Australian Journal of Ecology (1998) **23**, 158–167

Partitioning the variation among spatial, temporal and environmental components in a multivariate data set

M. J. ANDERSON[1] AND N. A. GRIBBLE[2]

[1]*Centre for Research on Ecological Impacts of Coastal Cities and School of Biological Sciences, Marine Ecology Laboratories A 11, University of Sydney, Sydney, NSW 2006 and* [2]*Queensland Department of Primary Industry Fisheries, Northern Fisheries Centre, PO Box 5396, Cairns, QLD 4870, Australia*

Abstract We propose a method of partitioning the variation in a multivariate set of data according to (i) environmental variables, (ii) variables describing the spatial structure in the data and (iii) temporal variables. This method is an extension of an existing method for partialling out the spatial component of environmental variation, using canonical analysis. Our proposed method extends this approach by including temporal variables in the analysis. Thus, the partitioning of variation for a data matrix of species' abundances or biomass can include, by our methodology, the following components: (1) pure environmental, (2) pure spatial, (3) pure temporal, (4) pure spatial component of environmental, (5) pure temporal component of environmental, (6) pure combined spatial/temporal component, (7) combined spatial/temporal component of environmental and (8) unexplained. In addition, permutation testing accompanying the analyses allows tests of significance for the relationship between the different components and the species data. We illustrate the method with a set of survey data of penaeid species (prawns) obtained on the far northern Great Barrier Reef, Australia. This extension is a useful tool for multivariate analysis of ecological data from surveys, where space, time and environment commonly overlap and are important influences on observed variation.

Key words: canonical correspondence analysis, multivariate analysis, prawn biomass, survey data.

INTRODUCTION

Spatial and temporal variation in the distribution and abundance of organisms is an inherent property of ecological systems. Ecology could be considered to be a study not of how things are but rather of how things change in space and time. An assessment of the relative importance of biotic or abiotic factors influencing species should consider, whenever possible, the inherent spatial and temporal variation of assemblages (Legendre & Fortin 1989; Butler & Chesson 1990; Underwood 1991, 1992, 1996; Borcard *et al.* 1992; Dutilleul 1993).

Borcard *et al.* (1992), followed by Borcard and Legendre (1994) and Legendre and Borcard (1994), demonstrated a new method for partitioning the variation among environmental and spatial components in ecological data sets using canonical ordination techniques. Their demonstrations highlighted the importance of considering the potential overlap in spatial and environmental components. Thus, they identified four different components in their partitioning scheme: pure spatial, pure environmental, spatial component of environmental and undetermined. A recent application of their proposed technique is found in Økland and Eilertsen (1994).

We propose here that their methodology be extended to include a third matrix of explanatory variables which correspond to temporal variation. Although partitioning among more than two independent matrices has been done previously (e.g. Pinel-Alloul *et al.* 1995), the concept of the third matrix containing temporal variables is new. This extension results in eight different components of variation (Fig. 1). These components are as follows.

(1) Pure environmental, *E*: Environmental variation that is neither spatially nor temporally structured (i.e. that fraction of the variation which can be explained by the environmental descriptors independently of any spatial or temporal structure).

(2) Pure spatial, *S*: Spatial patterns in the species data that are independent of any temporal or environmental variables included in the analysis.

(3) Pure temporal, *T*: Temporal patterns in the

Accepted for publication June 1997.

species data that are independent of any spatial or environmental variables included in the analysis.

(4) Pure spatial component of environmental, *SE*: As described in Borcard *et al.* (1992), this component is the overlap in the (non-temporal) variation explained by space and environmental variables. It can be considered as the spatially structured component of environmental variation and/or the component of spatial variation that is linked to some environmental variable(s).

(5) Pure temporal component of environmental, *TE*: Similar to (4), this component is that fraction of (non-spatial) variation explained by temporal and environmental variables, or the temporally structured environmental variation in the species data.

(6) Pure combined spatial/temporal component, *ST*: The fraction of the variation in the species data that is not related to the environmental variables, but which can be attributed to pure combined spatiotemporal patterns.

(7) Combined spatial/temporal component of environmental, *STE*: The fraction of the variation that can be explained by the combined action of spatial, temporal and environmental variables.

(8) Unexplained, *U*: This component is the remainder in the analysis; that is, the fraction that can not be explained by either the spatial co-ordinates (or a polynomial function of them), the temporal variables or the environmental variables included in the analysis.

Components 1, 2, 4 and 8 were previously described by Borcard *et al.* (1992). Of the other components, 3 and 5 are easily interpreted as the 'time' counterparts to the 'space' components described in 2 and 4. Components 6 and 7, however, are conceptually different from those that have been presented previously. Component 6 includes the variation attributable to combined spatial and temporal influences in the analysis. For example, a study of organisms inhabiting marine rocky intertidal zones might determine that their behaviours (in terms of foraging or movements of organisms) were related to their spatial location within the study area (i.e. high on the shore *vs* low on the shore). However, these patterns could also be related to the submergence of the organisms according to the temporal cycles of the tides. Thus, the combined influence of temporal and spatial factors would be important in explaining variations in their behaviour. This would be reflected in the analysis by the relative size of component 6. If, in such an intertidal study, a particular environmental variable had been measured that was also related to this spatial/temporal phenomenon, such as 'submergence time', the behaviours might be explained by the combined action of spatial, temporal and environmental variables measured. This would be an example of component 7. It should be noted, as was indicated by Borcard *et al.* (1992), that this general proposed methodology does not show causal relationships, but only indicates the quantitative overlap of the different components of variation for variables that are included in any particular study.

