CCAMLR Science, Vol. 7 (2000): 101–117

UTILISING DATA FROM ECOSYSTEM MONITORING FOR MANAGING FISHERIES: DEVELOPMENT OF STATISTICAL SUMMARIES OF INDICES ARISING FROM THE CCAMLR ECOSYSTEM MONITORING PROGRAM

W.K. de la Mare
Current Address: Marine and Ecological Research
17 Igluna Street, Kenmore 4069
Queensland, Australia

A.J. Constable
(Corresponding author)
Australian Antarctic Division
Channel Highway, Kingston 7050
Tasmania, Australia

Abstract

A potential method is presented for combining data collected as part of the CCAMLR ecosystem monitoring program (CEMP) into a single index for each of predator, prey and environmental parameters. The paper is divided into four main parts. The first part develops the proposed method of forming summary indices, which is based on the usual theory of multivariate statistics and takes into account the covariance between parameters. The second part reports on a Monte Carlo simulation study that examines the robustness of the indices to missing data and the degree of correlation between parameters. These trials show that missing values were unlikely to be a problem for time series of parameters that are highly correlated (>0.6). Criteria for inclusion of parameters in the indices are discussed when parameters are moderately or poorly correlated. The third part uses further simulation tests to examine the power of the statistical procedure adopted by WG-EMM in 1996 for identifying anomalies in CEMP parameters. The power of the procedure to detect anomalies was found to fall to low levels once more than a few anomalous values have appeared in the data. An alternative procedure, using estimates of the mean and variance of baseline time series, was found to have consistently better statistical power regardless of the accumulation of anomalies. The last section outlines an approach to the further development of CEMP indices for application in CCAMLR.

Résumé

Description d'une méthode susceptible de combiner les données collectées dans le cadre du Programme de contrôle de l'écosystème de la CCAMLR (CEMP) en un indice unique pour chacun des paramètres, à savoir prédateurs, proies et environnement. Ce document se divise en quatre parties. La première expose la méthode proposée pour former les indices récapitulatifs, laquelle est fondée sur la théorie habituelle des statistiques à variables multiples et tient compte de la covariance entre les paramètres. La seconde fait le compte rendu d'une étude de simulation de Monte Carlo qui examine la robustesse des indices face aux données manquantes et le degré de corrélation entre les paramètres. Ces expériences indiquent que les valeurs manquantes sont peu susceptibles de présenter de difficultés dans les séries chronologiques de paramètres grandement corrélés (>0,6). Les critères d'inclusion des paramètres dans les indices sont discutés pour les cas où les paramètres sont moyennement ou peu corrélés. La troisième partie a recours à de nouveaux tests par simulation pour examiner la puissance de la procédure statistique adoptée par le Groupe de travail sur le contrôle et la gestion de l'écosystème (WG-EMM) en 1996 pour identifier les anomalies dans les paramètres du CEMP. Au-delà de quelques valeurs anormales dans les données, il s'avère que la puissance de la procédure de détection des anomalies décroît fortement. Par contre, une autre procédure reposant sur les estimations de la moyenne et de la variance des séries chronologiques de base, fait en permanence preuve d'une meilleure puissance statistique quelle que soit l'accumulation des anomalies. La dernière section décrit les grandes lignes d'une approche de la mise au point d'indices du CEMP à l'intention de la CCAMLR.

Резюме

В статье предлагается метод комбинирования данных, собранных в рамках программы АНТКОМа по мониторингу экосистемы (CEMP), в единые индексы отдельно для хищников, добычи и окружающей среды. Статья разделена на 4 основных части. Первая часть дает описание предлагаемого метода расчета суммарных индексов, основанного на обычной теории многомерной статистики с учетом ковариации между параметрами. Во второй части сообщается о моделировании по методу Монте-Карло для выявления устойчивости индексов к отсутствию данных и степени корреляции между параметрами. В ходе анализа было обнаружено, что если временные ряды параметров сильно коррелированы (>0.6), то отсутствие значений обычно не создает проблемы. Если корреляция между параметрами средняя или низкая, то обсуждаются критерии включения параметров в индексы. В третьей части дополнительные модельные расчеты используются для исследования статистической мощности методики, принятой в 1996 г. Рабочей группой по экосистемному мониторингу и управлению (WG-EMM) для идентификации аномалий в параметрах CEMP. Было обнаружено, что способность методики выявлять аномалии снижается, если данные содержат больше, чем несколько аномальных значений. Было также показано, что альтернативная методика, использующая оценки среднего и дисперсии временных рядов за базовый период, имеет бóльшую статистическую мощность вне зависимости от накопления аномалий. В последней части намечен подход к дальнейшей разработке индексов CEMP для применения АНТКОМом.

Resumen

Se presenta un posible método para combinar los datos recopilados por el programa de seguimiento del ecosistema de la CCRVMA (CEMP) para cada uno de los parámetros, en un índice único para los depredadores, presas y medio ambiente. El documento consta de cuatro partes principales. La primera parte desarrolla el método propuesto de formular índices resumidos, basado en la teoría general de las estadísticas de múltiples variables, y toma en cuenta la covariancia entre los parámetros. La segunda parte presenta un estudio de simulación Monte Carlo donde se examina la validez estadística de los índices con respecto a los datos omitidos y el grado de correlación entre los parámetros. Mediante estas pruebas se demostró que era muy poco probable que los valores omitidos afectaran negativamente a las series cronológicas de los parámetros que tienen una correlación muy alta (>0,6). Se considera el criterio de inclusión de parámetros a los índices cuando la correlación entre éstos es moderada o baja. La tercera parte se vale de pruebas de simulación adicionales para examinar la potencia estadística del procedimiento adoptado por el grupo de trabajo para el seguimiento y ordenación del ecosistema (WG-EMM) en 1996 para identificar anomalías en los parámetros del CEMP. Se detectó una baja capacidad del procedimiento para detectar anomalías cuando aparecieron unos cuantos valores anómalos en los datos. Otro procedimiento, que utiliza estimaciones del promedio y la variancia de las series cronológicas básicas demostró más consistencia estadística a pesar de una acumulación de valores anómalos. La última sección describe un enfoque para seguir desarrollando los índices CEMP para su aplicación en la CCRVMA.

Keywords: monitoring, anomaly, statistical power, robustness, management strategy evaluation, feedback management, CCAMLR

INTRODUCTION

The CCAMLR Ecosystem Monitoring Program (CEMP) involves a diverse range of measures designed to indicate the status of predator and prey populations along with features of the physical environment. One of the tasks of the CCAMLR Working Group on Ecosystem Monitoring and Management (WG-EMM) is to examine this mass of monitoring data and to advise the Scientific Committee on the current state of Antarctic ecosystems. However, this task is made difficult because of:

(i) the large quantity of data over a large number of parameters; and

(ii) the heterogeneous composition of the parameters included in the data over years.

