

A bootstrap method for adjusting Adélie penguin counts to correct for estimated sampling biases and uncertainty

John McKinlay

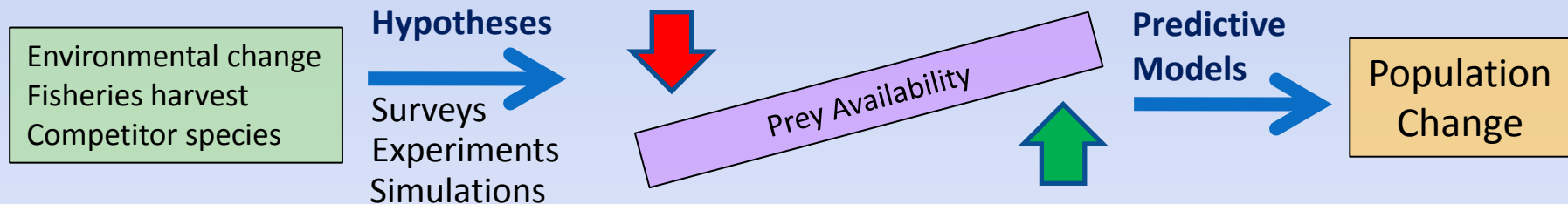
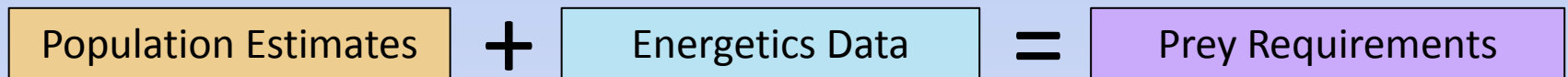
Southern Ocean Ecosystems

Acknowledgements:

Louise Emmerson, Kim Newbery, Colin Southwell, Rowan Trebilco

What is the Goal?

- To make use of current and historical penguin count data to estimate:
 - Time-series of abundance (individual sites) to examine if change is occurring over time
 - Regional estimates of abundance (many sites)
 - This will form an important input to help determine what might be driving change. How?





What are the problems to be overcome?

- Historical counts are generally not comparable due to differences in:
 - survey methodologies
 - methods of counting, measures of precision, how to deal with issues of detectability and sample fraction
 - population objects counted
 - adult, nest or chick counts
 - timing of the survey
 - affects availability of animals to sampling methods
- Both a goal and a problem: develop a method that can be easily used and understood by ecologists



A quick review of sampling concepts

- Availability
 - Not all animals in the target population may be available to the sampling methodology at the time the count is taken
- Detectability
 - When you count, not all animals that are present will be observed.
 - e.g. obscured by other animals or topography
- Sample fraction
 - Some colonies are too big to count, so we only count a known area or fraction and extrapolate to obtain a count for the total area

Abundance Estimator

$$\hat{N} = \frac{\hat{C}}{\hat{P}_{area} \cdot \hat{P}_a \cdot \hat{P}_{da}}$$

Where:

C count

P_{area} adjustment for sample fraction

P_a adjustment for availability

P_{da} adjustment for detection (given availability)