Further, the approach of Borcard *et al.* (1992), including the extension of their method proposed here, is not strictly a partitioning of *variance* among orthogonal components, such as in analysis of variance (ANOVA). In a structured multifactorial experimental design (with independent replicates sampled within levels of particular factors), the analysis of main effects and their interactions can be done by partitioning of orthogonal (non-overlapping) components. Such a structured design (with replication) allows unbiased estimates of variances for effects in the design to be calculated.

In the initial stages of some ecological research, where little is known *a priori* about the ecosystem being studied, large, relatively unstructured surveys are sometimes used. In these situations, the hypotheses of interest centre on broad correlative patterns among groups of variables. In addition, ecologists have hypotheses about the relative importance of particular groups of predictor variables in explaining variation in the response variables of interest. For example, a researcher might hypothesize that the distribution of a particular assemblage of animals is positively correlated with certain features of the habitat. Data are then gathered across the (usually large) area of interest, at different times and different places. Such data are inextricably confounded in space and time, with no replication through time at a particular place and no more than one place sampled at any particular time. This lack of replication makes it impossible to calculate unbiased estimates of variance for particular effects (whether spatial, temporal, environmental or some interaction of these). As a result, the analysis of the data should take into account such intrinsic confounding. This is analogous to the consideration of correlation among explanatory variables (X_1, X_2, X_3 etc.) in multiple regression of a single response variable (Y).

The analysis can be represented conceptually by a Venn diagram, as is commonly used in set theory (e.g. Kemeny *et al.* 1958; Fig. 1). Here, each component of variation can be drawn in proportion to the area that it covers in the diagram. The individual spatial, temporal and environmental components are easily visualized as overlapping circles that we can think of as 'areas of influence' of each set of variables.

To illustrate the proposed method, we present data from surveys of the cross-shelf distribution of penaeid species (prawns or shrimp) on the far northern Great Barrier Reef (GBR), Australia. The surveys were carried out by the fisheries divisions of CSIRO and the Queensland Department of Primary Industries between 1992 and 1994. This work was part of a collaborative project investigating the environmental effects of trawling on the GBR and provided a unique opportunity to study relatively undisturbed demersal assemblages of penaeid species in heterogeneous cross-shelf habitats across an area of *ca* 10 000 square nautical miles of the GBR.

The distribution and abundance of penaeids are known to be affected by factors such as substrate type and water depth (Somers 1987, 1994). The complexity of the habitat has also been shown to affect the distribution of prawn species (Dall *et al.* 1990; Poiner *et al.* 1997). Thus, measurement of such environmental influences and consideration of their spatial structuring were considered important for the survey.

In addition, abundances of prawns may be subject to seasonal, lunar, or diel cycles (Dall *et al.* 1990). Trainor *et al.* (1991) documented that prawn catches on the northern Queensland coast were seasonal, that they vary between years and are influenced by the lunar cycle. Furthermore, anecdotal reports from prawn trawler operators suggest that prawns are less easy to catch after moonrise than before moonrise on any particular night (Gribble, unpubl. data). Changes through time in the availability or susceptibility of prawns to capture (catchability) during the surveys may have affected the number sampled, independently of (or in combination with) any underlying spatial or environmental influences. Thus, it was considered that temporal variation across three scales (between years, between months and within days according to moonrise) was an important component to include in an analysis of the data from the surveys.

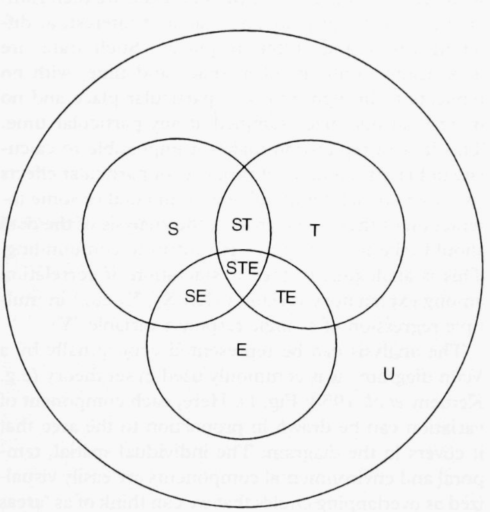

Fig. 1. A Venn diagram showing the partitioning of variation according to three sets of independent variables, spatial (S), temporal (T) and environmental (E). The largest circle is the set corresponding to the total variation in the dependent (species) data. Each area of overlap of the three smaller circles is representative of the intersection of the three sets, S, T and/or E in terms of their explained variation. For example, the intersection of circles S and T, but where the circle E does not intersect, represents the spatially and temporally structured explained variation (ST). The area of the largest circle (total variation) that the three smaller circles do not cover represents the unexplained variation (U).