With regard to (i), the approach so far has been to attempt a 'human integration' of the mass of data to give a scientific judgement of the status of the various ecosystems being monitored. However, 'human integration' is very difficult once more than about five variables are included. Therefore, it would be highly desirable to condense the diverse parameters into a much smaller set of indices. As indicated above, three indices of interest to CCAMLR would comprise, in the first instance, one each for the status of the predators, their prey and the environment. With regard to (ii) a method is required that allows for the three indices to be easily interpretable, that is comparable from year to year, even though different sets of parameters may contribute to the indices in different years.

Consideration needs to be given to the effects on this index of the addition of new parameters over time and the possibility that some parameters may not be measured in a given year due to logistic constraints and the difficulties of conducting field work in Antarctica. In future, new parameters may be added and others phased out. These possibilities may limit the types of parameters that could be included in an index. The CEMP indices are designed to help determine when fishing for a prey species, notably krill, is affecting predators, allowing for changes to predators that may arise from natural changes or variation in the environment. In this case, the status of the predator index needs to reflect the response of predators to the status of prey or the status of the environment. If the temporal trends in a single index are to be used as a reliable indicator of the status of that component of the system (predators, prey or physical environment) then the index needs to be robust against the loss of data (missing values), either as a series of years in which a parameter is not measured (prior to or after its use) or as random occurrences. Ideally the value of the index in a given year should be independent of the specific parameters present in that year. In this respect, the robustness of the index is dependent on that subset of parameters being representative of all the parameters for that year and on the underlying estimates of parameters in the statistic being unbiased.

Currently, WG-EMM is reviewing how land-based marine predators respond to variations in abundance of krill. In this context, the robustness of the index can be considered in two parts: (i) how well the index of predator parameters relates to krill abundance; and (ii) how well the index reflects the overall status of the predator parameters when data are missing. The examination of robustness presented here only investigates the second part.

This paper develops a possible approach for combining CEMP variables into a single index that may assist CCAMLR in making decisions on the effects of fishing on predator populations and reports on a Monte Carlo simulation study that investigates the robustness of the statistic to missing data. The third section of the paper assesses approaches to identifying when values (years) of the index in a temporal sequence are divergent from a baseline norm; and years in which causes for such divergence need to be determined.

Finally, steps for the further development of the CEMP indices are identified.

A POSSIBLE APPROACH FOR COMBINING CEMP VARIABLES

The standard theory of multivariate statistics provides one possible method of combining the various CEMP parameters into the three indices suggested above. The method would involve transforming and standardising the various parameters along the lines adopted by WG-EMM in 1996. That is, transforming each parameter in order to obtain roughly a standard normal distribution. The parameter values could then be simply added together and restandardised using the estimated standard deviation for the sum. Clearly, the values should be standardised also with respect to sign, for example, positive values indicating better-than-average conditions for the predator.

The statistical formalism is easily represented as matrix algebra (a clear exposition of the theory is given in Mardia et al., 1979). From the data we need to estimate the covariance matrix. If the multivariate data have already been standardised, the covariance matrix of the standardised parameters is identical to the correlation matrix. The covariance matrix needs to include estimates of the covariance between all the parameters that are to be combined into the index. The sum of the various parameters in a given year is given by:

$$y_t = \mathbf{a}'\mathbf{x}_t \tag{1}$$

where $\mathbf{x}_t$ is the vector of observed values of the parameters in year t, and $\mathbf{a}$ is a vector which has a unit value for each of the variables in $\mathbf{x}_t$ which were actually observed in year t (boldface symbols represent vectors or matrices, lower case for vectors, upper case for matrices). Both $\mathbf{x}_t$ and $\mathbf{a}$ are of length n, where n is the total number of variables that are available to be combined into an index. The sum is a scalar value.

The variance of the sum is also a scalar given by:

$$S_t = \mathbf{a}\mathbf{S}\mathbf{a}' \tag{2}$$

Table 1: Means and coefficients of variation for each of eight parameters used in all simulation trials to examine criteria for the inclusion of parameters in the index.

Predator Parameter	Mean	CV
1	3.06	0.18
2	9.92	0.50
3	5.68	0.41
4	43.52	0.41
5	20.67	0.37
6	42.37	0.29
7	37.99	0.34
8	29.90	0.16

Table 2: Correlation matrices used in simulation trials to examine criteria for the inclusion of parameters in the index.

Parameter	1	2	3	4	5	6	7	8
(i) Highly correlated ($0.6 < r < 0.95$)								
1	1	0.7301	0.8454	0.6797	0.6642	0.6613	0.8399	0.7626
2	0.7301	1	0.8467	0.6096	0.8438	0.6728	0.6103	0.8585
3	0.8454	0.8467	1	0.7385	0.8181	0.6929	0.86	0.7966
4	0.6797	0.6096	0.7385	1	0.7001	0.6534	0.6324	0.7641
5	0.6642	0.8438	0.8181	0.7001	1	0.8566	0.7484	0.8461
6	0.6613	0.6728	0.6929	0.6534	0.8566	1	0.6302	0.7903
7	0.8399	0.6103	0.86	0.6324	0.7484	0.6302	1	0.6956
8	0.7626	0.8585	0.7966	0.7641	0.8461	0.7903	0.6956	1
(ii) Moderately correlated ($0.3 < r < 0.6$)								
1	1	0.3703	0.5549	0.5116	0.3158	0.5696	0.4021	0.5233
2	0.3703	1	0.3413	0.4594	0.3702	0.3293	0.5273	0.5536
3	0.5549	0.3413	1	0.465	0.4073	0.5912	0.5699	0.5
4	0.5116	0.4594	0.465	1	0.4766	0.4004	0.4523	0.3032
5	0.3158	0.3702	0.4073	0.4766	1	0.4651	0.5738	0.3822
6	0.5696	0.3293	0.5912	0.4004	0.4651	1	0.4127	0.4635
7	0.4021	0.5273	0.5699	0.4523	0.5738	0.4127	1	0.5703
8	0.5233	0.5536	0.5	0.3032	0.3822	0.4635	0.5703	1
(iii) Poorly correlated ($0 < r < 0.3$)								
1	1	0.07639	0.2585	0.2671	0.265	0.07214	0.1329	0.004856
2	0.07639	1	0.2127	0.2007	0.04753	0.05666	0.2996	0.04783
3	0.2585	0.2127	1	0.2786	0.2572	0.08118	0.2015	0.2993
4	0.2671	0.2007	0.2786	1	0.1881	0.2092	0.0877	0.2992
5	0.265	0.04753	0.2572	0.1881	1	0.2567	0.1448	0.1936
6	0.07214	0.05666	0.08118	0.2092	0.2567	1	0.1924	0.2637
7	0.1329	0.2996	0.2015	0.0877	0.1448	0.1924	1	0.2597
8	0.004856	0.04783	0.2993	0.2992	0.1936	0.2637	0.2597	1
(iv) A mixture of four highly correlated parameters combined with four poorly correlated parameters.								
1	1	0.7301	0.8454	0.6797	0.1642	0.2613	0.1399	0.3626
2	0.7301	1	0.8467	0.6096	0.2438	0.2728	0.1103	0.2585
3	0.8454	0.8467	1	0.7385	0.2181	0.1929	0.0600	0.1966
4	0.6797	0.6096	0.7385	1	0.0001	0.1534	0.1324	0.0641
5	0.1642	0.2438	0.2181	0.0001	1	0.2566	0.2484	0.1461
6	0.2613	0.2728	0.1929	0.1534	0.2566	1	0.1302	0.3903
7	0.1399	0.1103	0.0600	0.1324	0.2484	0.1302	1	0.1956
8	0.3626	0.2585	0.1966	0.0641	0.1461	0.3903	0.1956	1
(v) Highly correlated parameters, two of which are however highly negatively correlated with the rest.								
1	1	0.7301	0.8454	0.6797	0.6642	-0.6613	-0.8399	-0.7626
2	0.7301	1	0.8467	0.6096	0.8438	-0.6728	-0.6103	-0.8585
3	0.8454	0.8467	1	0.7385	0.8181	-0.6929	-0.86	-0.7966
4	0.6797	0.6096	0.7385	1	0.7001	-0.6534	-0.6324	-0.7641
5	0.6642	0.8438	0.8181	0.7001	1	-0.8566	-0.7484	-0.8461
6	-0.6613	-0.6728	-0.6929	-0.6534	-0.8566	1	0.6302	0.7903
7	-0.8399	-0.6103	-0.86	-0.6324	-0.7484	0.6302	1	0.6956
8	-0.7626	-0.8585	-0.7966	-0.7641	-0.8461	0.7903	0.6956	1

