USING ECOSYSTEM MONITORING DATA TO DETECT IMPACTS

S.L. Hill✉, J. Forcada, P.N. Trathan and C.M. Waluda
British Antarctic Survey
Natural Environment Research Council
High Cross, Madingley Road
Cambridge CB3 0ET
United Kingdom
Email – sih@bas.ac.uk

Abstract

The purpose of ecosystem monitoring programs is to indicate the state of ecosystems and whether they have been impacted by activities such as fishing. This paper discusses a range of methods for inferring such impacts using monitoring data with no control sites. These methods assess either (i) the expected probability of an observed value in an unimpacted system, or (ii) the frequency of values below a fixed reference point. The second approach allows inference criteria based on changes in this frequency rather than by reference to a critical probability. All methods would have provided a sustained indication of a significant decline in Antarctic fur seal (*Arctocephalus gazella*) pup production at South Georgia from the early 1990s within a few years of its onset, but a fixed reference point method could have provided this sustained indication from the onset. Furthermore, simulation of all methods suggests that the total probability of error (false positives and false negatives combined) is lowest with fixed reference point methods. The probabilities of Type I and Type II error can be evaluated analytically for these methods, which facilitates decision-making based on attitudes to risk.

Résumé

Les programmes de suivi des écosystèmes ont pour objet d'indiquer l'état des écosystèmes et s'ils ont subi l'impact d'activités telles que la pêche. Le présent document examine diverses méthodes visant à inférer ces impacts au moyen de données de suivi sans sites de contrôle. Ces méthodes évaluent soit i) la probabilité prévue d'une valeur observée dans un système non touché, soit ii) la fréquence des valeurs situées en dessous d'un point de référence fixe. La seconde approche tient compte de critères d'inférence fondés sur les changements de cette fréquence plutôt que par référence à une probabilité critique. Toutes les méthodes auraient permis de révéler, dans les premières années, un déclin important prolongé de la production de jeunes chez l'otarie de Kerguelen (*Arctocephalus gazella*) en Géorgie du Sud depuis le début des années 1990, mais la méthode du point de référence fixe aurait pu l'indiquer dès le début. De plus, la simulation de toutes les méthodes semble indiquer que la probabilité totale d'erreur (faux positifs et faux négatifs confondus) est plus faible avec la méthode du point de référence fixe. Les probabilités d'erreurs de Type I ou de Type II de ces méthodes peuvent être évaluées analytiquement, ce qui facilite la prise de décision basée sur les attitudes face au risque.

Резюме

Цель программ экосистемного мониторинга заключается в том, чтобы выявить состояние экосистем и то, подвергаются ли они воздействию таких видов деятельности, как промысел. В данном документе рассматривается ряд методов для определения таких воздействий с использованием данных мониторинга без контрольных участков. Эти методы оценивают либо (i) предполагаемую вероятность наблюдавшейся величины в не подвергшейся воздействию экосистеме, либо (ii) частоту величин ниже фиксированного контрольного значения. Второй подход позволяет использовать критерии определения, основанные на изменениях этой частоты, вместо того, чтобы обращаться к критической вероятности. Значительное сокращение производства щенков южного морского котика (*Arctocephalus gazella*) в районе Южной Георгии с начала 1990-х гг. было бы подтверждено всеми методами через несколько лет после его начала, но метод фиксированного контрольного значения мог бы указать на это с самого начала. Кроме того, моделирование всех методов говорит о том, что общая вероятность ошибки (ложноположительные и ложноотрицательные результаты) является наиболее низкой при использовании методов фиксированного контрольного значения. Вероятности ошибки первого и второго рода для этих методов можно оценить аналитически, что упростит принятие решений на основе отношения к риску.

Resumen

El propósito de los programas de seguimiento de ecosistemas es indicar el estado de los mismos y si han sido afectados por actividades como la pesca. Este documento discute varios métodos para inferir este impacto utilizando datos de seguimiento pero sin sitios de control. Estos métodos evalúan (i) la probabilidad esperada de un valor observado en un ecosistema que no ha sufrido un impacto, o (ii) la frecuencia de valores por debajo de un punto fijo de referencia. El segundo enfoque permite adoptar criterios para la inferencia en base a cambios en dicha frecuencia, y no con referencia a una probabilidad crítica. Todos los métodos habrían proporcionado, a pocos años de iniciarse el suceso, una indicación consecuente de una disminución significativa en la producción de cachorros de las poblaciones del lobo fino antártico (*Arctocephalus gazella*) en Georgia del Sur a principios de los 90, pero un método con un punto de referencia fijo habría indicado esta disminución desde el comienzo. Más aún, las simulaciones de todos los métodos sugieren que la más baja probabilidad total de error (falsos valores positivos y negativos combinados) se obtiene con los métodos con punto fijo de referencia. La probabilidad de que ocurra un error Tipo I o Tipo II con estos métodos puede ser evaluada analíticamente, lo que facilita la toma de decisiones en base a la posición adoptada con respecto al riesgo.

Keywords: CCAMLR ecosystem monitoring program (CEMP), ecosystem management, reference points, risk, Antarctic fur seal, *Arctocephalus gazella*, pup production, CCAMLR

Introduction

The objectives of the ecosystem approach to fisheries management include limiting the impacts of fishing on natural ecosystems. The management of active fisheries should therefore include monitoring to indicate the state of the natural ecosystem and strategies for responding to potential impacts. Monitoring programs often focus on the average health or performance of individuals within indicator populations. Such parameters do not directly indicate overall population size, which is usually the tacit or explicit focus of conservation objectives for animal populations. Furthermore, the state of indicator populations cannot always be assessed by reference to a control population. Developing management strategies that use uncontrolled performance data to meet conservation objectives remains a significant challenge.

CCAMLR oversees the management of fisheries in the Southern Ocean and follows a set of principles that recognise the need to manage fishery impacts on the wider ecosystem (Constable et al., 2000; Miller, 2002). One of the ongoing challenges faced by CCAMLR is to develop a management strategy for the krill fishery which allows rational use of krill (including expansion beyond current catch levels) but which meets conservation objectives for 'dependent and related species' (Constable, 2002; Hewitt et al., 2004). These species include, but are not limited to, the predators of the harvested species. There is, therefore, a need to define operational conservation objectives for such species (Butterworth and Thomson, 1995; Constable, 2001).

The CCAMLR Ecosystem Monitoring Program (CEMP) is intended to detect fisheries impacts on the ecosystem, focusing on three main integrated study areas within the Southern Ocean (Agnew, 1997). At Bird Island, South Georgia, a total of 17 biological parameters (including diet, breeding success, weight, growth rate and foraging trip duration) are monitored from four key predator species as part of CEMP, along with a number of non-CEMP parameters (see Reid et al., 2005 for details). Most of the predator monitoring parameters reflect the local availability of the primary prey species of the monitored predator (Reid and Croxall, 2001; Barlow et al., 2002; Reid et al., 2005).