We wish to emphasize that the sampling design of this survey, as outlined below, was not done in a particularly structured way with regard to temporal, environmental and spatial variables. None of the potential explanatory variables were orthogonal to one another, in the sense of a structured ANOVA design. For example, the data were not collected during the same time periods from one cruise to the next, they were not collected at the same locations from one time to the next etc. Thus, traditional approaches of variance partitioning using multivariate ANOVA could not be considered. The sets of explanatory (or predictor) variables were inextricably confounded with one another.

The nature of our hypotheses is correlative rather than causative. The first (H_1) is that there is a significant relationship between each of the three sets of explanatory variables (spatial, temporal and environmental) and the species data, and the second (H_2) is that there is some overlap in the amount of variation explained by each of the three components, which reduces their independent importance in explaining variation in the species data (i.e. the sets of explanatory variables are themselves related). There is no way of attributing the causes of patterns in the species data with such an unstructured design. However, a general additive multiple regression approach can be used to test hypotheses about the relationships among the sets of variables (e.g. Whittaker 1984). For ecological surveys, this may be particularly important for determining which variables should be considered as factors for the design of mensurative or manipulative experiments.

The purpose of this paper is to demonstrate our proposed extension of the method introduced by Borcard *et al.* (1992). Quantification of explained variation and its partitioning is achieved, as in Borcard *et al.* (1992), by constrained and partial canonical correspondence analysis (CCA; ter Braak 1986, 1987). The nature of this extension is previously undescribed and we consider the technique to be generally applicable to many more situations involving complex spatiotemporal ecological data sets. In addition, we highlight the fact that the computer program CANOCO™ (ter Braak 1988) can be used for permutation tests of the significance of individual components in explaining the variation in the species data.

METHODS

Survey and data

The far northern section of the Great Barrier Reef Marine Park extends from 10°41′S, 145°00′E to

15°00′S, 145°00′E. Within the park, a 10 000 km^2 cross-shelf area was closed to fishing and set aside for research. Adjacent areas to the north and south of the cross-shelf closure are open to fishing and heavily trawled (Poiner *et al.* 1997). The cross-shelf study area included both the closed and open zones and consisted of four habitat types, coded as a semi-quantitative ranked variable (with ranks increasing from inshore to offshore portions of the reef) as: (i) inshore lagoon, (ii) inshore reefs/islands, (iii) mid-shelf shoal/reef, and (iv) offshore lagoon extending to the outer barrier ribbon reefs. The study area could be broadly categorized according to these habitat types but, within these, microhabitats made up a complex mosaic according to depth, sediment type, and three-dimensional structure (corals).

Selection of sites for sampling was done by three computerised steps. First, the whole study area was subdivided into a grid of squares, each measuring one square nautical mile, resulting in 21 600 squares. Second, 10–15 contiguous squares were assigned randomly into a group. These groups were spread across all habitat zones, but no group could overlap with another group, a reef or an island (although groups could abut). Finally, within each group, a site was randomly chosen to be sampled. In total, 126 sites (out of a possible 200) were sampled across the study area during two research cruises. The cruises were carried out at the same time of year for two consecutive years. The latitude and longitude at each site were recorded.

The first cruise lasted from 5 April to 1 May 1992. The second cruise (1993) was more complex, caused by the failure of engines on the boat. It lasted for 9 days initially (7–16 March), followed by two further periods (24 March–1 April and 25 April–7 May), each after intervening periods required for repair of the vessel.

Samples of prawns were collected by a trawler towing two 4-fathom 'Florida Flyer' nets over the stern in dual Otter trawl configuration. Each haul ($n = 126$; there was one haul at each site) was approximately 30 min in duration and covered 2.78 ($\pm$ 0.05) km. The hauls were done at night, but did not occur on all consecutive nights during the cruises, due to occasional bad weather (cyclones) or the need to repair gear. Depth profiles were recorded simultaneously by a depth sounder. These profiles yielded both mean depth and the standard error about the mean, which was used as a measure of the complexity (rugosity) of the sea bottom. The area swept by each haul was estimated to be 28 836 m^2, assuming a 70% spread of each net (Sparre *et al.* 1989). Prawns were sorted from both nets, bagged separately and snap frozen for transport to the Northern Fisheries Centre in Cairns for further analysis.

Twenty species of penaeid prawns were identified from the samples. The biomass of prawns of each species was recorded for each site and calculated as weight of prawns per hectare (Sparre *et al.* 1989). The 14 most abundant species were included in the analysis described here. The other six species identified in the study occurred sporadically (in only one or two of the 126 hauls) and in each case their numbers per swept area were very low (i.e. less than 0.5 animals per hectare), always making up less than 5% (in abundance) of the hauls in which they occurred.

