Filippo Galimberti [1] and **Simona Sanvito** [1,2]
1 Elephant Seal Research Group, Sea Lion Island, Falkland Islands
2 Dept. of Biology, Memorial University of Newfoundland, St. John’s NF A1B 3X9, Canada

ESRG technical report no. 2

Power analysis of population trends in southern elephant seals of the Falkland Islands

Results and potential applications

Milano, 12th August 2000

Address for correspondence: Dr. Filippo Galimberti – Via Buonarroti 35 – 20145 Milano – Italy
Phone +39 02 4980504 – Fax +39 02 48008145 – Email fil_esrg@tiscali.it

Introduction

Abundance monitoring is the basic tool of conservation biology and a main source of information on population dynamics (Bart et al. 1998). It represents a strong priority of research on Antarctic and sub-Antarctic seals (Hindell et al. 1994). The main goal of an abundance-monitoring program is to determine the changes in population size over time, with long-term trends as the focus of interest. Small populations present specific problems of trend detection, due to the intrinsic low power of analyses carried out on small samples (Forcada 2000). Hence, an accurate assessment of the status of small populations requires the estimation of the power of trend analysis.

From a statistical point of view, power is defined as 1 – β, with β being the probability of wrongly accepting the null hypothesis when it is actually false, *i.e.*, the probability to make a type II error. Power is, therefore, the likelihood to correctly rejecting the null hypothesis. The power of a monitoring program is influenced by many factors (Gerrodette 1987), including count error and variability, sample size, survey length, magnitude of trend to be detected (*i.e.*, in statistical terminology, the effect size), and statistical level of significance (*i.e.*, α, the probability of erroneously rejecting the null hypothesis when it is actually true, or type I error). Moreover, trends estimation is complicated by the variability of the results obtained using different models of population trend and different fitting methods (James et al. 1996; Thomas 1996).

The elephant seal (*Mirounga leonina*) population of Sea Lion Island (Falkland Islands) is a small and localized population (Galimberti and Boitani 1999). It has a particular conservation value, both locally, being the only notable breeding colony of this species in the Falklands, and globally, representing a potential conduit for gene flow between the two main populations of the South Georgia stock, South Georgia and the Valdés Peninsula (Hoelzel et al. 1993). From detailed counts carried out in 1995 and 1996, and from sparse counts carried out from 1989 to 1994, I provisionally concluded (Galimberti and Boitani 1999) that the population was almost stable. In this note, I re-analyze the trend including counts for the 1997-1999 period and using robust fitting methods, and I carried out a *post-hoc* power analysis. I then run a *a priori* power analysis to evaluate the trend detection capability of a long-term monitoring program based on different survey length and non-trend variability of counts.

Methods

The estimation of population size in elephant seals is complicated by the fact that, at any time, just a part of the individuals is hauled out, while the rest is at sea. Therefore, population size is estimated using an indirect method (McCann 1985) based on the number of females hauled out during the breeding season. This number may be accurately estimated from sparse daily counts by applying a mathematical model of the haul out process (Rothery and McCann 1987; Galimberti and Sanvito submitted).

Field data was collected intensively on Sea Lion Island during the period 1995-1999. Sparse counts were available for the period 1989-1994. Details of the censusing protocol and the haul out process modeling are available elsewhere (Galimberti and Boitani 1999; Galimberti and Sanvito submitted). The short length of the time series makes unlikely the presence of complex non-linearities. Therefore, to estimate the population trend, I fitted just two simple models, linear and exponential regression of number of females versus the year (Bart et al. 1998). I run ordinary least squares (OLS) regression models with permutation test (Manly 1991) using RT 2.0 software (Manly 1992; trial version available from http://www.west-inc.com/west-inc.htm). I checked the robustness of OLS regression by applying last absolute deviation (LAD) regression, a regression method that is much less affected by outliers than OLS regression (Cade and Richards 1996). I run LAD regression models using BLOSSOM software (Slauson et al. 1994; freely available for download from ftp://ftp.mesc.usgs.gov/webdl).

I ran the power analysis using the program MONITOR 6.2 (Gibbs 1995; http://WWW.MP1-PWRC.USGS.GOV/powcase/monitor.html), which uses a Monte Carlo simulation approach. I compared these results with the ones obtained with the program TRENDS (Gerrodette 1993; http://mmdshare.ucsd.edu/Trends.html), which uses a parametric analytical approach. Results from MONITOR and TRENDS were very similar, although the first produced slightly more conservative estimates of power and showed a larger difference among power to detect positive and negative trends of the same magnitude.