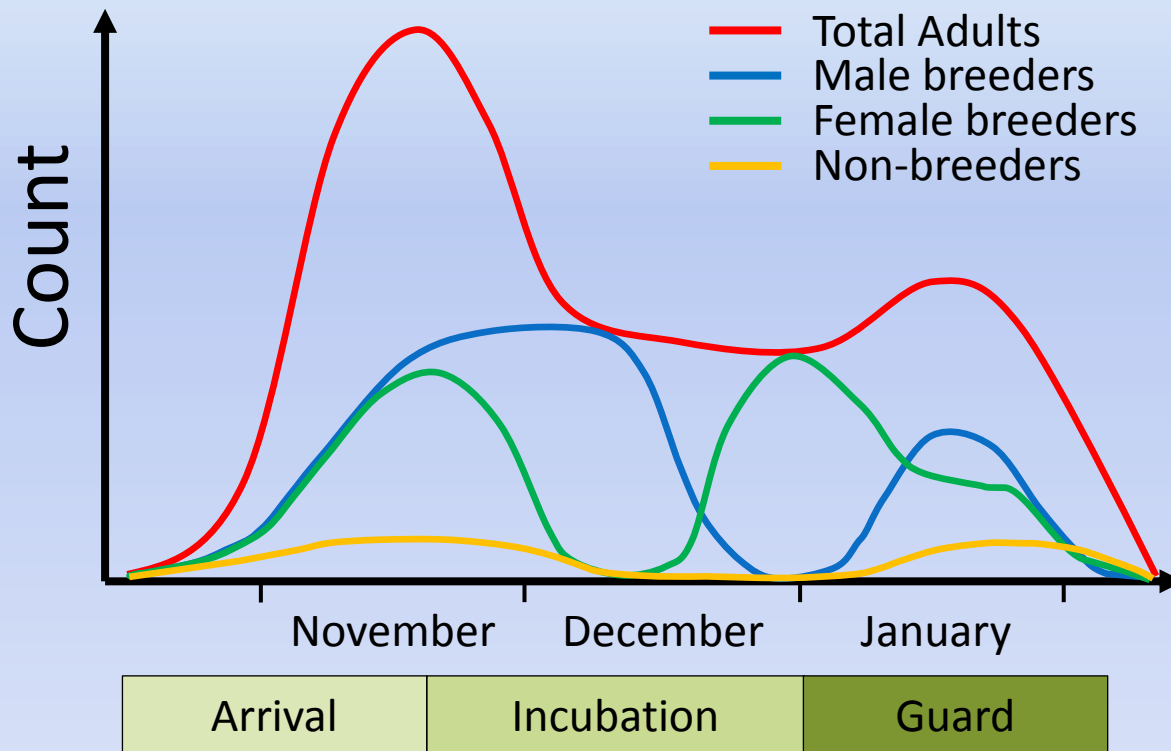
N Total population

- Some things to note

- quantities C, P_{area} , P_a and P_{da} are **estimates**, often with an **associated measure of precision**
- We can **resample** from postulated distributions of these variables to **propagate our level of uncertainty** through to the adjusted population estimate, N

Availability – what's that all about?

- Adélie penguins have a predictable pattern of attendance at breeding sites



- The date of the count will greatly influence your result. So how do we compare counts taken at different times?
- Ans: we measure the pattern of availability, and standardise counts taken at different times in a season so that they reflect a common point of breeding chronology

Availability – how do we measure it?

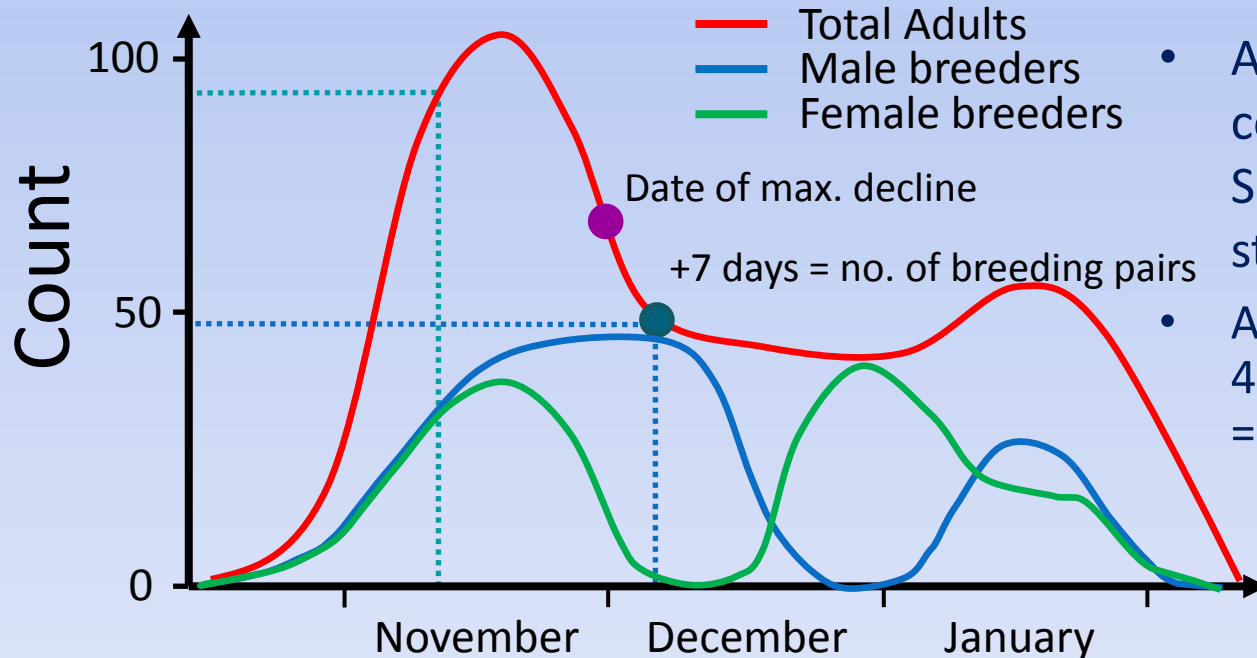
- Could count all the penguins in a colony at various points in the breeding season - but impractical