Sediment type was mapped at each site as percentage of mud (i.e. that portion of the sediment sample that is below 63 μm in particle size; Somers 1987), along with water depth and rugosity. These were considered likely to be the major abiotic factors affecting the spatial distribution of a given species (Dall *et al.* 1990; Somers 1994). A further consideration was the possible effect of trawling (both in the open zones and, illegally, in the cross-shelf closure) on prawn distribution and abundance (Poiner *et al.* 1997). For the current analysis, fishing effort at each site was summarized by a rank-ordered, semi-quantitative variable as (i) light, (ii) medium or (iii) heavy. Thus, five environmental variables were included in the analysis: habitat type, percentage of mud, mean water depth, rugosity and fishing effort.

Temporal effects on prawn abundance and catchability were considered at three scales with three temporal variables: (i) variation between the two years of sampling, using the Julian date (expressed as total number of days since 1 January 1900); (ii) lunar periodicity using the proportion of the lunar cycle (29.53 days) at the time of each haul; and (iii) diel periodicity using the time between the beginning of the haul and moonrise (this variable could be negative or positive as some hauls occurred before moonrise and others occurred after moonrise, respectively).

The first of these (variation between years) requires some additional explanation. The greatest variation in this variable is the distance in time (i.e. the number of days) between the first cruise (year 1) and the second set of cruises (year 2), which is why it is considered as a variable for the between-year variation. Day-to-day differences (e.g. from the beginning of a cruise to the end of a cruise) are also recorded by this variable, but are relatively unimportant compared to the inter-annual difference. It is quite simply a quantitative variable recording 'time' (in days) in a real, intuitive sense.

Greater details of the sampling methods and data sets for the study outlined here are available on request and can be found in Poiner *et al.* (1997).

Data analysis

Canonical correspondence analysis was used for data analysis. This technique was deemed appropriate because it assumes a unimodal distribution for the abundances of species across environmental gradients (e.g. ter Braak 1986, 1987), as opposed to a linear model. For the multivariate data from the prawn survey, it was reasonable to consider that species would show such a non-linear response across the entire area being sampled (ter Braak 1985).

First, the geographical co-ordinates (latitude and longitude) were centred on their means to reduce collinearity in the spatial data. The raw values for latitude and longitude were used and were not converted to great-circle distances because (i) the distance across one degree of latitude is about the same as that for longitude in this region (i.e. one degree ≈ 60 nautical miles) and (ii) the size of the study area was not large enough to result in large distortions due to global sphericity. The matrix of spatial variables was then calculated by including all terms for a cubic trend surface regression, with x = longitude (centred) and y = latitude (centred) (i.e. terms included were: x, y, x^2, xy, y^2, x^3, x^2y, xy^2 and y^3; see Legendre 1990; Borcard *et al.* 1992). These nine terms were then included in a procedure of 'forward selection' using CANOCO (ver. 3.1; ter Braak 1988). This procedure allows the selection of variables by sequential testing and helps to reduce inflation of explained variation due to pure chance by the retention of redundant parameters in the model (Borcard *et al.* 1992).

In general, forward selection of environmental or temporal variables can be done, although for the present example this did not result in any elimination of variables. Four fundamental matrices were, therefore, used for the analysis: species variables, environmental variables, spatial variables and temporal variables. Special matrices were also constructed for: spatial + temporal variables, environmental + spatial variables and environmental + temporal variables.

In a CCA of a species matrix, where an environmental (or other) matrix is used to constrain the analysis, the sum of canonical eigenvalues corresponds to the amount of variation in the species data explained by variables in the environmental (or other) matrix. If the analysis includes a matrix of covariables, then these variables are partialled out of the analysis. The total trace, or sum of all eigenvalues, obtained by an unconstrained correspondence analysis (CA) of the species data provides a measure of the total variation in the species data. The sum of canonical eigenvalues obtained by any CCA is a proportion of the total trace obtained by CA on that data (Borcard *et al.* 1992).

A series of steps, involving constrained and/or partial CCA, were done using CANOCO (Table 1). For each step, the value of the sum of canonical eigenvalues for the analysis was recorded. The proportion of the total variation that this sum represented was then calculated and multiplied by 100 to obtain a value for the percentage explained variation for each step. In addition, for each step, unrestricted permutation tests (with 1000 permutations) were done of the overall trace statistic (for details of this statistic and the permutation method used in CANOCO see ter Braak 1990, 1992). The permutation test for this statistic indicates the significance of the effects of constraining variables on the species variables (removing the effects of covariables, when present).

