Statistical Ecology Seminar, Canterbury 2014

Around the models in 60 minutes
a whirlwind tour of capture-recapture

Rachel McCrea

NERC SCIENCE OF THE ENVIRONMENT

EPSRC Engineering and Physical Sciences Research Council

ncse National Centre for Statistical Ecology

University of Kent

Overview

- Introduction
- Model development
- Case studies
- Computer software
- Recent developments (at Kent)

Historical Context

- Estimation of the population of France (1802)
- Human demography
  - What are we interested in?
- Marking of birds started ~1899 in Denmark
- Capture-recapture techniques widely applied:
  - Medical
  - Epidemiological
  - Sociological
  - Ecological

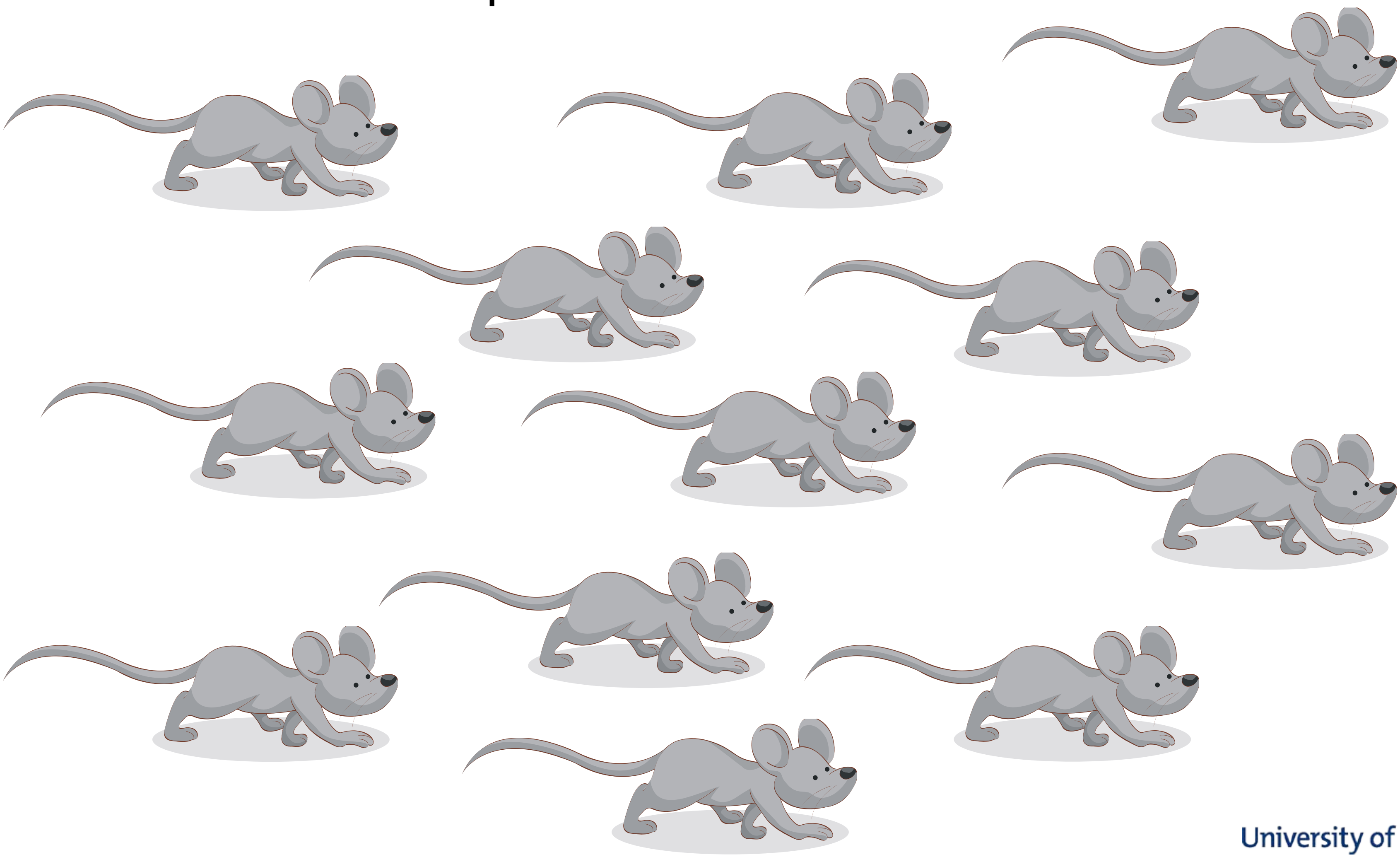
MODEL DEVELOPMENT

Lincoln-Petersen Estimate

- Two sample experiment
- Interested in estimating population size, **N**

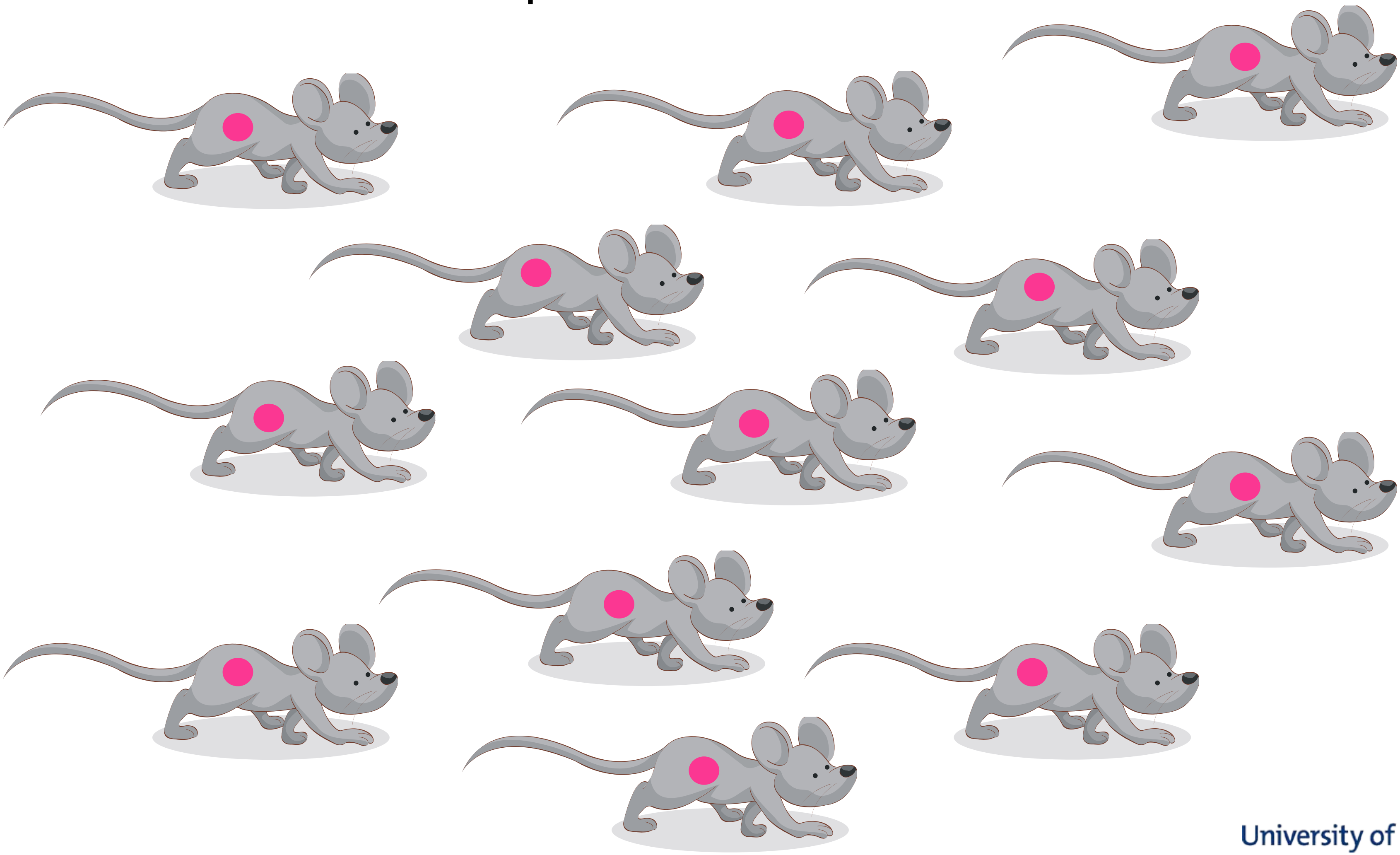
Lincoln-Petersen Estimate

- First sample



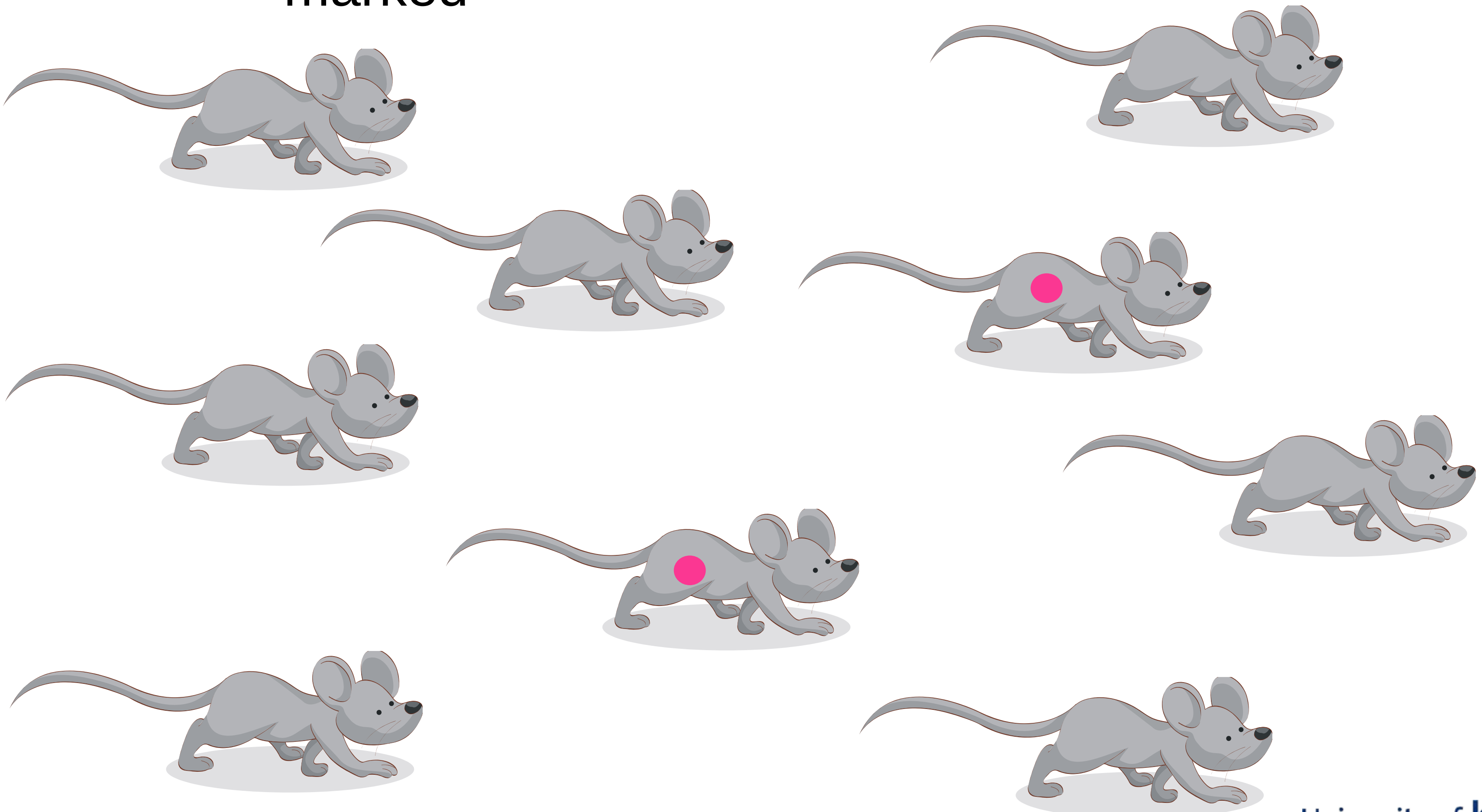
Lincoln-Petersen Estimate

- Mark all captured individuals, K



Lincoln-Petersen Estimate

- Second sample, n animals captured, k already marked



Lincoln-Petersen Estimate

- The proportion of marked individuals caught (k/K) should equal the proportion of the total population caught (n/N)

$$\widehat{N} = \frac{Kn}{k}$$

- Biased for small sample sizes, adapted Chapman estimator:

$$\widehat{N} = \frac{(K+1)(n+1)}{k+1} - 1$$

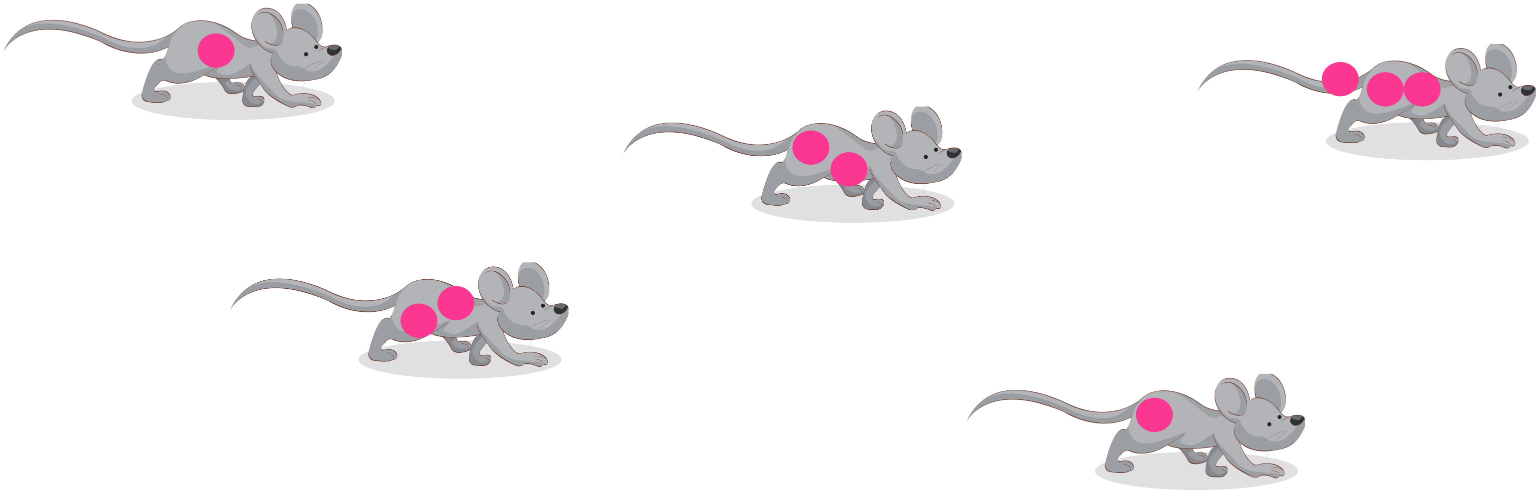
Lincoln-Petersen Estimate

- Model assumptions:
  - Population is closed (no births, deaths, immigration or emigration)
  - Marks do not affect catchability of animals
  - Marked animals mix with general population
  - No mark loss
  - All animals equally likely to be captured

Lincoln-Petersen

Schnabel census

- Extended number of sampling occasions, T
- f_t: number of individuals captured t times



Schnabel census

- Extended number of sampling occasions, T
- f_t: number of individuals captured t times

$$L(N,p) \propto \frac{N!}{(N-D)!} \prod_{t=0}^{T} g(t)^{f_t}$$

- D is the total number of distinct individuals caught, $D = \sum_{t=0}^{T} f_t$
- g(t) is the probability an animal is caught t times

$$g(t) = (1-p)^{T-t} p^t$$

- p: probability of capturing an individual

Schnabel census

- Extended number of sampling occasions, T
- f_t: number of individuals captured t times

$$L(N,p) \propto \frac{N!}{(N-D)!}\prod_{t=0}^{T} g(t)^{f_t}$$

- D is the total number of distinct individuals caught, $D = \sum_{t=0}^{T} f_t$
- g(t) is the probability an animal is caught t times

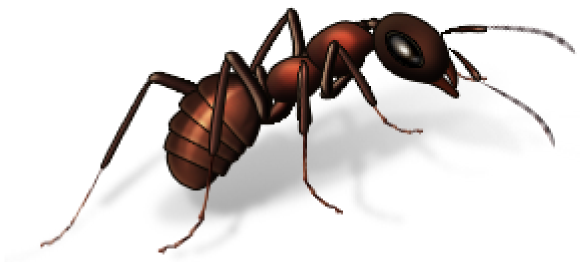
M_0

$$g(t) = (1-p)^{T-t}p^t$$

- p: probability of capturing an individual

Lincoln-Petersen

Schnabel census, M_0



- Ayre (1962) estimated an anthill population to be 109 when there were known to be 3,000 ants in it.

Ayre, L.G. 1962. Problems in using the Lincoln index for estimating the size of ant colonies (*Hymenopter formicidae*). *J.N.Y. Ent. Soc.* **70**: 159-166.