- Only need the shape to calculate an adjustment, so collect data on a subsample using remote camera technology

Availability – standardisation

- All counts are standardised to estimate the number of occupied nests, or equivalently breeding pairs, at their maximum value in a season
 - Defined by CCAMLR Ecosystem Monitoring Program Standard Method A3 as occurring 7 days after the peak in egg laying



- An example: we have a count of 5000 adult birds at Site X measured on 10 Nov, standardise this count.

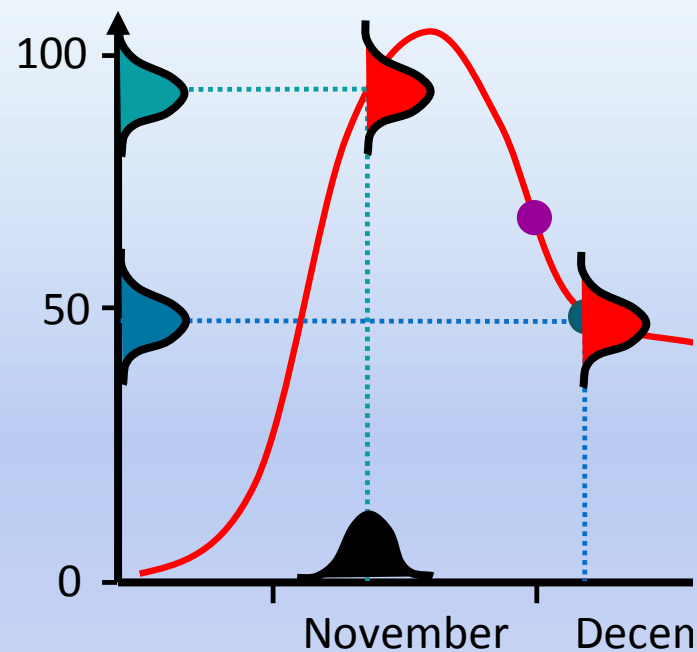
- Ans:
$$\frac{48}{94} * 5000$$

$$= 2553 \text{ breeding pairs}$$

(if only it were that simple!)