where **S** is the covariance matrix of the standardised data. The standardised index then for year t is simply:

$$z_t = \frac{y_t}{\sqrt{S_t}} \quad (3)$$

If the standardising transformations applied to the raw data produced observations with an expected value of zero, then this index also has an expected value of 0 with a standard deviation of 1. Thus, provided the covariance matrix can be estimated (the covariance matrix, **S**, must be a positive, semi-definite matrix) and that the standardisation is reasonable, the index does not require all variables to be measured in each year. Each index can then be screened for anomalies by an appropriate statistical procedure. The three indices could easily be presented graphically.

ROBUSTNESS OF THE INDEX TO MISSING DATA

Monte Carlo simulations were used to examine a number of criteria for including parameters in the index. These were undertaken using the Mathcad computer package. In this analysis, the test of robustness of the index is determined by how much a time series of this index based on a dataset in which values are missing might potentially deviate from that same dataset but with all the values recorded, i.e. ideally the values of the index based on a dataset with missing data would be the same as the index from the full dataset. A measure of this is the correlation of the index from the full dataset with that from the subset. Monte Carlo simulations are used to determine the range of correlations that might be expected under a given scenario of missing values; the range indicates the robustness of the index. Results of each trial consisting of 1 000 replicates are presented graphically, showing the median and 5th and 95th percentiles of these correlations.

Eight predator parameters were used in every trial in this simulation exercise. This is the number currently used by WG-EMM to generate an index for Bird Island in the South Atlantic. Their means and standard deviations were kept the same in each trial. Each parameter was normally distributed. The means and coefficients of variation for these were randomly chosen from uniform distributions to reflect the variability that might be observed in these parameters in practice. These are shown in Table 1.

The robustness of the index was examined using five different scenarios for the correlation between parameters:

(i) highly correlated ($0.6 < r < 0.95$);

(ii) moderately correlated ($0.3 < r < 0.6$);

(iii) poorly correlated ($0 < r < 0.3$);

(iv) a mixture of four highly correlated parameters combined with four poorly correlated parameters; and

(v) highly correlated variables but with two being highly negatively correlated with the rest.

The correlation matrices are shown in Table 2. Each matrix was generated by randomly selecting values from within the range specified for the trial with a single condition that their resulting eigenvalues were all positive.

Within a trial, a set of values for each parameter in each year was randomly generated using the true means, standard deviations and correlation matrix.

Calculation of the index depends on the estimates of the correlations between parameters. Each of the estimated correlations within the estimated correlation matrix will be influenced by the number of years in which a respective pair of parameters are monitored together. Fewer pairwise occurrences in the data will potentially reduce the precision of the corresponding element in the correlation matrix.

This paper examines a simple conceptual form of this problem by having a period when all the parameters are present (i.e. a full dataset) and determining the effect on the index of different lengths of time that the full dataset is available relative to the overall time span of the index. Three proportions of full dataset availability were considered – 0.25, 0.5 and 0.75 of the full time series.

Similarly, the proportion of parameters missing in a given year will influence the estimate of the index for that year. In this respect, trials were undertaken with 0.25, 0.5 and 0.75 of the parameters missing in the years when data were missing.

Two models of missing data were investigated. The first was to have the same parameters missing from the years in which data were missing, representing the phasing in or out of parameters being monitored. The second was to have randomly selected parameters missing in the years in which data were missing.

The length of the time series of missing values may also be important. Consistent with the time frame of the current CEMP Program, two lengths of time were investigated: 10 and 20 years.

The correlation matrix was estimated in two ways. The first method was to use all the available data to generate pairwise correlations. The second method was to restrict the pairwise correlations to the period when all parameters were present, i.e. all pairwise correlations had the same number of observations. This was undertaken to examine whether the robustness of the index was diminished by reducing the dataset available for generating the correlation matrix.

Lastly, some estimates of the pairwise correlations can result in a correlation matrix that is not positive and semi-definite, i.e. some eigenvalues of the matrix are negative. To overcome this, the procedure described in Huseby et al. (1980) was applied to provide the closest positive semi-definite matrix to the indefinite matrix estimated from the data. This smoothing procedure requires that a correlation matrix be checked initially for being a positive semi-definite matrix. In cases where eigenvalues are less than or equal to zero, then these values are made to be a small positive (10^{-5}) and a modified correlation matrix be reconstructed using the formula

$$\sum_{\lambda_h > 0} \lambda_h \gamma_h \gamma'_h \qquad (4)$$

where λ_h is the adjusted eigenvalues and γ_h represents the respective eigenvectors from the S matrix above. The modified correlation matrix was then used in place of S to estimate the index time series. The number of times this arose during the simulations was recorded.

The results of these trials are presented in Figures 1 to 4.

As would be expected, the index is relatively insensitive to these scenarios of missing data when the positive correlation between all parameters is high (between 0.6 and 0.9). In all cases presented here, 95% of observed correlations between the time series of the index with all parameters present compared to the same series with missing data were greater than 0.85. Notably, a mixture of highly correlated and poorly correlated parameters can diminish the performance of the index.