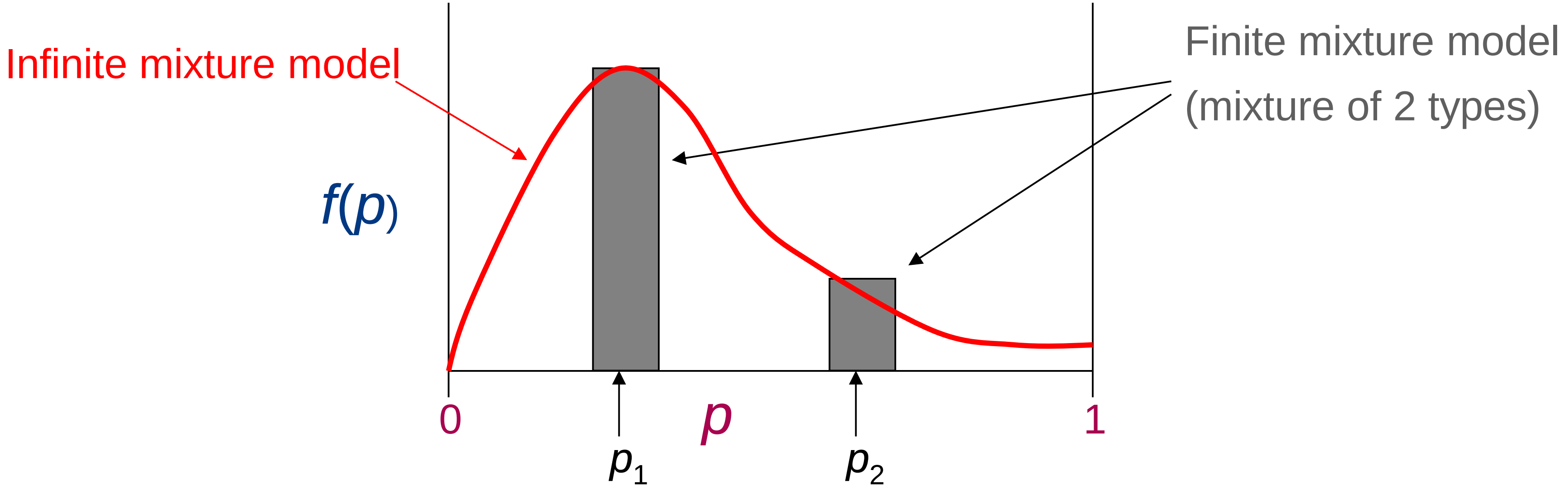
Heterogeneity models, M_h

- Not all animals have the same capture probability
- Finite Mixture models[1]: approximate the distribution of p using a mixture of a few ps (conceptually from a few sub-populations).
- Continuous (or infinite mixture) models[2]: Model the distribution of p using some flexible continuous distribution (e.g. Beta)
- Combinations of both[3].

[1] Pledger, (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**: 434-442.

[2]. Dorazio, and Royle, (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**: 352-364.

[3] Morgan and Ridout (2008) A new mixture model for capture heterogeneity. *JRSS-C* **57**, 433-446.



Infinite mixture model
Finite mixture model
(mixture of 2 types)
f(p)
0
p
1
p_1
p_2

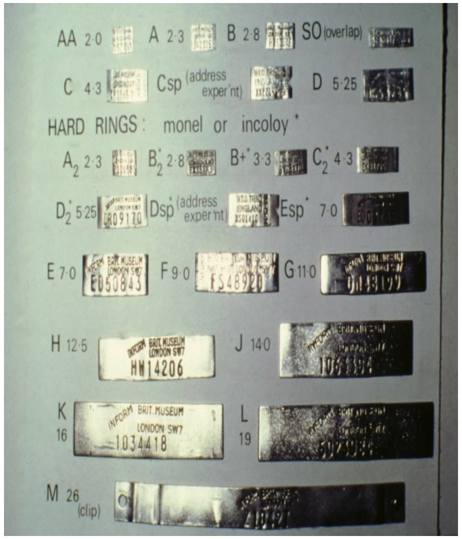
Lincoln-Petersen

Schnabel census, M_0

Heterogeneity model, M_h

Individual marking

- Capture-recapture data
  - 10010
  - 11011
  - 00101
  - …
- Retain individual and temporal information









Closed population model, M_t

- p_t: probability an individual is captured at occasion t

- Capture-recapture data and probabilities

  - 10010 $p_1(1-p_2)(1-p_3)p_4(1-p_5)$
  - 11011 $p_1p_2(1-p_3)p_4p_5$
  - 00101 $(1-p_1)(1-p_2)p_3(1-p_4)p_5$
  - …

Closed population models

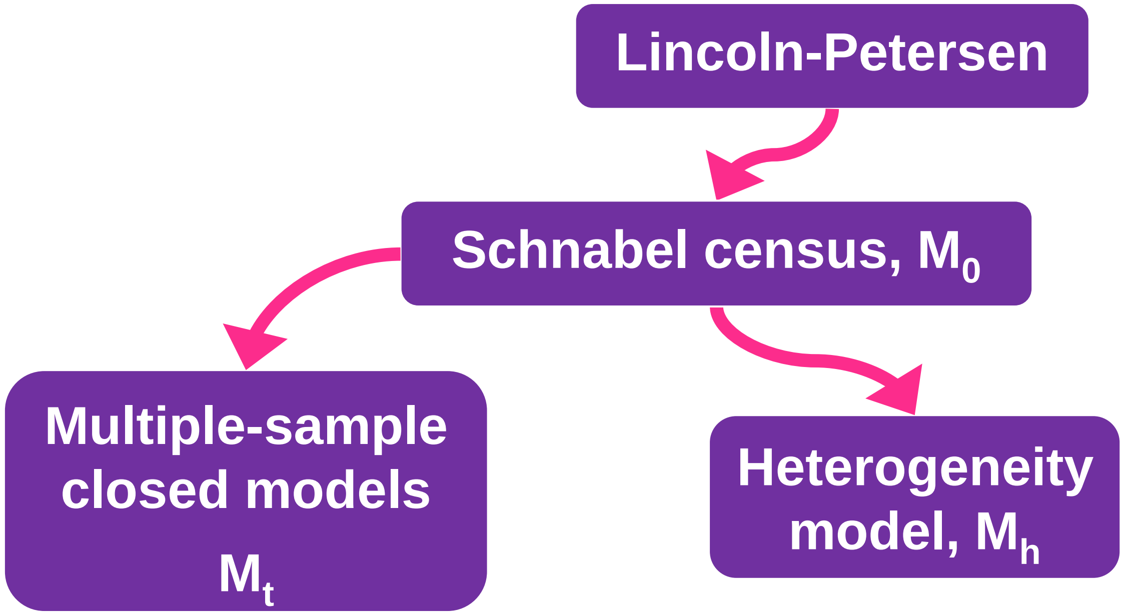
- Some individuals will not be captured at all during the study
- The encounter history for these individuals is given by
  - 00000 $(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5)$
- It is the number of individuals who are never captured that we need to estimate to estimate in order to estimate the total population size.

Closed population models

- Form a likelihood function as

$$L \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \Pr(h_i) \times \Pr(h_0)^{N-D}$$

- h_i: observed encounter history for individual i
- h_0: encounter history of never captured
- N: population size
- D: observed number of individuals
- Maximise L, to obtain *maximum-likelihood estimates* $\widehat{p}_t$ and $\widehat{N}$.

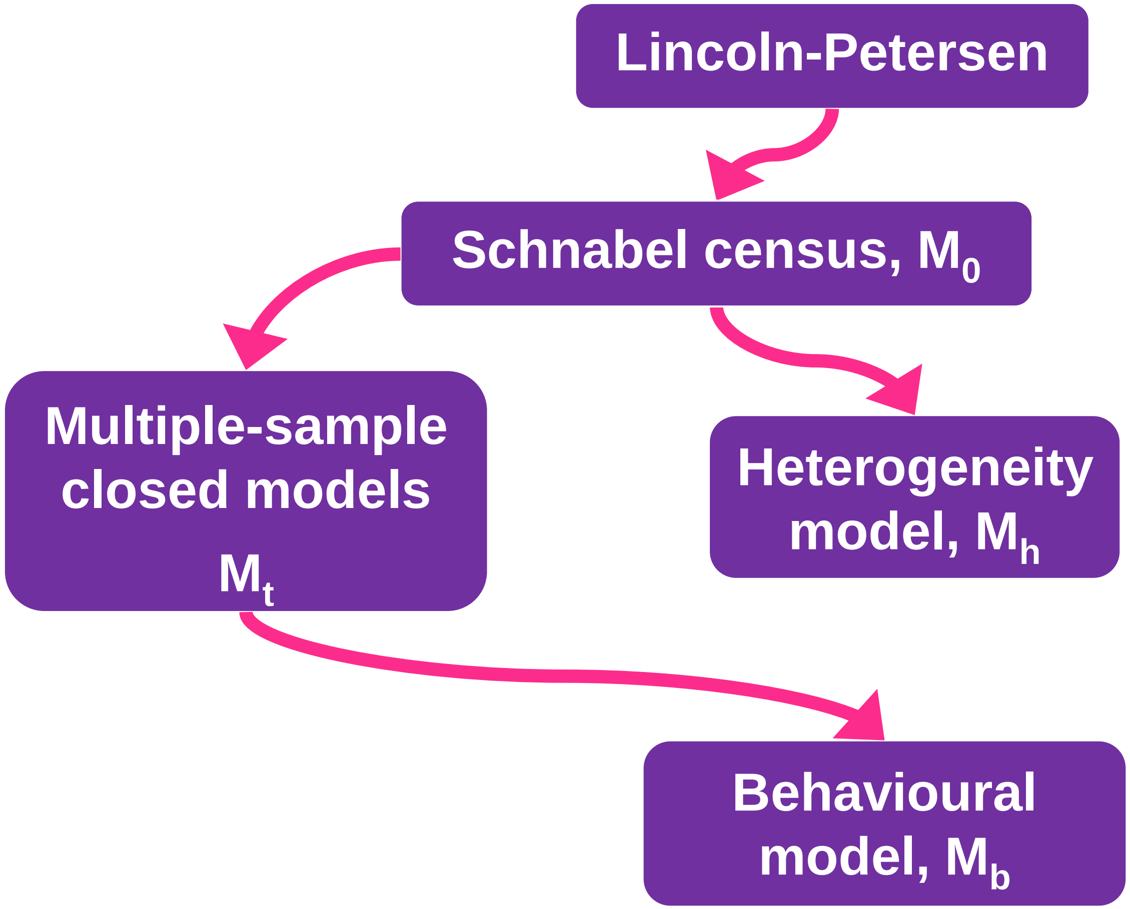

Lincoln-Petersen
Schnabel census, M_0
Multiple-sample closed models M_t
Heterogeneity model, M_h

Behavioural model, M_b

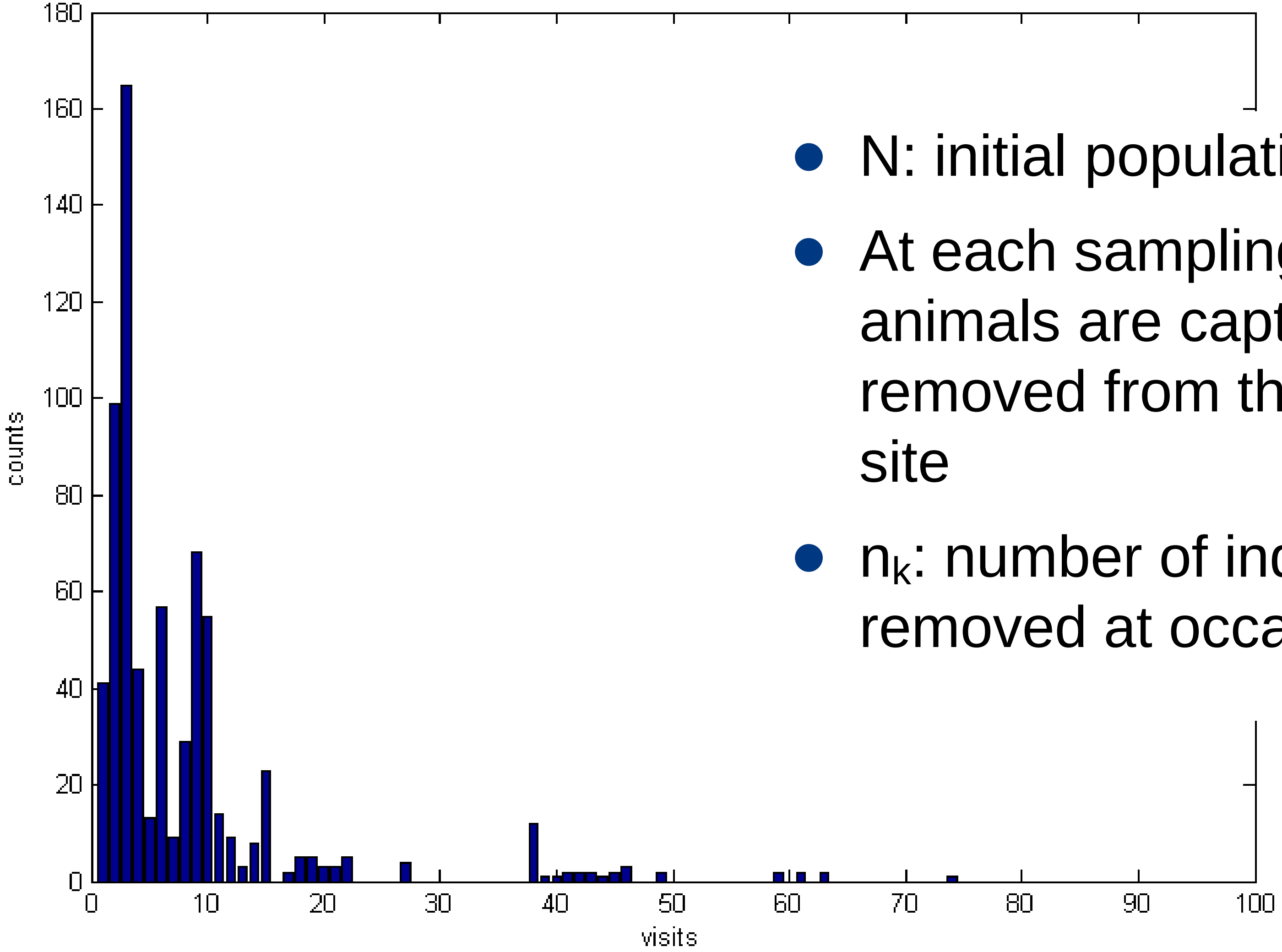
- p: probability an individual is captured at a given occasion
- Capture-recapture data and probabilities
  - 10010 p(1-p)(1-p)p(1-p)
  - 11011 pp(1-p)pp
  - 00101 (1-p)(1-p)p(1-p)p
  - …

Behavioural model, M_b

- c: probability of initial capture
- p: probability an individual is subsequently captured
- Capture-recapture data and probabilities
  - 10010 $c(1-p)(1-p)p(1-p)$
  - 11011 $cp(1-p)pp$
  - 00101 $(1-c)(1-c)c(1-p)p$
  - …

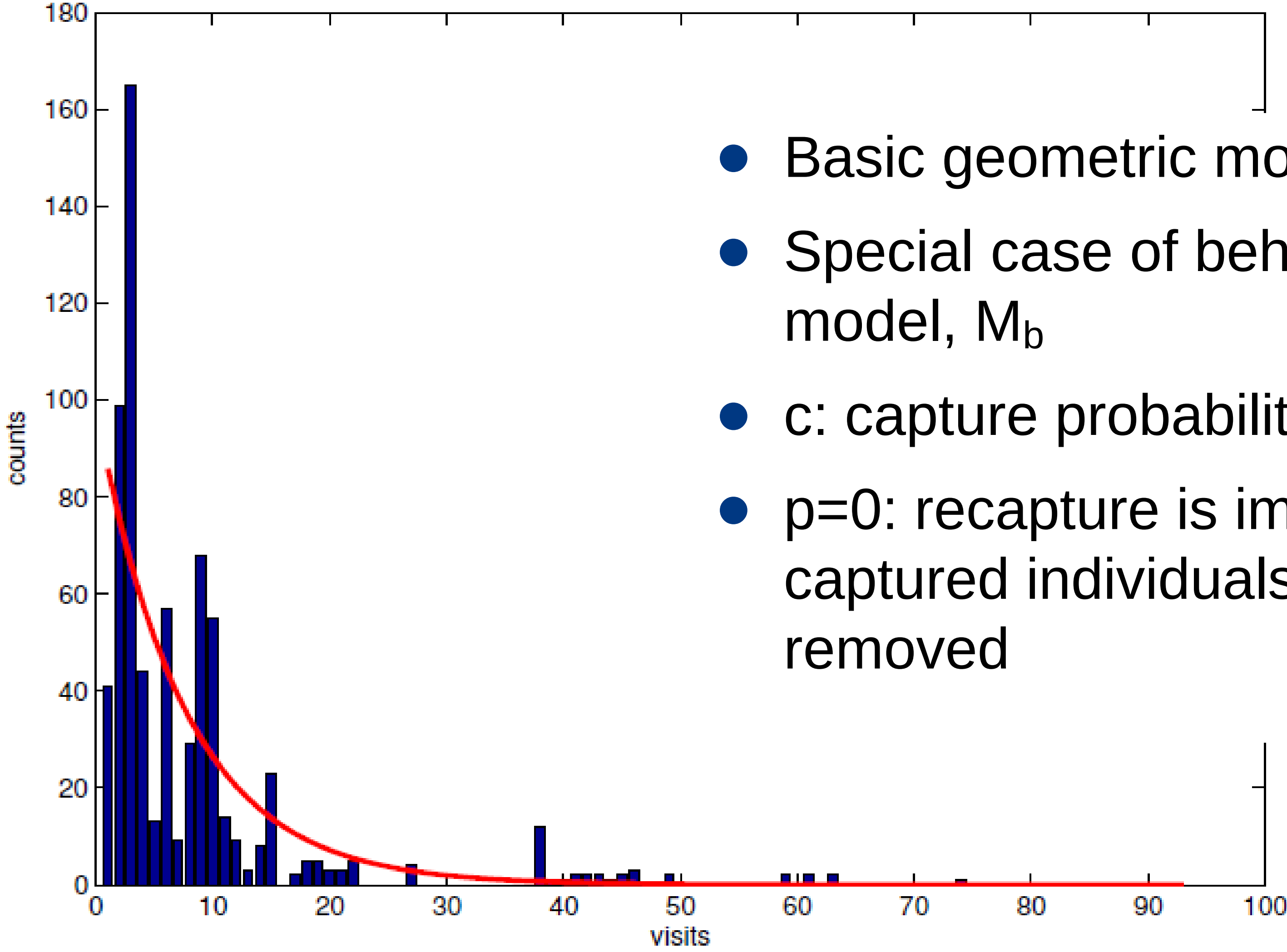

Lincoln-Petersen
Schnabel census, M_0
Multiple-sample closed models M_t
Heterogeneity model, M_h
Behavioural model, M_b

