In Collaboration with
the Netherlands Institute for Sea Research

JOURNAL OF SEA RESEARCH

ELSEVIER Journal of Sea Research 37 (1997) 167–179

Sampling design of monitoring programmes for marine benthos: a comparison between the use of fixed versus randomly selected stations

Jaap Van der Meer

Netherlands Institute for Sea Research, P.O. Box 59, 1790 AB, Den Burg, Netherlands

Accepted 10 February 1997

Abstract

Three possible sampling designs for monitoring programmes were compared for both bias and variance of the associated estimators of year-to-year change. The statistical power of the accompanying univariate analysis of variance models was also analysed. The analysis was based on species abundance data from monitoring programmes for soft-bottom marine benthos in The Netherlands. The first design, with random selection of stations in each year, revealed larger variance and smaller power than the second design, in which stations are sampled randomly in the first year and revisited in succeeding years. The third design, with fixed non-randomly selected stations, yielded the largest power, but gives biased estimates of the year-to-year changes. These results suggest the following rule-of-thumb for monitoring programmes focusing on the abundance of marine benthic species: revisit many randomly selected stations, and make little effort per station.

Keywords: long-term monitoring; macrozoobenthos; statistical power; finite population sampling

1. Introduction

Long-term monitoring programmes of marine benthos focus on the change over years in species abundance in an area. The sampling design will to a large extent determine the interpretation and accuracy of the annual abundance estimates, and of the changes therein. Habitually, benthic ecologists choose a number of sampling stations in a non-random, more or less systematic fashion (Buchanan and Moore, 1986; Desprez et al., 1986; Dörjes et al., 1986; Jensen, 1986; Pearson et al., 1986). Each year a few (usually 3–5) replicate samples are taken within each station (Downing, 1989). In succeeding years the same stations are revisited.

The present article examines the pros and cons of this sampling design, hereafter called design 3, in comparison to two alternative designs. In both alternatives the stations are randomly chosen. In the first alternative (design 1) the stations are selected at random each year. In the second alternative (design 2) the stations are selected at random in the first year. In later years these stations are revisited.

Two different approaches will be taken to evaluate the three sampling designs. First, finite population sampling theory (Cochran, 1963) will be used to examine the estimators of year-to-year change in abundance in terms of bias and variance. The advantage of using finite population sampling theory is that no model assumptions are needed. The basis for inference is entirely provided by the probability sampling design. A similar approach has been taken by Nicholson et al. (1991), who used finite population sampling theory to compare design 1 and design 3

PII S1385-1101(97)00007-5

in the context of fisheries studies. Warren (1994) extended the work of Nicholson et al. (1991) by taking into account partial replacement of stations. Neither of the two papers considered design 2 explicitly.

The second approach places the three designs in the framework of univariate analysis of variance models. The assumption of an analysis of variance model enables an appropriate null hypothesis, such as all yearly population means are equal, to be tested. The power of this test, i.e. the probability of rejecting the null hypothesis when the yearly population means are not equal in favour of a true alternative hypothesis, will be examined for the three designs. A disadvantage of the analysis of variance approach is that the underlying assumptions of the analysis of variance model might be unduly restrictive.

The analysis of variance has been widely used in the context of environmental impact studies (Green, 1979, 1993; Bernstein and Zalinski, 1983; Millard and Lettenmaier, 1986; Stewart-Oaten et al., 1986, 1992; Underwood, 1991, 1994). Most of these impact studies used a so-called BACI (Before and After Impact Control) design, where ideally multiple randomly chosen impact sites and multiple randomly chosen control sites are observed at multiple times before the impact takes place and at multiple times after the impact. Green (1993) gives an excellent overview of various ways to analyse data from such BACI studies. The present study is not focused on impact studies; where the major goal is testing the hypothesis that a well-defined impact had no effects. In the monitoring programmes discussed here, a single area which is visited once a year is considered. There is no talk of an a-priori known impact. Rather, the aim is to test whether any differences between years occurred. Consequently, the analysis of variance models to be discussed here are simpler than those examined by Green (1993).

Several recent papers championed the consideration of power in biological monitoring (Millard and Lettenmaier, 1986; Slob, 1987; Peterman, 1990; Fairweather, 1991; Nicholson and Fryer, 1992), but applications are rare. In the present article a power analysis is performed using three data sets concerning benthos monitoring programmes in The Netherlands. Besides, the variance of the difference between yearly sample means was studied by using finite population sampling theory. The study focused on the abundance of separate species as the variables of interest. Community attributes, such as diversity indices and multivariate measures, were not taken into account.

2. Finite population sampling

First, a short review of finite population sampling theory in the context of a benthos monitoring programme will be given. Suppose the study area is visited once a year for A years. Suppose further that the study area can be subdivided into B stations. Let each station be N times the size of the sampling device. Each station can be thought of as comprising N possible sampling elements. Let Y_{ijk} be some quantity of interest (such as the number of a species) in element k, station j, of year i, which it is assumed can be measured without error. For convenience (and without the requirement of any assumptions), the measurements Y_{ijk} can be rewritten in terms of a linear model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

where the overall mean μ is given by $Y_{...}$, the year factors α_i by $Y_{i..} - Y_{...}$ the station factors β_j by $Y_{.j.} - Y_{...}$, the year by station interaction factors $(\alpha\beta)_{ij}$ by $Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...}$, and the 'error', due to within station variation ε_{ijk} by $Y_{ijk} - Y_{ij.}$. A dot indicates that the mean has been taken. Thus, for example,

$$Y_{i..} = \frac{1}{BN}\sum_{j=1}^{B}\sum_{k=1}^{N} Y_{ijk},$$

The following variance components (Cornfield and Tukey, 1956) can be defined (and written in terms of the linear model)

$$\sigma_A^2 = \frac{1}{A-1}\sum_{i=1}^{A}(Y_{i..} - Y_{...})^2 = \frac{1}{A-1}\sum_{i=1}^{A}\alpha_i^2,$$

$$\sigma_B^2 = \frac{1}{B-1}\sum_{j=1}^{B}\left(Y_{.j.} - Y_{...}\right)^2 = \frac{1}{B-1}\sum_{j=1}^{B}\beta_j^2,$$

$$\sigma_{AB}^2 = \frac{1}{(A-1)(B-1)} \times \sum_{i=1}^{A}\sum_{j=1}^{B}\left(Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...}\right)^2 = \frac{1}{(A-1)(B-1)}\sum_{i=1}^{A}\sum_{j=1}^{B}(\alpha\beta)_{ij}^2,$$