The performance of the index was better when more parameters were present in years when data were missing. Also, longer time series improved the performance of the index in all cases. This was not related to the actual minimum number of years in which parameters were measured but to the total length of the time series (20 years was better than 10) and the proportion of the time series in which all parameters were estimated (0.75 of the total length was better than 0.25). However, the increase from 10- to 20-year time series resulted in little change in the median performance but in a reduction in the range between the confidence intervals.

The index was more sensitive to random missing data than it was to having the same parameter missing each year (i.e. phasing in new parameters or phasing out old ones). In these circumstances, a longer time series of a few parameters is likely to better reflect the interannual variability in the index than having a number of time series broken at random by missing years. The difference between such cases is only very small, even when there is a mixture of highly correlated and poorly correlated parameters.

The effect of the presence of negatively correlated parameters in the dataset was of greater significance. In the scenario examined here, two parameters were highly positively correlated and each of these was highly negatively correlated with the other six parameters that were highly positively correlated. The presence of negatively correlated parameters reduced the performance of the index substantially. This could be resolved by inverting the sign of negatively correlated parameters such that all the correlations became positive.

Smoothing of the correlation matrix from an indefinite to a positive semi-definite matrix occurred in most treatments. Most of the treatments involving 10-year time series required smoothing in more than 90% of the replicates. Smoothing occurred in less than 50% of replicates in some of the 20-year time series for which few data were missing.

The trials in which the correlation matrix was generated from only the years in which all parameters were present shows that the method of generating the correlation matrix is of less importance than choice of parameters to be included in the index when data are missing from the time series.

If the index based on a chosen set of parameters, but with missing data in the time series, is considered robust when the index has a 95% chance of having a 0.9 correlation with the underlying index of all parameters, then the following characteristics of the dataset may need to apply. If all the parameters are highly correlated (>0.6) then the index would be considered robust in all treatments of missing data presented here for which at least 50% (4) of the parameters are present for the full time series. In most treatments, presence of 25% (2) of the parameters in the full time series would be sufficient to satisfy this test. This means that missing data would be of little concern in an index based on highly correlated parameters.

In reality, most parameters are not highly correlated. If the parameters have a correlation of at least 0.3, then the time series needs to be greater than 10 years before an index would satisfy the test for robustness indicated above. For a dataset that extends to 20 years, at least half the parameters need to be present for the full time series and the remainder should be present for not less than 50% of the time.

Further work is required to assess the effects of sampling variability on the performance of the index. Also, this work needs to account for the characteristics of the current CEMP database and to assess which of the available parameters might be included in an index. An important task is to identify which characteristics of the system need to be summarised to facilitate decisions governing a fishery. Currently, there is a recognised need to identify how predators may change with respect to changes in krill abundance. This is the basis for choosing many of the parameters being monitored at present. If parameters are poorly correlated then it could be argued they are responding to factors other than those related to krill abundance, environment or otherwise. If that is the case, work needs to be done to identify which krill and environmental factors are important in the management context (e.g. small- or large-scale estimates of krill abundance) and, for each factor considered important, which predator parameters could be used in an index to monitor effectively a predator response to changes in the factor. Finally, the latitude given to accepting some parameters in the index needs to be evaluated in the wider context of making correct decisions based on the index despite uncertainties or variability in the behaviour of those parameters.

IDENTIFICATION OF ANOMALIES

At its 1996 meeting, WG-EMM adopted a procedure for identifying anomalies using the following statistic:

$$z_i = \frac{(x_i - M)}{s} \tag{5}$$

where x_i is the observed value of an index variable in year i, M and s are the sample mean and variance respectively of the time series of x_i, denoted $\mathbf{x}$. The values of z_i are considered to be anomalies if they exceed a critical value that depends on the length of $\mathbf{x}$. The critical values were calculated by simulation from the value such that 5% of observations in a series would exceed the critical value (SC-CAMLR, 1996). An important characteristic of any statistic to examine is its statistical power. In the terminology of hypothesis testing the critical values used have been fixed in terms of a probability of making a type I error $\alpha = 0.05$. A type I error is rejecting the null hypothesis when it is true. The other important consideration is the probability of making a type II error, i.e. the probability β of accepting the null hypothesis when it is false. The evaluation of β is carried out with respect to a specific alternative hypothesis. The power of a statistical test is defined as $1 - \beta$.

The power of the statistical test used by WG-EMM is evaluated by simulation for two forms of alternative hypothesis with data series of up to 20 and 30 years. In both forms, the index is drawn from a random normal distribution initially with mean = 1.0, which shifts to a new value after either 10 or 20 years to have value = 2.0. The variance of the distribution with mean = 1.0 is 0.3 and for the distribution with mean = 2.0 is in one case 0.3 and in the second case is 0.6. That is, one case is a constant variance:

$$\begin{aligned} X &= \mathrm{N}(1,0.3) \quad ;t < t_{shift} \\ &= \mathrm{N}(2,0.3) \quad ;t \geq t_{shift} \end{aligned}$$

and the other is a constant coefficient of variation (CV):

$$\begin{aligned} X &= \mathrm{N}(1,0.3) \quad ;t < t_{shift} \\ &= \mathrm{N}(2,0.6) \quad ;t \geq t_{shift} \end{aligned}$$

where $\mathrm{N}(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ. The two sets of trials consist of one in which the value of t_{shift} is fixed at 10 years and another in which it is fixed at 20 years. The data series is continued with additional shifted values for a further series

Table 3: Power of the statistical procedure for identifying anomalies adopted by WG-EMM in 1996. The length of the data without anomalies is fixed at either 10 or 20 years. Between 1 and 10 years of anomalous data are added to the series when calculating the statistical power.

Number of Additional Years with Anomalies	Power			
	Anomaly-free Length = 10 Years		Anomaly-free Length = 20 Years	
	Constant Variance	Constant CV	Constant Variance	Constant CV
1	0.8036	0.6764	0.8724	0.7234
2	0.4297	0.4535	0.6952	0.5954
3	0.2277	0.3103	0.5353	0.4927
4	0.1252	0.2281	0.4022	0.4065
5	0.0712	0.1758	0.2963	0.3480
6	0.0454	0.1428	0.2230	0.2979
7	0.0306	0.1153	0.1669	0.2586
8	0.0212	0.0966	0.1301	0.2261
9	0.0162	0.0862	0.0994	0.2011
10	0.0142	0.0778	0.0755	0.1730

Table 4: Power of a statistical procedure for the identification of anomalies using baseline mean and variance estimates.

Years Post-baseline	Critical z Value	H_0 True	Power	
			Constant Variance	Constant CV
Baseline length = 10 years				
1	2.4007	0.0484	0.8004	0.6845
2	2.3584	0.0550	0.8140	0.7002
5	2.3502	0.0540	0.8085	0.6945
Baseline length = 20 years				
1	2.1578	0.0472	0.8634	0.7142
2	2.1506	0.0515	0.8655	0.7301
5	2.1416	0.0502	0.8705	0.7264

of years between 1 and 10. The conclusions about the power of the procedure do not depend on the fact that the shifted values are accumulated at the end of the data series because the statistic in equation (5) does not depend on the order of the data. The power is calculated as the probability of detecting each of the shifted values in those years where they are present.