Depletion/removal model

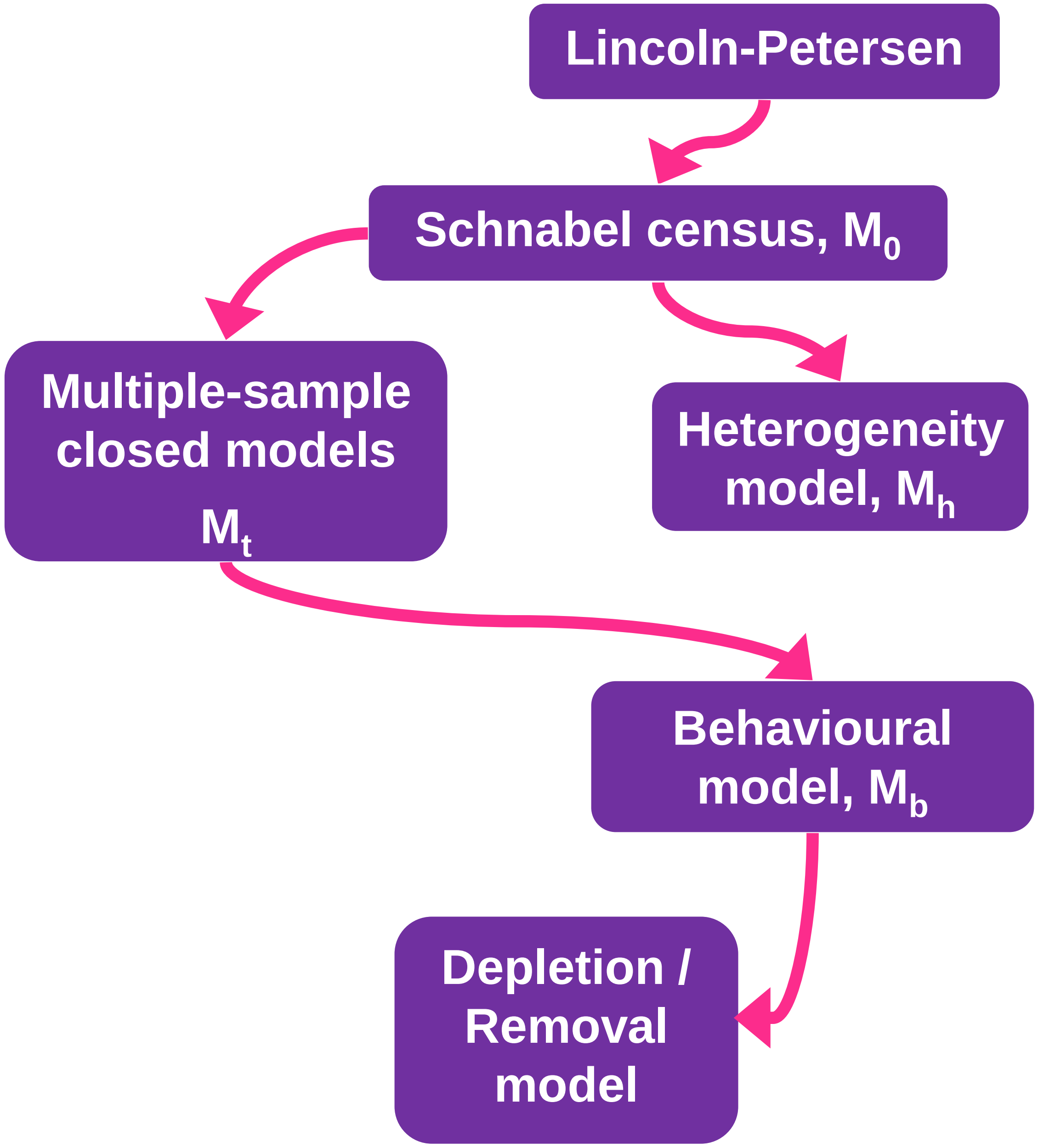


- N: initial population size
- At each sampling occasion animals are captured and removed from the study site
- n_k: number of individuals removed at occasion k

Depletion/removal model



- Basic geometric model
- Special case of behavioural model, M_b
- c: capture probability
- p=0: recapture is impossible as captured individuals are removed


Lincoln-Petersen
Schnabel census, M_0
Multiple-sample closed models M_t
Heterogeneity model, M_h
Behavioural model, M_b
Depletion / Removal model

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Entry time into the study population

University of Kent

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Entry time into the study population

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Entry time into the study population

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Departure time out of the study population

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Departure time out of the study population

Jolly-Seber model

- Studied population might not be closed, but still want to be able to estimate population size
- Jolly-Seber model
  - POPAN/Schwarz-Arnason formulation

0 0 1 0 0

- Departure time out of the study population

Jolly-Seber model

- If you assume the population is closed when it is not, the parameter estimates will be biased.

- Parameters for the Jolly-Seber model
  - N: **population size**
  - β_t: proportion of individuals first **available for capture** at occasion t+1, $\sum_{j=0}^{T-1} \beta_j$
  - p_t: probability an individual is **captured** at occasion t
  - ϕ_t: probability an individual present in the study area at occasion t **remains in the study area** until occasion t+1

Jolly-Seber model

- When forming the probability of an observed encounter history we need to sum over possible entry and departure times.
  - Suppose individual is first captured at occasion f_i and last captured at occasion l_i
  - $x_{ij}=1$ if individual i is captured at occasion j, $x_{ij}=0$ otherwise

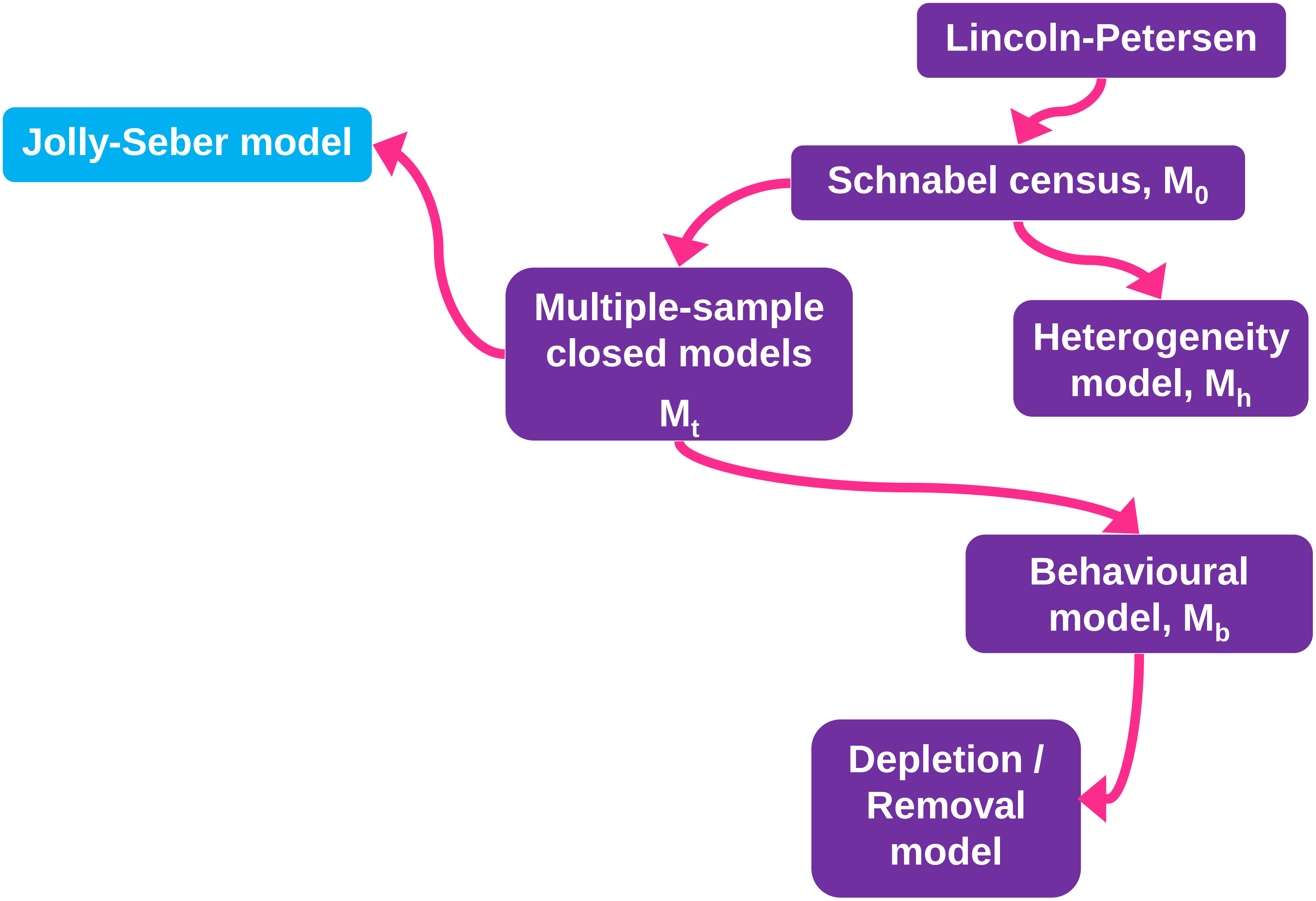
$$\Pr(h_i) = \sum_{b=1}^{f_i} \sum_{d=l_i}^{T} \beta_{b-1} \left(\prod_{j=b}^{d-1} \phi_j \right) (1-\phi_d) \left\{ \prod_{j=b}^{d} {p_j}^{x_{ij}} (1-p_j)^{1-x_{ij}} \right\}$$

Jolly-Seber model

- Corresponding probability of an individual not captured during the study

$$\Pr(h_0) = \sum_{b=1}^{T}\sum_{d=1}^{T}\beta_{b-1}\left(\prod_{j=b}^{d-1}\phi_j\right)(1-\phi_d)\left\{\prod_{j=b}^{d}(1-p_j)\right\}$$

$$L \propto \frac{N!}{(N-D)!}\prod_{i=1}^{D}\Pr(h_i) \times \Pr(h_0)^{N-D}$$


Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Multiple-sample closed models M_t
Heterogeneity model, M_h
Behavioural model, M_b
Depletion / Removal model

Stopover model

- Generalised version of the Jolly-Seber model
- Parameters are defined to be age-dependent
  - **age** corresponds to the time spent in the study area
  - arrival times and departure times are not independent

0 0 1 0 0

a

Stopover model

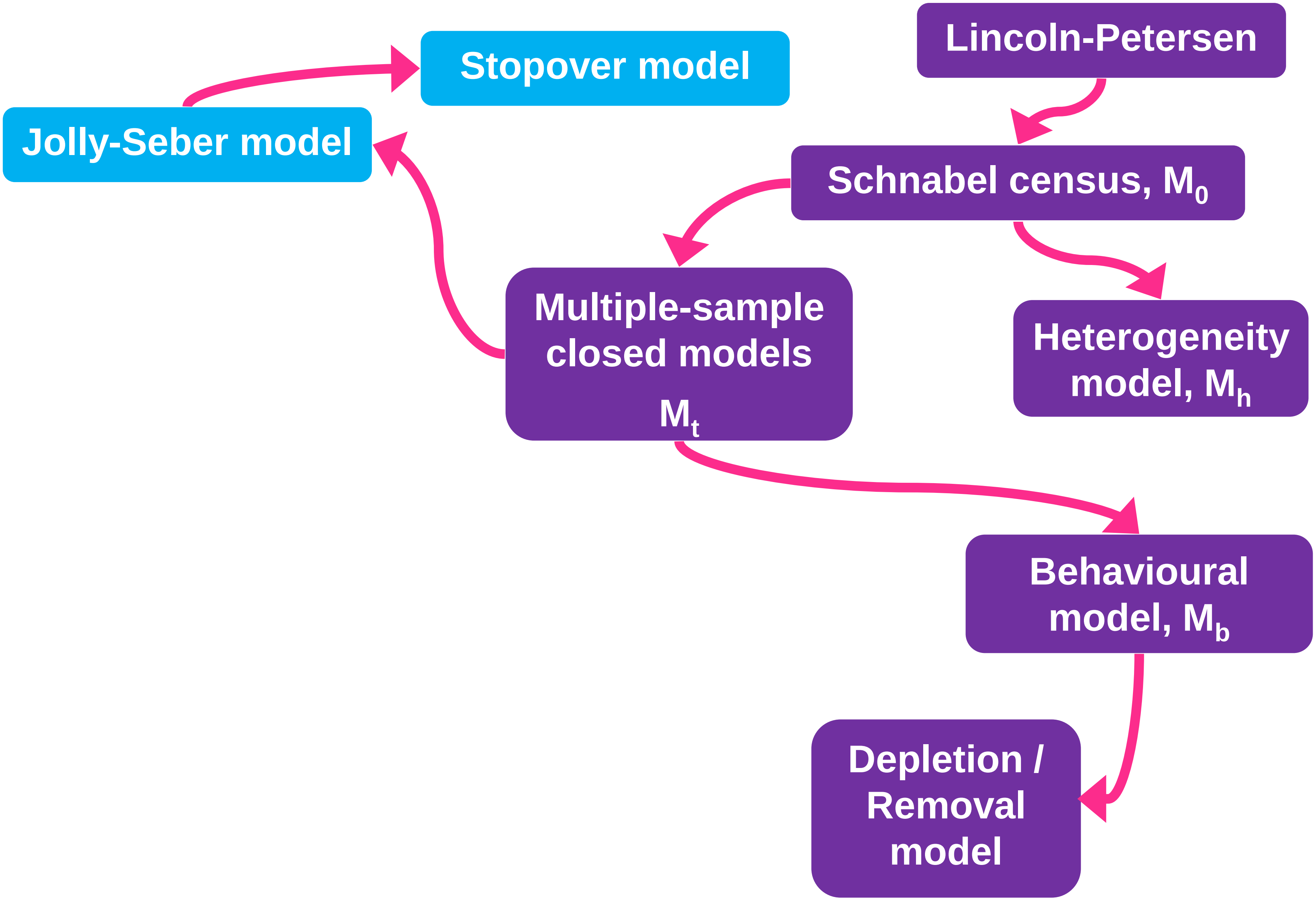
- Parameters
  - N: **population size**; this represents the total number of individuals who have been available for capture on at least one occasion
  - β_t: proportion of individuals first **available for capture** at occasion t+1, $\sum_{j=0}^{T-1} \beta_j = 1$
  - $p_t(a)$: probability an individual which entered the study a occasions previously is **captured** at occasion t
  - $\phi_t(a)$: probability an individual present in the study area at occasion t, which entered the study a occasions previously, **remains in the study area** until occasion t+1.

Stopover model

$$L \propto \frac{N!}{(N-D)!} \prod_{i=1}^{D} \Pr(h_i) \times \Pr(h_0)^{N-D}$$

$$\Pr(h_0) = \sum_{b=1}^{T} \sum_{d=b}^{T} \left(\beta_{b-1} \left\{ \prod_{j=b}^{d-1} \phi_j(a) \right\} \{1 - \phi_d(d-b+1)\} \left[\prod_{j=b}^{d} \{1 - p_j(a)\} \right] \right)$$

$$\Pr(h_i) = \sum_{b=1}^{f_i} \sum_{d=l_i}^{T} \left(\beta_{b-1} \left\{ \prod_{j=b}^{d-1} \phi_j(a) \right\} \{1 - \phi_d(d-b+1)\} \times \left[\prod_{j=b}^{d} p_j(a)^{x_{ij}} \{1 - p_j(a)\}^{1-x_{ij}} \right] \right),$$



University of Kent

Cormack-Jolly-Seber model

- Population is open – births and death can occur within the study period
- Population size is no longer the parameter of interest
- ϕ: survival probability
- Condition on first capture

Cormack-Jolly-Seber model

- Capture-recapture data and probabilities
  - 10010 $\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4 \chi_4$
  - 11011 $\phi_1 p_2 \phi_2 (1-p_3)\phi_3 p_4 \phi_4 p_5$
  - 00101 $\phi_3(1-p_4)\phi_4 p_5$
  - …
- ϕ_t: probability an individual survives from occasion t to t+1
- χ_t: probability an individual is not seen after occasion t

Capture-recapture m-array

Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	m_{12}	m_{13}	m_{14}	R_1-m_{12}-m_{13}-m_{14}
R_2		m_{23}	m_{24}	R_2-m_{23}-m_{24}
R_3			m_{34}	R_3-m_{34}

- R_i: Number of individuals released in year i
- m_{ij}: Number of individuals released in year i and next recaptured in year j

Encounter histories to m-arrays

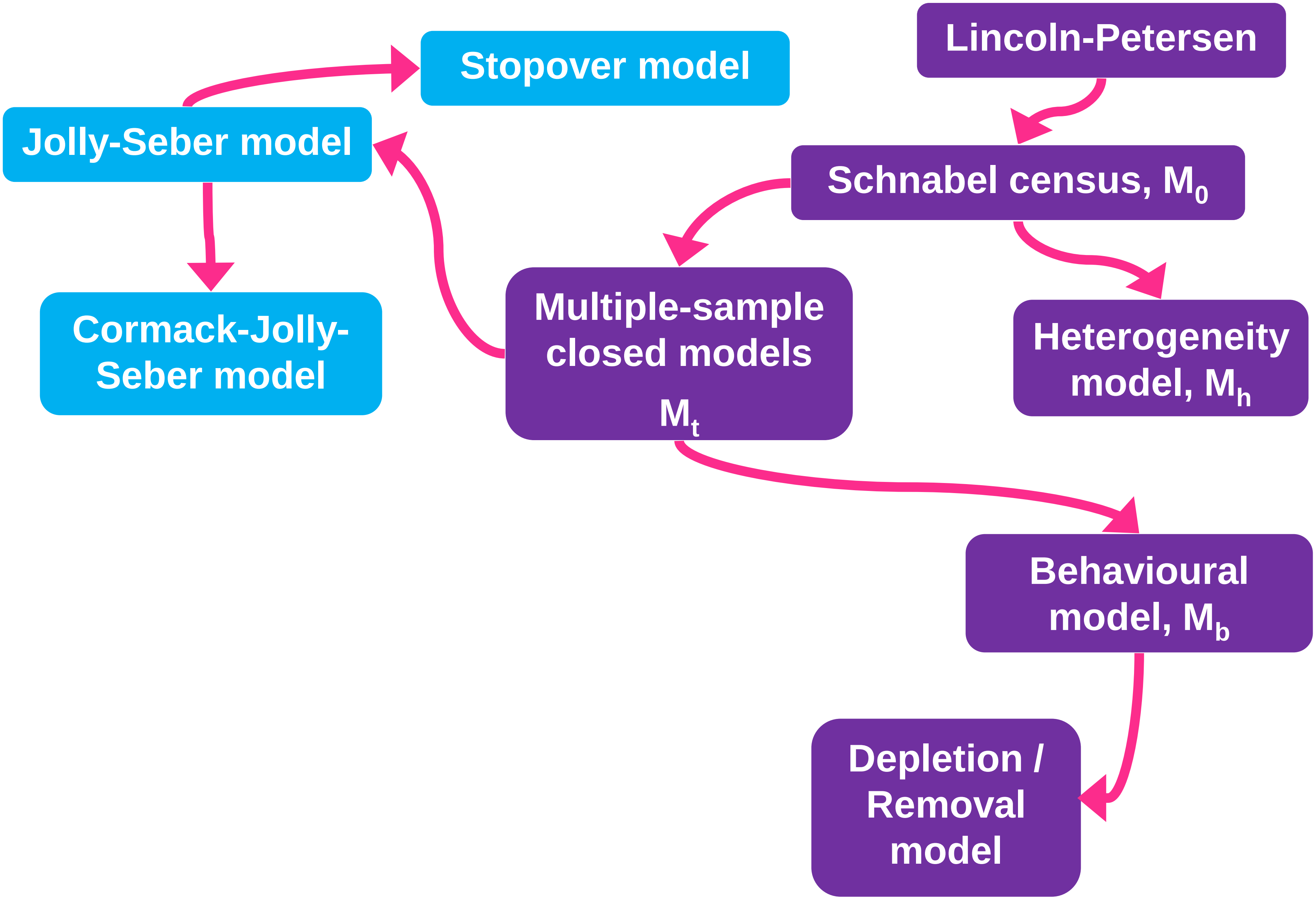
Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	m_{12}	m_{13}	m_{14}	R_1-m_{12}-m_{13}-m_{14}
R_2		m_{23}	m_{24}	R_2-m_{23}-m_{24}
R_3			m_{34}	R_3-m_{34}