The values corresponding to the percentage variation for each component of variation of interest in Fig. 1 were then calculated (numbers in square brackets refer to percentages of variation obtained in the corresponding steps shown in Table 1). The method of calculation is a straightforward extension of the Borcard *et al.* (1992) method. For the general mathematical form of such calculations, refer to Whittaker (1984). The percentage of total explained variation (Ω) is equal to [1] + [7] + [12], or [2] + [4] + [12] or [3] + [5] + [9]. Thus, the unexplained variation, $U = (100\% - \Omega)$. The pure environmental component, E, is equal to [6]; the pure spatial component, S, is equal to [9]; and the

Table 1. Steps in the analysis done using CANOCO

Step	Description
[1]	CCA of species matrix, constrained by the environmental matrix
[2]	CCA of species matrix, constrained by the spatial matrix
[3]	CCA of species matrix, constrained by the temporal matrix
[4]	CCA of species matrix, constrained by the environmental matrix, with spatial variables treated as covariables
[5]	CCA of species matrix, constrained by the environmental matrix, with temporal variables treated as covariables
[6]	CCA of species matrix, constrained by the environmental matrix, with spatial + temporal variables treated as covariables
[7]	CCA of species matrix, constrained by the spatial matrix, with environmental variables treated as covariables
[8]	CCA of species matrix, constrained by the spatial matrix, with temporal variables treated as covariables
[9]	CCA of species matrix, constrained by the spatial matrix, with environmental + temporal variables treated as covariables
[10]	CCA of species matrix, constrained by the temporal matrix, with environmental variables treated as covariables
[11]	CCA of species matrix, constrained by the temporal matrix, with spatial variables treated as covariables
[12]	CCA of species matrix, constrained by the temporal matrix, with environmental + spatial variables treated as covariables

pure temporal component, T, is equal to [12]. For the other components, it was necessary to consider some intermediate values.

First, the total spatially structured environmental variation is calculated as [1] – [4], or [2] – [7]. It is important to note that this estimate still contains some temporal component. This value will be referred to as SE_T to denote that a temporal component is still in the estimate. One can see from Fig. 1 that this value corresponds to $STE + SE$ and is the total overlap of the 'spatial' and 'environmental' circles of influence. Similarly, we consider the total temporally structured environmental variation, which is $TE_S = STE + TE$ and still contains some spatial component. TE_S can be calculated as [1] – [5] or [3] – [10]. Finally, there is the total spatial/temporal variation, which we denote by $ST_E = STE + ST$ and still contains some environmental component. It can be calculated as [2] – [8] or [3] – [11].

The central value that we need to obtain the real estimates of SE, TE and ST is the spatially and temporally structured environmental variation, STE (shown at the centre of our Venn diagram; Fig. 1). It is clear from the geometry of the areas identified in our diagram that the total variation explained by the spatial variables (Ω_S) is equal to $S + SE + ST + STE$. We know that $ST_E = STE + ST$ and that $SE_T = STE + SE$ therefore:

$$STE = S + ST_E + SE_T - \Omega_S$$

Similarly, from the total variation explained by the environmental variables (Ω_E) or the total variation explained by the temporal variables (Ω_T) we can obtain

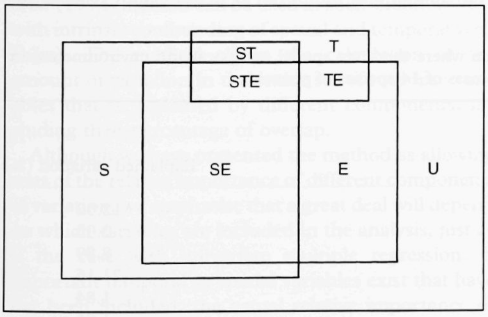

Fig. 2. Graphical representation of the partitioning of variation for the penaeid prawn data as in Table 3. The area of the largest rectangle is directly proportional to the total variation in the species data. The three smaller rectangles are directly proportional to the variation in the data explained by the spatial (S), temporal (T) or environmental (E) variables. Areas of overlap are also directly proportional to the percentages of explained variation of particular components as described in the text and labelled in Table 3. The area of the largest rectangle which is not covered by any of the other rectangles is directly proportional to the unexplained variation (U).

independent calculations for STE as $STE = E + SE_T + TE_S - \Omega_E$, or $STE = T + ST_E + TE_S - \Omega_T$, respectively. The calculations for these three equations in terms of the steps outlined in Table 1 (all of which should give the same results, apart from rounding error) are, therefore:

$$STE = [9] + ([2] - [7]) + ([2] - [8]) - [2], \text{ or}$$

$$STE = [6] + ([1] - [4]) + ([1] - [5]) - [1], \text{ or}$$

$$STE = [12] + ([3] - [10]) + ([3] - [11]) - [3]$$

Once the value for STE has been obtained, then calculations for the remaining three components are straightforward:

$$ST = ST_E - STE = ([2] - [8]) - STE$$

$$SE = SE_T - STE = ([1] - [4]) - STE$$

$$TE = TE_S - STE = ([1] - [5]) - STE$$

This finishes the analysis, partitioning the variation into its eight components, and allows a complete Venn diagram to be drawn for the data.

RESULTS

For the 14 species of penaeid prawns, forward selection of spatial variables resulted in the following six terms of the spatial polynomial being included as variables in the spatial matrix: x, y, xy, y^2, xy^2 and y^3. The total variation in the species data (sum of all eigenvalues in the CA) was 1.901. The percentage of total explained variation (Ω) was calculated as 50.5%. Consequently, the unexplained variation, U, was 49.5%. The sum of canonical eigenvalues and corresponding percentage of explained variation for each of the CCAs indicated by steps [1]–[12] are shown in Table 2. This table also shows the calculated probabilities (P) obtained from Monte Carlo tests of significance for each CCA, with 1000 permutations done for each test.