It has been suggested that data such as that collected by CEMP and associated long-term monitoring studies could be integrated into the management of krill fisheries (Constable, 2001, 2002), for example by triggering tactical adjustments to catch limits in relatively small spatial subdivisions of the fishery (Hewitt et al., 2004). Predator monitoring should provide information about the state of the monitored populations and, therefore, whether conservation objectives are being met. However, this requires a clear link between the conservation objectives and the state being monitored. Several previous analyses have assumed that such objectives would be defined in terms of population size reference points (Butterworth and Thomson, 1995; Plagányi and Butterworth, 2007; Watters et al., 2006), although Constable (2001) presents a number of alternatives. Population size reference points can be problematic if the population is not monitored at the relevant scale. Nonetheless, the CEMP parameters are indicative of population health which is a legitimate focus for conservation objectives.

This paper discusses ways of using data from long-term monitoring programs, such as CEMP, within management strategies to infer whether the state of the ecosystem is being adversely impacted and to trigger management responses. The nature and goals of these programs often preclude standard environmental impact assessment methods which examine the differences between the sites of putative impacts and unaffected control sites (Stewart-Oaten et al., 1986). Instead, tractable approaches rely on the timely detection of states whose probability of occurrence in the unimpacted ecosystem is low. Such states could be characterised by 'changing variability (range); trends; shifts; [or] changes in the frequency of anomalies [in monitoring data]' (de la Mare and Constable, 2000). It is necessary to distinguish between anomalies and impacts. Anomalies are rare observations which can occur in an unimpacted system. Indeed, statistical definitions of anomalies tend to be based on their probability of occurrence in baseline data (e.g. de la Mare and Constable, 2000). Although an impact might be characterised by a change in the frequency of anomalies, the mere presence of an anomaly is not necessarily diagnostic of an impact.

There are well-known limitations associated with statistical inference which, in this context, translate directly into risk. The probability of falsely inferring an impact (Type I error) carries the risk of unnecessarily disrupting the fishery, which must be balanced against the risk of failing to detect a real impact (Type II error) and, therefore, not acting to protect the ecosystem. There is an additional risk of detrimental delays in the management response associated with methods that take a long time to detect an impact. This paper introduces inference approaches and evaluates the probability of Type I and Type II error analytically where possible and through simulation using time series based on Antarctic fur seal pup production data from Bird Island. This data series includes a known non-fisheries impact (Forcada et al., 2005, 2008) characterised by frequent low values from 1991 onwards.

Materials and methods

Approaches to inferring an impact are illustrated using example data based on annual fur seal pup production at Bird Island, South Georgia (54°00'S 38°03'W). This is a long-term performance measure which has been well studied and which has very low sampling error (Forcada et al., 2005). There is no evidence of a fishery-induced impact on these data during the monitoring period. However, both pup production and population growth rate at Bird Island fell in 1991 and remain depressed. This was an apparent response to environmental conditions indicated by a high frequency of positive sea-surface temperature anomalies linked to the El Niño Southern Oscillation (Forcada et al., 2005, 2008).

Uninterrupted annual pup production data are available for Bird Island from 1984 to present. This was divided into 'baseline' (1984 to 1990) and 'impacted' (1991 to 2006) periods (Forcada et al., 2005, 2008). Each of the approaches described below was applied to the pup production time series, with critical probabilities established from a normal distribution fitted to the baseline data.

The approaches were also assessed using a range of simulated distributions based on the pup production data. In each case, the parameters of a baseline normal probability density function (PDF) were established from 20 observations drawn from a distribution representing the baseline period. The critical values identified using this baseline PDF were then used to test for impacts in 20 more observations drawn from the baseline period, representing an 'unimpacted' population, and 20 observations drawn from a distribution representing the impacted population. This process was repeated 1 000 times. The proportions of unimpacted observations falsely identified as impacted, and impacted observations falsely identified as unimpacted were then calculated. The simulated data were drawn from either the non-parametric distributions of pup production data in the baseline and impacted periods (sampling with replacement), or normal distributions representing the shifts in mean and variance between these periods. All parametric baseline and unimpacted distributions had the same mean (794) and standard deviation (66), while the parameters of the seven impacted parametric distributions represented shifts in the mean, standard deviation, or both that were equal to or half the magnitude of those observed in the pup production data.

Inferring an impact

The following discussion considers methods for inferring an impact through comparison with baseline data for a period which predates any impact. Random variable A describes the state of the unimpacted system. A monitoring program records a new observation, a_t, each year t. For ease of notation, monitoring is assumed to begin at $t = 1$ and low values of a_t indicate unfavourable conditions.

All the inference methods compare observations, a_t, with random variable A. The following equations for calculating the probabilities associated with the inference methods assume strict statistical independence of events.

Method 1

A simple approach is to assess the probability that a value $\leq a_t$ would have occurred by chance in the unimpacted system. This probability is obtained by firstly establishing $P(A \leq a_t)$, the marginal probability that any single observation of A will have a value $\leq a_t$, based on a cumulative distribution function fitted to the baseline data. The marginal probability is then adjusted to account for the number of observations. The probability of observing this outcome once in t years can be calculated using the binomial density function:

$$P(a_t) = t.p.(1-p)^t \tag{1}$$

where $p = P(A \leq a_t)$.

Method 2

Of course, values $\leq a_t$ might have been observed previously, so more insight can be gained by establishing $P(s \geq s_o)$, the expected probability that the number of observations, s, with value $\leq a_t$ is at least s_o, the actual number observed. The number of possible combinations of s_o events that could occur in t years is given by:

$$\binom{t}{s_o} = \frac{t!}{s_o!(t-s_o)!}\,. \tag{2}$$

The full version of the binomial density function, which accounts for $s > 1$ observations of random variable S with the relevant outcome, is:

$$P(S = s) = \binom{t}{s} p^s (1-p)^{t-s}\,. \tag{3}$$

Thus

$$P(s \geq s_o) = \sum_{\vartheta = s_o}^{t} \binom{t}{\vartheta} p^\vartheta (1-p)^{t-\vartheta}\,. \tag{4}$$

Method 3

A variation on Method 2 is to assess the probability of events observed within moving time windows (reference periods) rather than over the whole period of observation. The calculations are as for Method 2 but replacing t with $n \leq t$, the relevant reference period. This effectively weights recent observations and excludes those which occurred more than n years ago.

