Comparison of surfaces

Analysis of variance
Comparison of surfaces
Log-linear model
Sampling strategy
Trend surface analysis

Analyse de variance
Comparaison de surfaces
Modèle log-linéaire
Stratégie d'échantillonnage
Surface polynomiale

Pierre LEGENDRE [a] and Brian H. McARDLE [b]

[a] Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada.

[b] Biostatistics Unit, School of Biological Sciences, University of Auckland, Private Bag, Auckland, New Zealand.

Received in revised form 27/06/96, accepted 28/06/96.

ABSTRACT

Values of a variable observed over a delimited geographic area are called a "surface". The problem we investigate is how to compare measures of the same area observed at two or more sampling times. It is important to recognise that two different null hypotheses can be tested about the surface. The first is that the corresponding values in the surfaces differ only because of measurement or sampling error. The second is that the underlying process that generated the surfaces is the same. In the first case, the surfaces should have the same shape and values, while in the second case the surfaces might have different values, but if they still have the same means, variances and autocorrelation structures, the null hypothesis cannot be rejected. To prove the null hypothesis wrong would require stronger differences in the second case than in the first. We shall only be considering the first situation, primarily because we feel it is the most commonly investigated. The null hypothesis throughout the paper is therefore that the corresponding values in the surfaces differ only because of measurement or sampling error. The techniques used to test this will depend on the sampling strategy employed in the study. The paper is organised around the three major factors affecting the solutions: (1) Have the same sampling points been measured through time? (2) Were replicate sampling units measured or observed at each sampling location at each time? (3) How were the sampling locations distributed over the study site? The techniques we propose are in standard use and are widely available. The most important are analysis of variance, log-linear modelling, and multiple regression (in the form of trend surface analysis).

RÉSUMÉ

La comparaison de surfaces

On appelle « surface » un ensemble de valeurs prises par une variable dans un espace géographique. Dans ce travail, nous étudions le problème suivant: comment peut-on comparer des mesures prises à deux ou plusieurs occasions dans le même espace géographique ? Il importe de reconnaître que deux hypothèses nulles différentes peuvent être testées. La première affirme que les valeurs correspondantes des surfaces ne diffèrent qu'à cause d'erreurs de mesure ou d'échantillonnage. La seconde prétend plutôt que le processus sous-jacent qui les a générées est le même pour ces différentes surfaces. Dans le premier cas, les surfaces devraient présenter la même forme et les mêmes valeurs, alors que dans

le second cas les surfaces peuvent présenter des valeurs différentes, mais si elles ont la même moyenne, la même variance et la même structure d'autocorrélation, l'hypothèse nulle ne peut être rejetée. Des différences beaucoup plus importantes sont nécessaires pour montrer que la seconde hypothèse nulle est fausse. Seule la première situation sera considérée dans ce travail, principalement parce que nous croyons que c'est la plus couramment étudiée. L'hypothèse nulle dans cet article est donc que les valeurs correspondantes des surfaces ne diffèrent qu'à cause d'erreurs de mesure ou d'échantillonnage. Les techniques que nous utiliserons pour tester cette hypothèse dépendent de la stratégie d'échantillonnage ayant mené à la collecte des données. L'article est structuré en fonction de trois facteurs principaux touchant l'échantillonnage : (1) A-t-on mesuré la variable aux mêmes points géographiques, lors des différents échantillonnages ? (2) Lors de chaque échantillonnage, y a-t-il eu réplication à chaque station ? (3) Quel plan d'échantillonnage a été suivi pour distribuer les stations dans l'espace géographique à l'étude ? Les techniques que nous proposons sont d'utilisation courante. Les principales sont l'analyse de variance, les modèles log-linéaires et la régression multiple (pour l'analyse des surfaces polynomiales).

Oceanologica Acta, 1997, **20**, 1, 27-41.

INTRODUCTION

Values of a variable observed over a delimited geographic area are called a 'surface'. The problem we investigate in this paper is how to compare such surfaces, corresponding to the same area measured or observed at two or several sampling times. The problem is illustrated by Figure 6, near the end of this paper.

It is important to recognise that there are two different null hypotheses about the surface that can be tested. The first is that the corresponding values in the surfaces differ only because of measurement or sampling error. The second is that the underlying process that generated the surfaces is the same. In the first case, the surfaces should have the same shape and values, while in the second case the surfaces might have different values, but if they still have the same means, variances and autocorrelation structures, the null hypothesis cannot be rejected. This situation could arise when the spatial patterns observed are the result, at least in part, of random processes like migration, or mortality dependent on local densities of predators and prey. In such cases we might not expect to see identical surfaces in successive time periods (so long as the interval is large enough), even though there has been no structural change in the system being studied. To prove the null hypothesis wrong would require stronger differences in the second case than in the first, or else far more data.

In this paper we shall only be considering the first situation, primarily because we feel it is the most commonly investigated. The null hypothesis throughout the paper is therefore that the corresponding values in the surfaces differ only because of measurement or sampling error. The techniques used to test this will depend on the sampling strategy employed in the study.

The paper is organised around the three major factors affecting the solutions: (1) Have the same sampling points been measured through time? (I = yes, II = no). (2) Were replicate sampling units measured or observed at each sampling location and each time? (i = yes, ii = no). (3) How were the sampling locations distributed over the study site? We shall consider three sampling strategies in this study: a = stratified, b = systematic, c = random. The solutions are summarised in Table 1.

Table 1

Summary of the models to be used under different circumstances, to study values measured over an area during two or several sampling campaigns.

	(a) Stratified	**(b) Systematic**	**(c) Random**
I. Same locations			
i. With replication	— Model (3) $Y_{hjk} = \mu + l_j + (t_k + lt_{jk}) + \varepsilon$ is used with all designs —		
ii. Without replication	Model (4) $Y_{ijk} = \mu + s_i + l_{j(i)} + t_k + st_{ik} + \varepsilon$	Pair up neighbours and test using (4)	Polynomial model (5, 6, etc.) of appropriate order through time: $Y_{jk} = \mu + l_j + \beta_{1j}t + \beta_{2j}t^2 + \varepsilon$
II. Different locations			
i. With replication	Model (7) $Y_{hijk} = \mu + s_i + l_{j(ik)} + t_k + st_{ik} + \varepsilon$	Model (8) $Y_{hik} = \mu + b_i + (t_k + bt_{ik}) + \varepsilon$	Model (10) comparing trend surfaces
ii. Without replication	Model (11) $Y_{ijk} = \mu + s_i + t_k + st_{ik} + \varepsilon$	Pair up neighbouring boxes and test using (11)	Model (10) comparing trend surfaces

SAMPLING STRATEGIES

A sampling strategy comprises the following elements: first, *what*, or the variables that will be measured; and secondly, *where and when*, that is, the choice of the locations and/or sampling times. We shall assume here that the *what* is dictated by the question to be solved and is already known.

To establish the *where and when* part of a sampling strategy, one must consider the scales at which the biological processes to be studied occur. The size of the unit samples ("grain"), the distance among samples (interval, or "lag" in time series), and the size of the study area ("extent"), all determine the scale of the effects that are detectable (Bellehumeur and Legendre, submitted). If they are not imposed by ecological theory, these important parameters should be determined from a pilot study. Here are a few notes on sampling designs.

For a study to generate results that can be generalised to a statistical population, the sample of objects (locations, etc.) must be representative of its composition and complexity. Random sampling is the only way to achieve that. A sample is called random when each element in the population has a known and non-null probability of belonging to the sample; the most widely used strategy consists of giving every element in the population the same probability of belonging to the sample. Yet, there are many ways to achieve randomness; they are referred *to as sampling designs*. The most usual designs are simple random sampling, systematic sampling, sampling with unequal probabilities, stratified sampling, and sampling by degrees. They are described at length in Scherrer (1982, in French) and in Thompson (1992, in English) for example. The three following designs will be referred to in the present paper and are now briefly described (Fig. 1).

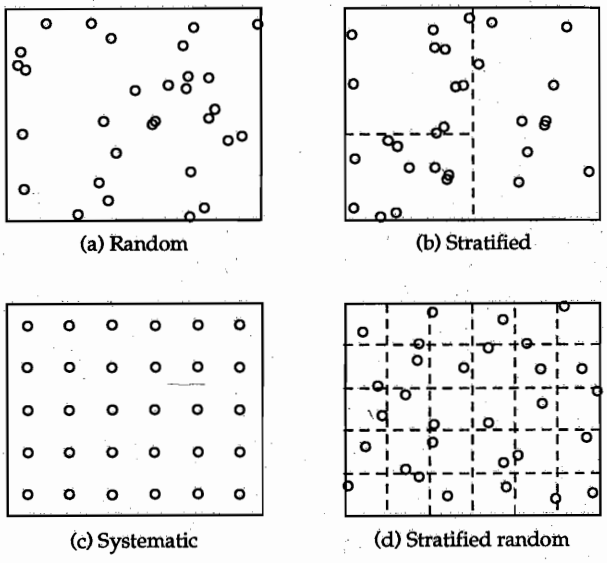

Figure 1

Illustration of the locations of 30 sampling stations, according to four sampling designs. There are three strata in (b), and 30 in (d); they are materialised by dashed lines. The sampling effort is the same (ten samples) in all three strata in (b).