The calculations of the values of each component of variation for these data are shown in Table 3. For the data described in Borcard *et al.* (1992), where there were only two sets of explanatory variables, it was possible to represent the results using a one-dimensional bar diagram (e.g. see figs 3 and 4 in Borcard *et al.* 1992). In the present application, it is not possible to show the partitioning of variation elegantly along one dimension, due to the four areas of overlap in explained variation (e.g. SE, ST, TE and STE; Fig. 1). At least two dimensions are required. Figure 2 shows the results for the penaeid prawn data. In Fig. 2, the areas of the individual rectangles correspond to the components of variation as presented in Table 3.

Despite wishing to be consistent with the circular Venn diagrams in Fig. 1, we have chosen to use rectangles instead of circles for Fig. 2 for two reasons. First, circles have only one linear dimension to them (their radius), so cannot be used to give proper representation of all areas required. Second, ellipses could have been used, but calculating their area and the areas of particular regions of overlap of various shapes is much more complex than the calculations required for rectangular representation. The reader should bear in mind that the particular choice of width versus length for the first few rectangles drawn will influence the positioning and shape of subsequent rectangles. So, some illusion as to the relative size of areas may result from these choices (e.g. the long thin 'arm' of the area corresponding to the 'T' makes it look smaller than it might have looked in a different representation; Fig. 2). We expect that any such illusion will not generally mask the patterns of the true values, which are calculated explicitly in any event.

Table 2. Summary of results of constrained and partial canonical correspondence analyses for data on biomass of 14 species of penaeids

Step in analysis	Value in calculations	Sum of canonical eigenvalues	Explained variation	P
[1]	Ω_E	0.659	34.67	0.001
[2]	Ω_S	0.717	37.72	0.001
[3]	Ω_T	0.151	7.94	0.001
[4]		0.204	10.73	0.001
[5]		0.571	30.04	0.001
[6]	E	0.169	8.89	0.001
[7]		0.262	13.78	0.001
[8]		0.640	33.67	0.001
[9]	S	0.238	12.52	0.001
[10]		0.064	3.37	0.003
[11]		0.074	3.89	0.001
[12]	T	0.039	2.05	0.058

The analyses corresponding to numbered steps are outlined in Table 1. The percentage of explained variation was calculated as a proportion of the total trace of the species data = 1.90077. The significance tests were done using the full model permutation method of ter Braak on the overall trace statistic (ter Braak 1990, 1992).

In general, the greatest variation in the prawn data was explained by the combined spatial/environmental component, with the temporal component explaining much less of the variation in the species data (Table 3, Fig. 2). Permutation tests showed that environmental, spatial and temporal variables, when considered separately, each contributed significantly to explaining variation in the species data (Table 2: [1], [2] and [3]). Although there was extensive spatial/environmental overlap, each of the pure spatial and environmental components explained significant portions of the variation, when the other components were partialled out of the analysis (Table 2: [6] and [9]). In contrast, when the spatial and environmental variables were partialled out of the analysis, temporal variables no longer explained a significant percentage of the variation (Table 2: [12]).

DISCUSSION

We present an extension of the method presented by Borcard *et al.* (1992), which increases the number of sets of independent variables, in a multiple regression approach using CCA, from two to three: environmental, spatial and temporal. It is important to note that

Table 3. Summary of calculations for the partitioning of variation where there are spatial, temporal and environmental explanatory variables, including results obtained for the data on biomass of 14 species of penaeids

Component	Calculation	Explained variation (%)
Spatial = S	[9]	12.52
Temporal = T	[12]	2.05
Environmental = E	[6]	8.89
SE	$SE_T - STE$ = [1] − [4] − STE	21.15
TE	$TE_S - STE$ = [1] − [5] − STE	1.84
ST	$ST_E - STE$ = [2] − [8] − STE	1.26
STE	[9] + ([2] − [7]) + ([2] − [8]) − [2], or [6] + ([1] − [4]) + ([1] − [5]) − [1], or [12] + ([3] − [10]) + ([3] − [11]) − [3]	2.79
Total explained, Ω	[1] + [7] + [12], or [2] + [4] + [12], or [3] + [5] + [9], or S + T + E + SE + TE + ST + STE	50.50
Unexplained	100% − Ω	49.50
Total variation	Total trace of CA = 1.90077	100.00

The numbers in square brackets refer to the numbered steps in the analysis described in Tables 1 and 2.

the methodology used here is not a linear ANOVA model, where manipulative experimental or structured sampling occurs and variances associated with particular effects are estimated. Rather, the percentages of explained variation for the individual components are purely correlative, and reflect also the choice of variables that happen to be measured and/or included in the analysis. The logical interpretation of results follows that of a univariate multiple regression, but for multivariate data. That is, different groups of explanatory variables (environmental, spatial or temporal) have different capacities to predict or explain patterns in the multivariate response variables (species data) and there is some overlap (correlation structure) among these groups of explanatory variables with regard to that capacity.