- **`1011`**
- R_i: newly marked individuals AND re-released individuals

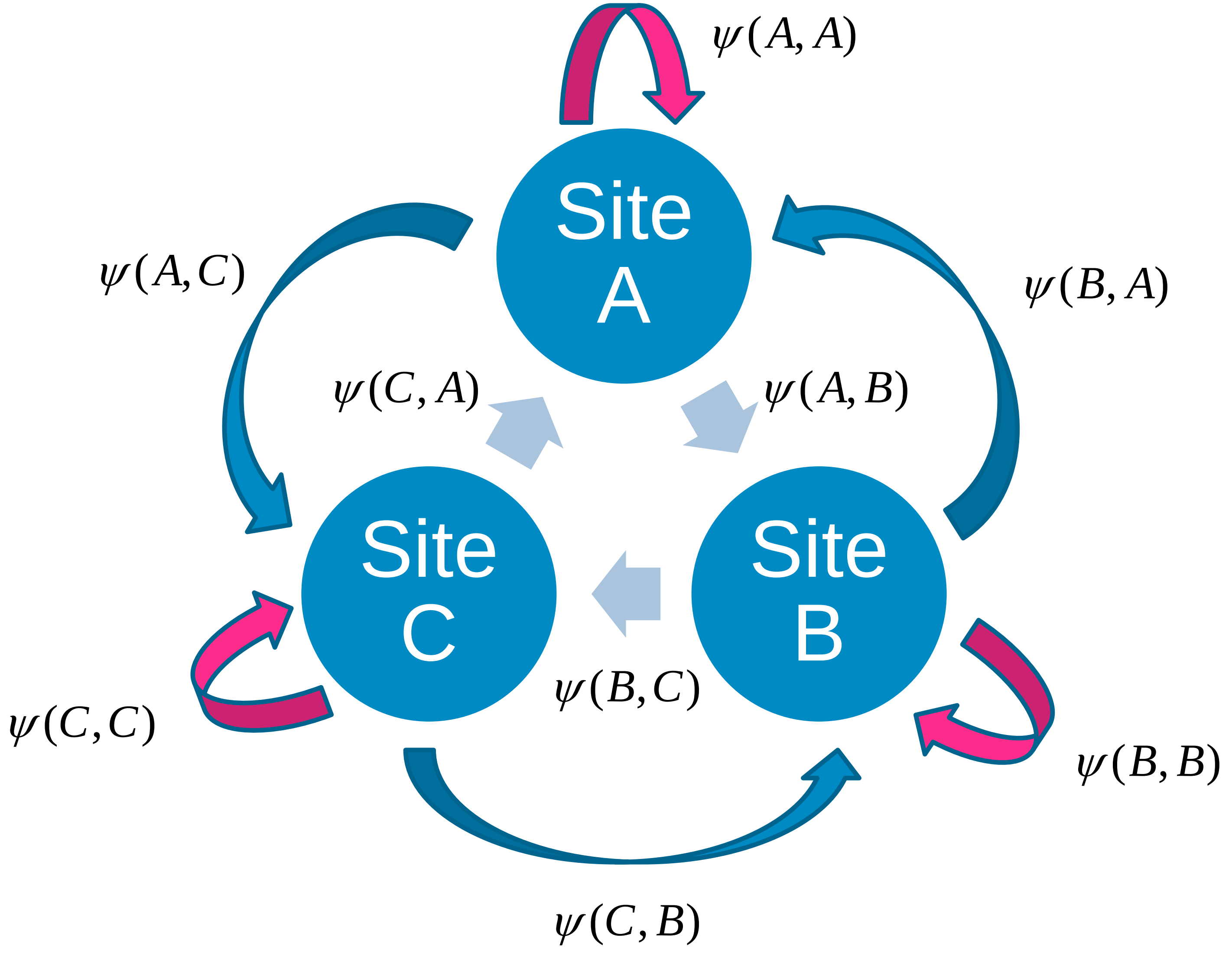
M-array probabilities

Number Released	Recaptured at 2	Recaptured at 3	Recaptured at 4	Never Recaptured
R_1	$\phi_1 p_2$	$\phi_1(1-p_2)\phi_2 p_3$	$\phi_1(1-p_2)\phi_2(1-p_3)\phi_3 p_4$	χ_1
R_2		$\phi_2 p_3$	$\phi_2(1-p_3)\phi_3 p_4$	χ_2
R_3			$\phi_3 p_4$	χ_3

- ϕ_t: probability an animal alive at time t survives to time t+1
- p_{t+1}: probability an animal alive at time t+1 is recaptured at time t+1
- Cormack-Jolly-Seber Model: $\phi(t)$,p(t)


Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Behavioural model, M_b
Depletion / Removal model

Multi-site/state models



Parameters

- $\phi_t(r)$: probability of an animal alive at time t in site r, survives until time t+1
- $p_{t+1}(s)$: probability of an animal alive in site s at time t+1 being recaptured
- $\psi_t(r,s)$: probability of an animal alive in site r at time t moving to site s by time t+1

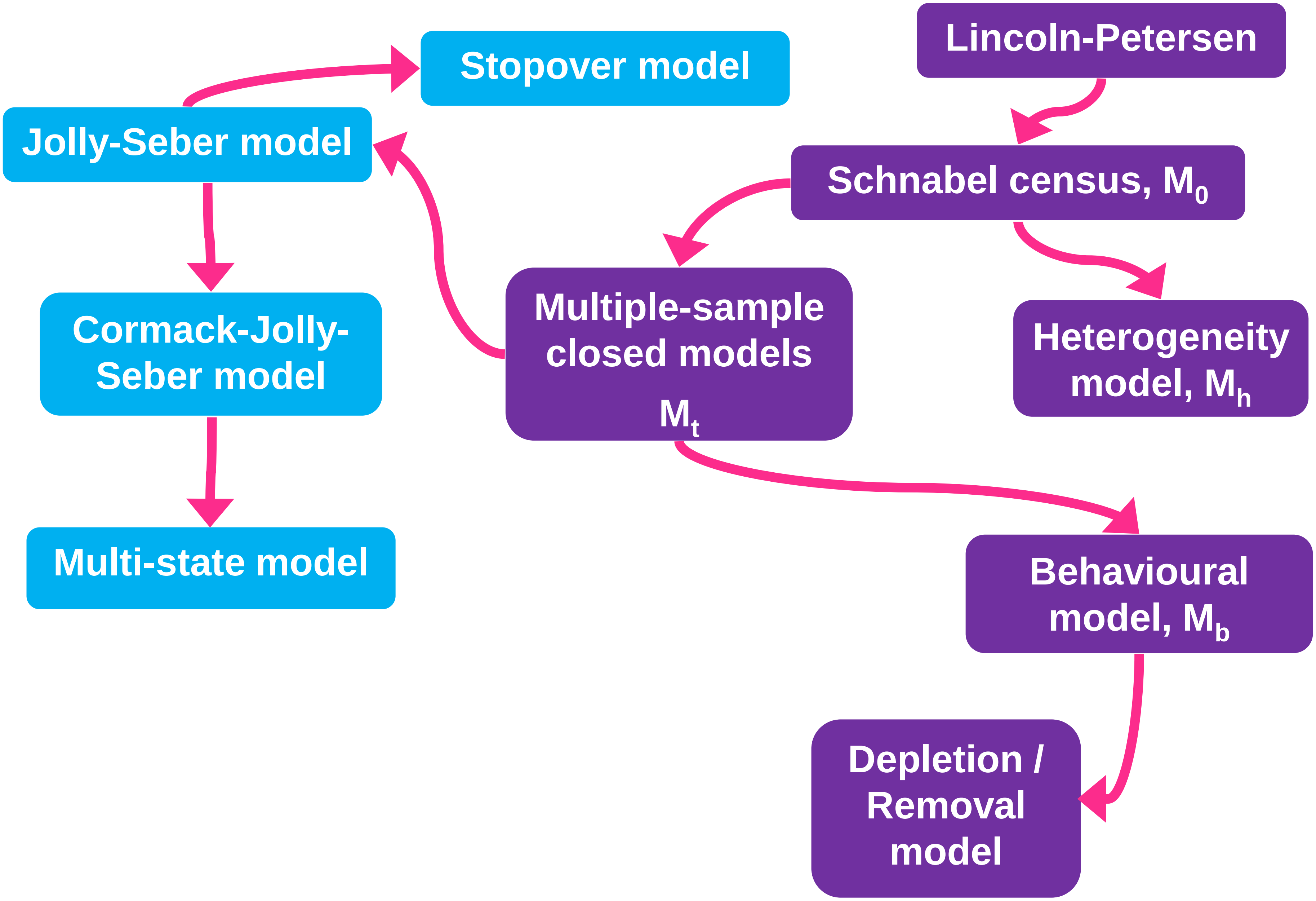
Constructing probabilities

- Two state system: A and B
- Observed encounter history: A 0 B

Constructing probabilities

- Two state system: A and B
- Observed encounter history: A 0 B
- A or B ? – sum over both possible options

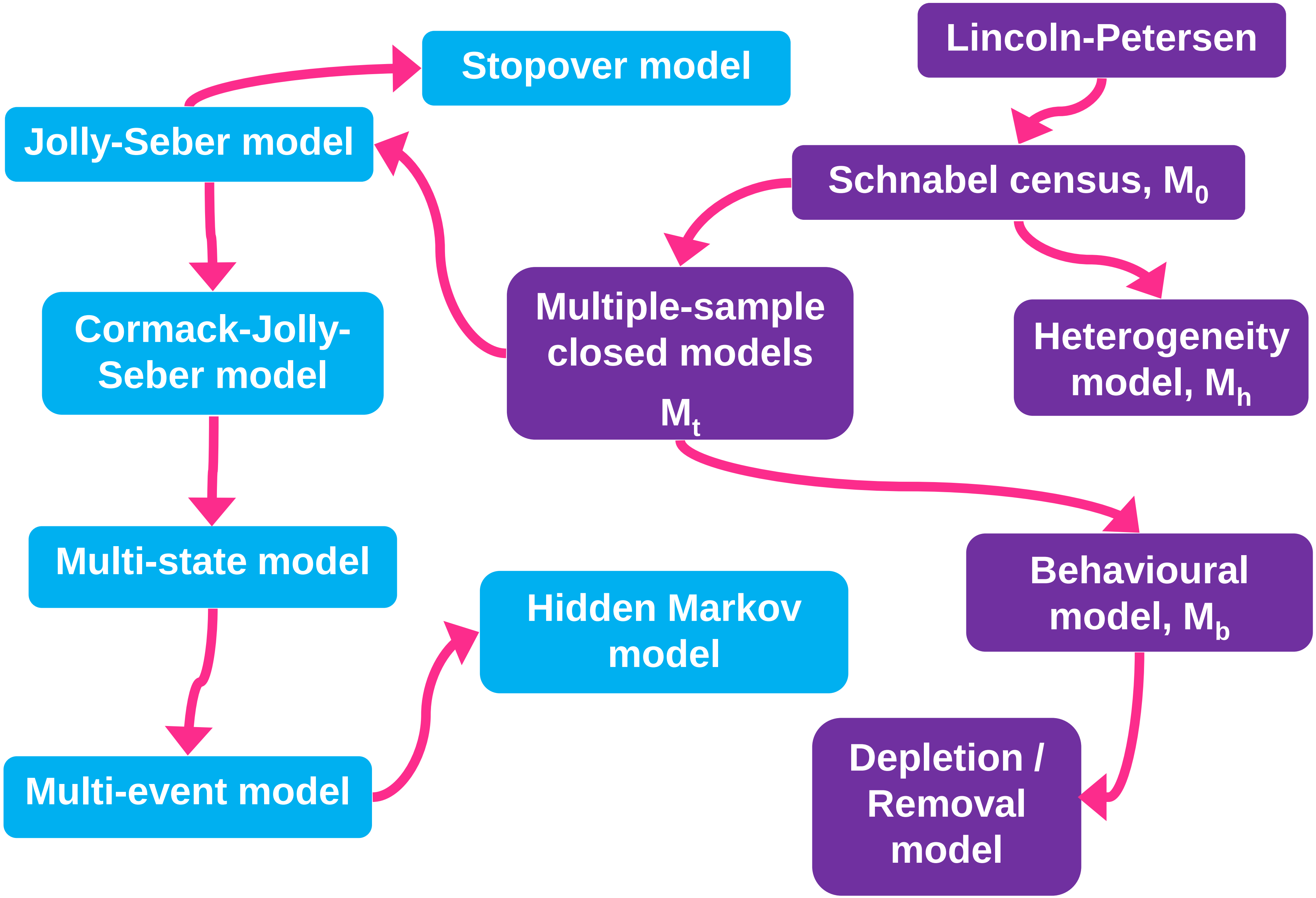
$$\phi_1(A)[\psi_1(A,A)\{1-p_2(A)\}\phi_2(A)\psi_2(A,B) \\ + \psi_1(A,B)\{1-p_2(B)\}\phi_2(B)\psi_2(B,B)]p_3(B)$$


Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Behavioural model, M_b
Depletion / Removal model

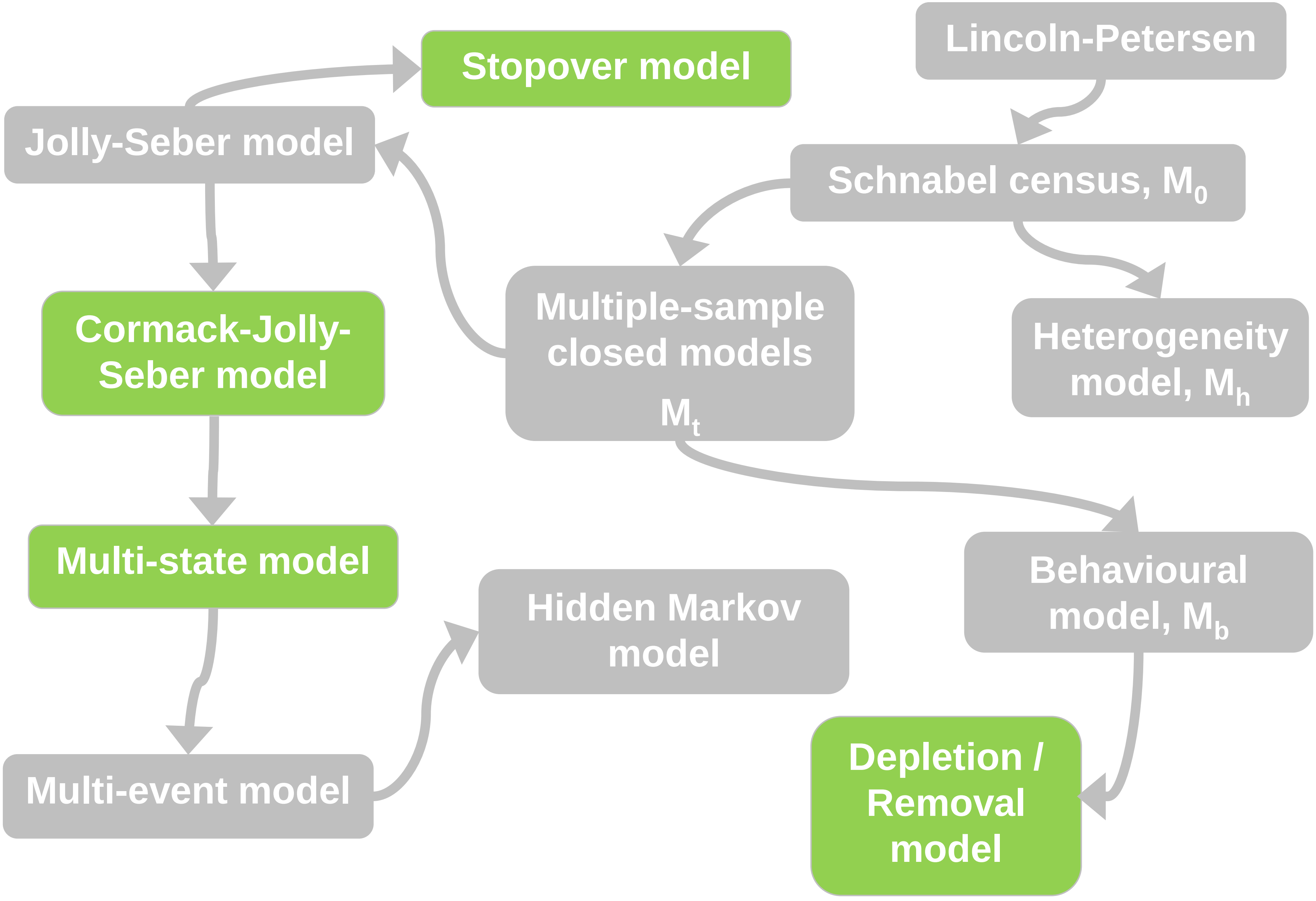
Multievent models

- Multistate models rely on the assumption that the state of an individual is perfectly assigned
- Reasonable assumption in some cases – geographical location
- In some cases there might be “mistakes” – breeding status
- General framework for many capture-recapture models
  - Multistate model
  - Memory models
  - Mixture models
- Hidden Markov model[1]

[1]King (2014) Statistical Ecology. *Annual Review of Statistics and its Application* **1** 401-426.


Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model

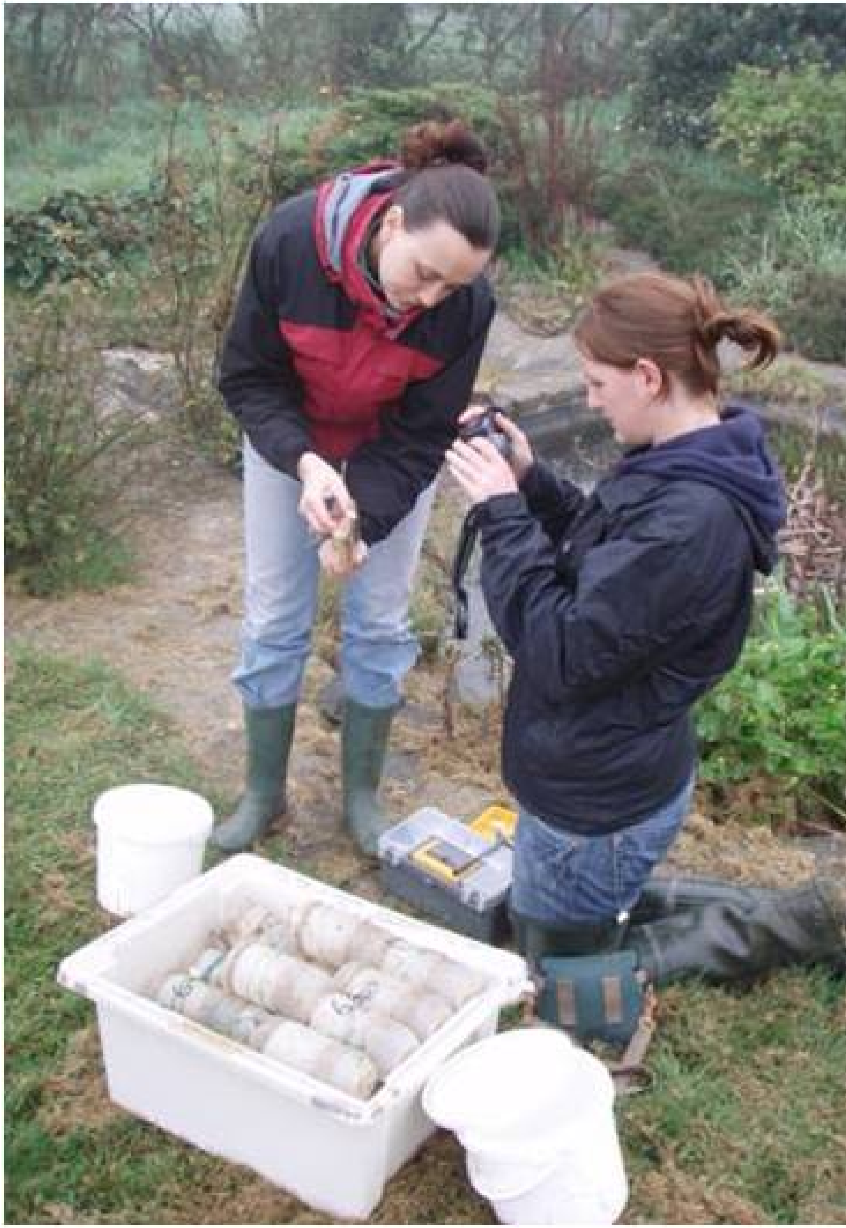
CASE STUDIES



Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model

Case study 1: Cormack-Jolly-Seber model

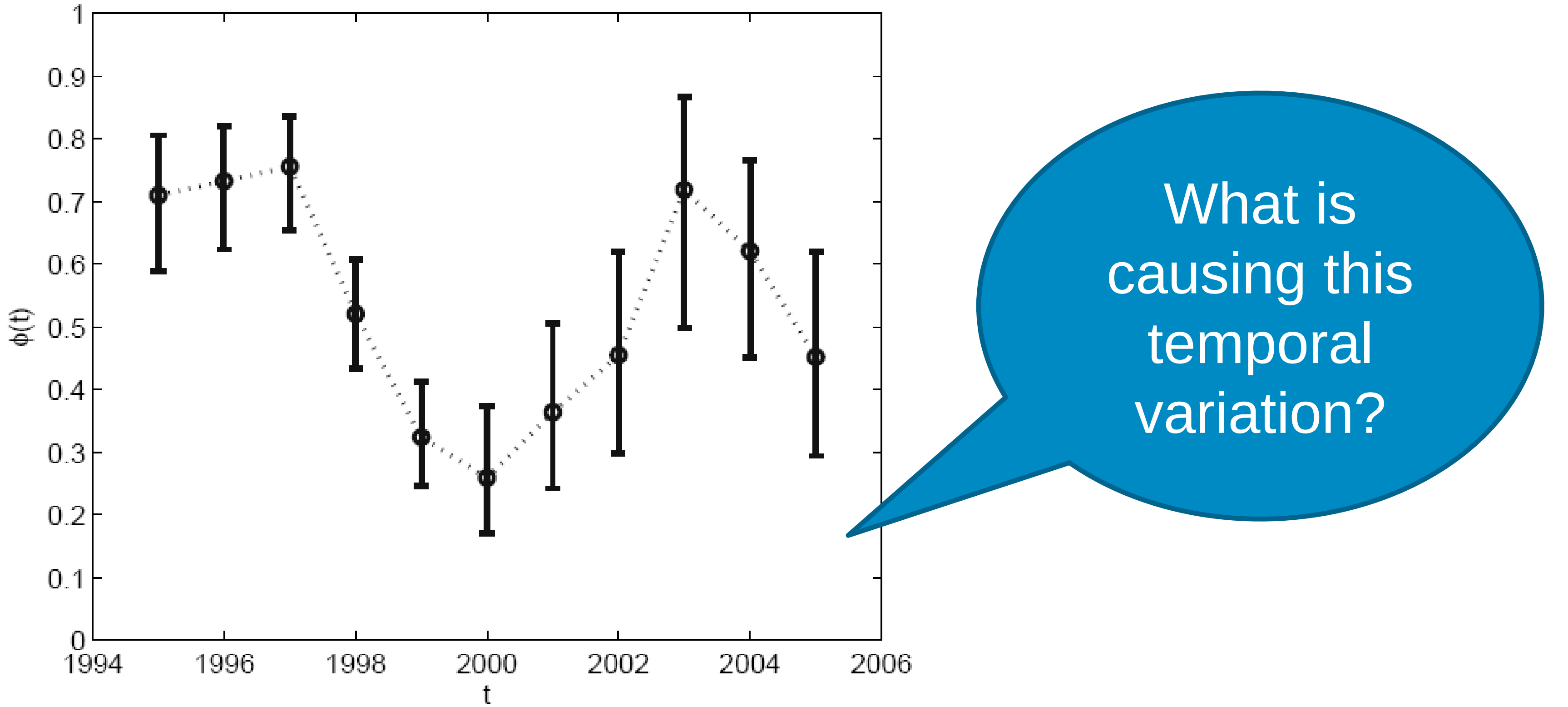
- Mark-recapture data collected between 1995 and 2006
- Population of Great crested newts, *Tristurus cristatus,* close to the University of Kent campus.
- Meta-population studied over four groups of ponds
- Best model: ϕ(t),p(pond)




WK2-1M-G1

Wellcourt study site

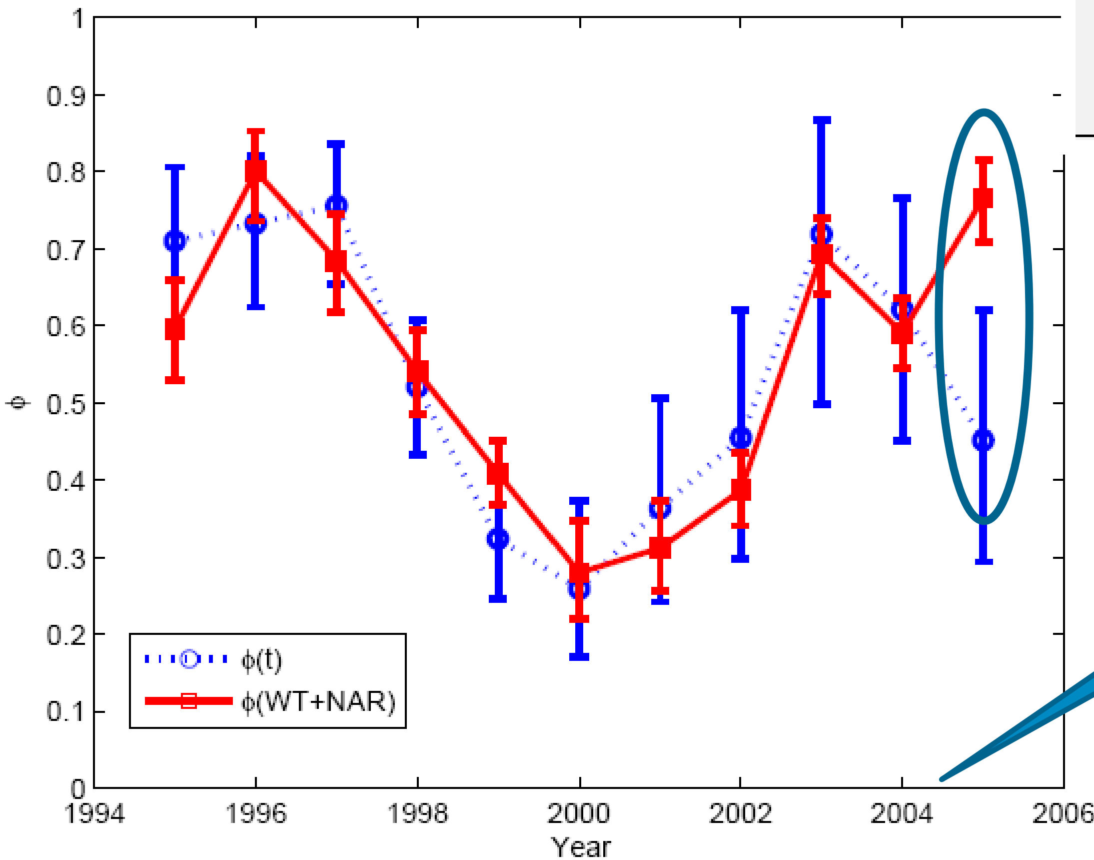
Great crested newts: ϕ(t),p(pond)



- **Garden Pond: P = 0.697 (0.629-0.757)**
- **Swimming Pool: p = 0.845 (0.726-0.919)**
- **Snake Pond: p = 0.294 (0.225-0.373)**
- **Pylon Pond: p = 0.233 (0.100-0.455)**

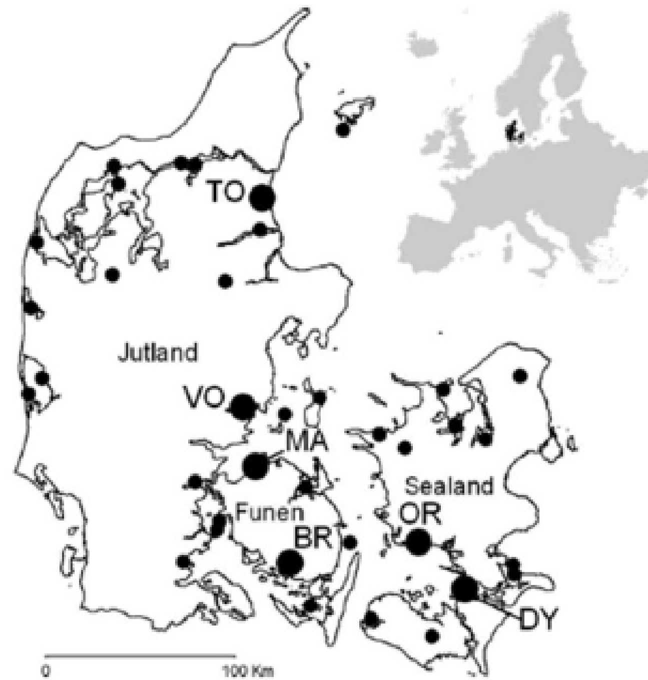
ϕ(t) versus ϕ(WT+NAR)

Model	QAICc	Delta QAICc
phi(WT + NAR + t2005), p(site)	1940.21	0.00
phi(time), p(site + sex)	1941.62	1.41
phi(time), p(site)	1941.86	1.65
phi(time), p(sex * site)	1942.60	2.39
phi(WT + SR + NAR + WGF + t2005), p(site)	1942.78	2.57
phi(time), p(site * time)	1945.99	5.78
phi(WT + SR + NAR + WGF), p(site)	1947.09	6.88
phi(WT + NAR + WGF), p(site)	1947.26	7.06
phi(WT + SR + NAR + WGF), p(site + sex)	1947.37	7.16
phi(WT + SR + NAR), p(site)	1948.04	7.84
phi(WT + SR + NAR + WGF), p(site * sex)	1948.48	8.27
phi(WT + NAR), p(site)	1949.16	8.95
phi(sex * time), p(site)	1950.53	10.33



Changes to ponds

Case study 2: Stopover model

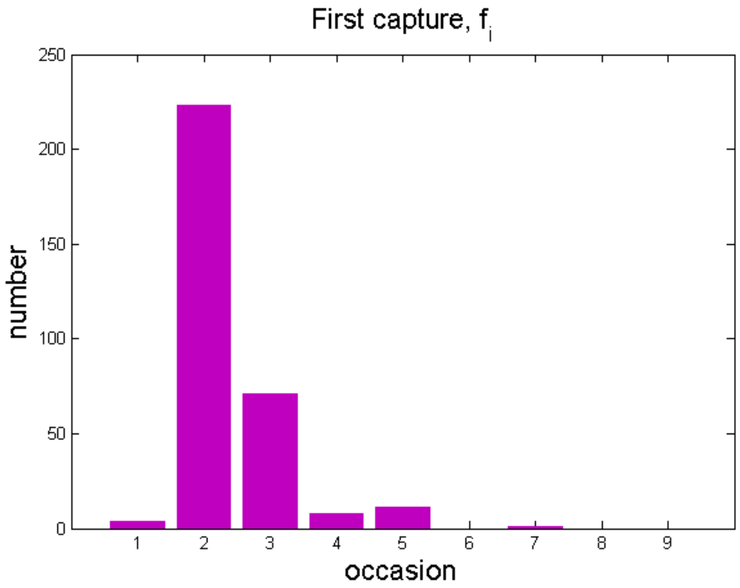


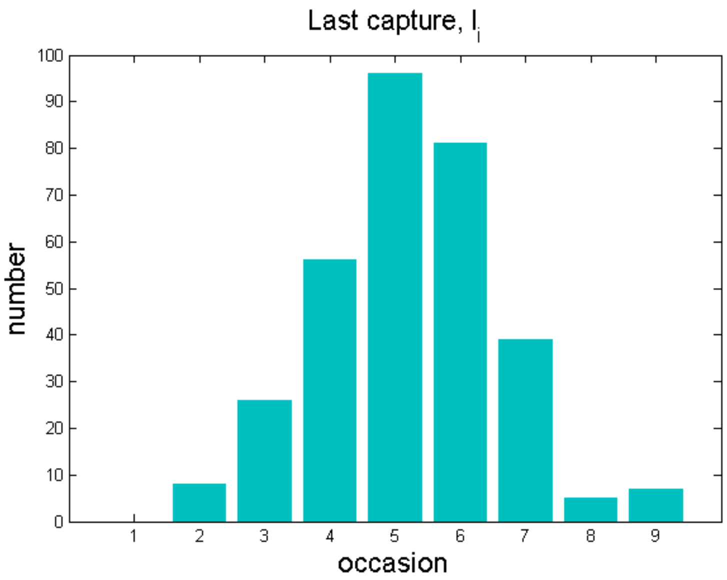
- Colony Vorsø
- Daily visits are made to the cormorant colonies during the breeding season, data summarised as monthly captures
- Records are made of whether or not the cormorants successfully breed