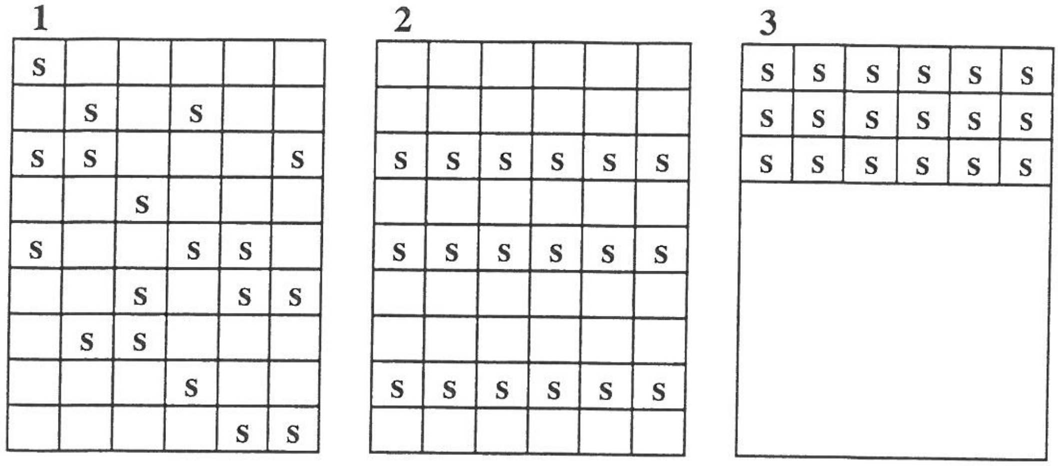

Fig. 1. In the first design the stations (rows) are each year (columns) sampled (indicated by s) randomly and independently. In the second design the stations are sampled randomly in the first year and revisited in succeeding years. The third design assumes fixed non-randomly selected stations and the data refer specifically to those stations.

$$\sigma_E^2 = \frac{1}{AB(N-1)} \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{N} \left(Y_{ijk} - Y_{ij.}\right)^2$$
$$= \frac{1}{AB(N-1)} \sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{k=1}^{N} \varepsilon_{ijk}^2.$$

These components express the variability among years, the variability among stations, the station by year interaction variability (the year-to-year changes may differ among stations), and the within station variability, respectively.

Next, the expectation and variance of the difference between two yearly sample averages will be considered for the three designs mentioned earlier (Fig. 1). For each design it is assumed that there is random selection of sampling elements within stations and all sampling is without replacement. For convenience only two years are considered, i.e. $A = 2$. In the remaining part of the paper, capital letters are used for all that has to do with the underlying array of ABN elements and lower case for all that has to do with the sample data. So b refers to the number of stations sampled each year, and n to the number of elements sampled per station.

The difference between two yearly sample averages $(y_{2..} - Y_{i..})$, where,

$$y_{i..} = \frac{1}{bn} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}$$

can be written in terms of the parameters of the linear model. For all three designs:

$$(y_{2..} - y_{1..}) = (\alpha_2 - \alpha_1) + \frac{1}{b}\left[\left(\sum_2 \beta_j - \sum_1 \beta_j\right) + \left(\sum_2 (\alpha\beta)_{2j} - \sum_1 (\alpha\beta)_{1j}\right) + \left(\sum_2 \varepsilon_{2j.} - \sum_1 \varepsilon_{1j.}\right)\right]$$

where $\sum_i$ means the summation over all elements sampled in year i. In design 1 and design 2 the stations (and sampling elements within stations) are selected at random with equal probability, implying that the expectation of the term between square brackets equals zero, because

$$\sum_{j=1}^{B} \beta_j = 0, \quad \sum_{j=1}^{B} (\alpha\beta)_{ij} = 0, \quad i = 1 \ldots A, \quad \text{and}$$
$$\sum_{k=1}^{N} \varepsilon_{ijk} = 0, \quad i = 1 \ldots A, \quad j = 1 \ldots B.$$

Thus, the expectation of the difference between two yearly sample averages equals

$$E(y_{2..} - y_{1..}) = \alpha_2 - \alpha_1.$$

For design 3 only the sampling elements within stations are sampled randomly, resulting in

$$E(y_{2..} - y_{1..}) = (\alpha_2 - \alpha_1) + \frac{1}{b}\left[\left(\sum_2 \beta_j - \sum_1 \beta_j\right) + \left(\sum_2 (\alpha\beta)_{2j} - \sum_1 (\alpha\beta)_{1j}\right)\right]$$

Since the same stations are revisited, it follows that

$$\sum_2 \beta_j - \sum_1 \beta_1 = \sum_{j=1}^{b} (\beta_j - \beta_j) = 0$$

and the expectation for design 3 simplifies to

$$E(y_{2..} - y_{1..}) = (\alpha_2 - \alpha_1) + \frac{1}{b}\sum_{j=1}^{b}\left((\alpha\beta)_{2j} - (\alpha\beta)_{1j}\right).$$

Generally, the second term on the right hand side will not be equal to zero. The expectation therefore will differ from $\alpha_2 - \alpha_1$, which is the true difference between years for the whole area. The bias will only be zero, whatever the selection of stations, when σ^2_{AB} is zero.

Appendix A gives the derivation for the variance of the three designs that have been used in the present article. The derivation resembles that of Warren (1994), who derived the variances for two designs that are quite similar to the first and third design presented in the present article. Only his assumption on the within-station variance differs from the one used here. For design 1, the variance (ignoring the finite population correction) of the difference between two yearly sample averages equals

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\left[\sigma^2_B + \frac{\sigma^2_{AB}}{2} + \frac{\sigma^2_E}{n}\right].$$

Next, for design 2,

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\left[\sigma^2_{AB} + \frac{\sigma^2_E}{n}\right].$$

For design 3 the only random components are the ε_{ijk} within the b non-randomly selected stations, and

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\frac{\sigma^2_e}{n},$$

where

$$\sigma^2_e = \frac{1}{Ab(N-1)}\sum_{i=1}^{A}\sum_{j=1}^{b}\sum_{k=1}^{N}\varepsilon^2_{ijk}.$$

Hence, if it is assumed that $\sigma^2_e = \sigma^2_E$ (the within station variability for the selected stations is just as large as for all stations), the variance for design 3 is never larger than the variance for the two other designs. The variance for design 2 is smaller than the variance for design 1 if

$$\sigma^2_B - \frac{1}{2}\sigma^2_{AB} > 0$$

The interpretation of this inequality is given below.