• Simple random sampling consists of selecting at random, and independently, a number of elements from a statistical population. In the spatial context, a simple random sample of localities can be obtained in a number of ways. For a rectangular study area, the simplest and most convenient way is to generate pairs of numbers drawn at random from a uniform [0, 1] distribution, scale these numbers to the size of the sides of the rectangle, and pick the points corresponding to these coordinates (Fig. 1*a*). For an irregularly-shaped surface, one technique consists of dividing the area (on paper) into elementary sampling units, attribute a number to each, and draw at random from among these numbers. Or, draw a rectangle larger than and covering the irregularly-shaped surface, and generate pairs of random coordinates until a sufficient number of them define points within the area under study; unsuitable locations, such as islands within a lake, can be eliminated in the same way, by discarding the randomly selected locations that fall on islands.

• Stratified sampling consists of subdividing a heterogeneous statistical population into more homogeneous subpopulations, called strata, that are mutually exclusive but cover, together, the original population (Fig. 1*b*). The strata do not have to be equal in size; the number of elementary sampling units each one contains can vary, and it often does. An independent sample is then obtained from each stratum, using a random sampling design that can vary from stratum to stratum. The stratification criteria depend on the objective of the study: to increase the precision of the estimates, to obtain some extra information about categories of interest, to over-represent small subpopulations, to facilitate data collection, etc.

• Systematic sampling, as applied to a geographic surface, uses a regular grid of points (nodes). The technique consists of selecting a starting element at random from within the surface, and placing the grid on the surface in such a way that one of the nodes of the grid is on the randomly selected location. This design is interesting in more than one respect; it is easy to carry out in the field, and it facilitates mapping of the observed data. When studying periodic phenomena, however, this technique may produce widely distorted results if the distance between nodes is a multiple of a natural periodic variation in the phenomenon under study. There are many variations on the theme of systematic design; see Quenouille (1949). Two types will be discussed in this paper. (1) Systematic sampling with regular spacing: the sampling units are centred at the nodes of a regular grid (Fig. 1*c*). (2) Stratified random design: the regular grid defines boxes, called strata; a sampling location is selected at random within each box (Fig. 1*d*). This design offers the advantage of good coverage of the study area (a good point for spatial analysis and mapping), although the distances among neighbouring units vary (which is useful to estimate the values of spatial structure functions, such as correlograms and variograms, that correspond to short distance classes); it also solves the problem of periodic structures incorrectly perceived by regular systematic sampling.

We complete this section with the following notes:

a) In systematic sampling, planning is made easier if the study site is regular in shape, so that all locations are

equally likely to get sampled by a randomly placed grid. For a given sampling effort, spacing should be large enough so that the entire site is covered. Place boxes and then decide how to place locations within boxes. Quenouille (1949) shows a number of ways of combining systematic sampling with stratified or random methods. We shall restrict ourselves at present to the special case of grid sampling, though many if not most of our suggestions will generalise to other systematic designs.

b) Replication can either be at a point, *i.e.* at a very local scale, or within the area defined by a grid box. The first gives the greatest power if the primary aim is to detect any changes including local ones; the second gives the greatest power if overall changes (changes in the average for the surface at each period) are wanted.

c) In most field surveys, replication is a view of the mind. At any one time, we cannot replicate the sampling at exactly the same locations; if we did that, we would obtain exactly the same measure (notwithstanding measurement error), provided that the measurement technique is non-destructive. What we may measure in fact are data points that are some distance apart, but that can be argued to effectively replicate measurements because this distance is too small to be of interest in the given study; so we forget about it and, for data analysis purposes, consider that we have "replicates". Even when measuring physical or chemical properties of water from a boat anchor position, one is measuring from volumes of water that were originally some distance apart. One can truly replicate only when conducting controlled experiments at the various geographic locations (e.g. measures of primary productivity from aliquots of the same, homogenised bucket of water).

d) Replicates obtained from neighbouring locations are, in most cases, positively autocorrelated, and this may generate overdispersion of the data values. The statistical problems associated with overdispersion are briefly alluded to below.

THE TECHNIQUES

The techniques we propose are in standard use and are widely available. The most important are analysis of variance, log-linear modelling, and multiple regression (in the form of trend surface analysis).

Analysis of variance (ANOVA) is so widely used as to require little introduction. It may however be relevant to recall that it estimates treatment and other effects by fitting a model which is actually a form of regression model. The significance tests compare the estimated parameter values of the model with zero. Thus a complete two-way factorial analysis fits the following model:

$$Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk} \qquad (1)$$

where Y_{ijk} is the k-th observation in the sample that received the i-th level of factor a and the j-th level of factor b, $(ab)_{ij}$ represents the interaction effect between the two factors, μ is the overall mean, and ε_{ijk} is the error term of the k-th observation in subgroup ij. The model is additive, which means that the effect of a is added to that of b and to the interaction (ab). If the factors do not operate additively in the biological system being modelled, then the statistical model will not be appropriate, the parameters will not have biological meaning, and in particular the interaction term will not imply a biological interaction; this additive model simply cannot be fitted to a non-additive reality. This is particularly important when working with organism count data. The processes that change organism density are seldom additive. Most organisms multiply or divide, they seldom add. As a result, ANOVA additive models are *a priori* inappropriate.

Since most changes in populations are conventionally assumed to be the result of birth and death processes, it is not unreasonable to suggest that most often the appropriate model is multiplicative; the processes operate on a proportional scale. The traditional response to this problem is to analyse population count data on a log scale; in this way, a multiplicative model can be fitted to the raw data by fitting an additive model to the log-transformed data. Regrettably, real animal abundance data often include zeros, so that instead of the desired log (Y) transformation, many, if not most workers, use log $(Y + 1)$. As has been shown elsewhere, this does not perform as the logarithmic transform for small values (McArdle *et al.*, 1989; Gaston and McArdle, 1994). For such situations, we should prefer a technique that operates on a strictly log scale (below), even when there are zeros in the data.

It is important to remember that ANOVA is also modelling the ε_{ijk} error term. Among other things, it is assumed to be independently and normally distributed with constant variance. This is often demonstrably not true, in particular with organism abundances. If numbers are low, even after a log transformation the distribution will not be normal, and the variance is seldom constant. At root these problems arise from trying to fit an inappropriate model to biological data. The biological situation we are modelling actually has two levels of error, while the ANOVA above allows only one. The count we take at a sampling location can be considered as an estimate of the number expected at that location due to the conditions that apply there, but conditions vary, often on a very small spatial scale. We cannot measure this expected density directly, however; we can only get a count of how many organisms are actually present. This count is constrained to be a whole number, usually following a Poisson or related distribution. Thus the predicted value from a model differs from the value observed for two reasons: the small-scale variation in the environmentally determined expected value (the model will not fit this exactly), and the observed count taken at that site (a Poisson or related distribution). A model exists for handling this type of error structure: log-linear, or Poisson, modelling. It fits a multiplicative model and assumes that the counts are Poisson rather than normally distributed. This allows zeros to be handled appropriately; they are treated as sampling estimates of a non-zero value, so no arbitrary constant needs to be added to them. Log-linear modelling methods are widely used for the analysis of count data (e.g. Crawley, 1993), and indeed are frequently used for the analysis of contingency tables since their underlying theory and resulting test statistics prove to be identical

with those of the traditional χ^2 and related methods. While at first sight these models may appear ideal, they too suffer from a disadvantage: in the real world, organisms are seldom Poisson-distributed at any spatial scale above the smallest. On the contrary, they generally tend to be aggregated; in statistical terms, the resulting probability distributions are overdispersed.