The results given in Table 3 show that, even though the effect size is large (>3 standard deviations), the power to identify anomalies is only substantial when there are a few anomalies. As the proportion of anomalies increases, the power of the procedure falls to quite low levels. This is because the sample variance is used in calculating the values of z_i. A shift in the data, with or without an increase in variability, results in an increase in s, and hence a decrease in the values of z_i. Values that had earlier been classified as anomalies may no longer be so classified with the addition of further anomalous data. Consequently, the probability of classifying an observation as an anomaly can fall below the $\alpha = 0.05$ probability level.

The obvious alternative for classifying anomalies is to calculate the mean and variance for use in equation (5) using baseline data. The rationale and procedure for developing the datasets in this test were the same as those used in the above test. Thus, the baseline data series were derived randomly from a normal distribution, N(1, 0.3). Table 4 shows the statistical power for identifying anomalies with a procedure using baseline data series with lengths of 10 years and 20 years. Data points are classified as anomalies in a data series that follows after the baseline data. The critical values for identifying anomalies were calculated by simulation from the value such that 5% of observations in a series would exceed the critical value. As before, the anomalies are generated from a shift in the distribution by +1.0, for the two cases of constant variance and constant CV. Only three lengths of post-baseline data series were used, 1, 2 and 5 years because it was clear that this procedure is unaffected by the accumulation of anomalies.

The power shown in Table 4 is the probability of identifying each anomaly in the post-baseline data. Clearly this is generally a more powerful statistical procedure than the one adopted by WG-EMM in 1996. Interestingly, even a 10 year baseline provides for quite good power. Increasing the baseline to 20 years leads to only a moderate increase in power. As would be expected, the powers of the two procedures are effectively identical when there is only one anomalous value.

These analyses are only illustrative, and do not necessarily reflect the power of the methods for actual CEMP indices. The important point however is that the baseline method is substantially more powerful than the method adopted in 1996. These examples show that it is important to evaluate the performance of methods in terms of the probability that they correctly detect the events of interest. The baseline can be selected either as comprising a series of contiguous years, or by excluding data with obvious anomalies.

SUGGESTED APPROACH FOR THE DEVELOPMENT OF CEMP INDICES

CEMP is a voluntary program involving many nations. The effort is spread over many species, a number of parameters for each species and many sites around Antarctica (Agnew, 1997). There is a need to identify methods that synthesise the data and incorporate it into advice to the CCAMLR Commission for decisions about managing fisheries, particularly the krill fishery (Constable, 1992). Notably, the Commission needs to use information on the target species, the physical environment and predators of the target species to help identify appropriate management action required to meet the objectives of the Convention. A consequence of the voluntary arrangement in CEMP is that many data are missing in these time series. Any methods used to summarise the data and to arrive at decisions need to be robust against missing data, such that the summaries are representative of the status of the system in a given year and that the decisions made are reliable.

This paper provides a method for summarising many parameters into a single statistic. The results of the simulations show how decisions should be made as to which parameters to consider for inclusion in the index, despite missing data in the overall time series. The method appears robust provided that the parameters have a reasonable degree of correlation, and this should focus attention on determining how much data is required to estimate a reliable correlation between parameters. Another question concerns the length of a baseline dataset which is necessary to determine when the system has changed in ecologically significant ways, either as an annual event or as a trend over time.

The overall value of the index is yet to be tested under a variety of plausible ecological scenarios to determine whether it would be useful to managers. Such testing needs to be undertaken prospectively, through simulation testing, before applying it in the field. In this way, the overall utility of CEMP can be evaluated in a simulation framework to identify which issues are important for its use in a formal management framework for Antarctic fisheries.

An outline of a number of propositions to be explored in the further development of CEMP indices is given below. In the beginning, effort would best be directed at the region that has the most comprehensive of the CEMP datasets. This dataset can then be used to examine the basic statistical properties of the data pertinent to building some form of combined index, either as suggested above, or such other schemes as may seem appropriate. The following tasks need to be undertaken:

(i) Definition of the classes of behaviour in the indices that we wish to detect.

The obvious candidates are: changing variability (range); trends; shifts; and changes in the frequency of anomalies.

(ii) Selection of the normalising transformations required for the various parameters.

This step is already in progress.

(iii) Selection of a baseline dataset.

This dataset will be used to estimate the centring matrix for the multivariate data and the variances to be used in transforming the data into an approximately standard multinormal distribution. From these data the covariance/correlation matrix can be estimated. As a stopgap, any missing correlation coefficients could be filled in from other data series if necessary. The parameters within an index should all be positively correlated. If they are not, we need to reconsider their role in the formation of an index. Examine the data for serial correlation.

(iv) Examination of the statistical properties of the proposed index including:

(a) detection of anomalies;

(b) effects of missing data in scenarios other than those presented here;

(c) effects on the index of variability due to sampling versus that due to intrinsic variability;

(d) effects of serial correlation;

(e) effects of non-linear correlations between parameters; and

(f) plotting of indices in the form of 'control charts'. Two types of charts could be examined:

- based on the index, with critical bounds – useful for displaying anomalies; and

- based on a renormalised cumulative sum of the indices – a 'cusum' chart, useful for detecting the effects of a systematic shift in mean level. A randomisation procedure could be tested for the identification of drift.

(v) Examination of the power of the indices to detect phenomena of interest:

(a) consider the appropriate levels of the probability of making type I and type II errors – type II errors may have more important consequences than type I;

(b) effect of the length and stability of baseline data;

(c) consider whether all parameters should have their normal range defined purely statistically, some parameters may have their anomalies defined on biological grounds;

(d) look for correlation between the three indices (predator–environment–prey);

(e) examine possible improvements in the design of the CEMP Program to increase the power of the indices. This would include the exploration of experimental designs such as before-after-control-impact designs (Constable, 1992); and

(f) examine how the indices could be included in the development of quantitative management advice (see Constable, 1992 for further discussion).

This represents a substantial program of work, but it should be feasible to make considerable progress over the next few years.

ACKNOWLEDGEMENTS

Thanks to the three anonymous referees for taking the time to comment on the manuscript and to Bryan Manly for highlighting the procedure for smoothing indefinite correlation matrices. Many thanks to Eugene Sabourenkov and Genevieve Tanner for their patience and editing. Thanks also to Ian Ball for constructive discussions in the final phase of this work. This paper has been updated from the original working papers submitted to WG-EMM, *WG-EMM-Stats-97/8* and *WG-EMM-98/45*.