For ecological surveys, such as that for the penaeid prawn data analysed here, often there will be a sampling design reflecting hypotheses about the relationships among groups of variables (e.g. species variables, in occurrences, abundances or biomass, are correlated with a set of environmental variables). Due to the sheer size of the areas of interest in such surveys, it is usually impossible to gather all of the data simultaneously. As a consequence, the response variables of interest are gathered at several different times. Similarly, the focus is usually not on a particular place, but on an entire area, which will have some intrinsic spatial gradients, whether of a physical or biological nature.

In these situations, the intrinsic confounding of space *and* time with the sets of variables of interest (species and environment) means the data cannot be examined using traditional methods of either spatial or temporal autocorrelation. Our expanded version of the Borcard *et al.* (1992) method can be used in such situations (i.e. with intrinsic confounding of spatial and temporal variables). The method allows the calculation of the amount of variation in the entire set of response variables that is explained by different components, including their percentage of overlap.

Although we have presented the method as allowing tests of the relative importance of different components of variation, we emphasise that a great deal will depend on which variables are included in the analysis, just as is the case with univariate multiple regression. If important temporal or spatial variables exist that have not been included, the actual relative importance of 'time' versus 'space' *per se* for the response variables is certainly not going to be addressed by the method. Furthermore, we are not suggesting that the proposed methodology of partitioning would allow the establishment of any causative effects, which would require proper experimental designs and analyses. The use of this method in the logic of scientific endeavours is analogous to that of univariate multiple regression. Included in this are all of the concomitant warnings concerning issues of causation, choice of variables and redundancy of parameters causing increases in explained variation due to chance alone (e.g. Sokal & Rohlf 1981).

For the present data, it was clear that many of the environmental variables included in the analysis were spatially structured, which was not surprising, given the features of the habitat recorded for the study and their apparent correspondence with differences across and along the continental shelf. Temporal variables did not appear to be very important at all in explaining the distribution of prawns. This may have been due to the fact that the sampling itself, as a result of previous trawling and other experience, already had some temporal structure by being organized at certain times of the year (see Trainor *et al.* 1991; Poiner *et al.* 1997). If the sampling had occurred completely randomly with regard to time, then we would expect that the temporal component, at the scale of seasons, would be very influential in explaining biomass of prawns. Nevertheless, the data do suggest that temporal influences at smaller scales are relatively unimportant in determining biomass.

The proposed methodology is primarily useful for providing information from data obtained in a broad survey to allow efficient planning of specific studies in the ensuing stages of a research programme. Whereas the hypothesis for some general or initial surveys centres on the correlation or relationship among sets of variables, hypotheses in the next stage of research may then focus on particular effects that are tested in a structured and/or manipulative experimental design. For example, the results of the prawn survey reported here were used in the development of hypotheses and the design of an experiment to test for effects of trawling on prawn biomass (Poiner *et al.* 1997). The design of this experiment was essentially a 'before-after/control-impact' or BACI design (Underwood 1991, 1992). Due to the relatively small amount of variation explained by short-term temporal variables in the survey (see Fig. 2, Table 3), especially when considering the influence of spatial variation, short-term temporal variation was not a grave concern for the BACI design. On the other hand, the significant spatial and environmental components of variation (and their overlap) in explaining distributions of prawns were used in identifying appropriate and comparable areas for sampling and for the experimental manipulations (i.e. trawling), with necessary replication. In particular, the experiment was restricted spatially to unfished areas of similar longitude in the offshore zone, which, from the survey data, had consistency in composition and biomass of prawns: environmental variables, such as depth, were included as factors in the experiment (for further details, see Poiner *et al.* 1997).

Although the method described here was limited to partitioning variation across three sets of independent variables for a multivariate data set, the general approach could be used to complete a similar partitioning of variation among greater numbers of matrices, using canonical techniques, if desired (e.g. Pinel-Alloul *et al.* 1995). A whole new set of explanatory variables could be added as a separate matrix in the analysis. For example, in addition to environmental (i.e. physical or chemical), spatial and temporal variables, a fourth matrix of other kinds of biological variables could be introduced. Such biological explanatory variables could include things like abundances of predators, competitors or prey. The complication of adding a fourth matrix of explanatory variables is a resulting increase in the number of possible regions of overlap. At least three dimensions would be required to graphically represent all of these regions with Venn diagrams.

By adding the temporal component to the Borcard *et al.* (1992) method of partitioning variation of multivariate data, the method is applicable to many situations where spatial and temporal factors are confounded in the sampling design. This confounding occurs for many ecological surveys of large areas in the initial stages of research programmes. By describing this method we do not hope to encourage the design of confounded survey investigations. Wherever possible, structured mensurative experimental designs (*sensu* Hurlbert 1984) with replication could be used for surveys, which would allow unbiased estimates of variance for particular factors (and at various scales) to be calculated and used for tests. When the initial hypotheses of a research programme are of a correlative nature, however, the approach described here will provide (i) quantification of the degree to which matrices of variables explain variation in multiple response variables, including the degree to which explained variation by different matrices overlaps, and (ii) tests of the significance of relationships between matrices of variables by permutation methods.