Method 4

The above methods assess the probability of observing values $\leq a_t$ in an unimpacted system. Therefore, the observed value a_t is the reference point by which an impact is judged. An alternative is to identify a fixed reference point, which might have relevance to the ecology of the monitored population or the management objectives, and to assess where observations lie relative to this reference point, which is denoted a_{crit}. Any observation a_t will have one of two possible states relative to a_{crit}. These states are α_1 ($a_t \leq a_{crit}$) and α_2 ($a_t > a_{crit}$). State α_1 occurs with marginal probability $P(\alpha_1) = p_1$ in the unimpacted system. An n-year reference period can be thought of as a series of Bernoulli trials in which the outcome can be either α_1 with marginal probability $P(\alpha_1)$ or α_2 with marginal probability 1 – $P(\alpha_1)$. The probability of observing outcome α_1 at least s_{crit} times in n years in the unimpacted system is obtained by replacing the relevant terms in equation (4):

$$P(s \geq s_{crit}) = \sum_{\vartheta = s_{crit}}^{n} \binom{n}{\vartheta} p^\vartheta (1-p)^{n-\vartheta} \tag{5}$$

where $p = p_1$.

An impact could be defined as an increase in $P(\alpha_1)$ to or above a critical marginal probability p_2 whose value reflects an acceptable risk of Type II error (see below). Thus, an impact is inferred if state α_1 is observed in $s \geq s_{crit}$ years where $s_{crit} = n.p_2$. The inference criterion is written (s_{crit},n). Thus, the criterion (2,4) means that an impact will be inferred if $a_t \leq a_{crit}$ in two or more years of a four-year reference period.

Summary of methods: anomalies versus impacts

Method 1 assesses whether or not a_t is an anomalous observation. This needs to be put into the context of other recent observations to establish whether the anomaly is likely to be a consequence of an impact or just part of the natural variability in the system. Methods 2 to 4, therefore, determine the probability in an unimpacted system of the observed frequency of anomalies. If this probability is below an arbitrary critical probability, P_{crit}, the hypothesis that no impact has occurred is rejected.

However, Methods 2 and 3 assess the probability of observations $\leq a_t$, and so only assess the frequency of anomalies when a_t is itself an anomaly. Method 4 resolves this problem by assessing the frequency of observations relative to a fixed reference point.

Reference points and periods, and inference criterion in simulations

According to the fitted normal distribution, 14% of observations in the baseline pup production data, and 73% of observations in the impacted data, had marginal probability ≤0.15. Therefore, in the simulations summarised in Tables 1 and 2, Method 4 identified an impact if an event with marginal probability ≤0.15 occurred in two out of three successive years. The reference period for Method 3 was also three years.

Results

Application of inference methods

Figure 1(a) shows the time series of pup production data at Bird Island, South Georgia. From 1991 onwards, observations were generally (with one exception) below the mean for the baseline period. There was one anomaly (below the 5th percentile of the baseline distribution) in the seven baseline observations compared to six in the 16 impacted observations. If P_{crit} is set at 0.05, which is common practice in statistical ecology, then Method 1 identifies each of the anomalies except 2005 (Figure 1b). Method 2 integrates information across years to detect impacts rather than isolated anomalies. With $P_{crit} = 0.05$, the impact is identified in 1991, but not in four of the subsequent years (Figure 1c). However, persistent observations below the baseline mean ensure that the impact is identified in each year from 1998.

Method 3 assesses probabilities within moving reference periods and so disregards earlier observations. With short reference periods, this method can be more sensitive to individual observations than Method 2 (Figure 2). Both 5- and 10-year reference periods give persistent indications of the impact a few years before Method 2 (1995 and 1997 respectively) with $P_{crit} = 0.05$. The figure illustrates two important characteristics of longer time windows: they are less sensitive to current observations (e.g. 2005) than short reference periods but they do not provide any information at all until the appropriate reference period has elapsed.

When Method 4 is used in conjunction with P_{crit}, there is an interaction between the choices of reference point, P_{crit} and reference period (Figure 3). With $P_{crit} = 0.05$, none of the variations illustrated detect the impact before 1995, and with $P_{crit} = 0.1$ the very low reference point (the 5th percentile of the baseline distribution) combined with the 'all years' time window detects the impact in 1991, but also identifies the 1985 anomaly as an impact (albeit based on only two observations). Even the higher reference points identify the anomaly within eight years of its onset. Indeed, the 25th percentile of the baseline distribution was the only reference point to provide a persistent indication from 1997 onwards.

Figure 4 illustrates changes over time in $\frac{s}{n}$ (scaled by p_1), which is an empirical estimate of $P(\alpha_1)$. Because of the 1985 anomaly, the 1988 and 1989 estimates are more than double the expected value in the unimpacted system $\left(\frac{s.p_1}{n} \geq 2\right)$ when the reference point is set at the 5th or 10th percentile of the baseline distribution. $\frac{s.p_1}{n} < 1$ for these years when the reference point is set at the 25th percentile of the baseline distribution. With all reference points and reference periods, there is an upward trend in $\frac{s.p_1}{n}$ after 1991, and it rapidly exceeds the 1988 to 1989 value.

Risks

Metrics of risk indicate both the probability and consequences of an undesirable outcome. It is beyond the scope of this contribution to assess the detailed consequences of inappropriate action or inaction, and the following analysis considers the probability component of these risks.

Each of our methods defines an impact in terms of observation a_t. However, these observations are imprecise indicators of the system and might fail to respond at the expected magnitude when a real impact occurs. If an impact is inferred when the probability of an observation is below a critical value, P_{crit}, then a model of the relationship between the state of the indicator and the state of the system is needed to assess the probability of Type II error. However, Method 4 infers an impact if the probability $P(\alpha_1)$ increases from p_1 in the unimpacted system to or above a level p_2. It follows that $P(s \geq s_{crit})$ for $p = P(\alpha_1) = p_1$ (see equation 5) is the probability of Type I error with detection criterion (s_{crit}, n). Furthermore, $1 - P(s \geq s_{crit})$ for $p = P(\alpha_1) = p_2$ is the probability of Type II error.

Figure 5 shows the probabilities of Type I and Type II error resulting from all possible Method 4 inference criteria with $n \leq 10$ and selected values of

Table 1: Proportional frequencies of false positives in simulations of each impact detection method (M1 to M4) with 'unimpacted' scenarios based on fur seal pup production data, and three different values of P_{crit} (the critical probability of an observation deemed to indicate an impact).

	P_{crit}	M1	M2	M3	M4
(1) Non-parametric: unimpacted sampled from years 1984–1990					
	0.05	0.01	0.35	0.48	0.15
	0.1	0.02	0.39	0.53	
	0.25	0.04	0.48	0.66	
(2) Parametric: unimpacted mean = 794, unimpacted SD = 66					
	0.05	0.00	0.03	0.34	0.13
	0.1	0.01	0.07	0.41	
	0.25	0.02	0.21	0.57	

Table 2: Proportional frequencies of false negatives in simulations of each impact detection method (M1 to M4) with 'impacted' scenarios based on fur seal pup production data, and three different values of P_{crit} (the critical probability of an observation deemed to indicate an impact). Impacts were either in the form of a 'shift', a sudden change to the parameters specified in the subheading, or a 'trend', a linear change to these parameters over the 20-year simulation period.