REFERENCES

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

Constable, A.J. 1992. CCAMLR ecosystem monitoring and a feedback management procedure for krill. In: *Selected Scientific Papers, 1992 (SC-CAMLR-SSP/9)*. CCAMLR, Hobart, Australia: 345–350.

Huseby, J.R., N.C. Schwertman and D.M. Allen. 1980. Computation of the mean vector and dispersion matrix for incomplete multivariate data. *Communications in Statistics – Simulation and Computation*, B9: 301–309.

Mardia, K.V., J.T. Kent and J.M. Bibby. 1979. *Multivariate Analysis*. Academic Press, London: 521 pp.

SC-CAMLR. 1996. Report of the Working Group on Ecosystem Monitoring and Management. In: *Report of the Fifteenth Meeting of the Scientific Committee (SC-CAMLR-XV)*, Annex 4. CCAMLR, Hobart, Australia: 123–299.

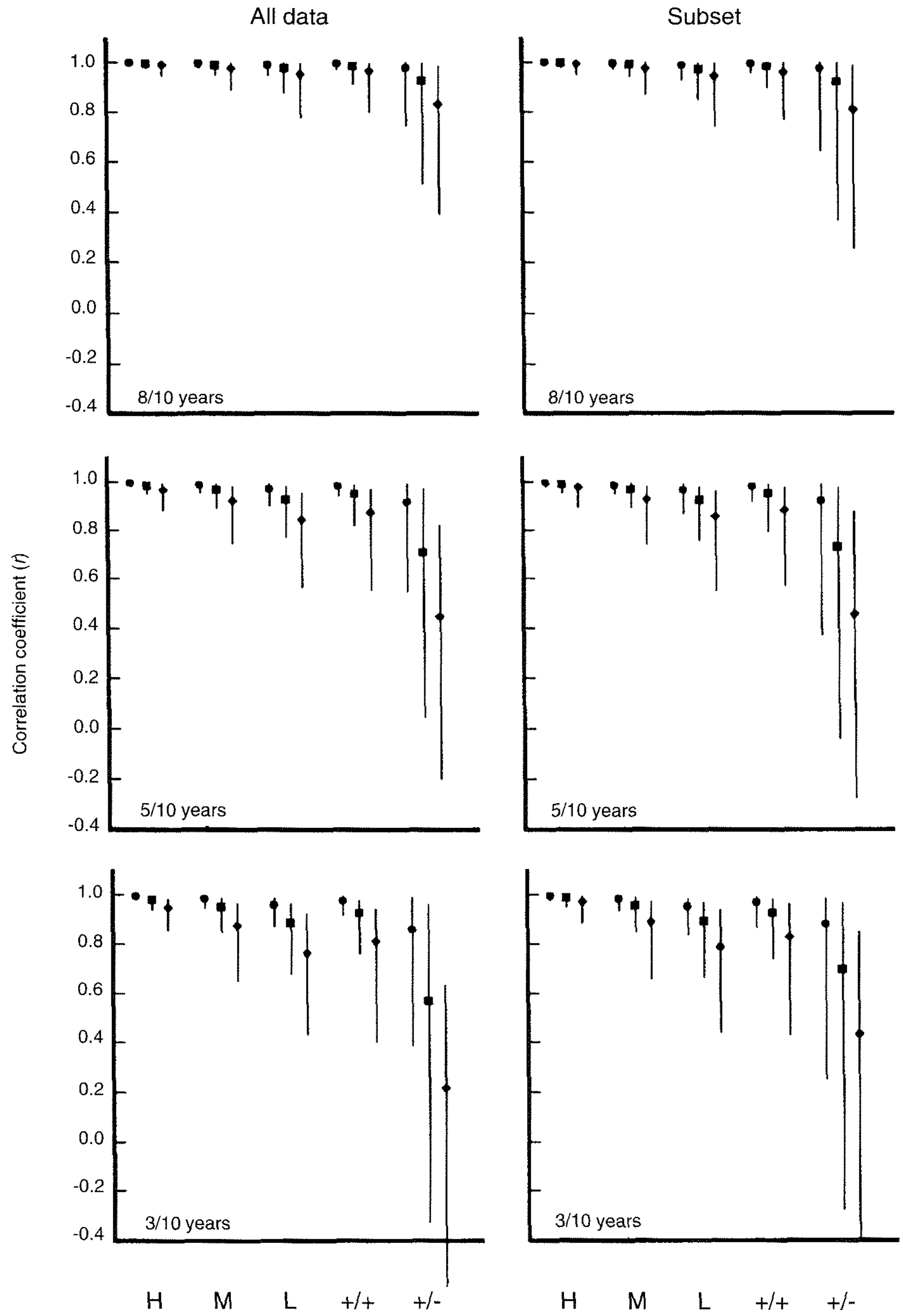

Figure 1: Summary plots of the distributions of correlation coefficients found in Monte Carlo simulation trials investigating the robustness of the index. These plots are for a model of missing data where the **same parameters were missing** in all the years when data were missing. The total length of the monitoring period is **10 years**. Each plot represents a different number of years in which the full set of eight parameters was measured – either 8, 5 or 3 years. These are then divided between the five models of the degree of correlation between parameters: H = all parameters were highly positively correlated; M = all parameters were moderately positively correlated; L = all parameters were poorly positively correlated; +/+ = four parameters were highly positively correlated while four were poorly positively correlated; and +/- = all parameters were highly positively correlated but two are negatively correlated with the rest. For each of these scenarios, the number of indices measured in years when data were missing was 6 (filled circles), 4 (filled squares) and 2 (filled diamonds). The symbols represent the median correlation coefficients from 1 000 runs in respective trials. The upper bars extend to the 95th percentile coefficients and the lower bars extend to the 5th percentile. Plots headed 'All data' are trials where all the data were used to estimate the correlation matrix. 'Subset' indicates trials where data used to estimate the correlation matrix were only derived from data for the years in which no data were missing. Lower bars intersecting the X-axis indicate lower confidence intervals of less than -0.4.