University of Kent

Great cormorants

- 318 individuals
- 9 occasions – monthly February-October

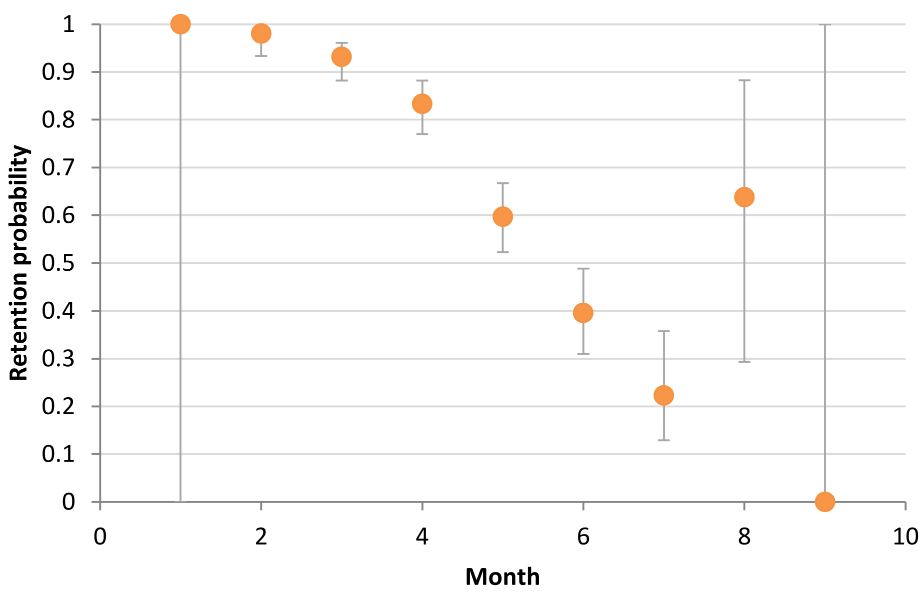




University of Kent

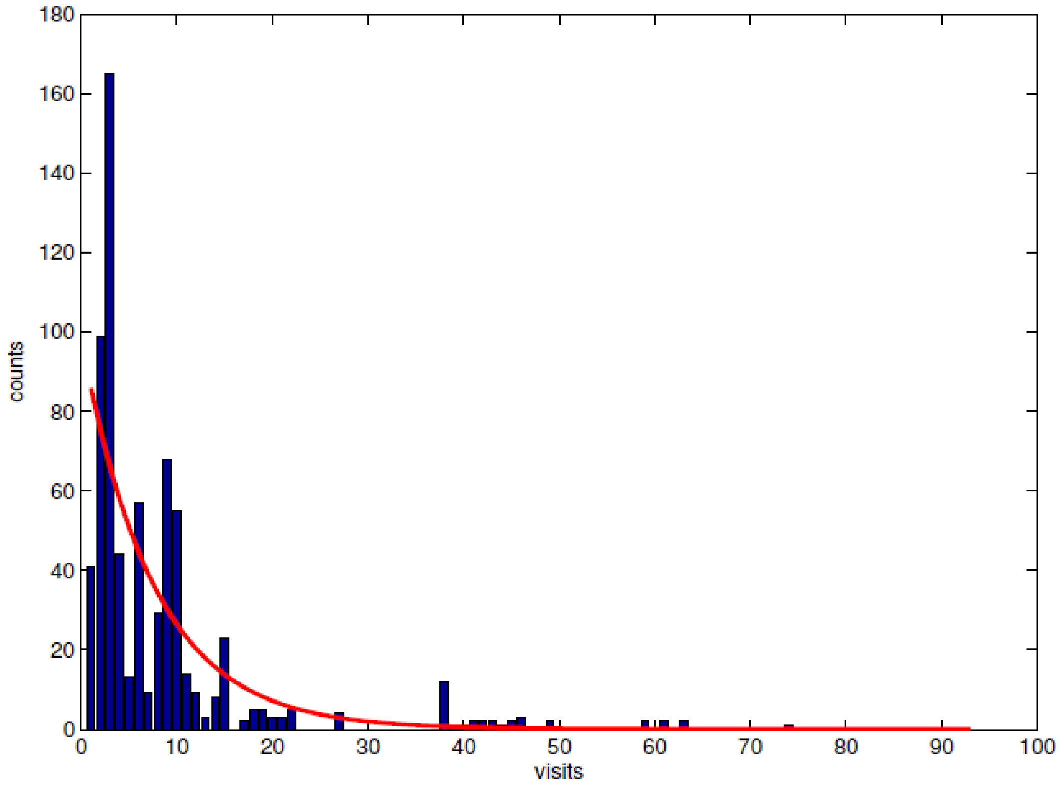
Case study 2

Model	AIC	ΔAIC	np	MLE of (N-D)
N,β(t),ϕ(.),p(.)	1459.38	332.32	6	11.6
N,β(t),ϕ(t),p(.)	1127.97	0.91	14	1.8
N,β(t),ϕ(a),p(.)	1156.24	29.18	14	1.2
N,β(t),ϕ(a+t),p(.)	1127.06	0.00	22	1.4



Case study 3: removal model

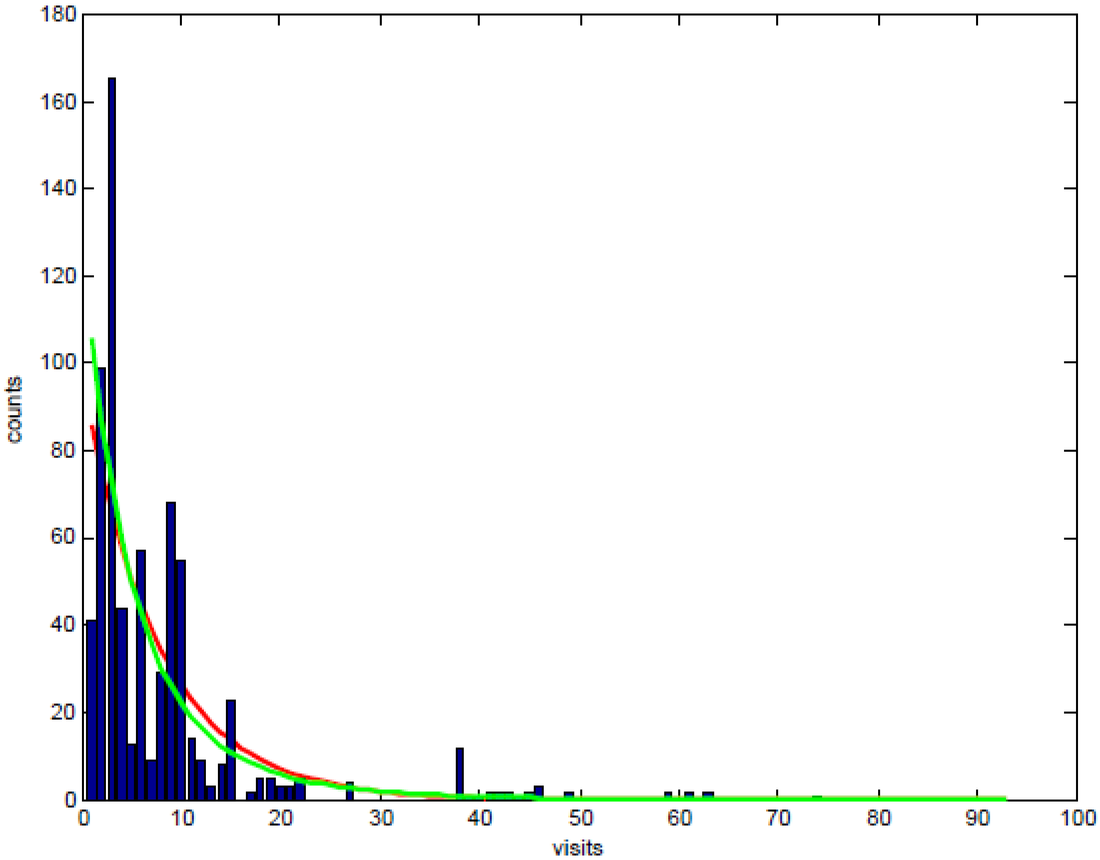
- Great crested newt removal data*



*data collected by Herpetologic Ltd

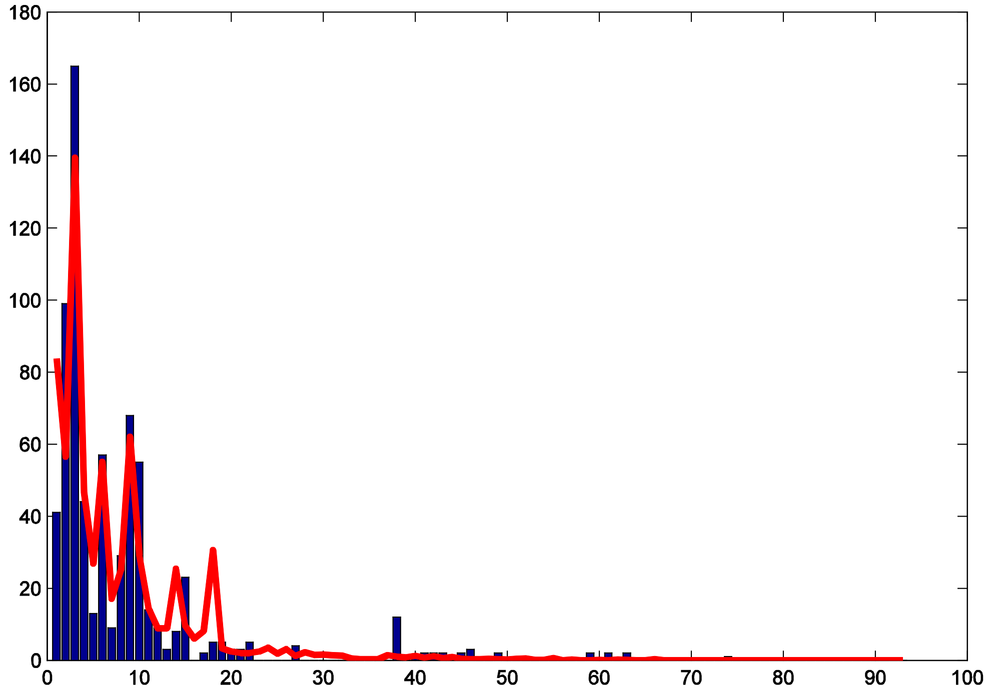
Case study 3: removal model

- Adding overdispersion – beta-geometric model

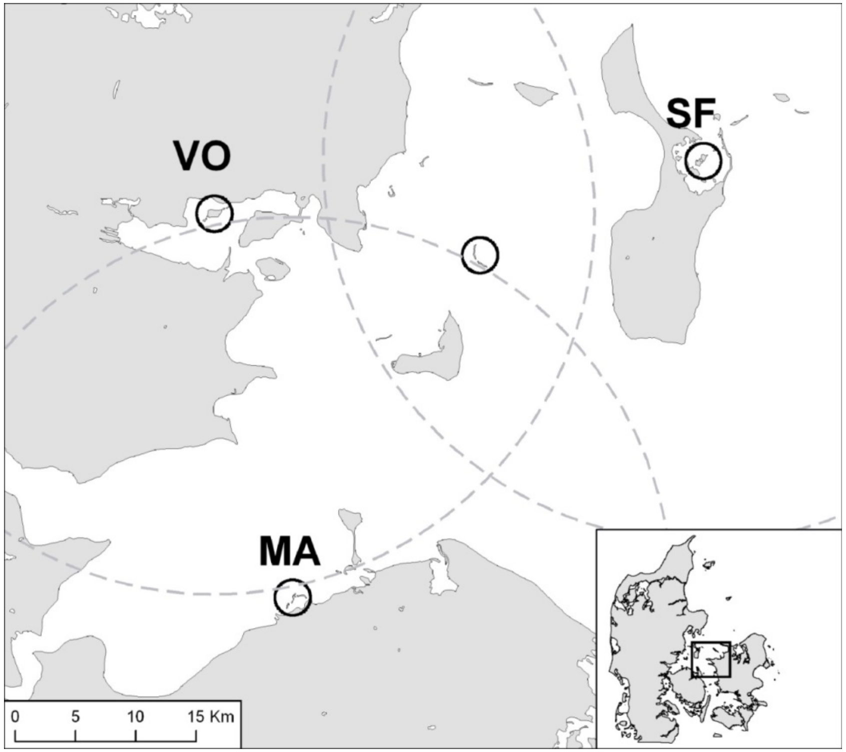


Case study 3: removal model

- Temporal variation – weather covariates
- Minimum air temperature



Case study 4: Multistate model



Phalacrocorax carbo sinensis

Recapture data

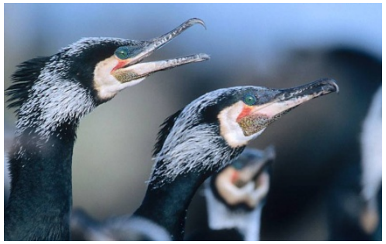
- Recapture data collected between 1989 and 1994
- Initial ringing was carried out on non-breeding individuals
- Birds were not recaptured until they had become breeding individuals
- Must model transition between non-breeding and breeding states as well as geographical movements between the colonies

Transitions

- We have a total of six states:
  - Non-breeders & Breeders at each of the sites
- Background biology:
  - Once an individual has attained a breeding status, it remains in the breeding state for the remainder of its life

Transitions

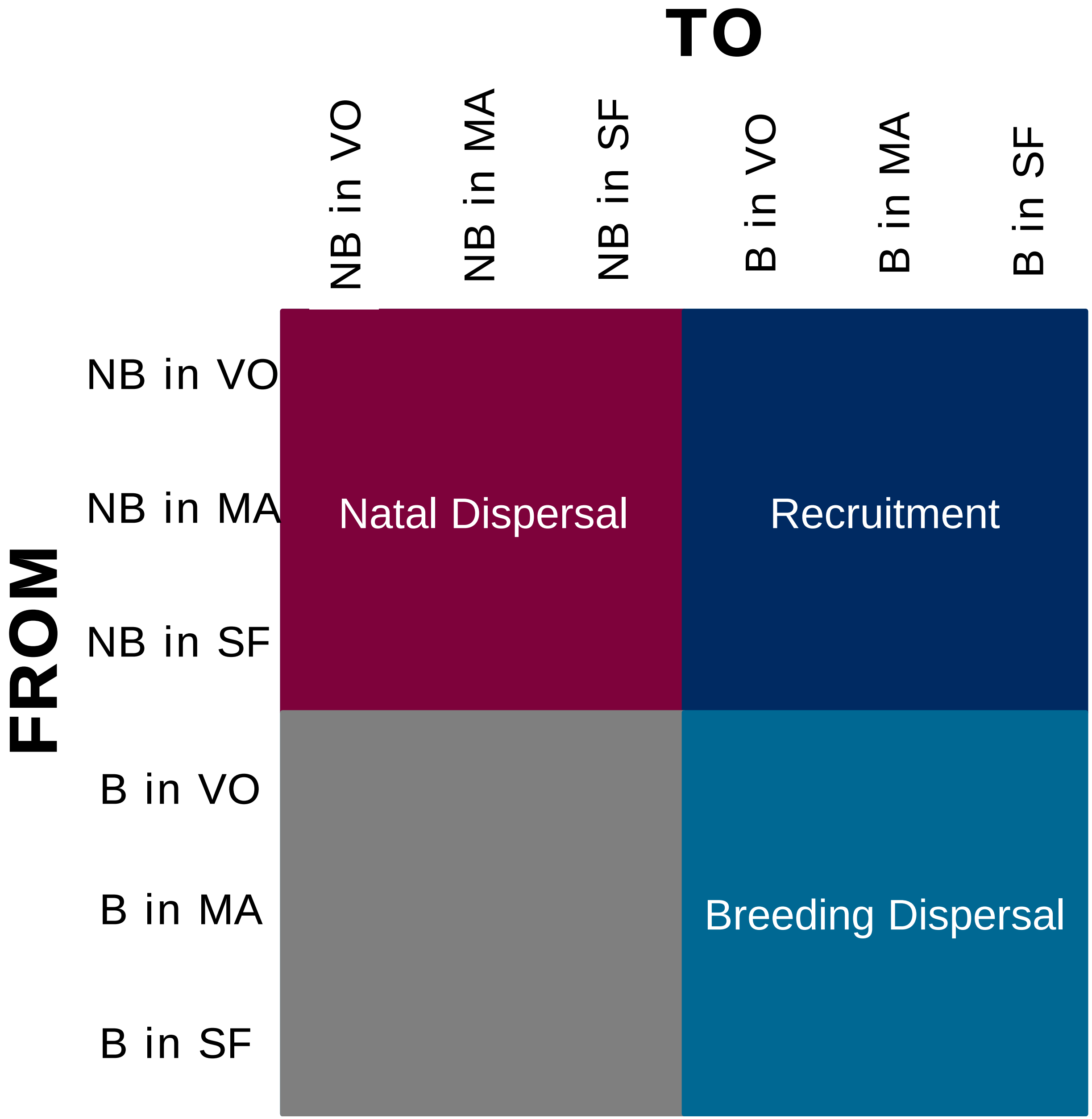
- **Natal Dispersal:** movement between the colonies whilst a non-breeder
- **Recruitment:** transition from non-breeding to breeding state
- **Breeding Dispersal:** movement between the colonies whilst a breeder



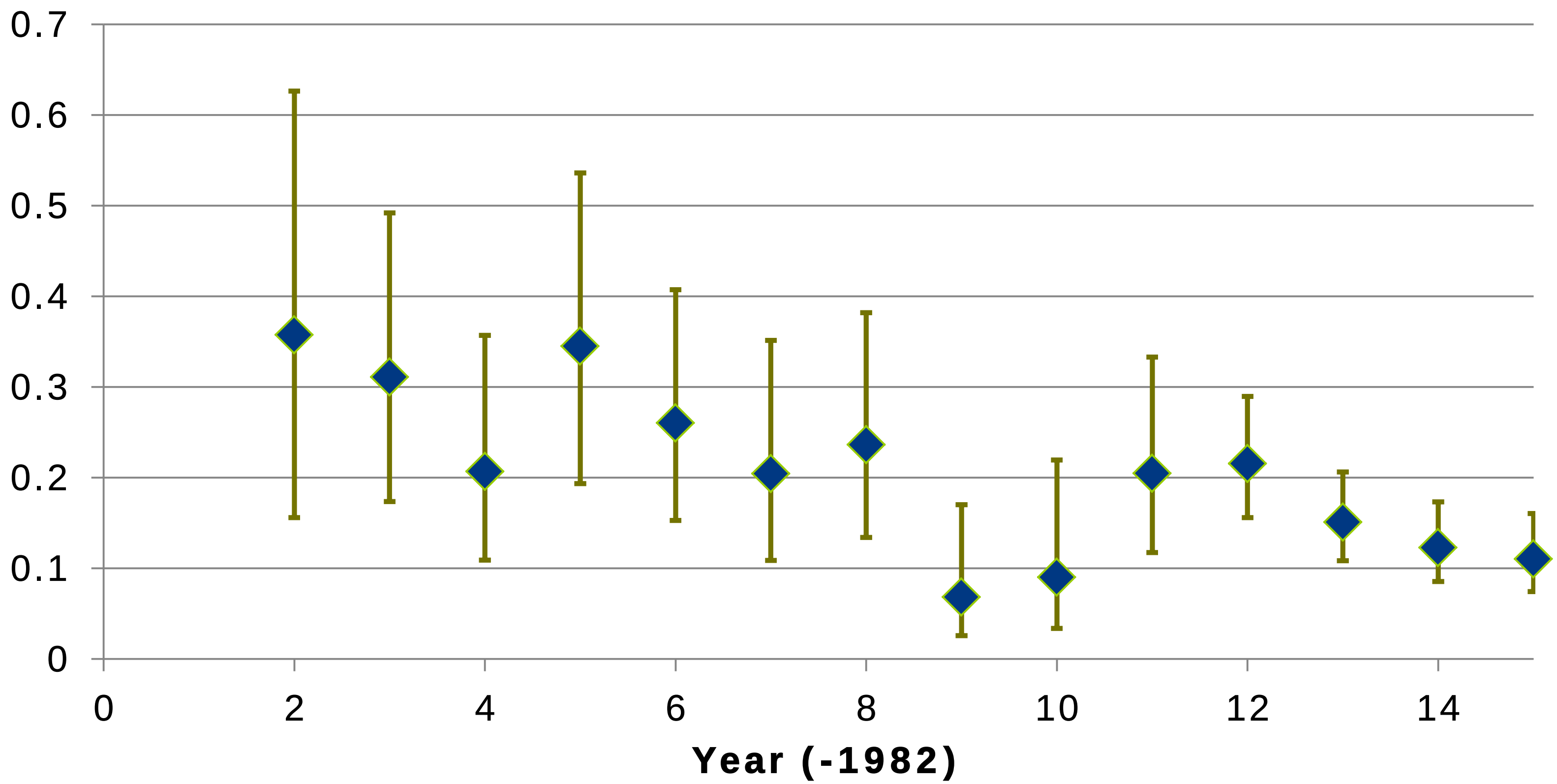
Natal Dispersal: movement between the colonies whilst a non-breeder