3. Univariate analysis of variance

In practice $b \ll B$ and $n \ll N$ and thus the assumption can be made that B and N are infinite. If it is further assumed that the ε values are independently and identically normally distributed, one arrives at the class of analysis of variance models.

The first design, with independent random selection of stations in each year, resembles a two-level nested model. The random factor, station, is nested in the fixed factor, year. It is also assumed that the β values are independently and identically normally distributed. In the second design randomly chosen stations are revisited in succeeding years. This design is related to a two-way mixed model, with a random factor, station, and a fixed factor, year. The third design, with fixed non-randomly selected stations, resembles a two-way fixed model. Both station and year are fixed factors.

For the nested model and fixed model the power of the appropriate F-test, i.e. the probability of rejecting the hypothesis that $\alpha_i = 0$ for all i, depends upon the significance level, the degrees of freedom and the non-centrality parameter Φ^2 (Scheffé, 1959). Once these values are available the power of the test can readily be obtained from (computerized) power charts (Pearson and Hartley, 1951; Dallal, 1988), which are derived from a non-central F-distribution (Johnson and Kotz, 1970). The non-centrality parameter depends upon the degrees of freedom and the expected mean squares. For the mixed model of Scheffé (1959) the appropriate F-value (i.e. MS_A/MS_{AB}, see below) does not in general have the F-distribution under the hypothesis that $\alpha_i = 0$ for all i (Scheffé, 1959). Use of the non-centrality parameter Φ^2 will only approximate the correct power.

Table 1
The expected mean squares for three analysis of variance models

	df	EMS
1. Two-level nested		
Among years	$A-1$	$\sigma_E^2 + n\left(\sigma_B^2 + \frac{A-1}{A}\sigma_{AB}^2\right) + nb\sigma_A^2$
Among stations within years	$A(b-1)$	$\sigma_E^2 + n\left(\sigma_B^2 + \frac{A-1}{A}\sigma_{AB}^2\right)$
Within stations	$Ab(n-1)$	σ_E^2
2. Two-way mixed		
Among years	$A-1$	$\sigma_E^2 + n\sigma_{AB}^2 + nb\sigma_A^2$
Among stations	$b-1$	$\sigma_E^2 + nA\sigma_B^2$
Year·station	$(A-1)(b-1)$	$\sigma_E^2 + n\sigma_{AB}^2$
Within stations	$Ab(n-1)$	σ_E^2
3. Two-way fixed		
Among years	$A-1$	$\sigma_e^2 + nb\sigma_A^2$
Among stations	$b-1$	$\sigma_e^2 + nA\sigma_b^2$
Year·station	$(A-1)(b-1)$	$\sigma_e^2 + n\sigma_{Ab}^2$
Within stations	$Ab(n-1)$	σ_e^2

The expected mean squares of the nested, mixed and fixed models can be given in terms of the variance components of the finite population model (see, for example, Cornfield and Tukey (1956), Scheffé (1959); Table 1). This is helpful in illustrating the similarity between the finite population sampling approach and the analysis of variance approach. Note that the two-way fixed model only refers to the b stations visited. In terms of the finite population model only AbN elements are involved. All variance components therefore differ by definition from the variance components of both other models. This difference is indicated by the use of lower case subscripts.

The non-centrality parameter Φ^2 (Scheffé, 1959) for the hypothesis that $\alpha_i = 0$ for all i is for the nested model (related to design 1):

$$\Phi_A^2 = \frac{bn\sum_{i=1}^{A}\alpha_i^2}{A\left(\sigma_E^2 + n\left(\sigma_B^2 + \frac{A-1}{A}\sigma_{AB}^2\right)\right)},$$

for the mixed model (design 2):

$$\Phi_A^2 = \frac{bn\sum_{i=1}^{A}\alpha_i^2}{A\left(\sigma_E^2 + n\sigma_{AB}^2\right)},$$

and for the fixed model (design 3):

$$\Phi_A^2 = \frac{bn\sum_{i=1}^{A}\alpha_i^2}{A\sigma_e^2}.$$

Note the similarity between the denominator of Φ^2 (divided by bn and with $A = 2$) and the variance of $(y_{2..} - y_{1..})$ under the finite population model. Generally, the power of the mixed model is greater than the power of the nested model if its non-centrality parameter is greater. This happens to be true if:

$$\sigma_B^2 - \frac{1}{A}\sigma_{AB}^2 > 0.$$

For $A = 2$, the same inequality was obtained using finite population sampling theory. If it is assumed that $\sigma_e^2 = \sigma_E^2$, the non-centrality parameter of the fixed model is always larger than that of both other models. Again, a similar result was obtained using finite population sampling theory.

Three things need some further attention. First, for the nested model usually a single among stations-within-years variance component is used. This component, in the notation of Sokal and Rohlf (1981) written as $\sigma_{B\subset A}^2$, can be defined in terms of the finite sampling scheme as follows:

$$\sigma^2_{B\subset A} = \frac{1}{A(B-1)} \sum_{i=1}^{A} \sum_{j=1}^{B} \left(Y_{ij.} - Y_{i..}\right)^2.$$

I have chosen to re-write it as

$$\begin{aligned}\sigma^2_{B\subset A} &= \frac{1}{A(B-1)} \sum_{i=1}^{A} \sum_{j=1}^{B} \left(\beta_j + (\alpha\beta)_{ij}\right)^2, \\ &= \frac{1}{(B-1)} \sum_{j=1}^{B} \beta_j^2 \\ &\quad + \frac{1}{A(B-1)} \sum_{i=1}^{A} \sum_{j=1}^{B} \left((\alpha\beta)_{ij}\right)^2 \\ &\quad + \frac{2}{A(B-1)} \sum_{j=1}^{B} \beta_j \sum_{i=1}^{A} (\alpha\beta)_{ij}, \\ &= \sigma^2_B + \frac{A-1}{A}\sigma^2_{AB}.\end{aligned}$$

When the nested model is discussed on its own, there is no need to introduce interaction terms (Scheffé, 1959). In the present context, however, I found it useful as it highlights the difference with the mixed model and the similarity with the finite sampling approach.