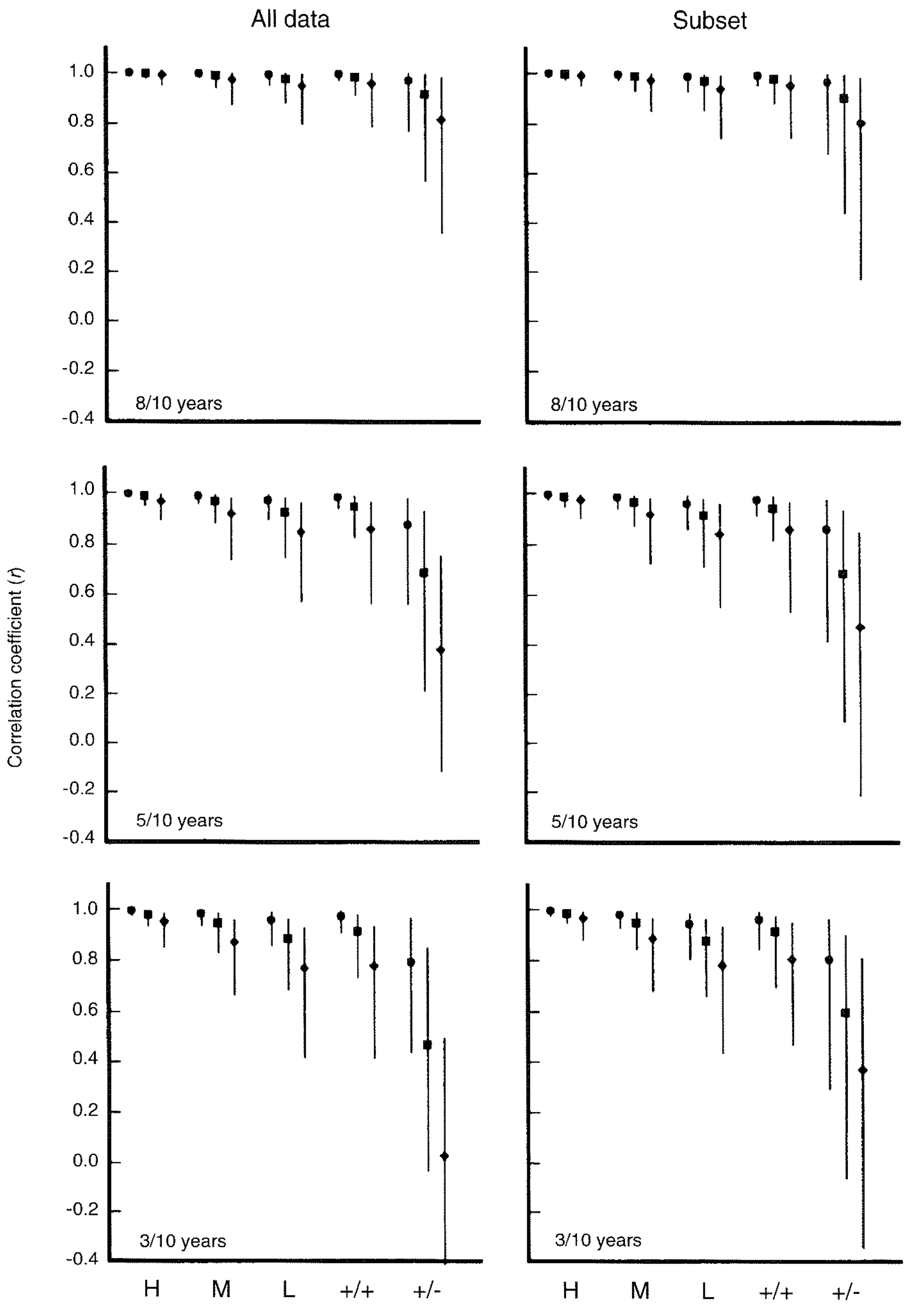

Figure 2: As for Figure 1, but these plots are for a model of missing data where the **parameters to be missing in a year were randomly selected** and the total length of the monitoring period is **10 years**.

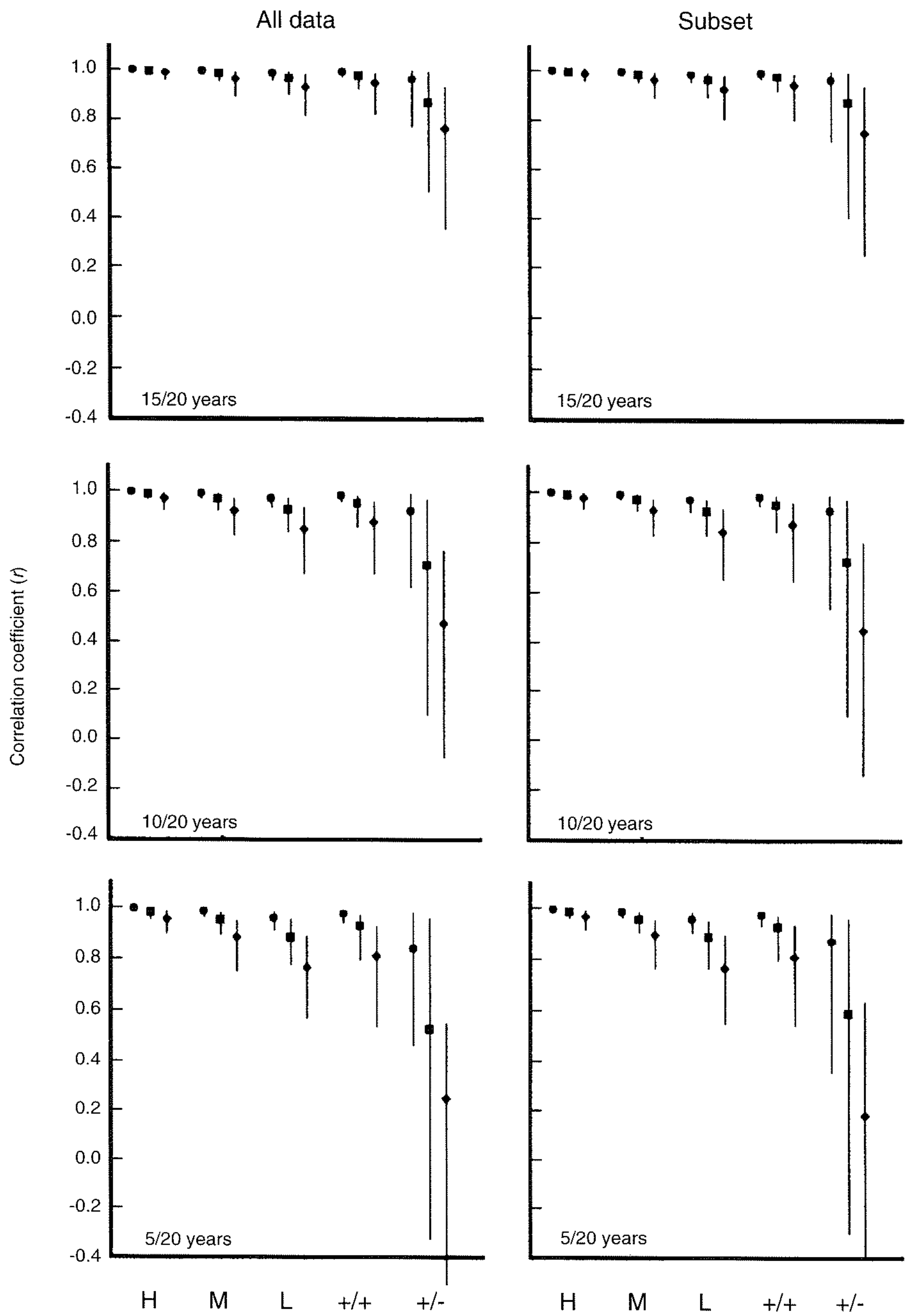

Figure 3: As for Figure 1, but these plots are for a model of missing data where the **same parameters were missing** in all the years when data were missing and the total length of the monitoring period is **20 years**. In this case, each plot represents a different number of years in which the full set of eight parameters were measured – either 15, 10 or 5 years.