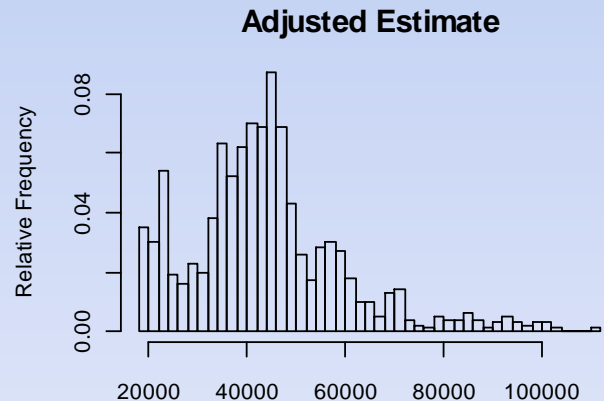
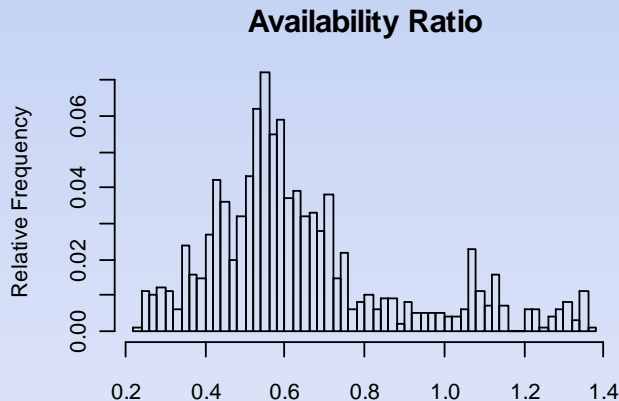
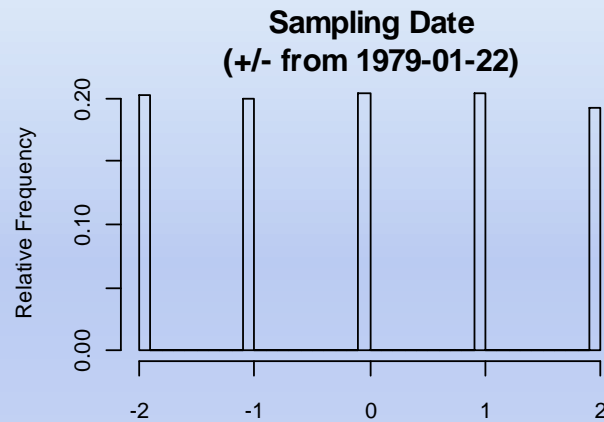
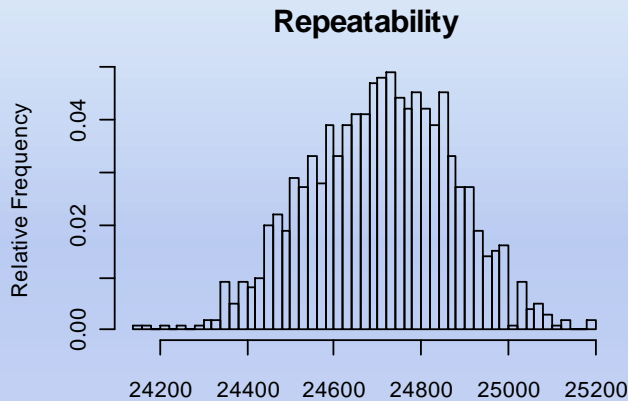
Availability – it's more complicated

- Date uncertainty
 - resample from a range of dates
- Curve uncertainty
 - resample taking the uncertainty of curve fit into consideration
- Different count objects
 - use object-specific availability functions for adults, chicks and nests
- No availability info for a site
 - Use surrogate availability curves from sites close in space or time
- Multiple availability curves for a site
 - Treated as equally weighted subsamples in the bootstrap process