The basic constraint of the simulations was a "one plot, one count per year" survey scheme imposed by the study of a single population of a species that presents a single optimal yearly estimate of the number of females. I used, as mean initial value for simulation, the mean of yearly counts, which is only slightly larger than the last count (1999), that could be a valid alternative choice as starting value (Gerrodette 1993). The estimation of variability in population size due to non-trend effects is complicated because, even in absence of estimation error, the yearly estimates would not lie exactly on the trend line due to stochastic factors and short term fluctuations (Gerrodette 1987; Link and Nichols 1994). Therefore, I calculated the residual standard deviation after trend fitting from the 1989-1999 period. I used this estimate as the "plot variance" requested by MONITOR (and not the variance in yearly counts as stated in the user manual, see also Thomas and Krebs 1997), and to calculate the initial coefficient of variation requested by TRENDS (Gerrodette 1993). I evaluated the effect of an increase in variability of counts (due to reduced accuracy of counts or increased short-term fluctuations) by increasing the standard deviation, with CV ranging from 1 to 20 %. All simulations were run using a two-tailed 0.05 α level, 10000 resamplings, and integer rounding of generated random numbers, as appropriate to count data (Gibbs 1995).

Results

For the 1989-1996 period the slope of the regression was close to 0 ($b = 0.3172 \pm 0.7775$, permutation test for slope $\neq 0$, 100000 resamplings: $P = 0.73$). In 1997, there was an increase of 5.7%, maintained in 1998 and followed by a decrease in 1999. For the period 1989-1999 the slope of the regression was greater but not significantly different from 0, with less than 1% yearly increase in the number of females ($b = 3.2503 \pm 1.7281$, $P = 0.0933$). Excluding 1997 and 1998 the increase in number of females was less than 1 female per year ($b = 0.9579$). LAD regression confirmed the presence of a very small, non-significant, increase trend ($b = 1.3333$, permutation test for slope $\neq 0$, 100000 resamplings: $P = 0.78$). The fitting of an exponential trend gave similar results, with an estimated, non-significant, 0.60% annual increase.

The residual standard deviation for the 1989-1999 counts after fitting of the trend was 15.5017 (CV = 2.90%). I run a Monte Carlo *post-hoc* power analysis using this value as spread of counts. Power was above 0.90 for trends of ± 2% or more ($1-\beta > 0.93$ for decrease trends and 0.97 for increase trends). On the contrary, it was very low for ± 1% trends ($1-\beta = 0.44$ for decrease trends and 0.49 for increase trend).

I ran two series of simulations to estimate *a priori* power for a long-term monitoring program. In the first series, I used the previously estimated 2.90% CV as measure variability of counts, and I simulated the effect of the length of the program by increasing the number of yearly surveys from 3 to 20. I fixed the target power at 0.90, which is a quite strict criterion (Cohen 1988), but seems a good balance between effectiveness and practical constraints for a small population with a potentially high extinction risk. Simulation results for a linear and an exponential trend were almost equal, and I present results of the former only (Table 1). The number of years required to achieve the target power depended on the effect size, i.e. the rate of the trend, in particular for small trends (Figure 1).