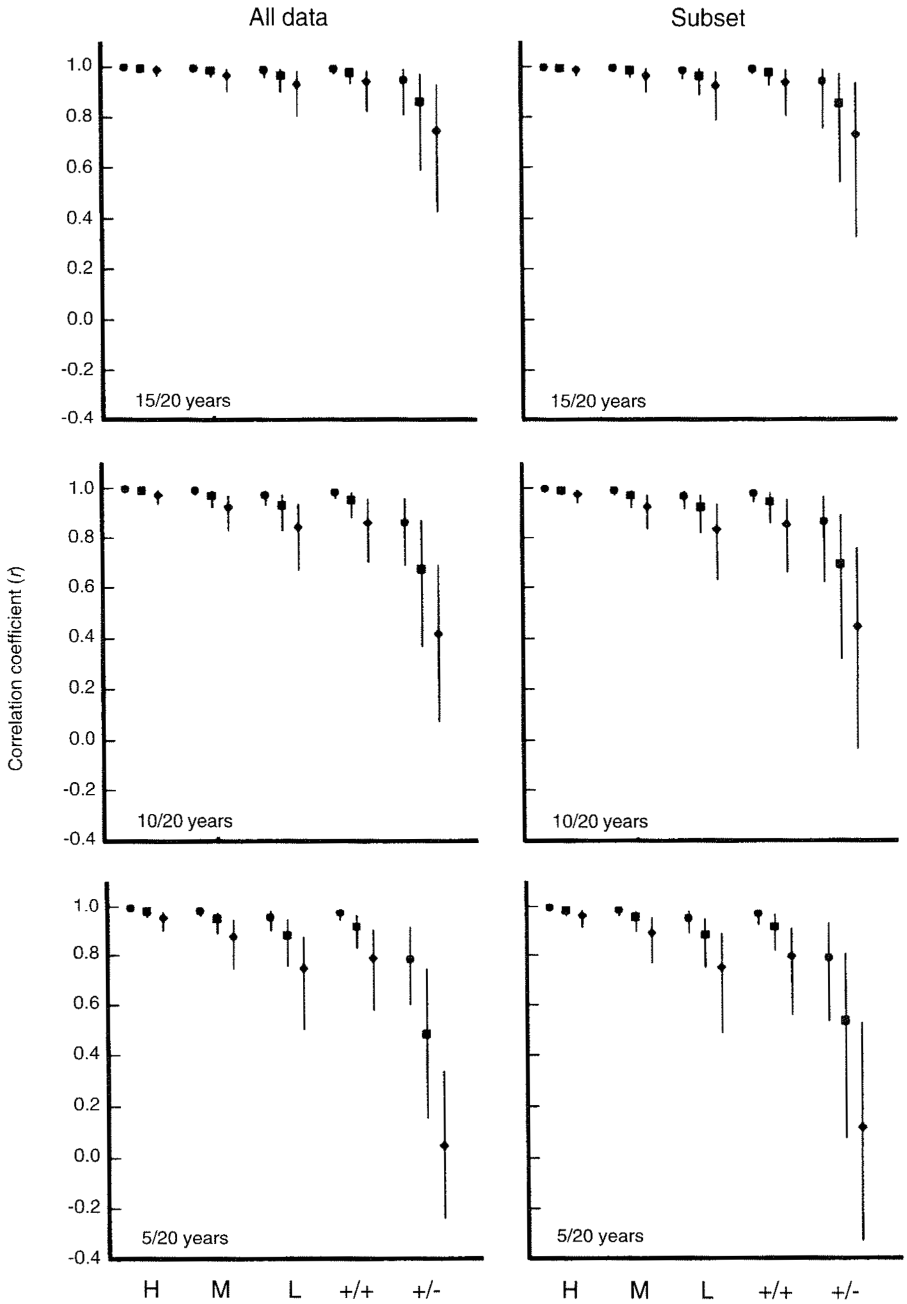

Figure 4: As for Figure 3, but these plots are for a model of missing data where the **parameters to be missing in a year were randomly selected** and the total length of the monitoring period is **20 years**.

Liste des tableaux

Liste des figures

Список таблиц

Список рисунков

Lista de las tablas

Lista de las figuras

modelo de los datos omitidos cuando **se habían omitido los mismos parámetros** en todos los años cuando faltaban datos. La duración total del período se seguimiento es de **10 años**. Cada gráfico representa un número diferente de años en que se midió el conjunto completo de ocho parámetros (8, 5 ó 3 años). Estos se dividen luego entre los cinco modelos del grado de correlación entre parámetros: H = todos los parámetros demostraron una alta correlación positiva; M = todos los parámetros estuvieron correlacionados moderadamente de manera positiva; L = todos los parámetros demostraron una baja correlación positiva; +/+ = cuatro parámetros mostraron una alta correlación positiva mientras que cuatro demostraron una baja correlación positiva; y +/- = todos los parámetros tuvieron una alta correlación positiva pero dos se correlacionan negativamente con el resto. Para cada una de estas situaciones, el número de índices medidos en los años cuando habían datos omitidos fue de 6 (círculos), 4 (cuatratos) y 2 (rombos). Los símbolos representan la mediana de los coeficientes de correlación de 1 000 pasadas en las pruebas correspondientes. Las barras superiores se extienden hasta el 95° percentil de los coeficientes de correlación y las barras inferiores hasta el 5° percentil. Los gráficos con el encabezamiento 'All data' corresponden a pruebas que utilizaron todos los datos para estimar la matriz de correlación. 'Subset' muestra las pruebas para estimar la matriz de correlación que utilizaron datos derivados solamente de años en que no faltaron datos. Las barras inferiores que cortan el eje X indican intervalos de confianza inferiores menores de -0,4.

Figura 2: Igual a la figura 1, pero estos gráficos corresponden a un modelo de datos omitidos donde los **parámetros que faltan en un año fueron escogidos al azar;** el período de seguimiento es de **10 años**.

Figura 3: Igual a la figura 1, pero estos gráficos corresponden a un modelo de datos omitidos donde **faltan los mismos parámetros** en todos los años donde se omiten datos; el período de seguimiento es de **20 años**. En este caso, cada gráfico representa un número distinto de años en que se ha medido el conjunto total de ocho parámetros (15, 10 ó 5 años).

Figura 4: Igual a la figura 3, pero estos gráficos corresponden a un modelo de datos omitidos donde **se escogieron al azar los parámetros que serían omitidos en un año;** el período de seguimiento es de **20 años**.