ACKNOWLEDGEMENTS

Data used in the analysis were collected as part of a collaborative CSIRO/QDPI project, co-ordinated through the GBRMPA as part of the 'Effects of Fishing Program,' and funded by the GBRMPA, QFMA and FRDC. Data analysis was completed during a visiting fellowship held by N. A. Gribble with the University of Sydney, Institute of Marine Ecology. We thank A. J. Underwood and the Institute of Marine Ecology for logistic support and P. Legendre for comments on a draft of the manuscript. M. J. Anderson also extends particular thanks to B. E. Wingett. This work represents an adjunct to a larger MSc project by N. A. Gribble through the Department of Tropical Environment Studies and Geography, James Cook University, Townsville.

REFERENCES

Borcard D. & Legendre P. (1994) Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environ. Ecol. Stat.* **1**, 37–53.

Borcard D., Legendre P. & Drapeau P. (1992) Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–55.

Butler A. J. & Chesson P. L. (1990) Ecology of sessile animals on sublittoral hard substrata: the need to measure variation. *Aust. J. Ecol.* **15**, 521–31.

Dall W., Hill B. J., Rothlisberg P. C. & Staples D. J. (1990) The biology of the Penaeidae. *Adv. Mar. Biol.* **27**, 1–489.

Dutilleul P. (1993) Spatial heterogeneity and the design of ecological field experiments. *Ecology* **74**, 1646–58.

Hurlbert S. H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**, 168–98.

Kemeny J. G., Mirkil H., Snell J. L. & Thompson G. L. (1958) *Finite Mathematical Structures*. Prentice-Hall, Englewood Cliffs, NJ.

Legendre P. (1990) Quantitative methods and biogeographic analysis. In: *Evolutionary Biogeography of the Marine Algae of the North Atlantic*, NATO ASI Series, Volume G 22 (eds D. J. Garbary & R. R. South) pp. 9–34. Springer-Verlag, Berlin.

Legendre P. & Borcard D. (1994) Rejoinder. *Environ. Ecol. Stat.* **1**, 57–61.

Legendre P. & Fortin M-J. (1989) Spatial pattern and ecological analysis. *Vegetatio* **80**, 107–38.

Økland R. H. & Eilertsen O. (1994) Canonical correspondence analysis with variation partitioning: some comments and an application. *J. Veg. Sci.* **5**, 117–26.

Pinel-Alloul B., Niyonsenga T. & Legendre P. (1995) Spatial and environmental components of freshwater zooplankton structure. *Écoscience* **2**, 1–19.

Poiner I. R., Blaber S. J. M., Brewer D. T. *et al.* (1997) *The Effects of Prawn Trawling on the Far Northern Section of the Great Barrier Reef.* Queensland Department of Primary Industries, Final Report to the Great Barrier Reef Marine Park Authority. CSIRO Publications, Melbourne.

Sokal R. R. & Rohlf F. J. (1981) *Biometry*, 2nd edn. W. H. Freeman, New York.

Somers I. F. (1987) Sediment type as a factor in the distribution of commercial prawn species in the western Gulf of Carpentaria. *Aust. J. Mar. Freshwat. Res.* **38**, 133–49.

Somers I. F. (1994) Species composition and distribution of commercial penaeid prawn catches in the Gulf of Carpentaria, Australia, in relation to depth and sediment type. *Aust. J. Mar. Freshwat. Res.* **45**, 317–35.

Sparre P., Ursin E. & Venema S. C. (1989) *Introduction to Tropical Fish Stock Assessment: Part 1, Manual.* FAO Fisheries Technical Paper 306/1. FAO, Rome.

ter Braak C. J. F. (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* **41**, 859–73.

ter Braak C. J. F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**, 1167–79.

ter Braak C. J. F. (1987) The analysis of vegetation–environment relationships by canonical correspondence analysis. *Vegetatio* **69**, 69–77.

ter Braak C. J. F. (1988) CANOCO: An extension of DECORANA to analyse species–environment relationships. *Vegetatio* **75**, 159–60.

ter Braak C. J. F. (1990) *Update Notes: CANOCO version 3.10.* Agricultural Mathematics Group, Wageningen.

ter Braak C. J. F. (1992) Permutation *vs* bootstrap significance tests in multiple regression and ANOVA. In: *Bootstrapping and Related Techniques* (eds K-H. Jöckel, G. Rothe & W. Sendler) pp. 79–85. Springer-Verlag, Berlin.

Trainor N., Gribble N. A. & Moxon A. (1991) Drop in catch and effort in Princess Charlotte Bay. *Aust. Fish.* 51, 38–42.

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**, 569–87.

Underwood A. J. (1992) Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. *J. Exp. Mar. Biol. Ecol.* **161**, 145–78.

Underwood A. J. (1996) Spatial patterns of variance in density of intertidal populations. In: *Frontiers of Population Ecology* (eds R. B. Floyd, A. W. Sheppard & P. J. de Barro) pp. 369–89. CSIRO Publishing, Collingwood, Vic.

Whittaker J. (1984) Model interpretation from the additive elements of the likelihood function. *Appl. Statist.* **33**, 52–64.