Second, in the statistical literature several mixed models for two factors have been proposed (Scheffé, 1959; Hocking, 1973; Schwarz, 1993). It is generally assumed that the β_j's and the $(\alpha\beta)_{ij}$'s are normally distributed. The models, however, differ in their assumptions on the covariances of β_j's and $(\alpha\beta)_{ij}$'s. The most general model is given by Scheffé (1959). A special case of this model assumes that the variance among station means is the same for each year, and that all covariances among station means from two different years are also equal for each pair of years (cf. model 3 of Hocking (1973), and formulation 2 of Schwarz (1993)). A covariance matrix that obeys this restrictive assumption is said to have compound symmetry (Winer, 1971; Crowder and Hand, 1990; Green, 1993). For the compound symmetry model the ratio MS_A/MS_{AB} has the F-distribution under the hypothesis that $\alpha_i = 0$ for all i, and use of the non-centrality parameter Φ^2 will reveal the exact power. For $A = 2$ the assumption of compound symmetry is, of course, always satisfied.

Finally, the equations for the non-centrality parameters point to the optimal allocation of sampling effort. Usually one assumes that the costs of a benthos sampling programme are directly related to the number of samples taken, i.e. to the product bn (Wildish, 1978; Downing, 1979, 1989; Riddle, 1989). Thus, for each value of the product bn, the choice of $n = 1$ yields the largest non-centrality parameter for both the nested and the mixed model, suggesting that little effort should be made per station. For the fixed model the choice $n = 1$ is less appropriate, as no error term can be estimated.

4. The correlation among years

When two years are considered ($A = 2$), the inequality $\sigma^2_B - \frac{1}{A}\sigma^2_{AB} > 0$ is equivalent to the requirement of a positive correlation (and covariance) among station means, $Y_{ij.}$, from the two different years. This follows easily from rewriting the covariance σ_{12} between $Y_{1j.}$ and $Y_{2j.}$, i.e.

$$\sigma_{12} = \frac{1}{(B-1)} \sum_{j=1}^{B} \left(Y_{1j.} - Y_{1..}\right)\left(Y_{2j.} - Y_{2..}\right),$$

in terms of the variance components

$$\begin{aligned}\sigma_{12} &= \frac{1}{B-1} \sum_{j=1}^{B} \left(\beta_j + (\alpha\beta)_{1j}\right)\left(\beta_j + (\alpha\beta)_{2j}\right) \\ &= \frac{1}{B-1} \sum_{j=1}^{B} \left[\beta_j^2 + \beta_j\left((\alpha\beta)_{1j} + (\alpha\beta)_{2j}\right)\right. \\ &\qquad \left. + (\alpha\beta)_{1j} \cdot (\alpha\beta)_{2j}\right] \\ &= \frac{1}{B-1} \sum_{j=1}^{B} \beta_j^2 - \frac{1}{B-1} \sum_{j=1}^{B} (\alpha\beta)^2_{1j} \\ &= \sigma^2_B - \frac{1}{2}\sigma^2_{AB}\end{aligned}$$

For $A > 2$ it can be shown that the condition $\sigma^2_B - \frac{1}{A}\sigma^2_{AB} > 0$ is satisfied if the sum of all (off-diagonal) covariances is greater than zero.

5. Application to some existing monitoring programmes

5.1. *Macrobenthic fauna of the Oyster Ground, North Sea*

A monitoring programme of the macrobenthic fauna of the Dutch sector of the North Sea started

Table 2
Back-transformed difference between the 1992 and 1991 sample means of the log-transformed density $\exp(y_{2..} - y_{1..})$ and the estimated variance components s_b^2, s_{Ab}^2, and s_e^2.

Species	$e^{(y_{2..}-y_{1..})}$	s_b^2	s_{Ab}^2	s_e^2
Amphiura filiformis	1.23	0.308	0.047	0.173
Callianassa subterranea	1.02	0.647	0	0.205
Harpinia antennaria	1.11	0.083	0.056	0.541
Mysella bidentata	1.42	1.256	0.062	0.546
Spiophanes bombyx	0.30	1.852	1.176	0.432

Oyster Ground; $A = 2$, $b = 5$, $n = 5$

in 1991 (Duineveld, 1992). The data analysed here is the 1991 and 1992 abundance data of five common species on the Oyster Ground, a muddy area between 53°40′ and 55°N and 3° and 5°E. Within this area (19,200 km^2) five stations (100 m^2 each) were non-randomly chosen. At each station five boxcore samples (0.068 m^2 each) were taken each year. Hence $B = 1.92 \times 10^8$ and $N = 1470$. A log(density + 1/0.068) transformation was chosen to satisfy the assumption of a normal distribution with constant variance within stations. This assumption was tested from plots of residuals and found to be reasonable. The data were analysed with a two-way fixed model analysis of variance. Table 2 gives the estimates of the among stations, year·station and within station variance components. These estimates were derived from the mean squares, for example:

$$s_b^2 = \frac{MS_b - MS_e}{nA}$$

Table 2 also gives the back-transformed difference among the two sample means, i.e. $\exp(y_{2..} - y_{1..})$. Hence the average density of *Amphiura filiformis* was estimated to be 23% higher in 1992 than in 1991. For all species $s_b^2 - \frac{1}{2}s_{Ab}^2 > 0$, and, except for *Spiophanes bombyx*, s_{Ab}^2 was much smaller than s_e^2. For each of the three designs the power of the F-test, i.e. the probability of rejecting the hypothesis that $\alpha_1 = \alpha_2 = 0$, is shown in Table 3a. It was supposed that the effect size $\exp(\alpha_1 - \alpha_2) = 1.5$, so the higher density exceeds the lower density by about 50%. The estimated variance components used for the assessment of the non-centrality parameter are s_b^2, s_{Ab}^2, and s_e^2. Hence for the power assessments of the nested and mixed model the implicit assumptions were made that $\sigma_B^2 = \sigma_b^2$ and $\sigma_{AB}^2 = \sigma_{Ab}^2$ and $\sigma_E^2 = \sigma_e^2$. In other words, it was assumed that the stations were randomly sampled. The sample size that is needed to detect an effect size of 1.5 with a probability of 0.80 is given in Table 3b. Finally, Table 3c gives the detectable effect size, which is the effect size that will be detected with a power of 0.80, while $bn = 25$. All calculations were done with the DESIGN computer programme (Dallal, 1988) with a probability level of 0.05. Generally, the nested design yielded the lowest power, and thus the largest detectable effect size and largest sample size needed. The difference in power between the mixed design, with $b = 5$ and $n = 5$, and the mixed design with $b = 25$ and $n = 1$, is striking. The latter design yielded, except for *Spiophanes bombyx*, a power that is only slightly smaller than the power of the fixed design.