		Shift				Trend			
	P_{crit}	M1	M2	M3	M4	M1	M2	M3	M4
(1) Non-parametric: impacted sampled from years 1991–2006									
	0.05	0.75	0.25	0.15	0.23				
	0.1	0.73	0.18	0.10					
	0.25	0.70	0.10	0.04					
(2a) Impacted mean = 657, impacted SD = 114									
	0.05	0.67	0.24	0.10	0.12	0.87	0.66	0.36	0.44
	0.1	0.63	0.19	0.07		0.85	0.59	0.28	
	0.25	0.56	0.12	0.03		0.79	0.43	0.17	
(2b) Impacted mean = 726, impacted SD = 90									
	0.05	0.9	0.55	0.3	0.36	0.96	0.86	0.51	0.64
	0.1	0.87	0.46	0.22		0.95	0.79	0.42	
	0.25	0.81	0.29	0.11		0.92	0.60	0.27	
(2c) Impacted mean = 657									
	0.05	0.77	0.17	0.07	0.07	0.92	0.69	0.38	0.44
	0.1	0.7	0.12	0.04		0.90	0.61	0.29	
	0.25	0.59	0.05	0.01		0.84	0.44	0.17	
(2d) Impacted mean = 726									
	0.05	0.95	0.55	0.33	0.36	0.98	0.86	0.53	0.65
	0.1	0.93	0.43	0.22		0.97	0.79	0.43	
	0.25	0.87	0.24	0.09		0.94	0.59	0.26	
(2e) Impacted SD = 114									
	0.05	0.95	0.86	0.53	0.71	0.97	0.93	0.59	0.79
	0.1	0.93	0.81	0.47		0.96	0.89	0.52	
	0.25	0.9	0.69	0.35		0.95	0.75	0.38	
(2f) Impacted SD = 90									
	0.05	0.97	0.91	0.58	0.76	0.99	0.95	0.62	0.83
	0.1	0.97	0.86	0.51		0.98	0.91	0.55	
	0.25	0.94	0.72	0.37		0.97	0.77	0.40	

p_1 and p_2. The general pattern is that the probability of Type I error starts low and increases with reference period, while the probability of Type II error starts high and decreases as reference period increases. Also, the probability of Type I error increases with p_1 while the probability of Type II error decreases as p_2 increases.

Overlaying the figures for the probabilities of Type I and Type II error identifies the inference criteria where these probabilities are balanced. In the example shown in Figure 6 (where $p_1 = 0.25$ and $p_2 = 0.50$), the probabilities are reasonably balanced in, for example, criteria (2,4), (3,7) and (4,10) and the actual probabilities fall as the reference period increases.

In all 'trend' simulations and all but one of the 'shift' simulations summarised in Tables 1 and 2, the total error (false positives plus false negatives) was lower with Method 4 than any other method. In non-parametric simulations, false positives were less common with Method 4 than any of the other impact detection methods (as distinct from Method 1 which is an anomaly detection method). In the parametric simulations, Method 2, with $P_{crit} \leq 0.1$ had a lower rate of false positives than Method 4. In all scenarios, the probability of a false negative was lowest with Method 3, followed by Method 4. These results are a function of the reference periods and reference points chosen. All methods performed better when the impact affected the mean than when it affected the variance only. There was a slight improvement in performance when the change affected both the mean and variance compared to the mean only. Performance was also better with higher magnitude changes and when the impact was a 'shift' rather than a 'trend', simply because the average magnitude of change was higher in the former.

Discussion

This paper presents several methods for inferring an impact using time series of observations without controls for comparison. It illustrates these methods using a range of real and simulated time series based on a protracted non-fisheries impact. de la Mare and Constable (2000) tested an anomaly detection method analogous to Method 1 above and noted the need to examine the power of indices based on monitoring data to detect changes in the state of the system. This paper assesses the trade-offs between statistical power and the probability of false positives for methods designed to detect such changes in the system. The simulation results presented above suggest that monitoring for increases in the frequency of observations below a reference point (Method 4) perform best in terms of the probability of overall (Type I plus Type II) error. With an appropriate inference criterion, Method 4 could have provided a sustained indication of the impact from 1991 onwards. However, the impact was so severe that all detection methods provided a sustained indication of the impact within a few years of its onset.

Table 2 shows that the probability of a false negative result in any individual year of an impact is often substantially above 10%. However, monitoring programs allow repeated observations as the years progress. The probability of observing only false negatives in, say, a 5-year period is 0.001 when the probability of a false negative in any year is 0.32. This must be balanced against the probability of observing a false positive at least once in five years (which is 0.556 when the probability of a false negative in any year is 0.15). So, while longer observation periods allow more time to confirm an impact, they also increase the probability of falsely diagnosing an impact and add to the risks associated with inaction.

Evaluation of the risks associated with Method 4 highlights the major trade-offs involved. The risk of Type I or Type II error can only be minimised at the expense of increasing the alternative type of risk. Also, the risk of Type II error falls with longer reference periods, while the risk of Type I error increases. Furthermore, there is a trade-off between the speed of response and the value of accumulated evidence, which is indicated by a fall in the point at which the two types of risk are equal, as the reference period increases.

An impact, in the sense used in this analysis, is defined in terms of the characteristics of the observed variable and says little about the causes of the impact or the state of the wider system, which might be the subject of management objectives. The relationship between indicator variables and management objectives, and the attribution of impacts to specific causes, are both issues that deserve wider consideration (Constable, 2002). The inference methods are sensitive to non-fisheries impacts, so there is a clear need to establish baseline, or 'no fishing' distributions of observed variables which include these non-fishery impacts. No fishing distributions (*sensu* Watters et al., 2008) are derived from model projections with appropriate representations of uncertainty and represent the possible states of the system in the absence of fishing but with other potential impacts, such as climate change. The assumption of statistical independence between events might be violated for many time series, as the performance of indicator species is often linked to autocorrelated environmental variables (Forcada et al., 2005; Trathan et al., 2006) and this issue also merits further exploration. However, the methods were effective in detecting the shift in fur seal pup production despite autocorrelation in the underlying environmental forcing.

