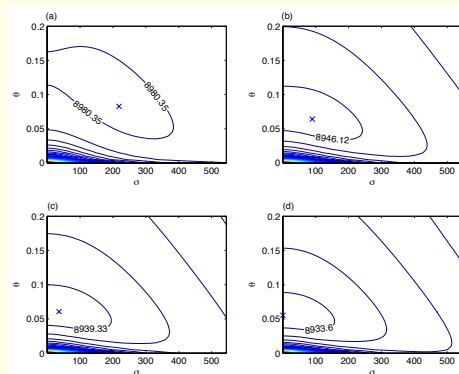
Pseudo replication
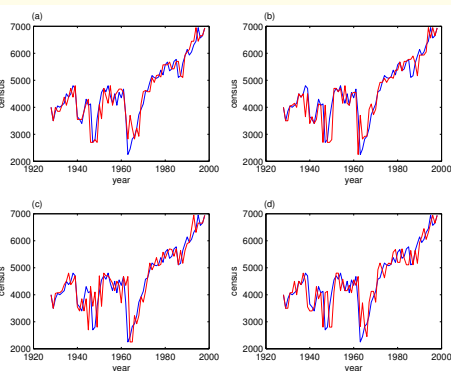
- Replicated time series provide good estimates of observation variance
- However it is hard to obtain such replication
- Differenced ecological time series have little structure
- We can obtain pseudo replicates by randomly sampling from neighbouring observations

## Penalised likelihood and plug-in method
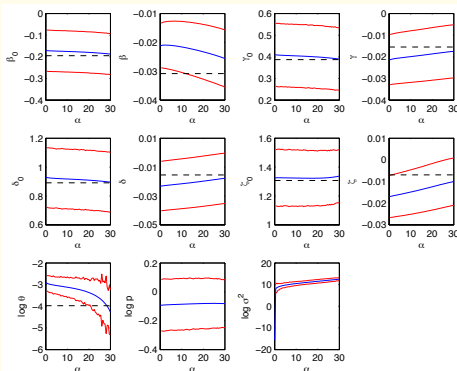
- We can penalise values of $\sigma = 0$:

 $$\log(L_j^p) = \log(L_r) + \log(L_c) + \alpha\log(\sigma).$$

- However the penalty makes little difference to parameters of interest
- Can estimate σ from a spline fit, and use that estimate
- A resulting plug-in method also does well in simulations
- There is apparently little effect on estimated standard errors

## 2. Model selection using AIC: determining the number of age classes
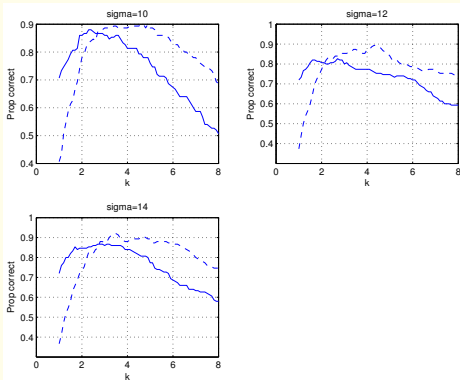
- $\mathrm{AIC} = -2 \log \mathrm{L} + \mathrm{k}d$, where d denotes the number of parameters, and $\mathrm{k} = 2$
- We vary k in a simulation study
- solid line = MRR only
- dashed line = IPM
- We see that using $k = 2$, MRR is better than IPM?
- AIC needs a larger penalty, $k > 2$
- We find that step-up likelihood-ratio tests perform better

A real example: Modelling senescence in ibex, *Capra ibex*
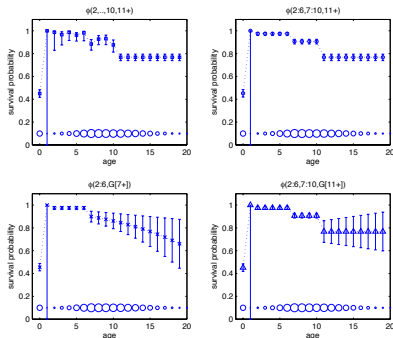
- Step-up likelihood-ratio tests combined with step-down approach
- Comparison with using Gompertz curve for describing senescence
- Best not to use Gompertz in this case.