5.2. *Macrobenthic fauna of the Oosterschelde, SW-Netherlands*

In 1984 and 1989 the macrobenthic fauna of three tidal-flat areas (31.1 km^2) in the Oosterschelde, SW Netherlands was sampled (Van der Meer et al., 1989; Van der Meer, 1991; Meire et al., 1994). A total of 300 randomly selected stations (100 m^2 each) were visited repeatedly. Within each station a compound sample, using 10 randomly taken small cores (0.00145 m^2 each and 0.1 m deep into the sediment), was obtained. Hence $B = 311000$ and $N = 6896$. For two large and deep-burrowing species, *Arenicola marina* and *Mya arenaria*, five randomly taken large cores (0.01767 m^2 and 0.3–0.4 m deep into the sediment) were used. A log(density + 1/0.0145) transformation was applied. The data for five common species were analysed with a two-way mixed model analysis of variance. As there was only a single compound sample, n was equal to one and σ_E^2 could not be estimated separately. A power analysis was therefore performed for the nested and mixed models only. The non-centrality parameter Φ_A^2 for the nested model can be estimated by

$$\Phi_A^2 = \frac{b\sum_{i=1}^{A}\alpha_i^2}{A\left(s_E^2 + s_B^2 + \frac{A-1}{A}s_{AB}^2\right)}$$

Table 3
Testing the hypothesis of no difference between the 1991 and 1992 means

Species	Nested		Mixed		Fixed	
	$n=5$	$n=1$	$n=5$	$n=1$	$b=5$	$b=1$
(a) Power						
Amphiura filiformis	0.16	0.51	0.40	0.83	0.92	0.92
Callianassa subterranea	0.11	0.33	0.66	0.86	0.87	0.87
Harpinia antennaria	0.23	0.41	0.23	0.43	0.48	0.48
Mysella bidentata	0.08	0.18	0.22	0.42	0.47	0.48
Spiophanes bombyx	0.06	0.13	0.05	0.19	0.57	0.57
(b) Sample size *bn*						
Amphiura filiformis	180	50	50	24	20	18
Callianassa subterranea	335	83	35	22	25	21
Harpinia antennaria	110	64	90	59	55	53
Mysella bidentata	675	177	95	61	55	54
Spiophanes bombyx	1215	276	615	156	45	43
(c) Detectable effect size						
Amphiura filiformis	3.40	1.78	1.97	1.47	1.40	1.40
Callianassa subterranea	5.36	2.11	1.62	1.45	1.44	1.44
Harpinia antennaria	2.58	1.92	2.62	1.89	1.82	1.81
Mysella bidentata	10.94	2.99	2.68	1.90	1.82	1.82
Spiophanes bombyx	24.98	3.94	14.48	2.85	1.71	1.70

(a) The power of testing the hypothesis of no difference between the 1991 and 1992 means; $bn = 25$ and the effect size equals 1.5; (b) the number of samples, bn, needed to achieve a power of 0.80 with an effect size of 1.5; (c) the detectable effect size, i.e. the effect size that will be detected with a power of 0.80 while $bn = 25$; Oyster Ground; $A = 2$.

$$= \frac{b\sum_{i=1}^{A}\alpha_i^2}{MS_B + (A-1)MS_{AB}},$$

and for the mixed model by

$$\Phi_A^2 = \frac{b\sum_{i=1}^{A}\alpha_i^2}{A\left(s_E^2 + s_{AB}^2\right)} = \frac{b\sum_{i=1}^{A}\alpha_i^2}{A\cdot MS_{AB}}.$$

Table 4 gives the among stations and year·station mean squares. Table 4 also gives the back-transformed difference among the two sample means, i.e. $\exp(y_{2..} - y_{1..})$. Clearly, $s_B^2 - \frac{1}{2}s_{AB}^2 > 0$ for all species, since $MS_B > MS_{AB}$.

Table 4
Back-transformed difference between the 1989 and 1984 sample means of the log-transformed density $\exp(y_{2..} - y_{1..})$ and the among stations and year·station mean squares

Species	$e^{(y_{2..}-y_{1..})}$	MS_B	MS_{AB}
Arenicola marina	1.08	0.118	0.050
Cerastoderma edule	8.14	2.984	1.536
Macoma balthica	3.02	1.160	0.600
Pygospio elegans	4.23	2.421	1.334
Scoloplos armiger	0.91	1.967	1.256

Oosterschelde; $A = 2$, $b = 300$, $n = 1$.

The power of the F-test, i.e. the probability of rejecting the hypothesis that $\alpha_1 = \alpha_2 = 0$, is shown in Table 5. The effect size was again supposed to be $\exp(\alpha_1 - \alpha_2) = 1.5$. The number of stations needed to obtain a power of 0.80, and the detectable effect size are also given. For all species the power was exceptionally high, indicating that at least for these species, whose rank numbers in terms of average density range from 2 to 21, too much work has been done. The mixed design performed better than the nested design.

5.3. Macrobenthic fauna of Balgzand, Wadden Sea

Since 1969 the macrozoobenthos of 15 randomly selected stations scattered over Balgzand, a 50-km^2 tidal-flat area in the westernmost part of the Wadden Sea, has been sampled (Beukema and Essink, 1986). At each station a compound sample of 1 m^2 was

Table 5
Testing the hypothesis of no difference between the 1989 and 1984 means

Species	Nested			Mixed		
	power	b	des	power	b	des
Arenicola marina	1.00	10	1.07	1.00	7	1.05
Cerastoderma edule	0.91	217	1.41	0.98	149	1.33
Macoma balthica	1.00	86	1.24	1.00	60	1.19
Pygospio elegans	0.95	181	1.37	0.99	130	1.30
Scoloplos armiger	0.97	155	1.34	0.99	122	1.29

The power of testing the hypothesis of no difference between the 1984 and 1989 means, when b = 300 and the effect size equals 1.5. The number of stations b needed to obtain a power of 0.80, when the effect size equals 1.5. The detectable effect size (des), when the power equals 0.80 and $b = 300$. Oosterschelde; $A = 2$, $n = 1$.