Example 1 – a single site

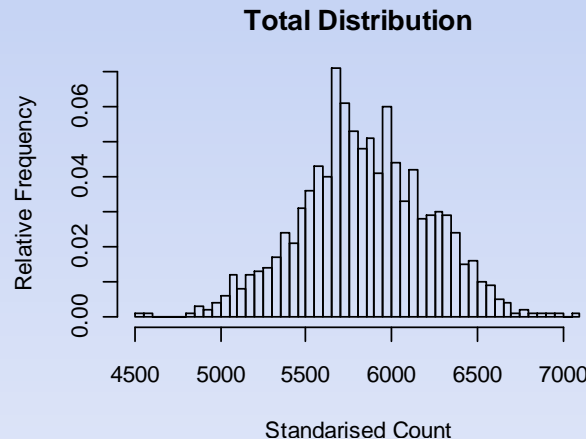
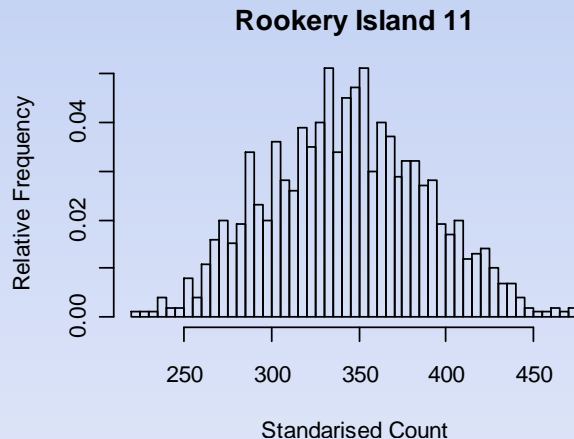
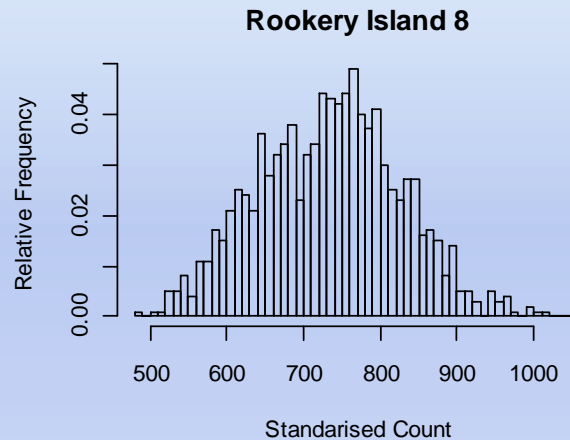
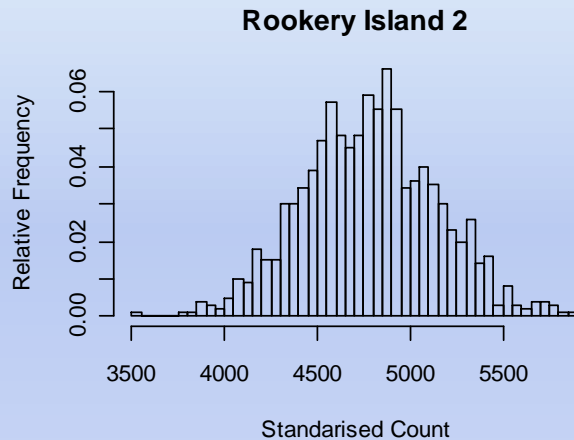
- Starck (1980) counted 24700 (+/- 346) adult penguins over 20-24 January 1979 at Haswell Island



- 1000 bootstrap replicates
- 5 possible dates
- 12 surrogate availability curves were used
- Sample fraction and detection fraction assumed 1
- Summary
mean 43 833
median 42 610
95% CI
(19 220, 86 358)