This overdispersion will tend to inflate the type I error rate of the tests and confidence intervals, and therefore often presents a more serious problem than if ANOVA had been performed on log $(Y + 1)$ transformed data. In this situation there are two main solutions. The first is to model the source of the overdispersion (e.g. Cormack and Skalski, 1992; Liang and McCullagh, 1993; Anderson *et al.*, 1994), or alter the error portion of the model to a more appropriate form, say by using a negative binomial rather than a Poisson. While these are attractive options, and are possible using the methods of generalised linear modelling, programs to perform such analyses are not widely available. The second solution is to modify the test statistics and standard errors to correct for the overdispersion. When there are error degrees of freedom available, this is possible using the *quasi-likelihood method* (McCullagh and Nelder, 1989). While the underlying theory appears complex, the actual correction is simply applied even by hand (it is also available in several of the major statistical packages). If the assumption of Poisson error is correct, then the model will fit and the χ^2 goodness-of-fit statistic will be non-significant. If however there is overdispersion, then the χ^2 statistic for the residuals will be large (there is more residual variation than expected under the Poisson model) and the model will be rejected as an adequate fit. In this case the χ^2 statistic itself becomes a measure of the degree of overdispersion and can be used to correct the other test statistics and standard errors. The correction is simple: divide all the test statistics by SQRT((error χ^2)/(error *d.f.*)). This simple correction can be applied by hand, and it corrects *a posteriori* for the overdispersion.

The third technique we suggest in this paper is *trend surface analysis*. In essence this is simply the fitting of a non-linear function of the spatial coordinates x and y to the biological response variable; most usually, a polynomial of x and y with cross-product terms is used. For example a relatively complex, but smooth surface might be fitted to a variable of interest by the third-order polynomial:

$$f(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3 \quad (2)$$

The biological variable Y is simply expressed as a function of the spatial coordinates of the sampling localities: $Y_{ij} = f(x_{ij}, y_{ij}) + \varepsilon_{ij}$. In this relatively old form of geographic modelling (Student, 1914), one implicitly assumes that the data are stationary over the study area; in other word, one must assume that a single "trend surface" equation is able to describe correctly the spatial variation of the data values in all parts of the study area. If tests of significance are to be performed, one must assume further that the observations are independent of one another. This technique does not produce accurate fine-grained maps of the spatial variation of a variable; it is useful to describe the overall spatial trend of data. Ripley (1981) gives a didactic account of the technique. Detrending a data set is easily achieved by computing the residuals from a trend surface equation of sufficient order; analyses that require stationarity of the data can be carried out on these residuals.

If there is replication at each point, it is possible to perform a test of goodness-of-fit (Draper and Smith, 1981). By comparing the observed error mean square after fitting the trend surface with the error mean square estimated by the among-replicate within-location variation (computed from the mean value at each location), we can test to see if the model fits the data properly. The among-replicate within-location variation is computed from the deviations from the mean at each location; it is actually the residual mean square of an ANOVA among locations. These two error mean squares are not much different if the trend surface goes through the expected values at the various locations, so that the F ratio of these two mean squares is not significant. On the other hand, if the fitted surface does not follow the major features of the spatial variation in the real surface, then the deviations of the data from the fitted values are likely to be larger than expected from our knowledge of the sampling error, and the F statistic becomes significantly larger than 1, indicating that the trend surface is misrepresenting the real (as opposed to "sampling") spatial information. Normally, model (2) is fitted assuming the errors are normal and of constant variance (*i.e.* as a standard multiple regression); there is no reason, however, why it should not be fitted using a Poisson regression, using the quasi-likelihood correction if necessary. Useless terms of the spatial polynomial can be dropped by backward elimination, as in standard multiple regression; it may happen, for instance, that the linear terms are not significant and should not be retained in the equation, while some of the quadratic and/or cubic terms are significant and describe real variation. The x and y geographic coordinates can be centred before computing the polynomial equation; this removes most or all of the collinearity between the linear and the quadratic terms, which in turn simplifies the interpretation of the coefficients, although it does not change the overall fit of the model.

EXAMPLE DATA SET

Illustrations of the methods will be presented for the various cases of sampling without replication, using the following data set.

Two powerful atmospheric nuclear tests, over one megaton each, were conducted by the *Direction des Centres d'Expérimentations Nucléaires* of France in 1968 and 1970 over Fangataufa, an atoll of French Polynesia, before France banned atmospheric testing in 1974. To study their effects on outer reef-dwelling gastropods, transects, established on the seaward side of the atoll at three

sites called Terme Sud, Hélène and Manchot, have been surveyed by B. Salvat over a 20-year period (1967 to 1987). The transects were 2 m wide and went from the supralittoral zone, through the reef flat, and up to the middle of the reef edge where ocean waves are breaking. They were divided into quadrats three meters long, thus creating strings of contiguous stations with 6 m^2 surface areas. In this form of systematic design, where there is no spacing between juxtaposed quadrats, each transect is random with respect to the seashore, in the area it is intended to represent. Gastropods were collected, identified to species and counted. The results are reported in Lanctôt *et al.* (1997). The Terme Sud transect (28 stations) will be used (with permission) to illustrate some of the methods discussed in the present paper. In some examples, we will use the whole transect comprising the supralittoral zone (6 stations), the reef flat (15 stations), and the reef edge (7 stations), while in other cases we will use only the 6 stations of the supralittoral zone. The supralittoral zone is located up the reef towards the lagoon of the atoll. Terme Sud was sampled in 1968, 1972, 1974, 1977 and 1987; the reef flat and reef edge of Terme Sud were also sampled in 1969, but these data will not be used here.

SCENARIOS

I.i. Same locations sampled with replication

In this situation, the choice of test for a change in surfaces is not influenced by whether the sampling was by grid, strata or simply completely random. Where possible, the following symbols and indices will be used consistently throughout the paper: replicate index h, stratum s_i or grid box b_i, location l_j, occasion t_k. The most powerful test of the hypothesis of "no change" is achieved by fitting the model:

$$Y_{hjk} = \mu + l_j + (t_k + lt_{jk}) + \varepsilon_{hjk} \quad (3)$$

where Y_{ijk} is the h-th replicate taken at the j-th sampling location on the k-th occasion. lt_{jk} represents the interaction between locations and times; significance of this term means that the temporal evolution is not the same at the various locations. The test of interest is the one for the composite null hypothesis that $(t_k + lt_{jk}) = 0$, *i.e.* that there have been no changes in the mean values at each sampling location, either due to uniform changes over the whole surface ($t_k = 0$) or effects restricted to individual locations ($lt_{jk} = 0$). If this hypothesis is rejected, the nature of the changes can then be explored using contour maps of the successive differences between the means at the various locations (below). Computer software especially designed for analysing complete block "repeated measures" offer a test of sphericity for the presence of temporal autocorrelation, and correction for its effect on the significance of the main factors in the analysis (Crowder and Hand, 1990).

More specific tests and contrasts involving the changes averaged over the entire surface (t_k) at each period would generally require the use of the local changes (lt_{jk}) as the error term in the ANOVA. The power of these more specific contrasts can be enhanced, at the cost of losing power for the composite hypothesis, if the sampling was stratified or systematic. In these cases, the variation among locations within strata or within grid boxes can be used as the error term.

Since the error term for the composite hypothesis represents variation at the very smallest spatial scale, it is possible that the error will be close to Poisson. Since there are replicates, this can be checked.

I.ii. Same locations sampled without replication at each location

One problem with spatial processes is that changes can only be detected at the scale of the replication by ANOVA-like models (*i.e.* including log-linear models for qualitative variables), as opposed to trend surfaces. Therefore, using these techniques, fully randomised and systematic designs can only detect, at first sight, differences at the scale of the study area as a whole, while a stratified design can detect strata-specific differences. In order to detect changes at a smaller scale, we must use a different approach for each situation. We will look at the stratified situation first since it is the simplest.

a) *Stratified sampling, no replication*

Because there is random sampling within each stratum, it is possible to fit the model:

$$Y_{ijk} = \mu + s_i + l_{j\,(i)} + t_k + st_{ik} + \varepsilon_{ijk} \quad (4)$$

where Y_{ijk} is the value of the j-th location from the i-th stratum on the k-th occasion; $l_{j(i)}$ measures the location-within-stratum effect. The null hypothesis of interest is the composite one, $(t_k + st_{ik}) = 0$ detecting any change that occurred either in individual strata or averaged over the surface as a whole. As we shall see below, in this situation there are advantages to making the strata as small as possible, so making many of them. In this way the spatial scale at which changes can be detected (that of the stratum) can be reduced. This is of course at the expense of error degrees of freedom and potentially therefore power.

If we can assume that the sampling error at a location (note: not among locations) is Poisson, then we can use a log-linear model and fit model (3) rather than the less sensitive model (4). This allows us to test the null hypothesis $(t_k + lt_{jk}) = 0$ that can detect differences at the scale of locations rather than merely strata. Because of overdispersion due to spatial differences within the locations, however, this assumption of Poisson error might be somewhat rash in many situations. It might be possible to modify the sampling programme so that replication at each location was performed on the first occasion, allowing overdispersion to be detected and estimated, but on subsequent occasions the locations might remain unreplicated. The quasi-likelihood correction term could then be used if necessary to correct the analysis of the unreplicated data.