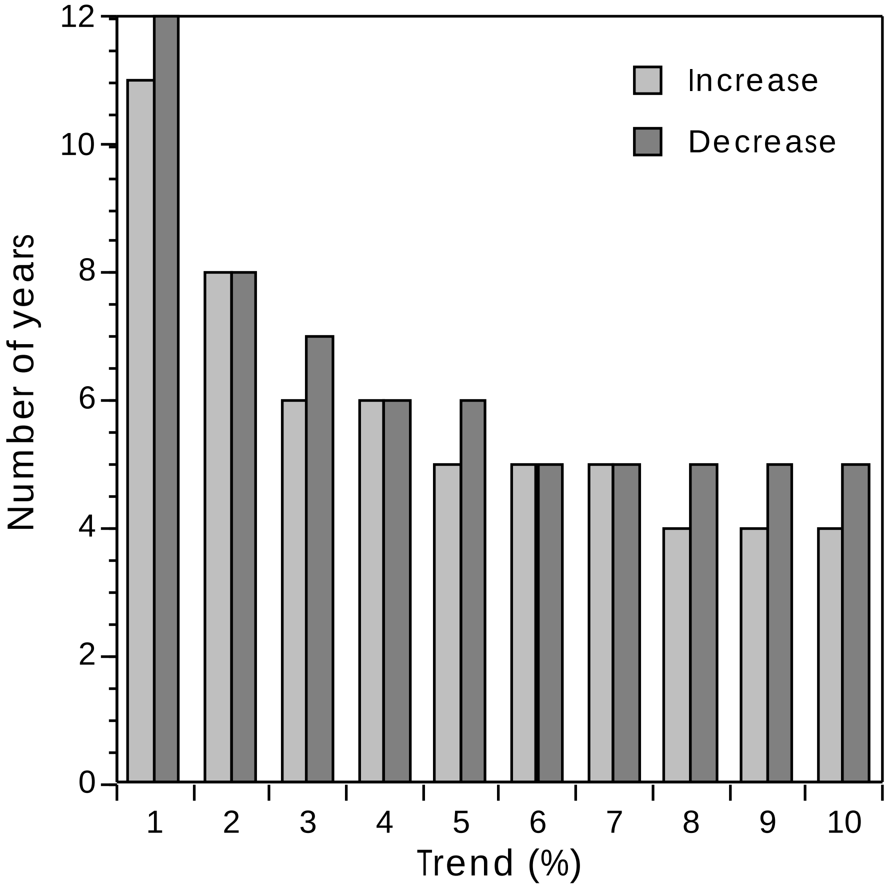

Figure 1 - Bar chart of the length of survey (number of yearly counts) required to achieve a 0.90 power for different trends (% increase or decrease), with the observed variation in counts (CV = 2.90%)

Trend	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	0.324	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0.293	0.955	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	0.261	0.914	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	0.224	0.838	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0.198	0.728	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.157	0.592	0.968	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.127	0.434	0.871	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.098	0.285	0.654	0.929	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.072	0.152	0.347	0.628	0.868	0.977	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	0.056	0.073	0.126	0.203	0.328	0.492	0.676	0.819	0.920	0.977	0.993	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0	0.043	0.047	0.050	0.052	0.049	0.053	0.052	0.051	0.048	0.050	0.051	0.046	0.056	0.052	0.052	0.049	0.049	0.051
-1	0.056	0.078	0.118	0.197	0.300	0.450	0.592	0.751	0.868	0.939	0.977	0.995	0.998	1.000	1.000	1.000	1.000	1.000
-2	0.070	0.152	0.303	0.548	0.789	0.932	0.987	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-3	0.099	0.247	0.543	0.844	0.973	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-4	0.122	0.369	0.758	0.965	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-5	0.147	0.496	0.891	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-6	0.170	0.600	0.958	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-7	0.194	0.697	0.984	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-8	0.227	0.777	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-9	0.241	0.848	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-10	0.272	0.886	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 1 – Power of trend detection with different size of the population trend (% increase or decrease) and different survey length.

A 1% increase required 11 years to be safely detected, and a 1% decrease 12 years. A 5-years long survey safely detected only larger trends, with 6% or more increase or decrease. The detection of a negative trend required in most cases one yearly count more than an equivalent positive trend.

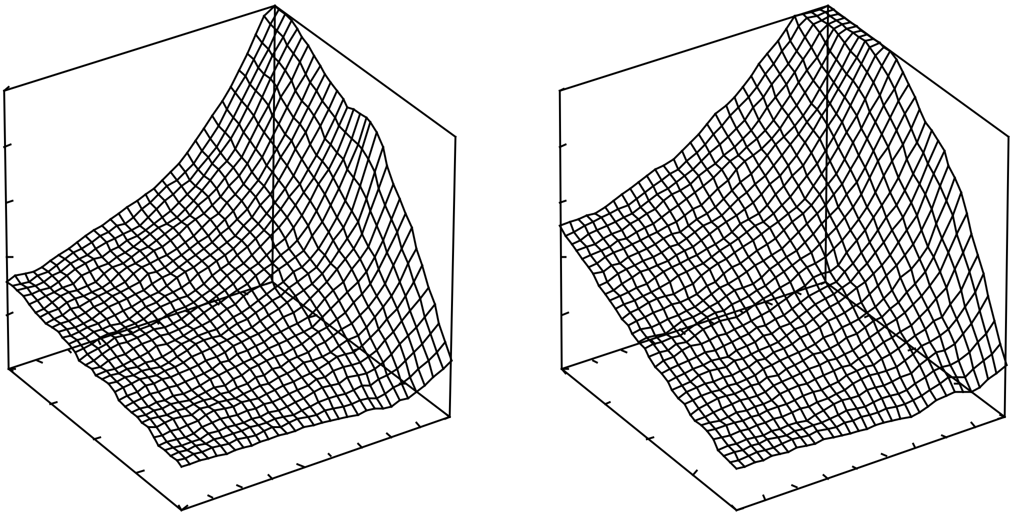

Figure 2 - Distance weighed last squares surface fitted to the number of years required to achieve 0.90 power with different trends (% increase, left, or % decrease, right) and different variability of counts.