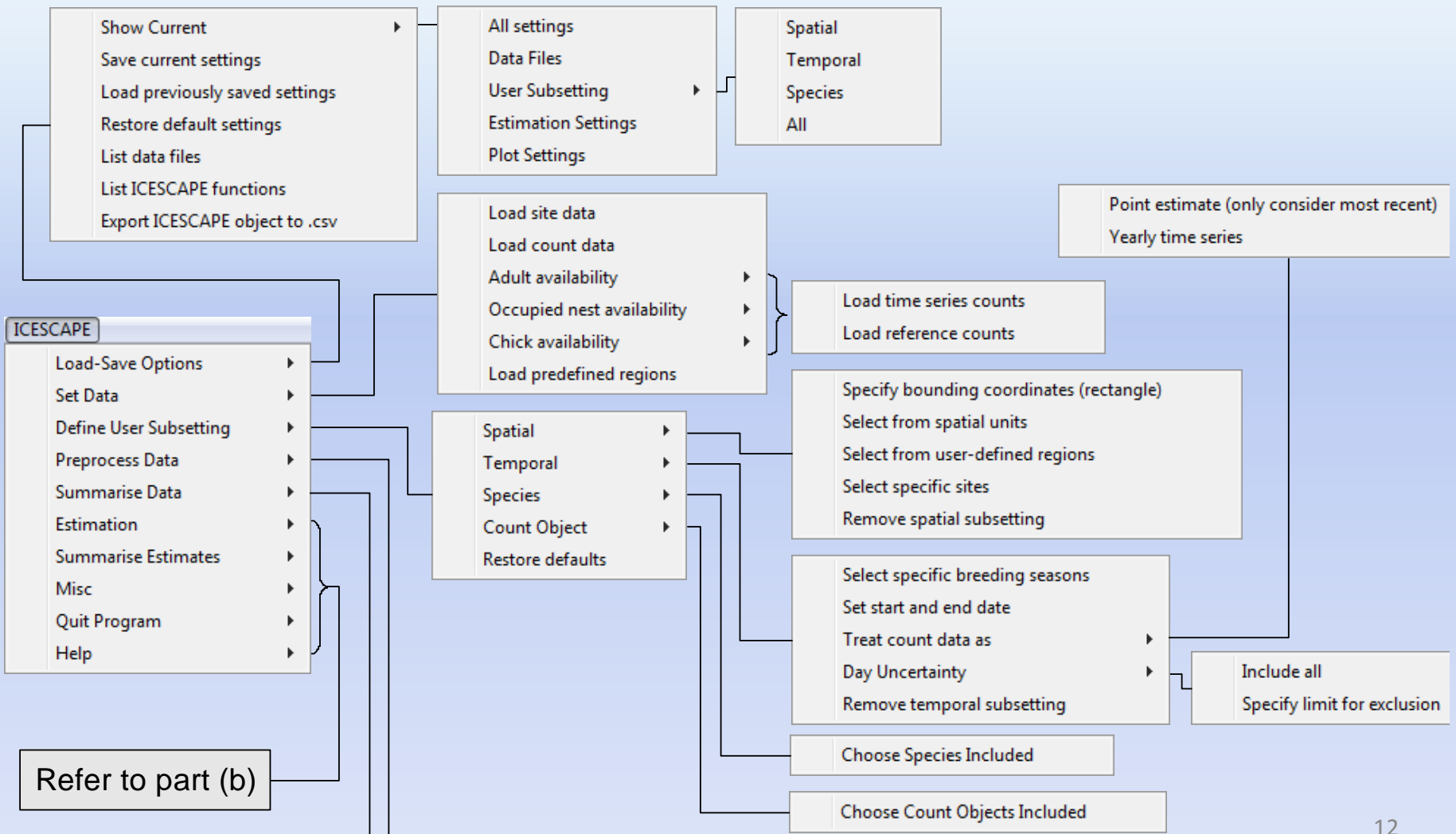
Example 2 – aggregating sites

- On 17 November 1972, Jones (reported in Horne (1983)) counted Adélie penguins at several sites in the Rookery Island group



- 1000 bootstrap replicates, 12 surrogate availability curves, sample and detection fractions assumed 1
- Site summation by independent draws without replacement from each contributing distribution
- Summary
mean 5 839
median 5 824
95% CI
(5 070, 6 551)

A menu-driven semi-GUI in R





Further Work

- ICESCAPE
 - Use covariates for selection of surrogate attendance curves
 - Formal tests for significant change (multiple comparisons, trend estimation)
 - Procedures for validating results
 - Account for non-breeding animals
 - Spatial display of results
 - Accommodate other colonial land-breeding species
 - Documentation, an R package for wider distribution
- Related
 - Sub-sampling of attendance function data
 - Automated methods of counting from photos



To finish, a personal perspective

- A difficult programming exercise; a complex, but not too difficult, statistical exercise
- Difficult to estimate timeframes for completion
 - Unbridled optimism, complexity of the problem, messy and sparse data, the unknowns
- Rewarding to provide an analysis tool, rather than simply an analysis
- Nice to be able to use existing historical data to their full potential

For those interested:

McKinlay, Southwell and Trebilco 2010 Integrating Count Effort by Seasonally Correcting Animal Population Estimates (ICESCAPE): A method for estimating abundance and its uncertainty from count data using Adélie penguins as a case study. CCAMLR Science, *In Press*.

Southwell, McKinlay, Emmerson, Trebilco and Newbery 2010 Improving estimates of Adélie penguin breeding population size: developing factors to adjust one-off population counts for availability bias. CCAMLR Science, *In Press*.