Table 6
The among stations and year·station mean squares

Species	MS_b	MS_{Ab}
Arenicola marina	21.58	0.290
Cerastoderma edule	42.47	1.307
Heteromastus filiformis	57.76	0.789
Macoma balthica	7.40	0.178
Nephthys hombergii	25.01	0.506
Scoloplos armiger	56.61	0.822

Balgzand; $A = 2$, $b = 15$, $n = 1$.

obtained each year. The log(density + 1) data for six common species until 1990 were used: 22 years. Here again σ_e^2 could not be estimated separately. The data were analysed with a mixed model. A power analysis was performed for the nested and mixed models. Table 6 gives the among stations and year·station mean squares.Clearly, $s_b^2 - \frac{1}{22}s_{Ab}^2 > 0$ for all species, since $MS_b > MS_{Ab}$.The power of the F-test, i.e. the probability of rejecting the hypothesis that $\sum \alpha_i^2 = 0$, is shown in Table 7. Two scenarios of possible patterns of change were examined: a sudden change after $A/2$ years, and a linear change in log-transformed density. Following Slob (1987) and Nicholson and Fryer (1992) $\sum \alpha_i^2$ was expressed in terms of the range k between the smallest and largest yearly mean. For the sudden-change scenario

$$\sum_{i=1}^{A} \alpha_i^2 = \frac{A}{4}k^2.$$

For the linear change scenario

$$\sum_{i=1}^{A} \alpha_i^2 = \frac{A(A+1)}{12(A-1)}k^2.$$

It was assumed $k = \log(1.5)$. Given k, the sudden change scenario yielded the largest power of all possible scenarios. Large differences in power among

Table 7
Testing the hypothesis of no difference among the yearly means

Species	Sudden change		Linear change	
	nested	mixed	nested	mixed
Arenicola marina	0.40	0.99	0.15	0.64
Cerastoderma edule	0.16	0.39	0.08	0.14
Heteromastus filiformis	0.15	0.65	0.08	0.22
Macoma balthica	0.88	1.00	0.36	0.90
Nephthys hombergii	0.31	0.88	0.12	0.36
Scoloplos armiger	0.15	0.62	0.08	0.22

The power of testing the hypothesis of no difference among the yearly means. The back-transformed difference between the minimum and maximum mean log-transformed abundance is supposed to be equal to 1.5. Two scenarios of possible patterns of change are examined: a sudden change after $A/2$ years, and a linear change. Balgzand; $A = 22$, $b = 15$, $n = 1$.

species and between scenarios occurred (Table 7). Again, the mixed design performed better than the nested design. Recall that the power calculations for the mixed model are approximate, because $A > 2$, and the assumption of compound symmetry was not met (Lawley's test for equal correlation structure, see Johnson and Wichern (1988), p. 364–366). Nevertheless, there was only a gradual decrease in the size of the correlation coefficient with increasing time lag.

6. Discussion

For all three datasets $s_B^2 - \frac{1}{A}s_{AB}^2 > 0$. Stated otherwise, there was on average a positive correlation between station means in one year and station means in another year. This is in agreement with Beukema and Essink (1986) and Essink and Beukema (1986), who reported synchronised patterns of change over years between stations on both Balgzand and the Groninger Wad (in the eastern part of the Dutch Wadden Sea). It follows that design 2, in which stations are selected at random in the first year and are revisited in succeeding years, generally will yield smaller variance of the estimators of year-to-year change in abundance than design 1, in which stations are selected at random each year.

A second result concerns the number of sampling elements that should be taken per station. Recall that if it is assumed that the costs of a benthos sampling programme are proportional to bn, the total number of samples taken (i.e. if the extra travel costs due to selecting extra stations can be ignored), then the variance of the estimator of year-to-year change in abundance for design 2

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\left[\sigma_{AB}^2 + \frac{\sigma_E^2}{n}\right]$$

shows that only a single haul per station should be taken.

It might be argued that the estimates of the mean and variance arising from design 3 are of severely limited scope, and even misleading. In fact, the results only refer to a few sampled stations per se, but the danger exists that the estimated mean and variance of the change in abundance will be referred to the whole area. In that case, as was shown above, the estimate of the mean for the whole area is biased and the variance is underestimated.

The results of the power calculations for the analysis of variance models point towards the same conclusions as those discussed above, which were based on comparing the variances of the estimators of year-to-year change in abundance. Yet, not all assumptions of the analysis of variance models were met in this study. Particularly, the assumption of compound symmetry for the Balgzand data, where $A > 2$, was not valid. To what extent violation of this assumption have biased the power calculations is not clear. Some authors indicate that violation of the compound symmetry assumption may not lead to serious error (Scheffé, 1959; Winer, 1971), while others state the opposite (Boick, 1981). See Crowder and Hand (1990) for an extensive discussion of the problem. Generally, the assumptions of the univariate analysis of variance models might be unduly restrictive. In that case one may turn to other more reliable models. One alternative for the compound symmetry mixed model is the model of Scheffé (1959), with no restrictions on the covariance matrix. In fact, Scheffé's model is developed from a multivariate normal model with an unstructured covariance matrix. If the number of stations sampled is greater than the number of years, i.e. if $b > A$, then Hotelling's T^2 statistic may be used to test the hypothesis that $\sum \alpha_i^2 = 0$. Yet, often $b < A$. There is, however, room between the compound symmetry model and the unstructured multivariate normal model. Diggle (1988), for example, proposes a model in which the correlation between the same units (stations) depends on their separation in time, and decreases as a monotone function. The covariance matrix is specified by using four parameters, compared to two for the compound symmetry model. Thus, it may describe the covariance structure more accurately than the compound symmetry model, while avoiding the overparametrization of Scheffé's model. For this model, and for others, in which, for example, the assumption of normality is not made, simulation studies might be helpful in determining power (Link and Hatfield, 1990). The assumption of normality can indeed be a nuisance, in particular for rarer species with many zero observations.

In short, a random selection of stations in the first year, which are revisited in succeeding years, seems to be the most appropriate design for a monitoring programme for soft-bottom marine benthos, where the primary objective is detection of change. This follows from the variance of the estimators of year-to-year change, as derived from finite population sampling theory. It should be emphasised that no models concerning the population or the sample had to be assumed in order to arrive at this result. It also follows from the power calculations, but these results are model dependent. The suitability of the models remains a point of concern. The suggested dictum 'many stations and little effort per station' appears to be supported by a tentative comparison between the Oosterschelde and Balgzand datasets. The Oosterschelde dataset, where 300 stations were used and the total area sampled by the small cores equalled 4.35 m^2, revealed smaller variance and larger power than the Balgzand dataset, with 15 stations and a sampled area of 15 m^2, for all species. The difference was most pronounced for *Cerastoderma edule* and *Scoloplos armiger*.