Example 1: stratified sampling without replication, same locations

The Terme Sud data can be seen as stratified. The transect is composed of three geomorphologically distinct zones (strata). While each zone was sampled systematically (with a random starting position), a simple random sampling design within each stratum can be mimicked; from each zone, three stations were selected at random using numbers obtained from a uniform pseudo-random number generator. Stations 1, 3 and 6 of the supralittoral zone, stations 1, 4 and 9 of the reef flat, and stations 3, 6 and 7 of the reef edge were thus selected; an equal number of locations is not required from each stratum, but it increases the robustness of the test for a "Stratum effect". Analysis of variance was carried out following model (4). Results are presented in Table 2.

So, for the hypothesis $(t_k + st_{ik}) = 0$, $p(H_0)=0.0709$. Results are similar using a $\ln(Y + 1)$ transformation of the gastropod count data. The test is not powerful enough to detect a temporal change that may have occurred either within individual strata, or over the transect as a whole. Of course, we could have increased the power of the test by selecting more stations per stratum; the same analysis, conducted over all 28 stations systematically sampled along the Terme Sud transect, finds all three main effects, as well as the interaction (stratum × year), to be very highly significant (detailed ANOVA table not presented here).

b) *Systematic sampling, no replication*

The method that allows changes to be detected at the smallest scale in this unreplicated situation is based on one of the standard methods for calculating a standard error for a mean from a grid sample (Ripley, 1981; McArdle and Blackwell, 1989). Pair up neighbouring grid points and treat each pair as a stratum, and then analyse as though it were a stratified situation with permanent sampling units, using model (4) above. This, in principle, allows changes to be detected at the scale of paired locations, a potentially powerful approach. This is not post-stratification, since we are not using knowledge about the observed values when creating the strata.

As above, if we could assume Poisson error or use a partial replication to test for overdispersion, then we could again test model 3 without any pairing of neighbouring points. This approach, where values are compared at each location separately rather than at the scale of the mean of two neighbouring points, will give us finer spatial resolution in the test.

Example 2: systematic sampling without replication, same locations

The 6 stations of the supralittoral zone of Terme Sud were paired into three groups of 2 stations each. ANOVA results using model 4 are presented in Table 3. The combined null hypothesis $(t_k + st_{ik}) = 0$ is rejected ($p=0.0017$). Figure 3 illustrates the fact that the pre-nuclear-test gastropod densities of 1968 dropped drastically in the supralittoral zone after the blasts, especially at the stations higher up the beach (stratum 3). This was followed by a slow recovery up to some intermediate level, but never back to the mollusc densities found before the nuclear tests.

c) *Random sampling, no replication*

Though it is simple to test for an average change over the entire surface using two-way ANOVA, attempting to

Table 2

Stratified sampling without replication. Analysis of variance (model 4) for "Number of gastropods" at nine stations of the Terme Sud transect, representing three geomorphological zones (strata).

Source	d.f.	Sum of squares	Mean square	F-value	p-value
Stratum (zone)	2	83.333	41.667	0.542	0.5884
Location within zone	6	587.067	97.844	1.273	0.3064
Year	4	565.467	141.367	1.840	0.1541
Stratum × year	8	1284.667	160.583	2.090	0.0778
Residual	24	1844.267	76.844		
Any change: $t_k + st_{ik}$	12	1850.133	154.178	2.006	0.0709

Table 3

Systematic sampling without replication. Analysis of variance (model 4) for "Number of gastropods" at the six stations of the supralittoral zone of the Terme Sud transect. The stations were paired in three "strata" of two stations each.

Source	d.f.	Sum of squares	Mean square	F-value	p-value
Stratum (neighbours)	2	485.000	242.500	2.558	0.1188
Location	3	140.900	46.967	0.495	0.6922
Year	4	4735.133	1183.783	12.487	0.0003
Stratum × year	8	2377.667	297.208	3.135	0.0368
Residual	12	1137.600	94.800		
Any change: $t_k + st_{ik}$	12	7112.800	592.733	6.252	0.0017

detect more local effects is more difficult. Probably the most powerful device is to model directly, by polynomial, the trajectories through time of the individual locations. This can work provided there are more than two sampling campaigns. The aim is to fit a polynomial of time to each trajectory, using time as the covariate in the ANCOVA, and test for differences among times. For this, one can fit a polynomial of, at most, degree two less than the number of sampling occasions. For example, if there were 3 sampling occasions, the trend ANCOVA model can be:

$$Y_{jk} = \mu + l_j + \beta_j t + \varepsilon_{jk} \quad (5)$$

where Y_{jk} is the value from the j-th location on the k-th occasion, β_j is the slope of the temporal trend (if any) in the changes at the j-th location, and l_j is the location effect, which sets different intercepts for the different locations in the model. The null hypothesis of interest is $\beta_j = 0$. This tests for the existence of any linear trend, unique to each location or uniform throughout locations, in the changes through time. If there were 4 sampling occasions then a quadratic term could be added, allowing more complex trends to be modelled and therefore detected.

$$Y_{jk} = \mu + l_j + \beta_{1j} t + \beta_{2j} t^2 + \varepsilon_{jk} \quad (6)$$

The null hypothesis would now be $(\beta_{1j} t + \beta_{2j} t^2) = 0$.

If we could assume a Poisson error (or use a partially replicated design as above to estimate the correction factor) then we could fit the same model as in the replicated case. The same techniques can actually be used for systematic sampling designs.

Example 3: random sampling without replication, same locations, ANCOVA

The trend ANCOVA technique (models 5 and 6), described above for simple random sampling, is a very general technique that can actually be used with any type of random sampling (systematic or stratified, for instance). For illustration purposes, we used it on the Terme Sud data for the supralittoral zone (6 successive stations, 5 years). First, a standard two-way ANOVA was conducted to establish that significant differences were present among sampling campaigns (Table 4, top).

To identify the shape of the change, a 3-rd order polynomial of time was used in the trend ANCOVA method (5 years –2 = order 3 for the polynomial). From the results in Table 4 (bottom), we conclude, as in the previous example, that there are significant differences among years, and that the evolution of gastropod counts at the various stations can be described by a cubic polynomial effect (the terms Year^2 and Year^3 are significant). This is supported by Figure 3, presented in the previous subsection. Of course, Table 4 (bottom) can also be obtained by multiple linear regression against 8 independent variables: the three "time" variables (Year, Year^2 and Year^3), and five of the six stations coded as dummy variables. Exactly the same probabilities would have been obtained for the various "time" components. Since temporal autocorrelation of the residuals would invalidate the significance tests in this example, it is important, after the polynomials have been fitted, to check that the residuals are not autocorrelated. This assumption is hard to test statistically since there are only five years of data; plots of the residuals did not show any evidence to that effect, however.

Example 4: random sampling without replication, same locations, Poisson error model (GLM)

A Poisson error model ANOVA was used to fit model (3) for unreplicated data at the same locations, using the same data as in the trend ANCOVA calculations of the previous example. Results are reported in Table 5. The fact that all effects are significant here, while in the previous example only "Year" was significant, shows the greater power of the GLM. Indeed, the ANOVA has to use the interaction term to estimate error, and this large term obscures the small location effect in the present case; this problem is not encountered with GLM.

II.i. Different locations at each time period, sampled with replication

a) *Stratified sampling with replication*

The locations are now randomised afresh at each time period. The model is therefore:

$$Y_{hijk} = \mu + s_i + l_{j(ik)} + t_k + st_{ik} + \varepsilon_{hijk} \quad (7)$$

where Y_{hijk} is the value of the h-th replicate from the j-th location within the i-th stratum sampled on the k-th occasion; note that this will not be the same location as the j-th location on any other occasion. The null hypothesis of interest is $(t_k + st_{ik}) = 0$, checking for any changes,

Table 4

Top: two-way ANOVA applied to systematic sampling without replication, for variable "Number of gastropods" at the six stations of the supralittoral zone of the Terme Sud transect. Bottom: trend ANOVA (model 6 with cubic time term). The three "year" terms are pooled to test the composite null hypothesis of interest.

Source	d.f.	Sum of squares	Mean square	F-value	p-value
Stration no.	5	625.900	125.180	0.712	0.6214
Year	4	4735.133	1183.783	6.735	0.0013
Residual	20	3515.267	175.763		
Station	5	625.900	125.180	0.710	0.6227
Year + Year^2 + Year^3	3	4547.311	1515.770	8.596	0.0006
Residual	21	3703.089	176.338		

whether they are average surface changes or those specific to individual strata means.