Recruitment: transition from non-breeding to breeding state

Breeding Dispersal: movement between the colonies whilst a breeder

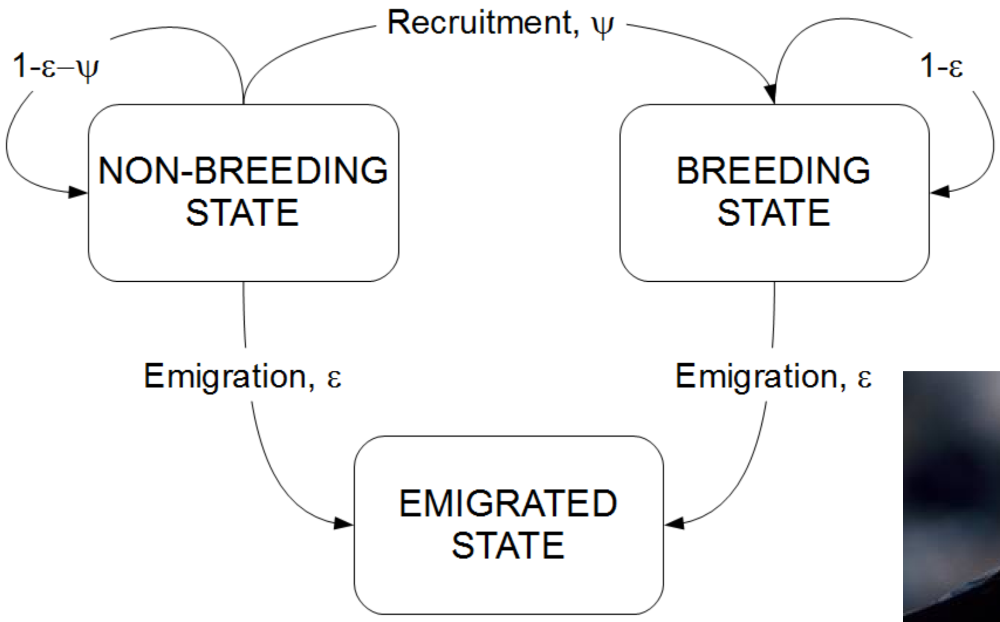


Recruitment probability

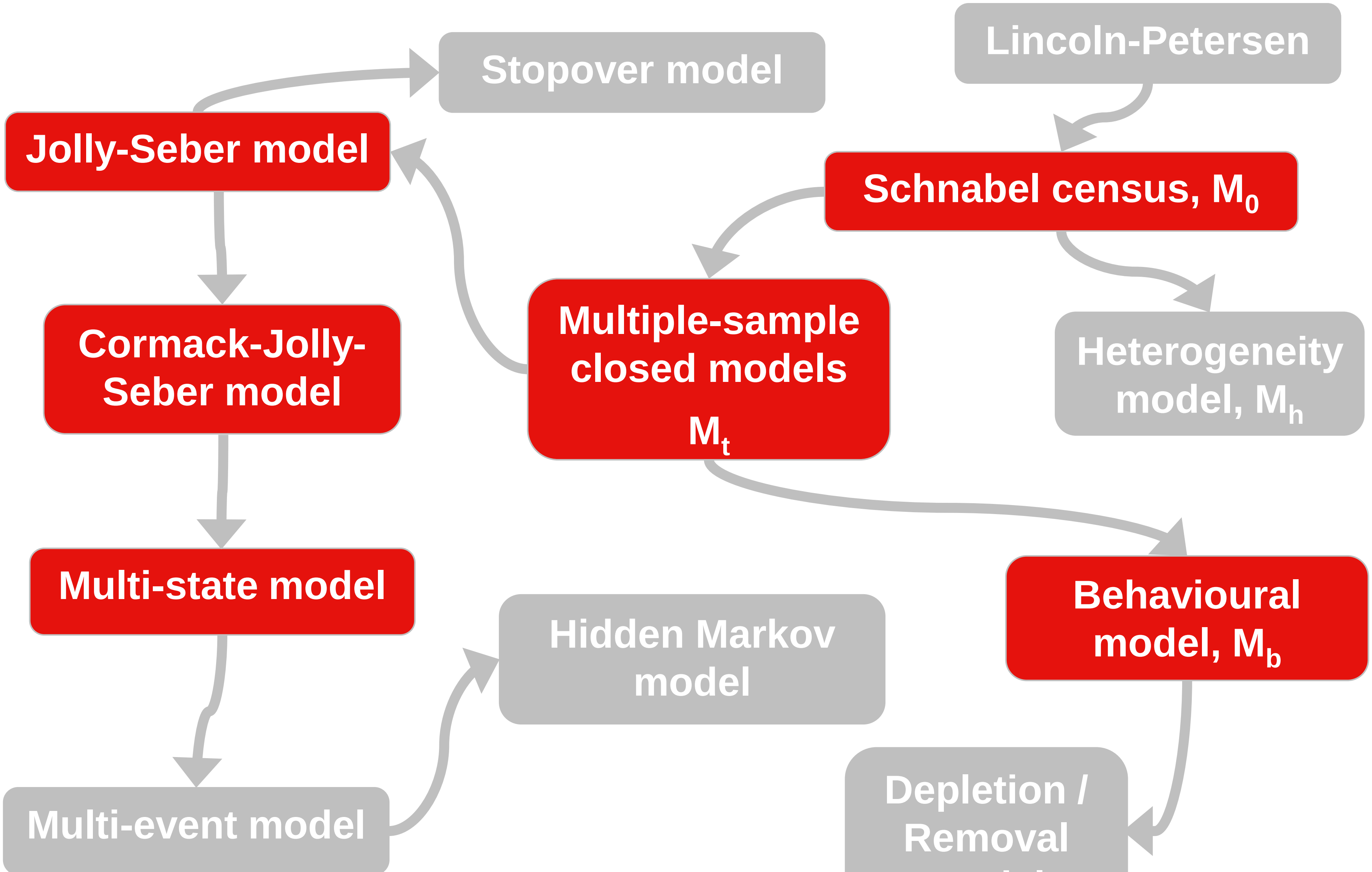


University of Kent

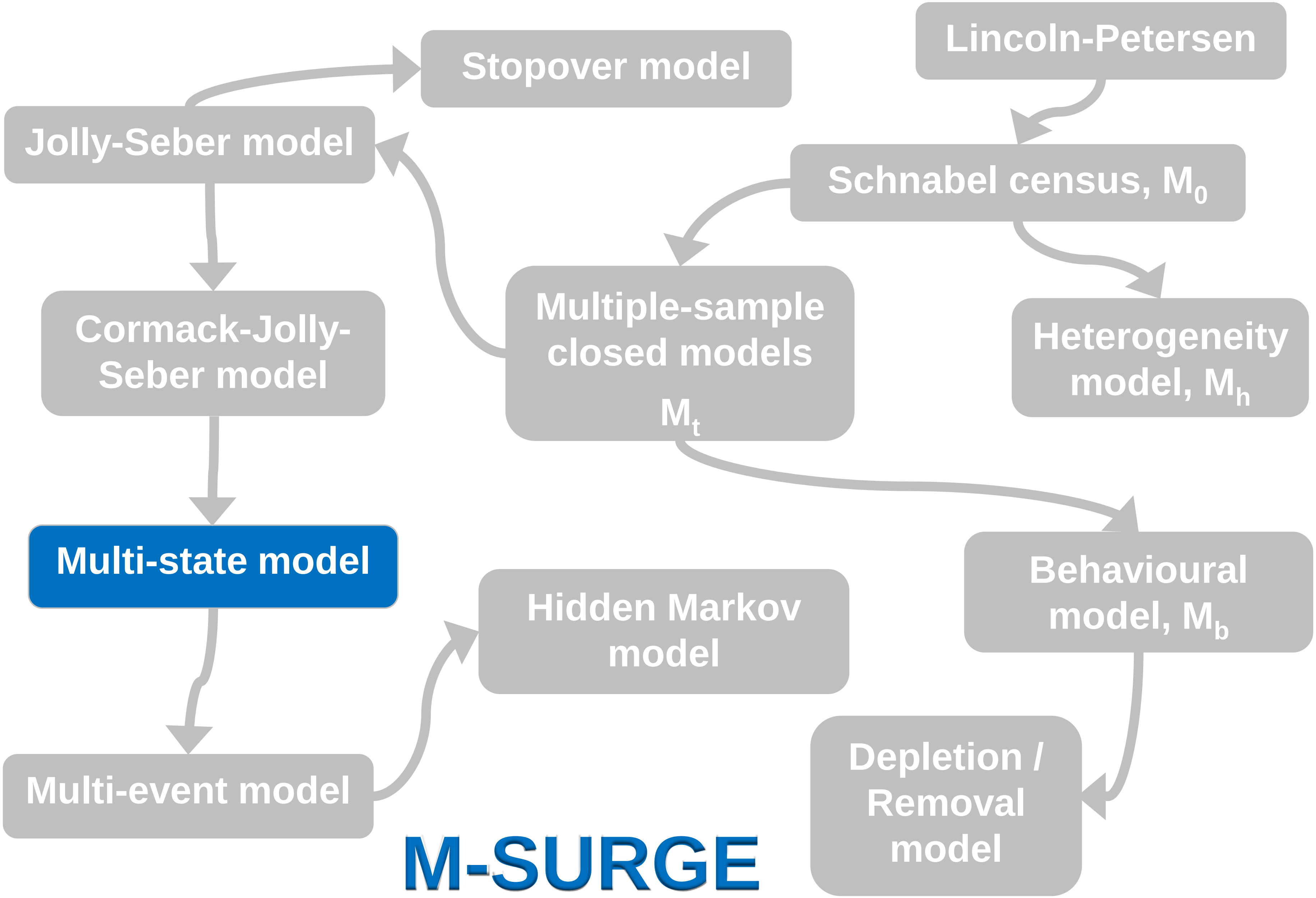
Emigration: adapted cormorant model

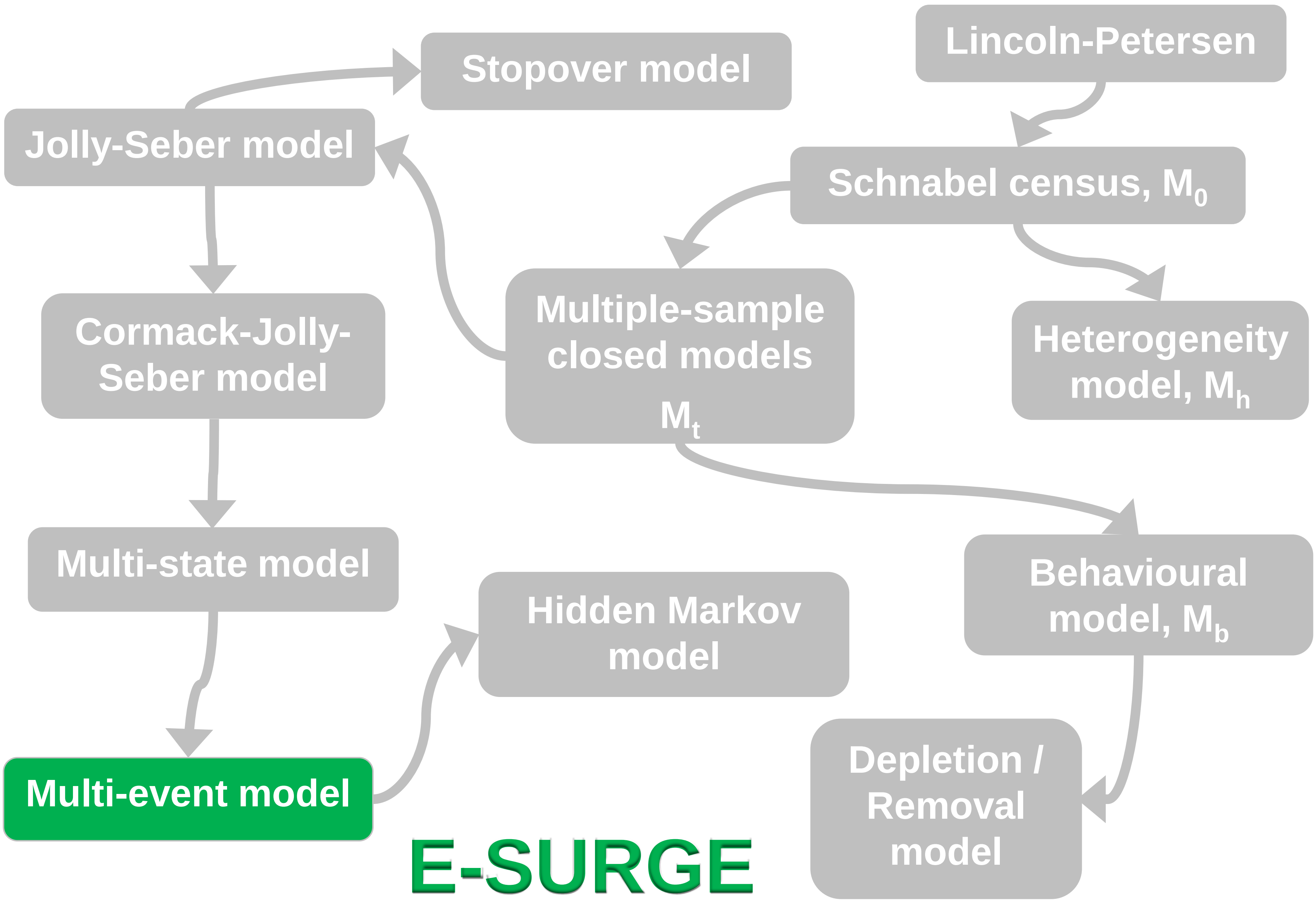


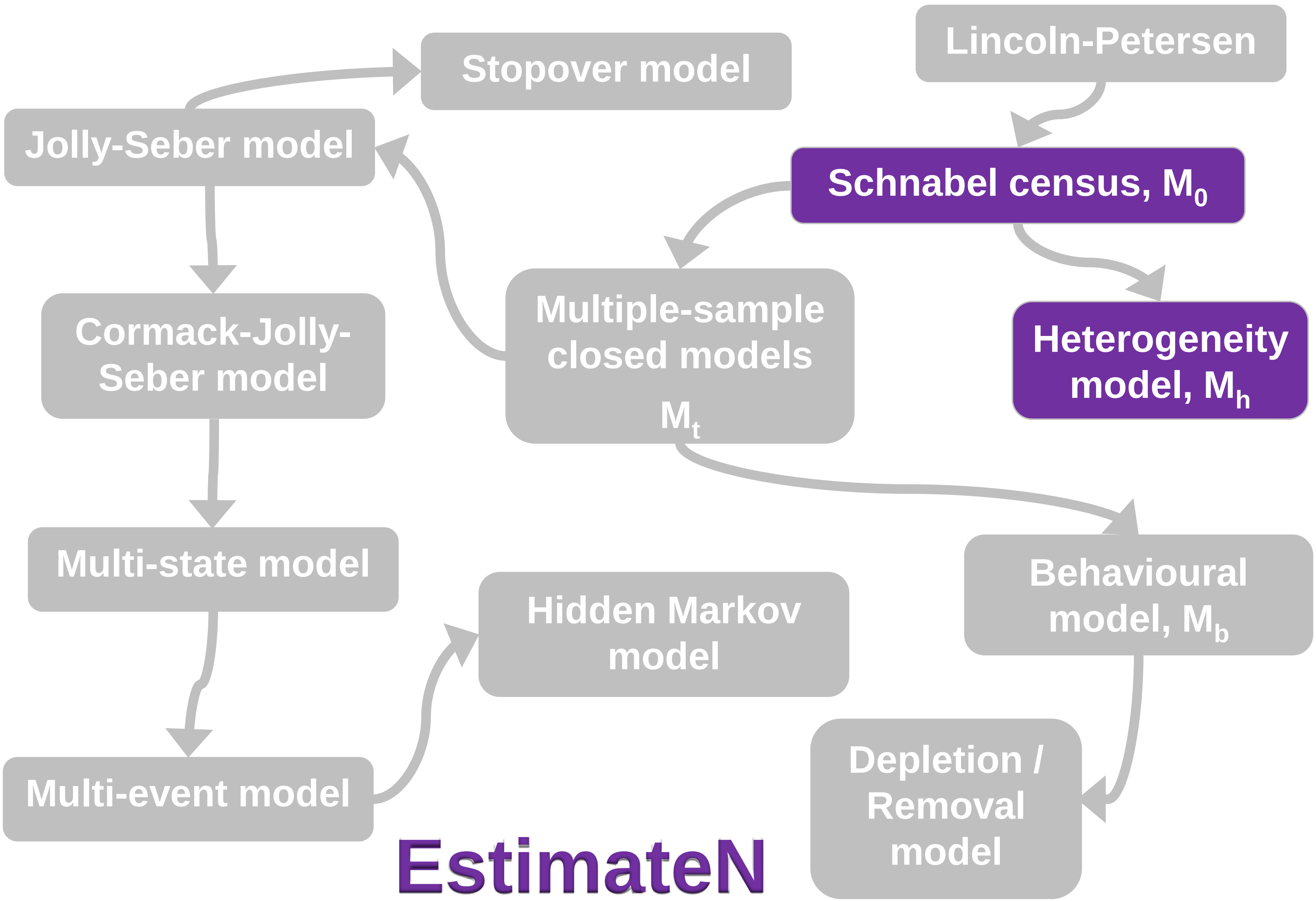
COMPUTER SOFTWARE

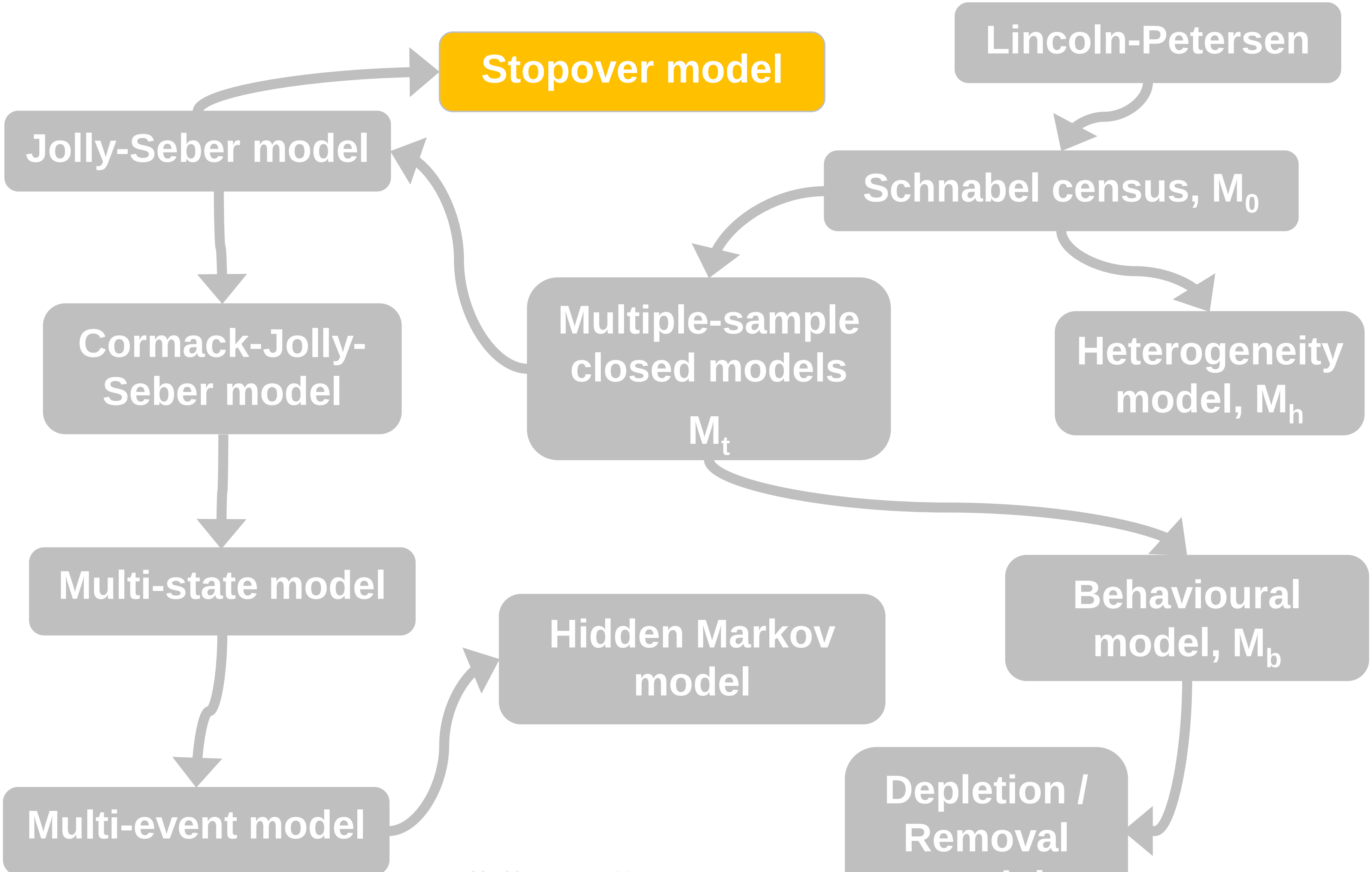

Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model

Mark/RMark


Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Multiple-sample closed models M_t
Cormack-Jolly-Seber model
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model
M-SURGE

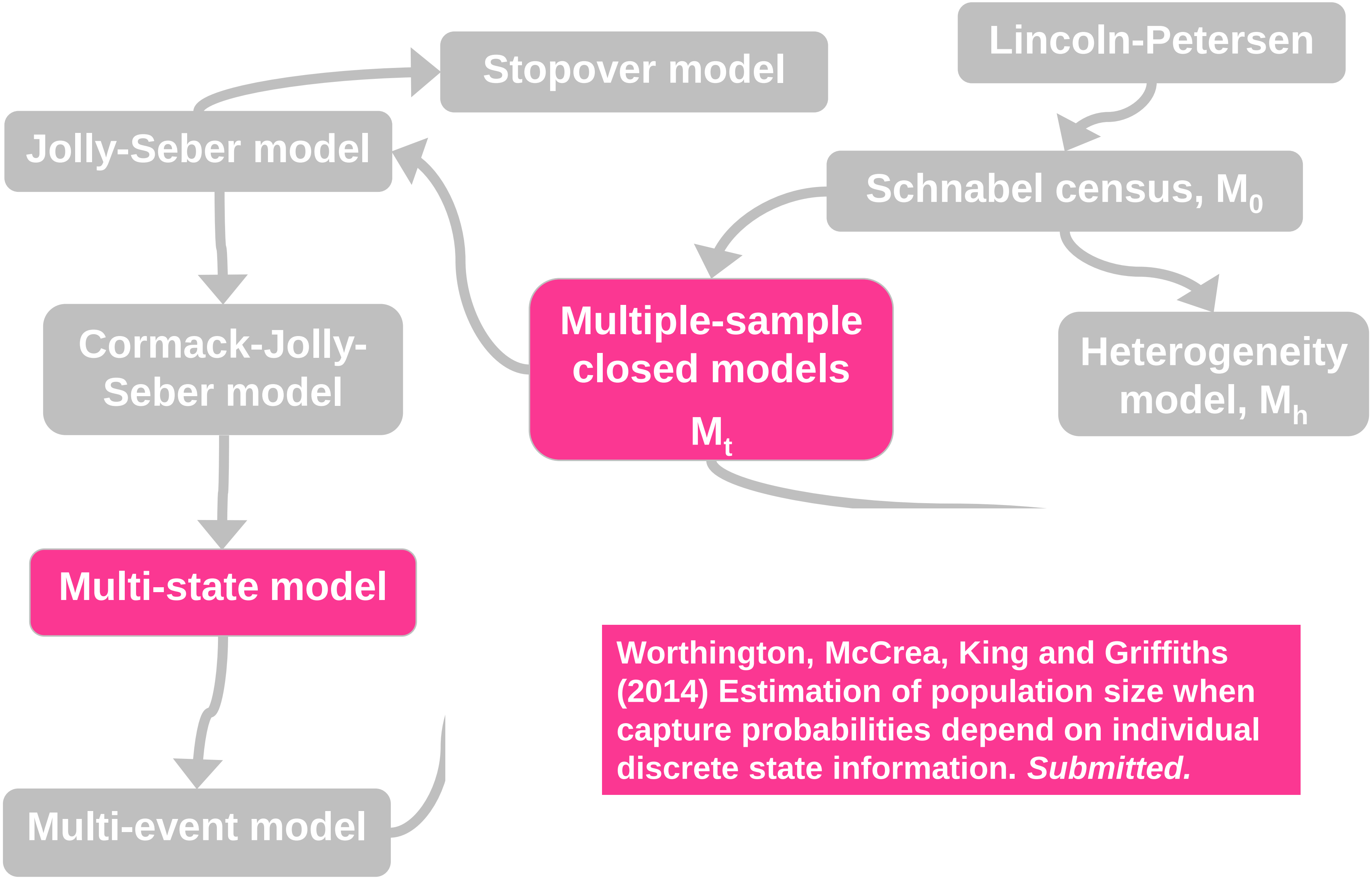

Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model
E-SURGE


Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model
EstimateN

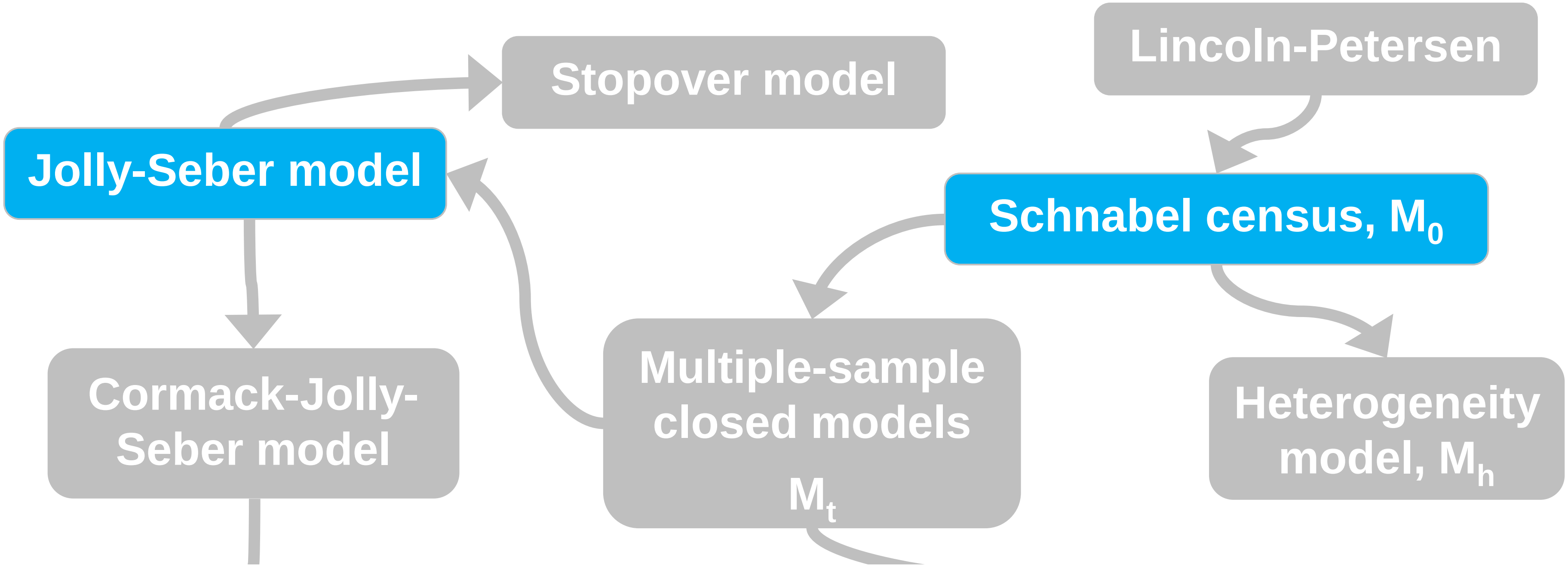

Stopover model
Lincoln-Petersen
Jolly-Seber model
Schnabel census, M_0
Cormack-Jolly-Seber model
Multiple-sample closed models M_t
Heterogeneity model, M_h
Multi-state model
Hidden Markov model
Behavioural model, M_b
Multi-event model
Depletion / Removal model

HetAge

RECENT DEVELOPMENTS