3. Goodness of fit
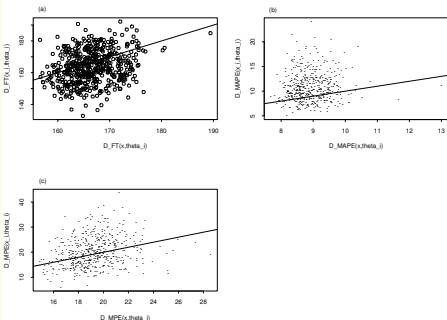
- model check using classical version of Bayesian p-values
- We use asymptotic multivariate normal distribution of maximum-likelihood estimators.
- For each simulated value we construct discrepancy values with real data and the model and a simulated data set and the model
- we use different discrepancy measures for capture-recapture data (Freeman-Tukey) and time-series (two measures used)
- p_c values obtained are: 0.39, 0.72, 0.52.
- We then use simulation to check these values.

3. Goodness of fit: use of calibrated simulation

- We use calibrated simulation to check the goodness-of-fit results
- This is done for a variety of models, for lapwing data
- We find that the p_c-values obtained are in agreement with the appropriate/best model
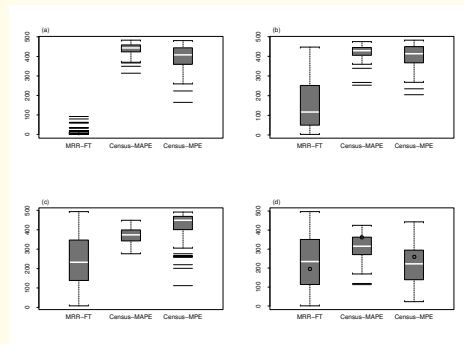
Summary

- Integrated population modelling is now a standard procedure in statistical ecology

- There are many examples of its use, in a variety of contexts

- Important features of the approach are still being elucidated.

References

- Besbeas *et al* (2002) *Biometrics*.
- Besbeas *et al* (2015) *Technical Report*.
- Besbeas and Morgan (2014) *Methods in Ecology & Evolution*.
- Besbeas and Morgan (2016) *AiSA, under review*.
- Fournier and Archibald 1982) *Can. J. Fish Aquat. Sci.*.
- Kéry and Schaub (2012) *Bayesian Population Analysis using WinBUGS: a hierarchical perspective,* Ac. Press.
- Lahoz-Monfort *et al* (2016) *Technical Report*.
- McCrea and King (2016) In preparation.
- McCrea and Morgan (2014) *Analysis of capture-recapture data*, CRC Chapman & Hall, Boca Raton.
- Maunder and Punt (2013) *Fisheries Research*.
- Newman *et al* (2014) *Modelling Population Dynamics*, Springer.
- Schaub and Abadi (2011) *J. Ornithology*.

Many processes are state-space models: eg CJS

- Here $x(i, t), i = 1, \ldots, n, t = 1, \ldots, T$ are Bernoulli indicator random variables for the elements of life histories obtained on n individuals, describing whether or not an individual is alive, when $x(i, t) = 1$, or dead, when $x(i, t) = 0$.

- The state process is given solely in terms of survival. For the ith individual we have

$$x(i, t)|x(i, t - 1) \sim \text{Bernoulli}(x(i, t - 1)\phi_{t-1})$$

- The observation equation is then given by

$$y(i, t)|x(i, t) \sim \text{Bernoulli}(p_t x(i, t))$$

  where the capture probability is p_t.

- Therefore the observed capture history for the ith individual is given by $\{y(i, t)\}_{t=f_i}^{T}$.

Smoothing

- The filtering algorithm evaluates the expected value of the state vector α_t conditional on the information available at time t, i.e. $\mathbb{E}(\alpha_t | \mathbf{y}_1, \ldots, \mathbf{y}_t)$.
- The next step of smoothing determines the expected value of α_t conditional on all available data, i.e. $\mathbb{E}(\alpha_t | \mathbf{y}_1, \ldots, \mathbf{y}_T)$.
- The most appropriate smoothing algorithm is fixed-interval smoothing. The fixed-interval smoothing algorithm commences with the final estimates $\mathbf{a}_T$ and $\mathbf{P}_T$ and iterates backwards.
- The recursions are defined by:

$$\mathbf{a}_{t|T} = \mathbf{a}_t + \mathbf{P}_t^*(\mathbf{a}_{t+1|T} - \mathbf{\Lambda}_{t+1}\mathbf{a}_t)$$

and

$$\mathbf{P}_{t|T} = \mathbf{P}_t + \mathbf{P}_t^*(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_t^{'*}$$

where $\mathbf{P}_t^* = \mathbf{P}_t \mathbf{\Lambda}_{t+1}' \mathbf{P}_{t+1|t}^{-1}$ for $t = T-1, \ldots, 1$ with $\mathbf{a}_{T|T} = \mathbf{a}_T$ and $\mathbf{P}_{T|T} = \mathbf{P}_T$. Computationally, this is a simple step of programming and just requires that α_t and $\mathbf{P}_t$ are stored for all values of t.