In this situation there would be little gain in power to be achieved by using a log-linear model instead of the usual normal error model. The test of interest would use the $l_{j(ik)}$ term as the error. In any case, it is quite awkward to analyse this kind of mixed model with the statistics packages currently available.

b) *Systematic sampling with replication*

This approach, replicated systematic samples where locations shift at each time period, only makes sense if replication is at the scale of the box rather than at the point. Replicating in the immediate vicinity of each location will add little to the unreplicated scenario discussed below, because of the autocorrelation in the data. For permanent sampling locations the information at a location can be used to characterise the changes in that point in space through time. Because the location of the sample in the current situation is being moved between sampling campaigns, it can only be taken as a replicate representing the box; we can only look at changes in the box through time. Extra replicates in the immediate vicinity of the sampling location tell us little more about the box than a single sample. If they were spread out, by say random sampling, treating the box as a stratum, we would get a much clearer picture of the changes in the box through time. If the replication is at the scale of the box, then the model is as in the permanent sampling location case (3), except that what we called the location effect is now at the scale of the box rather than at that of the point; the spatial scale of detectable differences is now larger. It is also the same model as the unreplicated stratified design model (4). The model is therefore:

$$Y_{hik} = \mu + b_i + (t_k + bt_{ik}) + \varepsilon_{hik} \qquad (8)$$

where Y_{hik} is the h-th replicate on the k-th occasion located in the i-th grid box. The test of interest is the composite one $(t_k + bt_{ik}) = 0$, *i.e.* that there have been no changes in the mean values at each box, either due to uniform changes over the whole surface ($t_k = 0$) or effects restricted to individual boxes ($bt_{ik} = 0$).

Once again there would be probably little gain in power to be achieved if a log-linear model were used instead of the usual normal error models. The fitted model would now be:

$$Y_{hijk} = \mu + b_i + l_{j(ik)} + t_k + bt_{ik} + \varepsilon_{hijk} \qquad (9)$$

and in this case $l_{j(ik)}$ would again be the relevant error term, as in the ANOVA model 7. Such a mixed model analysis with Poisson error is not simple to perform with current statistical packages.

c) *Random sampling with replication*

In this simple nested design (locations in time), local changes in the surface could not be detected using analysis of variance. ANOVA can only test for a change in the overall mean of the surface, not its shape, because there is no replication, and different locations have been sampled in each sampling campaign.

The most obvious approach here is to produce spatial trend surfaces, one for each time period, and test for differences among them. Depending on the order of the surface fitted, smaller and smaller scale changes can be detected; computation requires however that a large number of data points be available. Because of the potential complexity of the model being fitted, we will only show the model for a 3rd order surface. Symbols concern the h-th replicate for the j-th location sampled on the k-th occasion. Let us designate all individual samples by index i, and their geographic coordinates by x_i and y_i. Replicates should be assigned to the central "locality" coordinate they estimate.

$$\begin{aligned} Y_{ik} = \beta_0 + t_k + (&\beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i \\ &+ \beta_6 x_i^3 + \beta_7 y_i^3 + \beta_8 x_i^2 y_i + \beta_9 x_i y_i^2) \\ &+ (\beta_{1k} x_i + \beta_{2k} y_i + \beta_{3k} x_i^2 + \beta_{4k} y_i^2 + \beta_{5k} x_i y_i \\ &+ \beta_{6k} x_i^3 + \beta_{7k} y_i^3 + \beta_{8k} x_i^2 y_i + \beta_{9k} x_i y_i^2) + \varepsilon_{ik} \end{aligned} \qquad (10)$$

The trend surface parameters in the first parenthesis provide a "null" surface for all time periods, while t_k accounts for any change in the "height" (mean value) of the surfaces overall (averaged over all locations) from time to time. A separate series of trend surface parameters has been added (second parenthesis) for each time period (k) to allow the shape of the surface to change over time. So, this equation represents a model for a main effect "time" (t), a main effect "locations" (the terms in the first parenthesis), and the interaction between them (the terms in the second parenthesis). The hypotheses to be tested are nested.

- The first model to test is the simplest one:

$$\begin{aligned} Y_{ik} = \beta_0 + (&\beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i \\ &+ \beta_6 x_i^3 + \beta_7 y_i^3 + \beta_8 x_i^2 y_i + \beta_9 x_i y_i^2) + \varepsilon_{ik} \end{aligned}$$

This model will fit best if the surface has changed neither in mean nor in shape over time: a single response surface

Table 5

Poisson error model analysis of deviance using GLM, for the same data as in Table 4. Deviance statistics are chi-squares.

Source	d.f.	Chi-square	p-value
Location	5	46.26	<0.0001
Year	4	170.10	<0.0001
Year × Location	20	108.47	<0.0001
Any change	24	278.57	<0.0001

fits at all times. The corresponding hypothesis "H_0: there has been no change at all" is tested by the H_0:

$$t_k + (\beta_{1k}x_i + \beta_{2k}y_i + \beta_{3k}x_i^2 + \beta_{4k}y_i^2 + \beta_{5k}x_iy_i + \beta_{6k}x_i^3 + \beta_{7k}y_i^3 + \beta_{8k}x_i^2y_i + \beta_{9k}x_iy_i^2) = 0.$$

• If only the mean changed over time, the extra parameter t_k is required in the model:

$$Y_{ik} = \beta_0 + t_k + (\beta_1x_i + \beta_2y_i + \beta_3x_i^2 + \beta_4y_i^2 + \beta_5x_iy_i + \beta_6x_i^3 + \beta_7y_i^3 + \beta_8x_i^2y_i + \beta_9x_iy_i^2) + \varepsilon_{ik}$$

The corresponding test for no change in shape over time would be a test on the H_0:

$$(\beta_{1k}x_i + \beta_{2k}y_i + \beta_{3k}x_i^2 + \beta_{4k}y_i^2 + \beta_{5k}x_iy_i \beta_{6k}x_i^3 + \beta_{7k}y_i^3 + \beta_{8k}x_i^2y_i + \beta_{9k}x_iy_i^2) = 0.$$

• If only the shape, but not the mean changed over time, the model describing the variation would be:

$$Y_{ik} = \beta_0 + (\beta_{1k}x_i + \beta_{2k}y_i + \beta_{3k}x_i^2 + \beta_{4k}y_i^2 + \beta_{5k}x_iy_i + \beta_{6k}x_i^3 + \beta_{7k}y_i^3 + \beta_{8k}x_i^2y_i + \beta_{9k}x_iy_i^2) + \varepsilon_{ik}$$

The corresponding test of the "H_0: no change in the mean" would thus be a test of the $H_0 : t_k = 0$.

• Finally, if the surface has changed in mean and in shape, these models will be inadequate, the adequate model requiring the extra parameters found in model 10.

Replication at each randomly chosen point allows a test of goodness of fit to be performed which could assess the appropriateness of the model, as explained in the Techniques section. Example 7 will show how to compute this model using multiple regression instead of ANOVA technology.

II.ii. Different locations at each time period, sampled without replication at each location

a) *Stratified sampling, no replication*

Each stratum has had a random sample of locations taken at each time period. The model is:

$$Y_{ijk} = \mu + s_i + t_k + st_{ik} + \varepsilon_{ijk} \qquad (11)$$

Because there is no replication at the scale of locations, it is now only possible to detect changes at the spatial scale of the strata. The null hypothesis of interest is therefore $(t_k + st_{ik}) = 0$.

Example 5: stratified sampling without replication, different locations

Stations were selected at random from each stratum (zone) of the Terme Sud transect (which contains 28 stations altogether), and this independently for each year. Two stations were selected from the reef edge (that contains 7), 4 from the reef flat (that contains 15), and 2 from the supralittoral zone (that contains 6). The ANOVA results under model (11) for variable "Number of gastropods", presented in Table 6, indicate no significant difference among years, either within individual strata, or over the transect as a whole. Lack of pairing of the stations sampled from year to year gives low power to this test, however; the results of I.ii.a, where the same locations were used from year to year, were much better (*i.e.* closer to significance, for about the same number of stations as what we are using here (45, compared to 40).

b) *Systematic sampling, no replication*

Different sampling grids are obtained by selecting at random, within a box, the starting position of the regular grid for each sampling campaign. This has the effect of forming a grid of boxes within which the sampling locations will randomly fall (Fig. 2). Another strategy would be the stratified random approach, discussed earlier, which will give similar boxes that stay constant but where the locations vary from box to box. By pairing neighbouring boxes as in section II.i.b, we can still test for changes that occur at the spatial scale of paired boxes using (11). Paired boxes are now simply strata and the model and tests are the same as in II.ii.a.

A log-linear model could be used to fit model (11), treating each box as a stratum. Of course the spatial variation in each box will lead to overdispersed errors, but provided the small-scale variation is not too great, the overdispersion correction parameter should be able to cope.