Implementing the ecosystem approach will require a set of operational management objectives for dependent predators (Constable, 2001, 2002). Management performance can only be assessed in terms of what is known about the state of the system, which suggests that it is appropriate to define operational objectives in terms of the monitoring data, assuming that these data indicate the state of the system envisaged in the objectives of the relevant management organisation (e.g. CCAMLR). This analysis suggests that operational management objectives, which identify both a reference state of the indicator and a reference probability of observing that state, would be pragmatic (see also Constable, 2001). A management strategy should also include a criterion for inferring the state of the system relative to these reference points. No inference criterion will be completely reliable, and so the choice of criterion should be based on risk and detection probability. It is not sufficient to specify that the system should not be detectably different from baseline, it is also necessary to specify this detectable difference in terms of the ecosystem manager's tolerance to the various risks. This tolerance should, of course, be informed by stakeholder requirements.

There are a number of issues to consider when choosing reference states. The detection of the early stage of an impact allows greater opportunity to solve the problem than confirmation of a catastrophic impact. Also, the detection of subtle changes might allow equivalently subtle management responses. This might suggest that it is appropriate to choose a reference state which represents a subtle deviation from baseline conditions, such as the 25th percentile of the unimpacted distribution. However, any single reference state might have very different ecological significance for different variables (including the same parameter for different species). A system which oscillates between 'good' and 'very bad' might be more resilient to several very bad events than a system which responds more subtly to perturbation, as the dynamic 'exploitation phase' of Holling's (1986) adaptive cycle is more resilient to disturbance than the more stable 'conservation phase' (Gunderson, 2000).

In view of these considerations, it would be inadvisable to choose a single reference state unless the dynamics of the system were already well understood. It is, of course, relatively trivial to compare indicator data with several reference states simultaneously, as in Figure 4. This will inevitably increase the risk of Type I error, analogously with performing multiple statistical tests (Rice, 1989) which should also be accounted for in devising the management strategy.

Conclusions

Monitoring the frequency of observations below a reference state seems to be a useful way of identifying impacts in time series without control observations, suggesting that such data are appropriate for use in a feedback management system. As with other aspects of the ecosystem approach to fisheries, it is necessary to understand the limitations of, and risks associated with, the available methods. In this case, the risks can be readily evaluated, which should facilitate decision-making based on trade-offs. Because monitoring data are the main indication of the state of dependent species, it is appropriate to devise operational management objectives (including reference states and probabilities of observing them) in terms of these data.

Acknowledgements

This is a contribution to the British Antarctic Survey's *Ecosystems* programme, funded by the Natural Environment Research Council. The authors thank the Bird Island field assistants who collected the data on fur seal pup production, members of CCAMLR's Working Group on Ecosystem Monitoring and Management (WG-EMM) for useful discussion and Jefferson Hinke and Andrew Constable for helpful comments.

References

Agnew, D.J. 1997. Review: the CCAMLR Ecosystem Monitoring Program. *Ant. Sci.*, 9 (3): 235–242.

Barlow, K.E., I.L. Boyd, J.P. Croxall, K. Reid, I.J. Staniland and A.S. Brierley. 2002. Are penguins and seals in competition for Antarctic krill at South Georgia? *Mar. Biol.*, 140 (2): 205–213.

Butterworth, D.S. and R.B. Thomson. 1995. Possible effects of different levels of krill fishing on predators – some initial modelling attempts. *CCAMLR Science*, 2: 79–97.

Constable, A.J. 2001. The ecosystem approach to managing fisheries: achieving conservation objectives for predators of fished species. *CCAMLR Science*, 8: 37–64.

Constable, A.J. 2002. CCAMLR ecosystem management and monitoring: future work. *CCAMLR Science*, 9: 233–253.

Constable, A.J., W.K. de la Mare, D.J. Agnew, I. Everson and D. Miller. 2000. Managing fisheries to conserve the Antarctic marine ecosystem: practical implementation of the Convention on the conservation of Antarctic Marine Living Resources (CCAMLR). *ICES J. Mar. Sci.*, 57 (3): 778–791.

de la Mare, W.K. and A.J. Constable. 2000. Utilising data from ecosystem monitoring for managing fisheries: development of statistical summaries of indices arising from the CCAMLR ecosystem monitoring program. *CCAMLR Science*, 7: 101–117.

Forcada, J., P.N. Trathan, K. Reid. and E.J. Murphy. 2005. The effects of global climate variability in pup production of Antarctic fur seals. *Ecology*, 86 (9): 2408–2417.

Forcada, J., P.N. Trathan and E.J. Murphy. 2008. Life history buffering in Antarctic mammals and birds against changing patterns of climate and environmental variation. *Glob. Change Biol.*, 14 (11): 2473–2488.

Gunderson, L.H. 2000. Ecological resilience – in theory and application. *Annu. Rev. Ecol. Syst.*, 31: 425–439.

Hewitt, R.P., G. Watters, P.N. Trathan, J.P. Croxall, M.E. Goebel, D. Ramm, K. Reid, W.Z. Trivelpiece and J.L. Watkins. 2004. Options for allocating the precautionary catch limit of krill among small-scale management units in the Scotia Sea. *CCAMLR Science*, 11: 81–97.

Holling, C.S. 1986. Resilience of ecosystems; local surprise and global change. In: Clark, W.C. and R.E. Munn (Eds). *Sustainable Development of the Biosphere*. Cambridge University Press: 292–317.

Miller, D.G.M. 2002. Antarctic krill and ecosystem management – from Seattle to Siena. *CCAMLR Science*, 9: 175–212.

Plagányi, É.E. and D.S. Butterworth. 2007. A spatial multi-species operating model of the Antarctic Peninsula krill fishery and its impacts on land-breeding predators. Document *WG-SAM-07/12*. CCAMLR, Hobart, Australia.

Reid, K. and J.P. Croxall. 2001. Environmental response of upper trophic-level predators reveals a system change in an Antarctic marine ecosystem. *Proc. R. Soc. Lond. B*, 268: 377–384.

Reid, K, J.P. Croxall, D.R. Briggs and E.J. Murphy. 2005. Antarctic ecosystem monitoring: quantifying the response of ecosystem indicators to variability in Antarctic krill. *ICES J. Mar. Sci.*, 62 (3): 366–373.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution*, 43: 223–225.

Stewart-Oaten, A., W.W. Murdoch and K.R. Parker. 1986. Environmental impact assessment: 'pseudoreplication' in time? *Ecology*, 67 (4): 929–940.

Trathan, P.N., E.J. Murphy, J. Forcada, J.P. Croxall, K. Reid and S.E. Thorpe. 2006. Physical forcing in the southwest Atlantic: ecosystem control. In: Boyd, I.L., S. Wanless and C.J. Camphuysen (Eds). *Top Predators in Marine Ecosystems: their Role in Monitoring and Management*. Cambridge University Press: 28–45.