I then evaluated the effect of an increased variability of counts by systematically changing the non-trend variability in counts (CV ranging from 1 to 20%, at 1% steps). From these simulations I calculated the minimum number of yearly counts required to achieve a 0.90 power. The increase in variability of counts greatly increased the number of year required (Figure 2; Table 2). Negative 1% trends were not detectable in the 20 years survey span when CV of counts was higher than 6% (8% for 1% positive trends). Negative 2% trends were not detectable when CV of counts was higher than 12% (18% for 2% positive trends). Negative 3% trends were not detectable when CV of counts was higher than 17%.

Trend	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	7	8	8	8
9	4	4	4	5	5	5	6	6	6	6	7	7	7	7	7	8	8	8	8	8
8	4	4	5	5	5	5	6	6	6	7	7	7	7	8	8	8	8	9	9	9
7	4	4	5	5	5	6	6	7	7	7	7	8	8	8	9	9	9	9	10	10
6	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	10	10	10	10	11
5	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	11	11	11	12	12
4	4	5	6	6	7	7	8	9	9	10	10	11	11	11	12	12	13	13	13	14
3	4	5	6	7	8	9	10	10	10	11	12	13	13	14	14	15	15	16	16	17
2	5	7	8	9	10	11	12	13	14	15	16	16	17	18	18	19	20	20		
1	7	9	11	13	15	16	18	20												
0																				
-1	7	10	12	14	17	18														
-2	5	7	8	10	11	12	14	15	16	18	19	20								
-3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	19	20			
-4	4	5	6	7	8	8	10	10	10	12	13	14	14	15	16	17	17	18	19	20
-5	4	5	6	6	7	8	9	9	10	11	11	12	13	13	14	15	16	16	17	18
-6	4	5	5	6	7	7	8	8	9	10	10	11	12	12	13	13	14	15	15	16
-7	4	4	5	6	6	7	7	8	8	9	10	10	11	11	12	12	13	14	14	15
-8	4	4	5	5	6	6	7	7	8	9	9	9	10	11	11	12	12	13	13	14
-9	4	4	5	5	6	6	7	7	8	8	8	9	10	10	11	11	12	12	13	13
-10	4	4	5	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13

Table 2 – Survey length (number of yearly censuses) required to achieve a 0.90 power with different size of the population trend.

Discussion

There is an increasing appreciation of the importance of power analysis in conservation and wildlife biology (Taylor and Gerrodette 1993, Gibbs et al. 1998), although there is much less consensus about how to carry out it (Link and Hatfield 1991, Thomas 1997). The scope of trend analysis is to recognize sustained patterns in counts (the "signal" part of the time series), and to discriminate them from short-term fluctuations and variance due to estimation methodology (the "noise" part). The power of a monitoring program is the capability of its survey plan to detect population trends, and permit their assessment of their statistical significance (for an alternative view of the role of statistical inference see Johnson 1999).

The analysis of trends in small populations is complicated by the low power typical of small samples, due to the small number of plots and the large variation of counts. In such situations, small trends will require long series of data to be safely detected. This is particularly unfortunate when decreasing trends are involved, because the safe detection of a small trend will be granted only after a notable reduction in the total size of the population (see also Forcada 2000). The re-analysis of the Sea Lion Island data confirmed that the population is almost stable, but clearly showed that the current time series is not long enough to safely detect small trends, in the ± 1% range. There is some indication that the population is in fact increasing at a less than 1% rate, but the current data set is to small to permit safe statistical assessment of the trend. The importance of small trends should not be underestimated: for example, the notable reduction in size of the Macquarie Island population (57.5% decline in the period 1949-1990) was due to a mere 2.1% mean annual decrease (Hindell et al. 1994).

The coefficient of variation of the Sea Lion Island counts was very small if compared to values observed in surveys of large mammal populations (mean CV = 19.9%; database available at http://www.mp1-pwrc.usgs.gov/powcase/powcase.html). This was due in part to the small fluctuations observed in the population, but also to the very accurate counts repeated during the whole length of the breeding season. By modeling the houl out process (Galimberti and Sanvito submitted), we demonstrated that a very good estimate of the total number of breeding females could be obtained with a small number of daily counts. If counts are carried close to the day of peak presence on land of females (i.e., in a 1-week period centered on the peak) even a single count guarantees an estimation error for the total number of breeding females within ± 2%. Therefore, even with a relaxation of the censusing protocol, it should be possible to keep the total variation (count error plus fluctuations) with a 5% limit. This means that a ± 2% trends could be detected in 10 years-long survey, and that even a ±1% trends could be detected in 20 years.