Example 6: systematic sampling without replication, different locations

To illustrate the difference between models (4) and (11), we will now analyse data that actually come from the same locations, and pretend that they don't. Stations 1 to 6 of the reef edge of Terme Sud were paired into three groups

Table 6.

Stratified sampling without replication. Analysis of variance (model 11) for "Number of gastropods" at stations selected at random from each stratum (zone) of the Terme Sud transect, independently for each year.

Source	d.f.	Sum of squares	Mean square	F-value	p-value
Stratum (zone)	2	955.800	477.900	2.135	0.1393
Year	4	684.100	171.025	0.764	0.5586
Stratum × year	8	3394.700	424.338	1.895	0.1060
Residual	25	5597.000	223.880		
Any change: $t_k + st_{ik}$	12	4078.800	339.900	1.518	0.1828

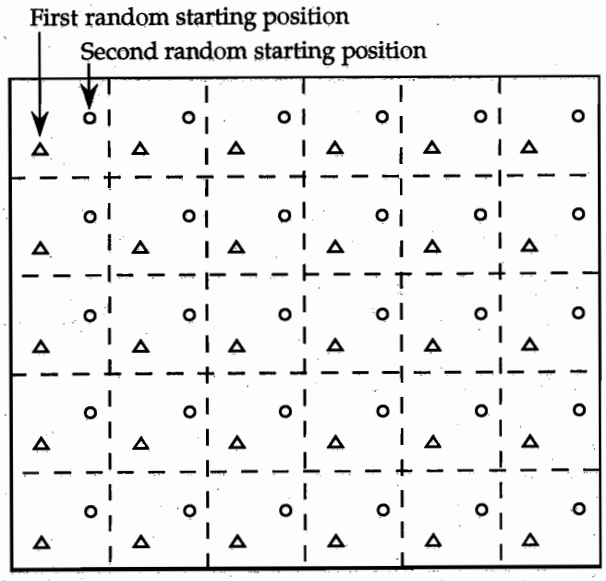

Figure 2

A location, selected at random from within the upper left-hand box, serves as starting position for each systematic sampling campaign (first: triangles; second: circles).

of 2 stations each. ANOVA results using model (4) are presented in Table 7 (top). We do not detect any significant difference among years; for the combined null hypothesis $(t_k + st_{ik}) = 0$, $p(H_0) = 0.2137$. We do, however, detect significant differences among strata and among locations within strata. In Table 7 (bottom), the same data are analysed using model (11); the residual now includes the "Location" effect, so that it is larger, which in turn makes the effects of interest (Stratum, Year, Stratum × year interaction, Any change) less significant. This clearly shows the interest of pulling out the "Location within stratum" effect from the residual, which can only be accomplished when the same locations have been sampled time after time.

c) *Random sampling, no replication*

The most obvious approach to employ here is again to produce trend surfaces and test for differences among them. Depending on the order of the surface fitted, smaller and smaller scale changes can be detected; computation requires however that a large number of data points be available. The model fitted is equation (10) and the null hypotheses to test are the same. However, in this case since there is no replication, no test for adequacy of fit is possible.

A log-linear model could be used. But unlike the ANOVA situations, where the error term was given by replication, here the error term could be due to lack of fit as well as overdispersion. Under these circumstances the overdispersion correction might be expected to be conservative.

Example 7: random sampling without replication, different locations

Ten stations were selected at random from the Terme Sud transect (28 stations), and this independently for each year, without consideration for the geomorphological zones of the transect. Using a transect here instead of a surface allows one to limit the computations to a few parameters. When analysing a surface, more parameters are involved in the model, and thus more locations are necessary. Multiple regression will be used for the computations instead of ANOVA. The following variables have to be constructed first. Each of the five years is represented by a dummy variable called "68", "72", "74", "77" and "87"; any four of them will be used in the multiple regression equation. The geographic position of the stations along the transect is given by variable X; the values were actually centred on their mean before the calculations, so that X_i =station number –14.6. The general trend surface equation is made of variables X, X^2 and X^3 representing a cubic polynomial of X. Each year's trend surface equation is obtained by constructing a product of the polynomial terms of variables X by that year's dummy variable. Thus, the specific shape component for year 1968 is made of the product variables $X \times$ "68", $X^2 \times$ "68" and $X^3 \times$ "68".

Results, computed by multiple regression, indicate that the simplest model, describing variation by a single general trend surface equation, is significant. Adding to this model the years' dummy variables (corresponding to a change

Table 7

Systematic sampling without replication. Analysis of variance (top: model 4; bottom: model 11) for "Number of gastropods" at the first six stations of the reef edge of the Terme Sud transect. The stations were paired in three "strata" of two stations each.

Source	d.f.	Sum of squares	Mean square	F-value	p-value
Stratum (neighbours)	2	3688.467	1844.233	5.388	0.0214
Location	3	3760.900	1253.633	3.662	0.0440
Year	4	3184.667	796.167	2.326	0.1156
Stratum × year	8	3388.533	423.567	1.237	0.3567
Residual	12	4107.600	342.300		
Any change: $t_k + st_{ik}$	12	6573.200	547.767	1.600	0.2137
Stratum (neighbours)	2	3688.467	1844.233	3.516	0.0560
Year	4	3184.667	796.167	1.518	0.2471
Stratum × year	8	3388.533	423.567	0.807	0.6065
Residual	15	7868.500	524.567		
Any change: $t_k + st_{ik}$	12	6573.200	547.767	1.044	0.4709

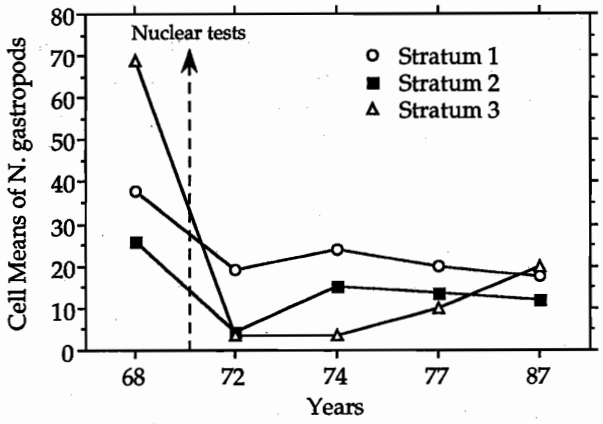

Figure 3

The temporal evolution of the "number of gastropods" variable clearly shows the impact of the nuclear tests in each of the three strata (pairs of stations) of the supralittoral zone.

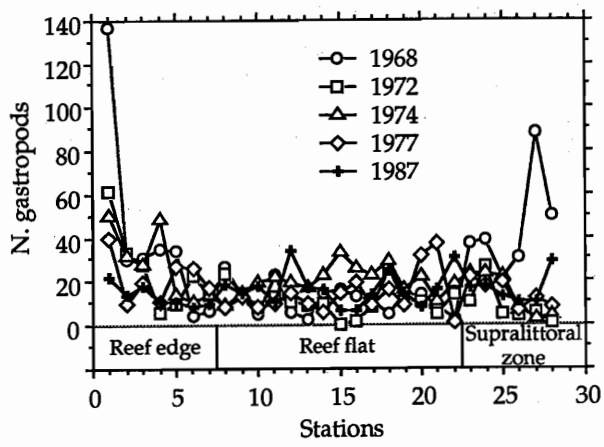

Figure 4

Differences among years for the "number of gastropods" along the transect. The open ocean is on the left, the lagoon on the right of the graph.

in mean over time) reveals, however, that the increase in explained variation is not significant; so, this added component is inadequate to describe the data. (As in stepwise regression, this is verified by an F test of the ratio of (1) the difference in residual sum of squares between the two models, divided by the difference in degrees of freedom of these residuals, on (2) the larger model's residual sum of squares divided by its degrees of freedom.) On the contrary, adding the parameters describing the change in shape over time, the increase in explained variation is significant, so that the model including changes in shape over time is adequate. Simplifying the equation by backwards elimination of nonsignificant variables gives the following model ($R^2 = 0.5418$, $p < 0.0001$), that contains significant terms only:

$$Y_{ik} = 15.4478 + 1.3919\,X - 0.0156\,X^3 + 0.0277\,X^3 \times \text{“68”} + \varepsilon_{ik}$$

This equation indicates that the shape of the curve for 1968 differs significantly from that of the other years, given the points that have been selected. There is no point testing the model with all parameters included since we have already shown that the years' dummy variables add nothing significant to the model.