Finally, only three obvious and relatively simple sampling designs were compared. The possibility that more advanced designs (such as stratified random sampling or sampling with partial replacement) may be preferred in some cases is of course not ruled out.

Acknowledgements

I thank Gerard Duineveld for allowing me to use the Oyster Ground data, Patrick Meire and Aad Smaal for the use of the Oosterschelde 1989 data, and Jan Beukema for the use of the Balgzand data. The Oosterschelde 1985 data were obtained in cooperation with Annet van den Dool, Anke Engelberts, Dick J. de Jong and Rob H.D. Lambeck. The 1985 sampling programme was funded by the Rijkswaterstaat and the former Delta Institute of Hydrobiological Research. I am grateful to Rob Fryer, Mike D. Nicholson, Stephen J. Smith, Cajo. J.F. ter Braak, William G. Warren, Alain Zuur and an anonymous referee for valuable discussions and/or their comments on an earlier draft. This is NIOZ publication no. 3160.

Appendix A

First, the variance of $(y_{2..} - y_{1..})$, the estimator of the change in abundance between two years, will be considered for design 1, with independent random selection of stations in both years. Recall that

$$(y_{2..} - y_{1..}) = \frac{1}{b}\left[b(\alpha_2 - \alpha_1 + \left(\sum_2 \beta_j - \sum_1 \beta_j\right)\right.$$
$$\left. + \left(\sum_2 (\alpha\beta)_{2j} - \sum_1 (\alpha\beta)_{1j}\right) + \left(\sum_2 \varepsilon_{2j.} - \sum_1 \varepsilon_{1j.}\right)\right]$$

The probability of occurrence within the two samples is equal for each pair of stations, and thus for each pair β_u, β_v, where $u = 1, B$, and $v = 1, B$. This implies that $\mathrm{Cov}\left(\sum_2 \beta_j, \sum_1 \beta_j\right)$ will be zero, which follows from $\sum_{u=1}^{B}\sum_{v=1}^{B} \beta_u\beta_v = \left[\sum_{j=1}^{B} \beta_j\right]^2$. Using theorem 2.2 from Cochran (1963) we have

$$\mathrm{Var}\left(\sum_2 \beta_j - \sum_1 \beta_j\right) = 2b\frac{B-b}{B}\frac{\sum_{j=1}^{B}\beta_j^2}{B-1} = 2b\frac{B-b}{B}\sigma_B^2$$

Proceeding as before, since $\mathrm{Cov}\left(\sum_2 (\alpha\beta)_{2j}, \sum_1 (\alpha\beta)_{1j}\right)$ and $\mathrm{Cov}\left(\sum_2 \varepsilon_{2j}, \sum_1 \varepsilon_{1j}\right)$ and are zero,

$$\mathrm{Var}\left(\sum_{j=1}^{b} (\alpha\beta)_{2j} - \sum_{j=1}^{b} (\alpha\beta)_{1j}\right) = b\frac{B-b}{B}\frac{\sum_{i=1}^{2}\sum_{j=1}^{B}(\alpha\beta)_{ij}^2}{B-1}$$
$$= b\frac{B-b}{B}\sigma_{AB}^2,$$

and

$$\mathrm{Var}\left(\sum_{j=1}^{b} \varepsilon_{2j.} - \sum_{j=1}^{b} \varepsilon_{1j.}\right) = b\frac{N-n}{N}\frac{\sum_{i=1}^{2}\sum_{j=1}^{B}\sum_{k=1}^{N}\varepsilon_{ijk}^2}{nB(N-1)}$$
$$= b\frac{N-n}{N}\frac{2\sigma_E^2}{n}.$$

Similarly, all the covariances such as $\mathrm{Cov}\left(\sum_2 \beta_j - \sum_1 \beta_j, \sum_2 (\alpha\beta)_{2j} - \sum_1 (\alpha\beta)_{1j}\right)$ are zero. In practice $b \ll B$ and $n \ll N$ and the finite population corrections $\frac{B-b}{B}$ and $\frac{N-n}{N}$ can be ignored. Combining the above we have that

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\left[\sigma_B^2 + \frac{\sigma_{AB}^2}{2} + \frac{\sigma_E^2}{n}\right].$$

Next, for design 2,

$$\mathrm{Var}(y_{2..} - y_{1..}) = \frac{1}{b^2}\mathrm{Var}\left[\sum_{j=1}^{b}\left((\alpha\beta)_{2j} - (\alpha\beta)_{1j}\right)\right.$$
$$\left. + \sum_{j=1}^{b}\left(\varepsilon_{2j.} - \varepsilon_{1j.}\right)\right].$$

Now

$$\mathrm{Var}\left[\sum_{j=1}^{b}\left((\alpha\beta)_{2j} - (\alpha\beta)_{1j}\right)\right] = \mathrm{Var}\left[\sum_{j=1}^{b}\left((\alpha\beta)_{2j} + (\alpha\beta)_{2j}\right)\right]$$
$$= 4\mathrm{Var}\left[\sum_{j=1}^{b} \alpha\beta)_{2j}\right]$$
$$= 2b\frac{B-b}{b}\sigma_{AB}^2$$

thus

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\left[\sigma_{AB}^2 + \frac{\sigma_E^2}{n}\right].$$

For design 3 the only random components are the ε_{ijk} within the b non-randomly selected stations. Defining (using a lower case subscript to emphasise the difference with σ_E^2)

$$\sigma_e^2 = \frac{1}{Ab(N-1)} \sum_{i=1}^{A} \sum_{j=1}^{b} \sum_{k=1}^{N} \varepsilon_{ijk}^2,$$

then

$$\mathrm{Var}\,(y_{2..} - y_{1..}) = \frac{2}{b}\frac{\sigma_e^2}{n}.$$

References

Bernstein, B.B., Zalinski, J., 1983. An optimum sampling design and power tests for environmental biologists. J. Envir. Manag. 16, 35–43.

Beukema, J.J., Essink, K., 1986. Common patterns in the fluctuations of macrozoobenthic species living at different places on tidal flats in the Wadden Sea. Hydrobiologia 142, 199–207.