Watters, G.M., J.T. Hinke, K. Reid and S. Hill. 2006. KPFM2, be careful what you ask for – you just might get it. Document *WG-EMM-06/22*. CCAMLR, Hobart, Australia.

Watters, G.M., J.T. Hinke and S.L. Hill. 2008. A risk assessment to advise on strategies for subdividing a precautionary catch limit among small-scale management units during stage 1 of the staged development of the krill fishery in Subareas 48.1, 48.2 and 48.3. Document *WG-EMM-08/30*. CCAMLR, Hobart, Australia.

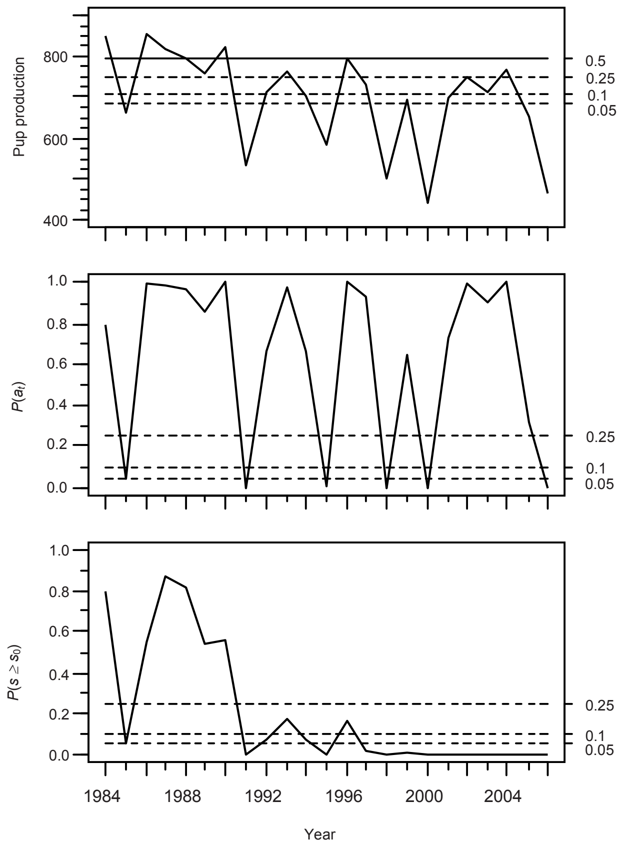

Figure 1: (a) Annual fur seal pup production at Bird Island, South Georgia, with the mean (solid horizontal line) and 25th, 10th and 5th percentiles (dashed horizontal lines) of the modelled unimpacted population. (b) Method 1: probability (in an unimpacted population) of each observation of pup production, given the cumulative number of observations. (c) Method 2: probability (in an unimpacted population) of the cumulative number of observations equal to or below the observed value, given the cumulative number of observations.

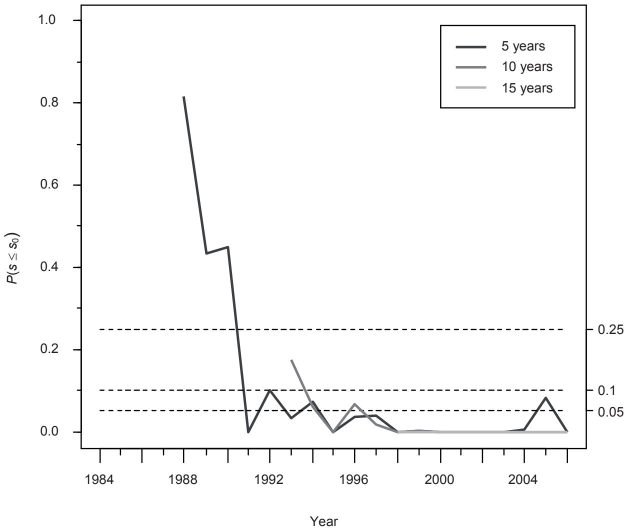

Figure 2: Method 3: probability (in an unimpacted population) of the cumulative number of observations equal to or below the observed pup production value, given the number of observations in the specified reference period.

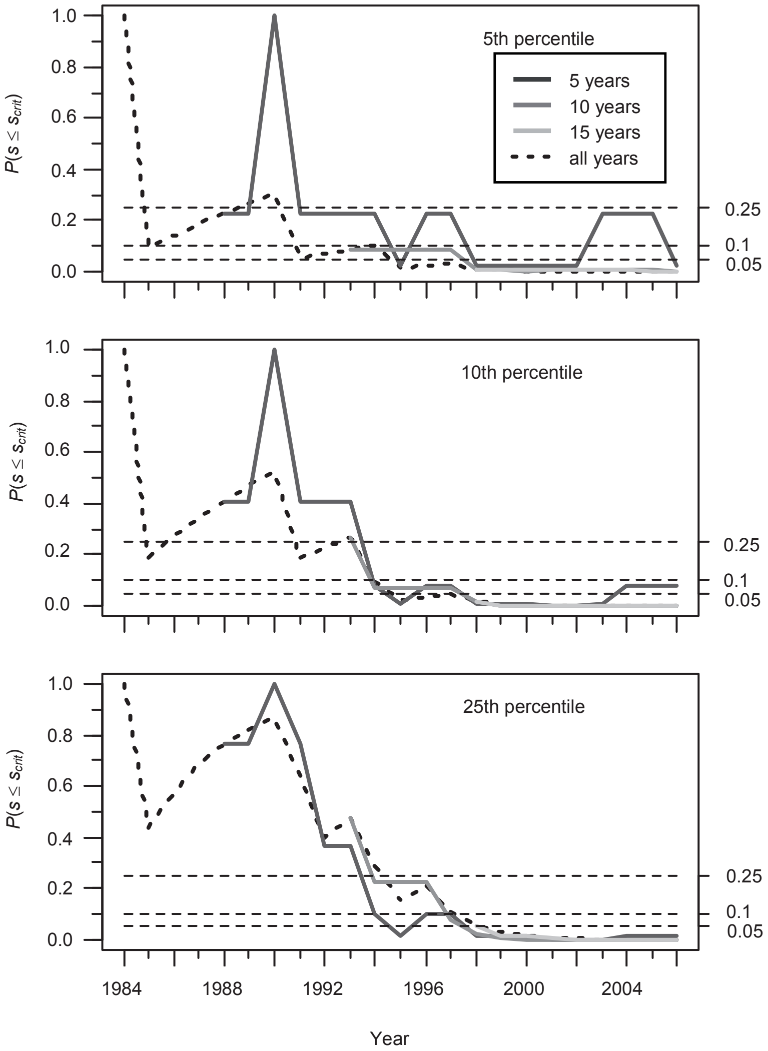

Figure 3: Method 4 with critical probabilities: probability (in an unimpacted population) of the cumulative number of observations equal to or below the specified percentile of the modelled unimpacted distribution of pup production values, given the number of observations in the specified reference period.