To check the influence of random sampling on the results, the same calculations were carried out again using all data (28 stations, 5 years). In addition to the differences in shape, the parameters describing differences in means also added a significant amount of explained variation to the simplest model. So, the general model (equation 10) is the one that describes best the variation. Simplifying the equation by backwards elimination of nonsignificant variables gives the following model ($R^2 = 0.5325$, $p < 0.0001$), that contains significant terms only:

$$Y_{ik} = 13.2410 + 0.8091\,X + 0.0322\,X^2 - 0.0077\,X^3 - 10.3276\ \text{“68”} + 5.6548\ \text{“74”} + 0.3172\,X^2 \times \text{“68”} - 0.0061\,X^3 \times \text{“72”} - 0.0052\,X^3 \times \text{“74”} + \varepsilon_{ik}$$

It indicates that significant differences in mean are found, involving years 1968 and 1974 compared to all others, and in shape for 1968, 1972 and 1974. We must then conclude that the first test, carried out over 10 randomly selected stations per year only, had reduced power, since it has been unable to find a significant increase in explained variation when adding to the model the parameters describing differences in mean among years. Figure 4 illustrates the data; the differences in mean and in shape should be obvious at least for 1968.

VISUALISING SPATIAL STRUCTURES

Results of the tests of significance for differences among surfaces, described in the previous sections, only tell if differences exist. The nature of the differences must be studied using other techniques of spatial analysis. Three of them are now briefly described.

1. Description of spatial structures by contouring

Contouring is the production of contour maps. Several techniques are available for doing so, as well as a variety of computer programs. Reviews and references to programs are found in Ripley (1981), Burrough (1987), Legendre and Fortin (1989), Isaaks and Srivastava (1989) and Legendre (1993). Commonly-used techniques include trend surface analysis (described above), interesting only to describe the large-scale trend in data; moving average interpolation, including the widely-used method of inverse-distance interpolation; and kriging, the geostatistical technique based on variograms (below). Maps can nowadays be produced in different graphical flavours, using either flat or raised contours, deformed mesh grids, shaded reliefs, and so on. Maps provide ecologists with a feeling for the spatial variability in their data and for the quality of the sampling design they have used. They are interesting for what they display, and sometimes even more so for what they don't show while the research team expected to

find it. For instance, small sampling units introduce larger variance and a smaller degree of precision than larger units (Bellehumeur *et al.*, submitted); so, the map is very noisy if the sampling units are too small. On the other hand, if the samples are too far apart, spatial structures that the investigator knows to exist in the field may not appear on the map. So, contouring is a good way of getting a feeling for scale problems in the data: does the expected kind of variability appear on the map? If not, it may be necessary to increase the extent of the study. Of course, some contouring methods smooth the data a great deal. Some of the methods of analysis, particularly the replicated ones, can detect changes at a smaller scale than some of the contouring methods might be able to show, particularly if the sampling density was high and the local variation large. Contouring methods perform best when the geographic extent of the study is well-covered by the sampled locations; systematic sampling is best here.

2. Structure functions

Correlograms and variograms are the most commonly used structure functions for studying the shape of the spatial structuring present in a surface. Structure functions are graphs of a statistic in ordinate (autocorrelation, or semi-variance), plotted against distance classes in abscissa. They indicate the scales at which the most important spatial variation occurs in a geographic surface. They can be computed for single variables, or for multivariate data sets if the problem calls for that. Reviews and lists of available programs are found in Burrough (1987), Legendre and Fortin (1989), Isaaks and Srivastava (1989) and Legendre (1993).

3. Comparison of successive surfaces

Here is a technique to display differences among sampling times, for a single variable. Let us consider the case where the sampling locations are the same; to compare surfaces with different sampling locations, some form of interpolation would be required to produce values at corresponding locations. One simply has to subtract one time's values from the other, and map (method 1 in the present section) or analyse (method 2) the differences. Presence of a spatial structuring in the "difference" indicates that the two surfaces do not have the same shape. (Beware: as noted in the Introduction, this does not mean that the processes that generated these surfaces differ). Two examples will be given of the technique.

Example 8: one-dimensional data

The first example (Fig. 5) is artificial and represents a spatial transect (20 equispaced locations). A first variable creates a perfect spatial gradient; the values (1 to 20) were standardised to form variable (1 to 20). A second variable is made of random values drawn from a uniform distribution; its standardised values form variable (random). The "difference" variable, in the right-hand panel, clearly displays a spatial gradient, materialised by the OLS regression line which can be construed as a trend surface "map" in one spatial dimension.

Example 9: two-dimensional data

The second example is drawn from a long-term sampling programme to study changes in a reef fish community and in their habitat (Galzin *et al.*, 1993). Ten sampling corridors, each (2 m × 50 m), have been established in a (100 m × 200 m) area of the barrier reef, in the north-west part of the high volcanic island of Moorea in French Polynesia. Corridors were used instead of square plots because this is the best way for divers to survey the fish. Counts within each corridor were attributed to the geographic centre of the corridor.

We will here examine (with permission from R. Galzin) whether there have been changes in the distribution of the % coverage of rubble over the sampled area between 1992 and 1993 (Fig. 6). The abiotic components of the substrate are reworked from year to year by hydrodynamic processes; the main hydrodynamic gradient, from the reef outwards, is in the "y" direction of the reference coordinate system. The surface of differences (1992 values – 1993 values) can be modelled by a polynomial ($R^2 = 0.84$, $p < 0.0077$) in x, x^2 and xy^2. The map describing this trend surface, presented in Figure 7, clearly shows that the difference between years in the south-to-north direction, in the left-hand part of the sampled area, is opposite in sign to that in the right-hand part (convex on the left, concave on the right).

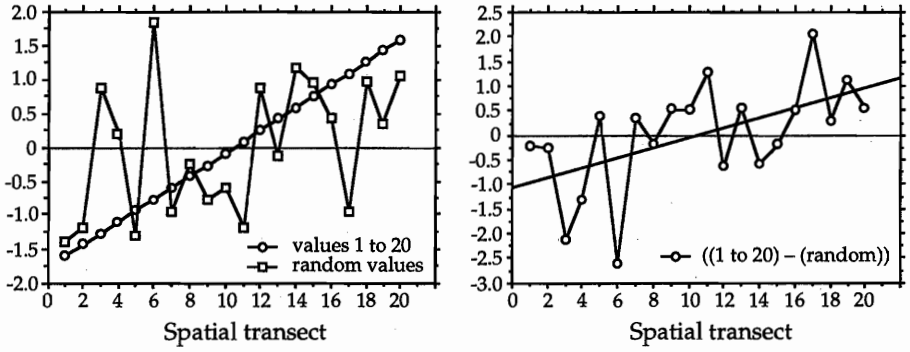

Figure 5

Artificial data set. Left: two variables (a gradient, and a set of random values) represent a variable sampled in two occasions. Right: the difference between them (broken line) can be modelled by a regression line.

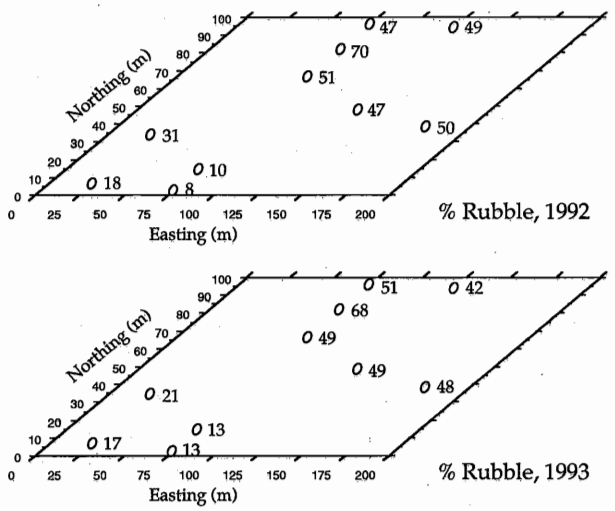

Figure 6

Percentage of rubble, in 1992 and 1993, at ten stations sampled by Galzin et al. (1993) in the Tiahura reef sector of the high volcanic island of Moorea, French Polynesia.

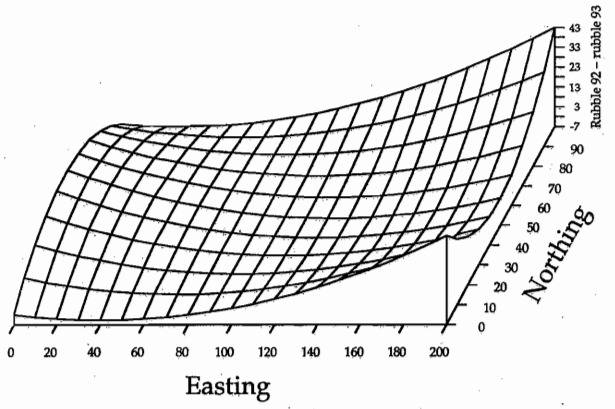

Figure 7

Trend-surface map of the difference in % rubble over the study area, between 1992 and 1993. The model for centred coordinates (x = easting -85, y = northing -48) is: (rubble 92 − rubble 93) = $-1.0117 - 0.1187\,x + 0.0016\,x^2 + 0.0001\,xy^2 + \varepsilon$.