Boick, R.J., 1981. A priori tests in repeated measures designs: effect of nonsphericity. Psychometrika 46, 241–255.

Buchanan, J.B., Moore, J.J., 1986. Long-term studies at a benthic station off the coast of Northumberland. Hydrobiologia 142, 121–127.

Cochran, W.G., 1963. Sampling Techniques, 2nd edn. Wiley, New York.

Cornfield, J., Tukey, J.W., 1956. Average values of mean squares in factorials. Ann. Math. Statist. 27, 907–949.

Crowder, M.J. and Hand, D.J., 1990. Analysis of Repeated Measures. Chapman and Hall, London.

Dallal, G.E., 1988. Design: a supplementary module for SYSTAT and SYGRAPH. SYSTAT, Inc., Evanston, Il.

Desprez, M., Ducrotoy, J.-P., Sylvand, B., 1986. Fluctuations naturelles et évolution artificielle des biocénoses macrozoobenthiques intertidales de trois estuaires des côtes françaises de la Manche. Hydrobiologia 142, 249–270.

Diggle, P.J., 1988. An approach to the analysis of repeated measurements. Biometrics 44, 959–971.

Dörjes, J., Michaelis, H., Rhode, B., 1986. Long-term studies of macrozoobenthos in intertidal and shallow subtidal habitats near the island of Norderney (East-Frisian coast Germany). Hydrobiologia 142, 217–232.

Downing, J.A., 1979. Aggregation, transformation, and the design of benthos sampling programmes. J. Fish. Res. Bd. Can. 36, 1454–1463.

Downing, J.A., 1989. Precision of the mean and the design of benthos sampling programmes: caution revised. Mar. Biol. 103, 231–234.

Duineveld, G., 1992. The macrobenthic fauna in the Dutch sector of the North Sea in 1991. NIOZ, NIOZ Report, 1992-6.

Essink, K., Beukema, J.J., 1986. Long-term changes in intertidal flat macrozoobenthic as an indicator of stress by organic pollution. Hydrobiologia 142, 209–215.

Fairweather, P.G., 1991. Statistical power and design requirements for environmental monitoring. Aust. J. Mar. Freshwat. Res. 42, 555–567.

Green, R.H., 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley, New York.

Green, R.H., 1993. Application of repeated measures designs in environmental impact and monitoring studies. Aust. J. Ecol. 18, 81–98.

Hocking, R.R., 1973. A discussion of the two-way mixed model. Am. Statist. 27, 148–152.

Jensen, K., 1986. Changes of the macrozoobenthos at 3 monitoring stations in the western Baltic sea and the Sound. Hydrobiologia 142, 129–136.

Johnson, N.L. and Kotz, S., 1970. Continuous Univariate Distributions. Wiley, New York.

Johnson, R.A. and Wichern, D.W., 1988. Applied Multivariate Statistical Analysis, 2nd edn. Prentice Hall, Englewood Cliffs, New Jersey.

Link, W.A., Hatfield, J.S., 1990. Power calculations and model selection for trend analysis: a comment. Ecology 71, 1217–1220.

Meire, P.M., Seys, J., Buijs, J., Coosen, J., 1994. Spatial and temporal patterns of intertidal macrobenthic populations in the Oosterschelde: are they influenced by the construction of a storm surge barrier? Hydrobiologia 282/283, 157–182.

Millard, S.P., Lettenmaier, D.P., 1986. Optimal design of biological sampling programs using the analysis of variance. Est. Coast. Shelf Sci. 22, 637–656.

Nicholson, M.D., Fryer, R., 1992. The statistical power of monitoring programmes. Mar. Poll. Bull. 24, 146–149.

Nicholson, M.D., Stokes, T.K. and Thompson, A.B., 1991. The interaction between fish distribution, survey design and analysis. ICES, Statistics Committee, C.M. 1991/D:11.

Pearson, E.S., Hartley, H.O., 1951. Charts of the power function for analysis of variance tests derived from the non-central F-distribution. Biometrika 38, 112–130.

Pearson, T.H., Duncan, G., Nuttall, J., 1986. Long term changes in the benthic communities of Loch Linnhe and Loch Eil (Scotland). Hydrobiologia 142, 113–120.

Peterman, R.M., 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47, 2–15.

Riddle, M.J., 1989. Precision of the mean and the design of benthos sampling programmes: caution advised. Mar. Biol. 103, 225–230.

Scheffé, H., 1959. The Analysis of Variance. Wiley, New York.

Schwarz, C.J., 1993. The mixed-model ANOVA: the truth, the computer packages, the books. Part I: balanced data. . Am. Statist. 47, 48–59.

Slob, W., 1987. Strategies in Applying Statistics in Ecological Research. Free University Press, Amsterdam.

Sokal, R.R. and Rohlf, F.J., 1981. Biometry, 2nd edn. Freeman, San Francisco.

Stewart-Oaten, A., Murdoch, W.R., Parker, K.R., 1986. Environmental impact assessment: pseudoreplication in time. Ecology 67, 929–940.

Stewart-Oaten, A., Bence, J.R., Osenberg, C.W., 1992. Assessing effects of unreplicated perturbations: no simple solutions. Ecology 73, 1396–1404.

Underwood, A.J., 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Aust. J. Mar. Freshwat. Res. 42, 569–587.

Underwood, A.J., 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecol. Appl. 41, 3–15.

Van der Meer, J., 1991. Exploring macrobenthos-environment relationship by canonical correlation analysis. J. Exp. Mar. Biol. Ecol. 148, 105–120.

Van der Meer, J., Van den Dool. A., Engelberts, A., De Jong, D.J., Lambeck, R.H.D. and Polderman, J.M., 1989. Het macrozoöbenthos van enkele intergetijdengebieden in de Oosterschelde. Resultaten van een inventarisatie in de nazomer van 1985. Dienst Getijdewateren, Internal report, GWAO 89.307.

Warren, W.G., 1994. The potential of sampling with partial replacement for fisheries surveys. ICES J. Mar. Sci. 51, 315–324.

Wildish, D.J., 1978. Sublittoral macro-infaunal grab sampling: reproducibility and cost. Tech. Rep. Fish. Mar. Serv. Can. 770, 1–14.

Winer, B.J., 1971. Statistical principles in experimental design, 2nd edn., McGraw-Hill, New York.