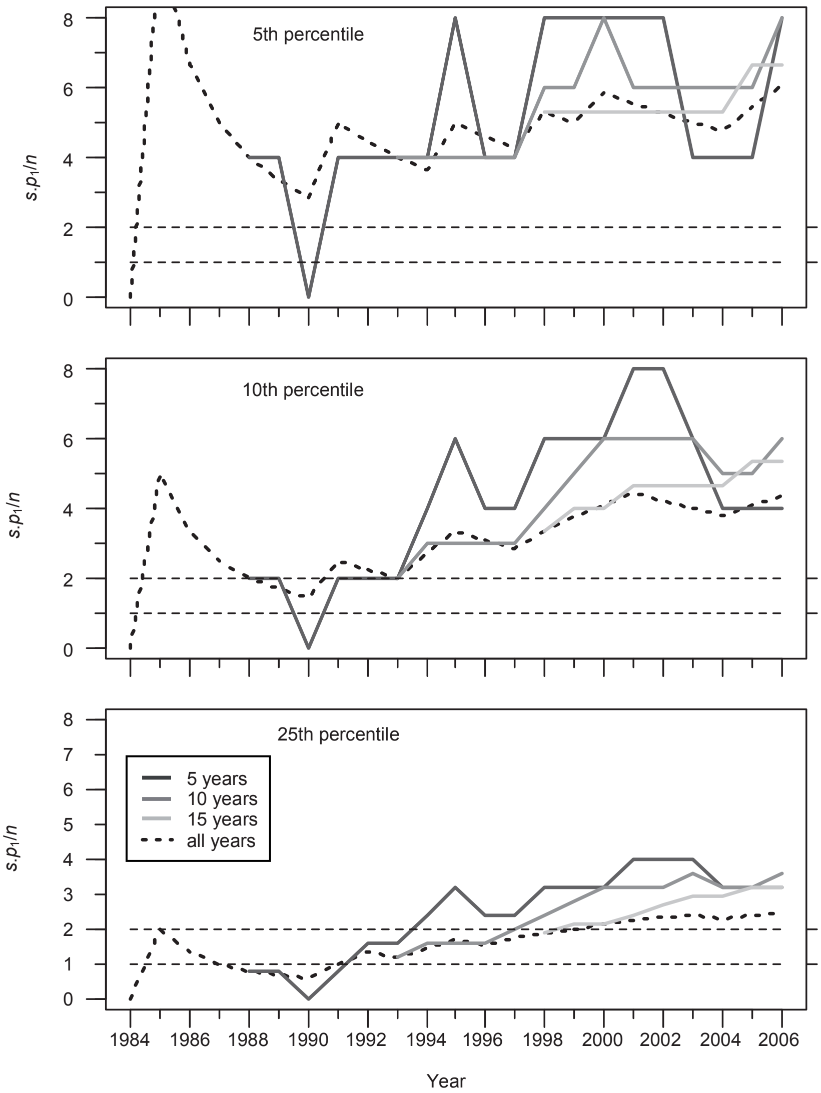

Figure 4: Method 4 with inference based on changes in frequency: the observed proportion of years in an n-year reference period in which pup production was equal to or below the specified percentile of the baseline distribution. Values are shown relative to p_1, the expected value in an unimpacted population. This figure uses the same data as Figure 3.

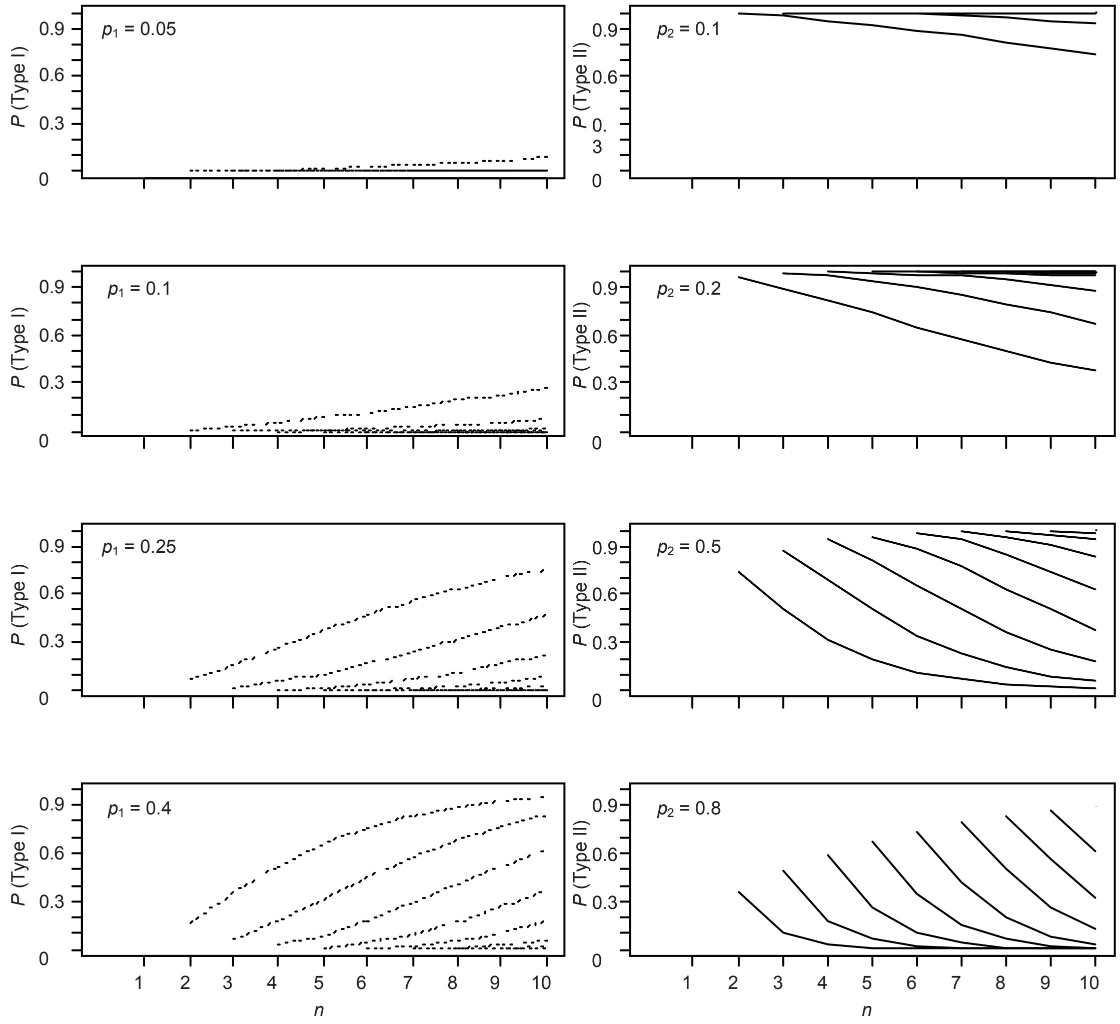

Figure 5: Probability of Type I and Type II error associated with inferring an increase in the probability of an event from p_1 to p_2 based on inference criteria consisting of an n-year reference period (x-axis) and a minimum number of years, s_{crit}, in n in which the event must be observed (lines in each panel left to right represent s_{crit} values of 2 to 10).