DISCUSSION

From the methods discussed above, the following recommendations can be derived about sampling designs. They are of prime importance when planning a data collection with the purpose of comparing surfaces.

1. The situation where locations are the same from campaign to campaign leads to more powerful analyses. From a sampling point of view, in general more information is obtained about the variation strictly attributable to time by resampling the same locations instead of different ones. From the analysis of variance point of view, the part of the variation corresponding to locations can be pulled out of the residual component of the ANOVA table (compare equations 4 and 11), and if there are differences among locations (as there usually are), this will increase the F statistics of the main effects. So, fixed locations through space usually lead to a major increase in power.

2. On the other hand, it is desirable to have replicate measurements. When available, replicate variation is usually used as the error mean square in the ANOVA table. Replicates, however, must be independent of one another and come from the same reference population (*i.e.* they must estimate the same population mean). Strictly speaking, true replicates are available only when controlled experiments are performed. There are difficulties with survey data, because samples that are too close to one another may not be independent, while if they are too far the local population means may differ. So, to obtain an estimate of the measurement error, we must measure the variable of interest within some local neighbourhood. Replicates randomly selected from within a defined study location (local neighbourhood) represent the best information available to estimate the error structure of the mean for that location. If locations are suitably well-defined, like coral heads on a reef for instance, then randomly-selected small quadrats allow to sensibly measure the error on the mean of the variable of interest. If the environment is more continuous, then the sampling unit representing a location must be as precisely defined in shape and size as the replicate sampling unit. In this way replicates can be taken randomly within a location sampling unit. The size of the location sampling unit defines the smallest spatial scale of the changes you are trying to detect. The practical rule in that situation is that if replicates are allocated randomly at the scale for which we want an error estimate, then they represent the best available information about the error they are trying to estimate.

3. Stratified sampling is good, but finer-scale changes can be detected by keeping the strata small.

So, in summary, when planning a sampling programme,

• Use the same locations from campaign to campaign.

• The best coverage of geographic space is obtained by systematic sampling, followed by stratified. Simple random sampling is the worst.

• Some level of replication is always desirable. It can be obtained by true replication, by pooling neighbouring locations of a systematic sampling grid, or by using small strata. A balance has to be found between power to detect changes (obtained from replication) and spatial resolution (obtained from more locations), as discussed further below. The extreme situations, all replicates at a single location, or no replicate at all, should be avoided.

A final point concerns the scales at which each design allows to detect change. In Table 1, the first row (same locations sampled with replication) is the only one that offers detection of effects at the scale of individual locations, as well as an overall time effect, unless an assumption of Poisson error can be justified. The scale of the effects that can be detected is given by the interaction terms; for instance, lt_{jk} and $\beta_j t_k$ shows that changes can be detected at the level of the location, st_{ik} at the level of the strata or pairs, and so on.

Among the remaining stratified designs, it is hard to determine whether model 4 is better than 7 or *vice versa*. In

any case, both seem more powerful at detecting a stratum effect than model 11, because the location-within-strata effect is pulled out of the residual mean square in both of these cases, becoming part of the hypothesis rather than part of the residual mean square. For systematic designs, and for similar reasons, models 4 and 8 seem more powerful than model 11 at detecting pseudo-strata or box effects.

With random sampling, models 5 or 6 seem more capable of detecting a locations effect than model 10, because they are explicitly modelling changes at each location, while 10 is smoothing over all locations. Power, however, depends on the situation. Models 5 and 6 will be more powerful if the changes through time can be modelled by low-order polynomials; this is the case when the surfaces are not random or chaotic, and there are few locations and many sampling periods. On the other hand, if the changes are widespread and smooth, and the spatial trend surfaces at each time period is well modelled by a low-order polynomial, then model 10 will be more powerful if there are many locations and few sampling times.

Everything we have described for ordinary ANOVA can also be computed by ordinary least squares regression (as explained for instance by Draper and Smith, 1981). As a consequence, these techniques can naturally be extended to tables of multivariate dependent variables using canonical analysis, and especially canonical correspondence analysis in the case where the dependent variables are multivariate tables of species abundance data (ter Braak, 1986, 1987, 1994).

Acknowledgements

The research that led to writing this paper was motivated by discussions between Dr. Rick D. Pridmore and the authors, during the workshop "Spatial Modelling of Benthic Communities: Matching Pattern to Process" organised and financed by the National Institute of Water and Atmospheric Research (NIWA) of New Zealand in Hamilton, NZ, from 3 January to 18 February 1995. The present paper is thus a contribution of that workshop.

REFERENCES

Anderson D.R., K.P. Burnham, G.C. White (1994). AIC model selection in overdispersed capture-recapture data. *Ecology* **75,** 1780-1793.

Bellehumeur C., P. Legendre. Multiscale sources of variations in ecological variables: modelling spatial dispersion, elaborating sampling designs. Submitted.

Bellehumeur C., P. Legendre, D. Marcotte. Variance and spatial scales in a tropical rain forest: changing the size of sampling units. Submitted.

Burrough P.A. (1987). Spatial aspects of ecological data, *In: Data analysis in community and landscape ecology*, R.H.G. Jongman, C.J.F. ter Braak, O.F.R. van Tongeren, eds. PUDOC, Wageningen, 213-251.

Cormack R., J.R. Skalski (1992). Analysis of coded wire tag returns from commercial catches. *Can. J. Fish. Aquat. Sci.* **49,** 1816-1825.

Crawley M.J. (1993). *GLIM for ecologists*. Blackwell, Oxford.

Crowder M.J., D.J. Hand (1990). *Analysis of repeated measures*. Chapman and Hall, London, vii + 257 p.

Draper N.R., H. Smith (1981). *Applied regression analysis, 2nd ed.* Wiley, New York, xiv + 709 p.

Galzin R., J.-P. Marfin, B. Salvat (1993). Long term coral reef monitoring program: heterogeneity of the Tiahura barrier reef (Moorea, French Polynesia). *Galaxea* **11,** 73-91.

Gaston K.J., B.H. McArdle (1994). The temporal variability of animal abundances: measures, methods and patterns. *Phil. Trans. R. Soc. Lond. B.* **345,** 335-358.

Isaaks E., R.M. Srivastava (1989). *An introduction to applied geostatistics*. Oxford University Press, New York, xix + 561 p.

Lanctôt J.-L., P. Legendre, B. Salvat (1997). How do coral reef Gastropods feel about nuclear blasts? A long-term study of the effects of man-made perturbations. *Oceanologica Acta* **20**, 1 (this issue).

Legendre P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology* **74,** 1659-1673.

Legendre P., M.-J. Fortin (1989). Spatial pattern and ecological analysis. *Vegetatio* **80,** 107-138.

Liang K.-Y., P. McCullagh (1993). Case studies in binary dispersion. *Biometrics* **49,** 623-630.

McArdle B.H., R.G. Blackwell (1989). Measurement of density variability in the bivalve *Chione stutchburyi* using spatial autocorrelation. *Mar. Ecol. Prog. Ser.* **52,** 245-252.

McArdle B.H., K.J. Gaston, J.H. Lawton (1990). Variation in the size of animal populations: patterns, problems and artifacts. *J. Anim. Ecol.* **59,** 439-454.

McCullagh P., J.A. Nelder (1989). *Generalized linear models, 2nd ed.* Chapman and Hall, London, xix + 511 p.

Quenouille M.H. (1949). Problems in plane sampling. *Ann. Math. Stat.* **20,** 355-375.

Ripley B.D. (1981). Spatial statistics. Wiley, New York, x + 252 p.

Scherrer B. (1982). Techniques de sondage en écologie, *In: Stratégies d'échantillonnage en écologie*, S. Frontier, éd. Collection d'Écologie n° 16, Masson, Paris, 63-162.

Student [W. S. Gosset] (1914). The elimination of spurious correlation due to position in time or space. *Biometrika* **10,** 179-180.

ter Braak C.J.F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67,** 1167-1179.

ter Braak C.J.F. (1987). Ordination, *In: Data analysis in community and landscape ecology*, R.H.G. Jongman, C.J.F. ter Braak, O.F.R. van Tongeren, eds. PUDOC, Wageningen, 91-173.

ter Braak C.J.F. (1994). Canonical community ordination. Part I: Basic theory and linear methods. *Écoscience* **1,** 127-140.

Thompson S.K. (1992). *Sampling*. Wiley, New York, 368 p.