In all, *a priori* power analysis seems a useful tool for the planning of monitoring programs in small population of seals. Power analysis, both with Monte Carlo and analytical methods, produces approximate results (Gerrodette 1991). Notwithstanding this, it permits a first assessment of the likelihood of trend detection, and a valuable comparison of the pros and cons of alternative survey designs.

Acknowledgements

We wish to thank Carla and Alberto Galimberti for their support of our research on elephant seals; Simona Sanvito, Anna Fabiani, Ilaria Marzetti, Alessia Gallastroni, Alice Camplani, and Daniela Marasco for their help with data collection; the Falkland Islands Government for allowing us to carry out our research in the Falklands; the Falkland Islands Development Corporation for allowing us to do the field work on Sea Lion Island; David Gray for his help with the logistics.

Literature cited

Bart, J., M. A. Fligner and W. I. Notz. 1998. Sampling and statistical methods for behavioral ecologists. Cambridge University Press, London.

Cade, B. S. and J. D. Richards. 1996. Permutation tests for least absolute deviation regression. Biometrics 52:886-902.

Cohen, J. E. 1988. Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale, NJ.

Forcada, J. 2000. Can population surveys show if the Mediterranean monk seal colony at Cap Blanc is declining in abundance ? Journal of Applied Ecology 37:171-181.

Galimberti, F. and L. Boitani. 1999. Demography and breeding biology of a small, localized population of southern elephant seals (*Mirounga leonina*). Marine Mammal Science 15:159-178.

Galimberti, F. and S. Sanvito. Modelling females haul out in southern elephant seals (*Mirounga leonina*). Submitted to Aquatic Mammals.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364-1372.

Gerrodette, T. 1991. Models of power of detecting trends - a reply to Link and Hatfield. Ecology 72:1889-1892.

Gerrodette, T. 1993. TRENDS: software for a power analysis of linear regression. Wildlife Society Bulletin 21:515-516.

Gibbs, J. P. 1995. MONITOR. User manual. Department of Biology, Yale University, New Haven, Connecticut.

Gibbs, J. P., S. Droege and P. C. Eagle. 1998. Monitoring local populations of plants and animals. Bioscience 48:935-940.

Hindell, M. A., D. J. Slip and H. R. Burton. 1994. Possible causes of the decline of southern elephant seal populations in the southern Pacific and southern Indian ocean. Pages 66-84 *in*
B. J. Le Boeuf and R.M. Laws. Elephant seals. Population ecology, behavior and physiology. 66-84. University of California Press, Berkeley, CA.

Hoelzel, A. R., J. Halley, S. J. O'Brien, C. Campagna, T. R. Arnbom, B. J. Le Boeuf, K. Ralls,K. and G. A. Dover. 1993. Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. Journal of Heredity 84:443-449.

James, F. C., C. E. McCulloch and D. A. Wiedenfeld. 1996. New approaches to the analysis of population trends in land birds. Ecology 77: 13-27.

Johnson, D. H. 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63:763-772.

Link, W. A. and J. S. Hatfield. 1991. Power calculations and model selection for trend analysis: a comment. Ecology 71:1217-1220.

Link, W. A. and J. D. Nichols. 1994. On the importance of sampling variance to investigations of temporal variation in animal population size. Oikos 69:539-544.

Manly, B. F. J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, London.

Manly, B. F. J. 1992. RT. A program for randomization testing. User manual. University of Otago, Otago, New Zealand.

McCann, T. S. 1985. Size, status and demography of Southern elephant seal (*Mirounga leonina*) populations. Pages 1-17 *in* J. K. Ling and M. M. Bryden. Studies of sea mammals in South Latitudes. South Australia Museum, Northfield.

Rothery, P. and T. S. McCann, T. S. 1987. Estimating pup production of elephant seals at South Georgia. Symposia of the Zoological Society of London 58:211-223.

Slauson, W. L., B. S. Cade and J. D. Richards. 1994. BLOSSOM Statistical Software. User manual. Midcontinent Ecological Science Center, National Biological Survey, Fort Collins, Colorado.

Taylor, B. L., and Gerrodette, T. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. Conservation Biology 7:489-500.

Thomas, L. 1996. Monitoring long-term population change: why are there so many analysis methods ? Ecology 77:49-58.

Thomas, L. 1997. Retrospective power analysis. Conservation Biology 11:276-280.

Thomas, L. and C. J. Krebs. 1997. A review of statistical power analysis software. Bulletin of the Ecological Society of America 78:128-139.