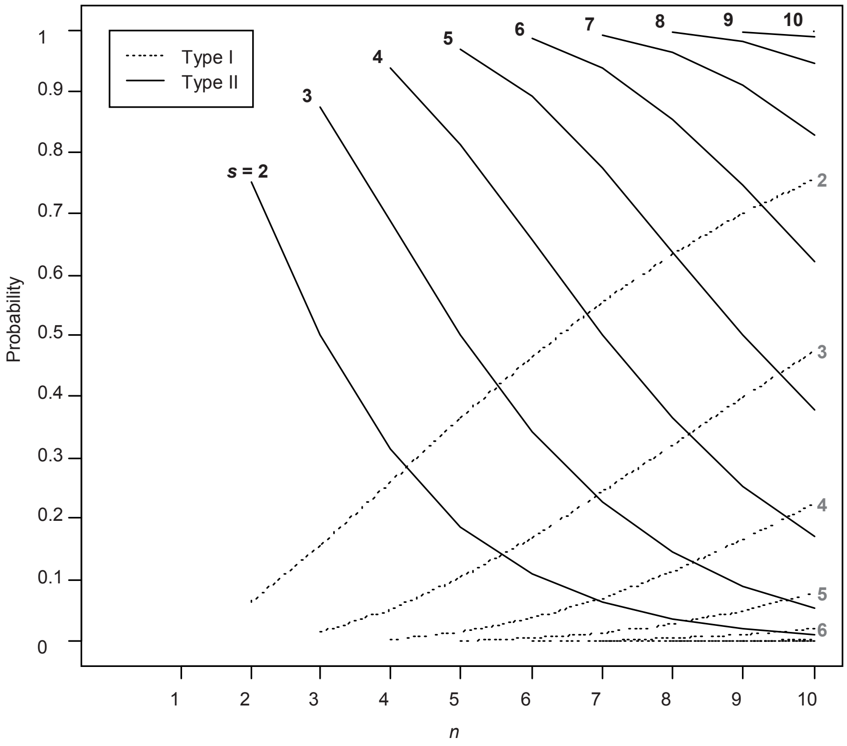

Figure 6: Trade-offs between the probabilities of Type I and Type II error when attempting to detect an increase in the probability of an event from 0.25 to 0.5 based on inference criteria consisting of an *n*-year reference period (x-axis) and a minimum number of years, s_{crit} (line labels), in *n* in which the event must be observed.

Liste des tableaux

Liste des figures

Список таблиц

Список рисунков

Рис. 5: Вероятности ошибок первого и второго рода, связанных с выводом об увеличении вероятности события с p_1 до p_2, на основе критериев выводов, включающих контрольный период n лет (ось x) и минимальное количество лет s_{crit}, в n, в течение которых данное событие должно наблюдаться (линии на каждом графике слева направо показывают значения s_{crit} от 2 до 10).

Рис. 6: Отрицательная корреляция между вероятностями ошибок первого и второго рода при попытке выявить возрастание вероятности события с 0.25 до 0.5 на основе критериев выводов, включающих контрольный период n (ось x) и минимальное количество лет s_{crit} (отмеченных на линиях) в n, в течение которых данное событие должно наблюдаться.

Lista de las tablas

Tabla 1: Frecuencia proporcional de falsos valores positivos en las simulaciones de cada método de detección del impacto (M1 a M4) en poblaciones "no afectadas" detectados mediante datos de la producción de cachorros de lobo fino, y tres valores diferentes de P_{crit} (la probabilidad crítica de una observación indicativa de un impacto).

Tabla 2: Frecuencia proporcional de falsos valores negativos en las simulaciones de cada método de detección del impacto (M1 a M4) en poblaciones "afectadas" detectados mediante datos de la producción de cachorros de lobo fino, y tres valores diferentes de P_{crit} (la probabilidad crítica de una observación indicativa de un impacto). El impacto se manifestó en la forma de un "cambio", una variación súbita en los parámetros especificados en el subtítulo, o en la forma de una "tendencia", un cambio lineal de estos parámetros a través de los 20 años del período de la simulación.

Lista de las figuras

Figura 1: (a) Producción anual de cachorros de lobo fino antártica en la Isla Bird, Georgia del Sur, mostrándose el promedio (línea horizontal sólida) y los percentilos 25, 10 y 5 (líneas horizontales entrecortadas) de la población simulada "no afectada". (b) Método 1: probabilidad (en una población no afectada) de cada observación de la producción de un cachorro, dado el número acumulativo de observaciones. (c) Método 2: probabilidad (en una población no afectada) del número acumulativo de observaciones igual o por debajo del valor observado, dado el número acumulativo de observaciones.

Figura 2: Método 3: probabilidad (en una población no afectada) del número acumulativo de observaciones igual o por debajo del valor observado de producción de cachorros, dado el número de observaciones en el período de referencia especificado.

Figura 3: Método 4 con probabilidades críticas: probabilidad (en una población no afectada) del número acumulativo de observaciones igual o por debajo del percentil especificado de la distribución de valores de la producción de cachorros en la población no afectada modelada, dado el número de observaciones en el período de referencia especificado.

Figura 4: Método 4 con inferencias basadas en cambios de la frecuencia: la proporción observada de años en un período de referencia de n-años en el cual la producción de cachorros fue igual o menor que el percentil especificado de la distribución básica. Se muestran valores relativos a p_1, el valor esperado en una población no afectada. En esta figura se utilizan los mismos datos que en la figura 3.

Figura 5: Probabilidad de un error Tipo I y Tipo II al inferir un aumento en la probabilidad de un suceso de p_1 a p_2 en base al criterio de un período de referencia de n-años (eje x) y un número mínimo de años (s_{crit}) de n en los cuales debe observarse el suceso (las líneas en cada cuadro de izquierda a derecha representan valores de s_{crit} de 2 a 10).

Figura 6: Equilibrio entre las probabilidades de un error Tipo I y Tipo II al tratar de detectar un aumento de la probabilidad de un suceso de 0.25 a 0.5 en base al criterio de un período de referencia de n-años (eje x) y un número mínimo de años (s_{crit}), indicado en cada línea) de n en los cuales debe observarse el evento.