Worthington, McCrea, King and Griffiths (2014) Estimation of population size when capture probabilities depend on individual discrete state information. *Submitted.*



Cowen, Besbeas, Morgan and Schwarz (2014) A comparison of abundance estimates from extended batch-marking and Jolly-Seber-type experiments. ***Ecology and Evolution. 4, 210-218.***

Stopover model

Lincoln-Petersen

Jolly-Seber model

Schnabel census, M_0

mple
dels

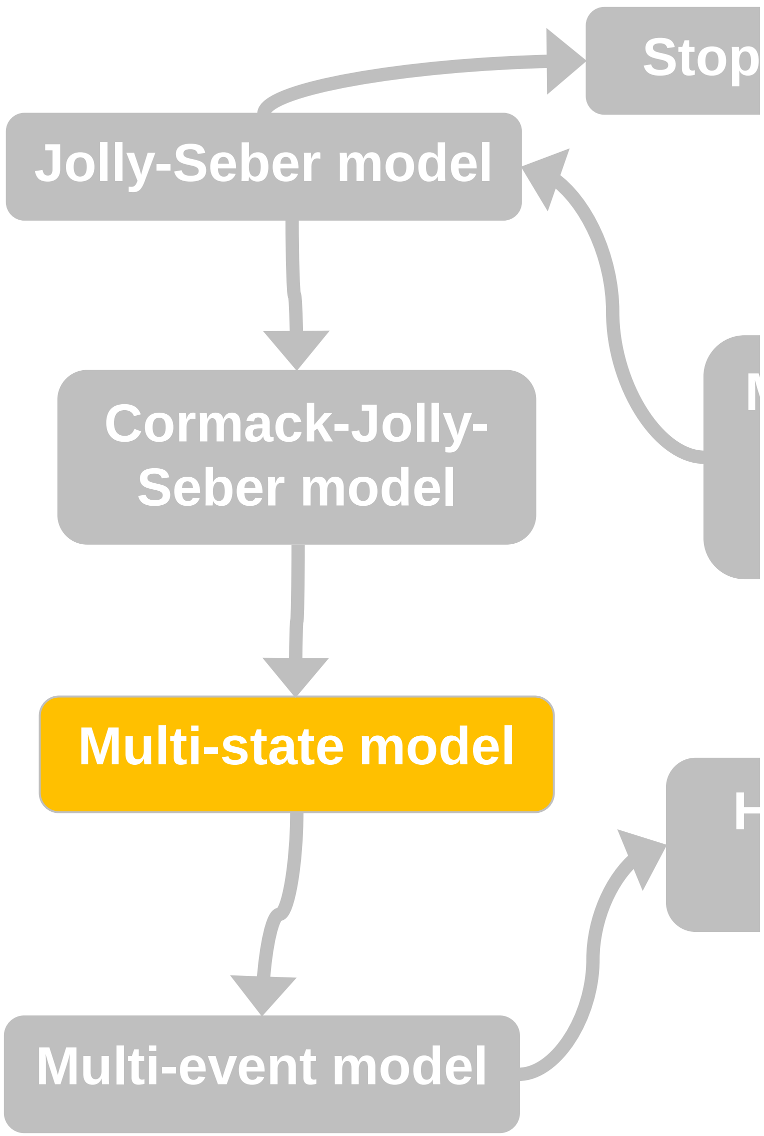
Heterogeneity model, M_h

Matechou, McCrea, Morgan, Sewell, Lewis and Griffiths (2014) New models for removal data: accounting for unknown arrival times of individuals. *In prep.*

kov

Behavioural model, M_b

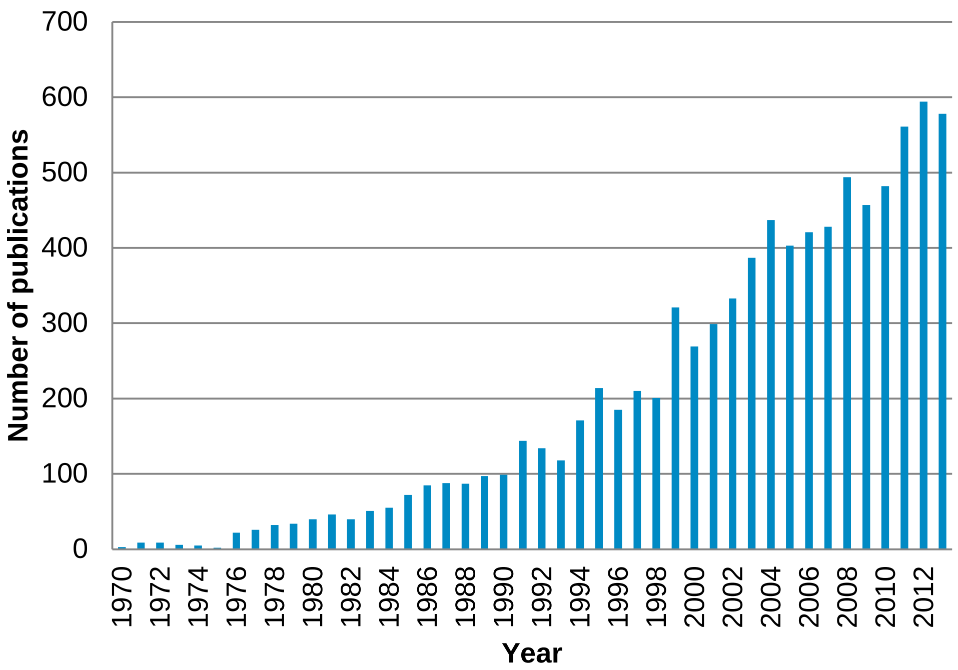
Depletion / Removal model



McCrea, Jeyam and Morgan (2014)
High dimensional multistate capture-recapture models for time-varying individual covariates.
In prep.

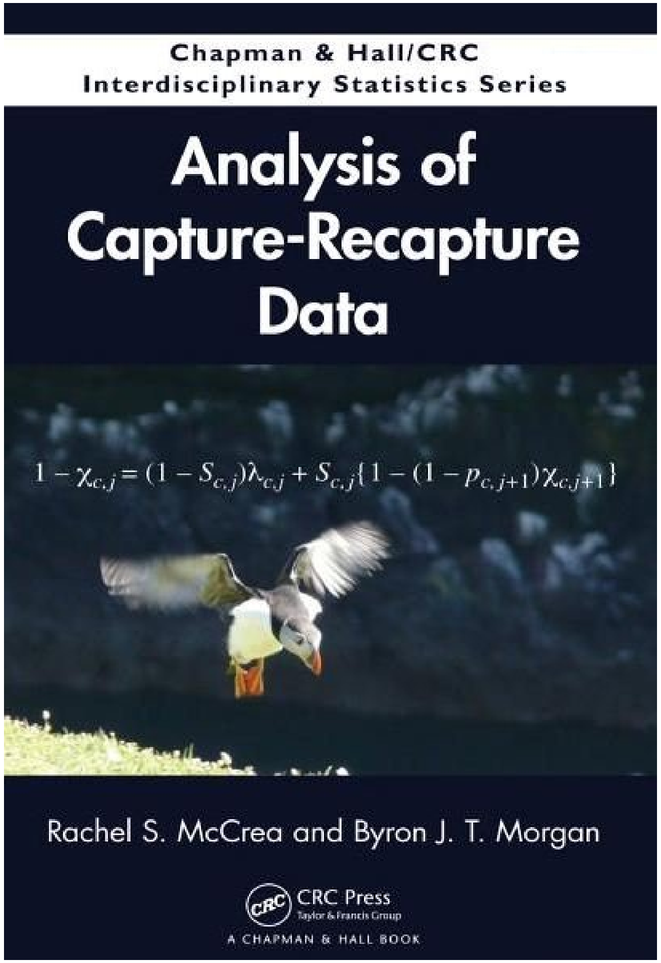
Conclusions

- Rapidly developing research area
- New data collection techniques are driving the statistical developments
- Collaborative research



Data collected from web of science using search terms: Capture-recapture; mark-recapture; mark-resight

Book, website and e-mail list



- www.capturerecapture.co.uk
- stats-ecology@kent.ac.uk
- R.S.McCrea@kent.ac.uk

University of Kent