Auckland
Regional Council
TE RAUHĪTANGA TAIAO

Modification of the Model to Assess Ecosystem Health Using Benthic Communities

April   TR 2009/047

This report is part of a series that were commissioned to characterised urban stormwater discharges. The reports were used to support the establishment and ongoing operation of the Regional Discharges Project.

Reviewed by:

Approved for ARC Publication by:

Name: Megan Stewart

Position: Project Leader Marine

Organisation: ARC

Date: 01/09/09

Name: Grant Barnes

Position: Group Manager Monitoring & Research

Organisation: ARC

Date: 01/09/09

Recommended Citation:

Anderson, M.J. (2009). Modification of the Model to Assess Ecosystem Health Using Benthic Communities. Prepared by University of Auckland for Auckland Regional Council. Auckland Regional Council Technical Report 2009/047.

Modification of the Model to Assess Ecosystem Health using Benthic Communities

Marti J Anderson

By

University of Auckland
Department of Statistics
Tamaki Campus
Private Bag 92019
Auckland, New Zealand
Tel: 64-9-373-7599 ext. 5052
Fax: 64-9-373-7000

Contents

1 Executive Summary

Two concerns were raised about the utility of the ecosystem health model for tidal creeks and harbours proposed by Anderson et al. (2002). The first was that too many of the sites used to develop the model that were identified as "healthy" in the case of the tidal creeks data set came from a single estuarine location (i.e. Whitford). Thus, effects that were deemed to be due to pollution would have been confounded with location-specific effects. The second concern was that the procedure for assessing the statistical significance of results for the model (i.e. the P-value from the CAP.exe computer program, Anderson 2003) has an inherent bias, due to the canonical analysis itself selecting the number of coordinate variables that would maximize the relationship.

The first concern was ameliorated by randomly selecting one site from each of the locations for which data were available, which resulted in a subset of 22 out of the original 58 observations. The group model (5 groups) and the gradient model (ranking from 1 = healthy to 5 = polluted) were each re-analysed using this reduced dataset. It was found that the group model no longer held for this reduced dataset. There was, instead, evidence for a single gradient rather than there being several distinct groups. Given this and the relative weakness of the group model for the existing harbour data set as well, it was decided to look towards the use of the gradient models alone instead. The gradient models, relating the faunal data with the linear rank pollution index directly, were found to be statistically significant and useful for the reduced tidal creek data set and for the harbours, even after correcting for the selection bias inherent in the original canonical procedure. The squared canonical correlation for the relationship was $\delta^2 = 0.823$ for tidal creeks and $\delta^2 = 0.632$ for harbours.

We recommend that, in the absence of further data, the method of assessment of the health of a new site be amended so that it be based only on the gradient relationship for each habitat type. The value of new observations along this canonical axis may be tracked individually through time as a measure of ecosystem health. When a greater number of observations from more sites and locations are obtained from across the region, this model may be re-run and modified in light of the new data. The assessment of the utility of any future model (i.e. the calculation of a P-value for the canonical correlation) must take into account the inherent selection bias of the canonical method, which is achieved by the new permutation technique described here.

2 Background

The ARC proposed the development of a model for ecosystem health, based on benthic community data from the Auckland region. This model was developed by Anderson et al. (2002) and allows a new observation (data for the benthic assemblage at a new site) to be placed within the context of other sites and so to be assigned a qualitative ranking of health, from 1 = "healthy" to 5 = "polluted". The model was considered in two parts: one was to model the 5 ranks as 5 separate groups ("the group model"), the other was to model the rankings as a single quantitative gradient ("the gradient model"). The model was also developed separately for two habitats: tidal creeks and harbours. Each data set, but particularly the one for tidal creeks, appeared to produce a fairly useful model, using either groups or a gradient approach (see details in Anderson et al. 2002).

2.1 Potential problems

2.1.1 Spatial confounding

Upon closer examination, a concern was raised about the utility of the ecosystem health model developed for tidal creeks (Shane Kelly, pers. comm.). In essence, many of the sites used to develop the model and identified as "healthy" occurred in a single estuarine location (i.e. Whitford). Thus, differences among sites seen in the model that were deemed to be due to differences in health status could actually simply have been caused by sites being in different specific locations.

Therefore, to avoid this problem of spatial confounding, one site was randomly chosen from each of the available locations from the original dataset. This resulted in there being 22 site observations, with the number of observations in each group being $n_1 = 5$, $n_2 = 4$, $n_3 = 5$, $n_4 = 4$, $n_5 = 4$. The group model and the gradient model were each then applied to this reduced set of data. The results are shown in Fig. 1.

The analysis of the group model on the reduced dataset was no longer convincing. Although a general pattern of "healthier" sites on the left, moving across to more "polluted" sites on the right of the plot appeared to occur (Fig. 1a), symbols corresponding to groups were not distinct from one another. This was also reflected in the leave-one-out cross-validation analysis (Table 1). In this procedure, an individual site is left out and the model applied to the remaining points without it. Then, the model is used to try to place the "left-out" point into the diagram to classify it. The allocation success rate of categorizing individual sites in this manner was poor. Only five out of 22 sites were correctly allocated, less than 23% correct, which is quite close to what would be expected merely by chance for 5 groups with a random set of points (i.e. 20%). Thus, the "group model" was no longer considered to be useful for the reduced tidal creek dataset.

Table 1. Results of "leave-one-out" classification of observations into each of 5 groups for the reduced tidal creek data set.

		Classified into groups...						
		1	2	3	4	5	Total	%correct
Original	1	2	2	0	1	0	5	40.0%
groups...	2	0	1	2	1	0	4	25.0%
	3	0	3	1	1	0	5	20.0%
	4	0	0	2	1	1	4	25.0%
	5	0	0	1	3	0	4	0.0%
							Overall	22.7%

On the other hand, the model of the assemblages in terms of a single rank pollution gradient was apparently much more successful (Fig. 1b). The squared canonical correlation (which can generally be interpreted as the correlation between the faunal assemblages as a whole and the rank pollution index) was quite high: $\delta^2 = 0.823$. However, allocation success is not possible to calculate in the absence of groups, so how can the adequacy of this gradient model be addressed?

2.1.2 Adequacy or statistical significance of the gradient model

Canonical analysis of principal coordinates (CAP), the method used to develop the models of community health, is described by Anderson and Willis (2003) and Anderson and Robinson (2003). The method essentially proceeds in two steps: (1) calculate principal coordinate axes from the dissimilarity matrix and (2) calculate canonical axes (based on specific hypotheses) from a subset of the PCO axes. How many PCO axes should be included in step 2? The size of the canonical correlation coefficient obtained is dependent to some extent on the choice of the number of PCO axes one chooses to use. The method suffers from potential arbitrariness in the choice made: the greater the number of axes chosen, the higher the canonical correlation value will be and the better the relationship will appear in the plot of the canonical axis. This is true, even if the original set of points is random. The suggestion by the authors to avoid arbitrariness in the results is to choose the number of PCO axes (we shall call this number m) that minimizes the residual sum of squares (or, in the case of groups, maximizes the allocation success). More specifically, the criterion used by the CAP.exe computer program to choose m is to find the value that minimizes the residual sum of squares (in the case of a gradient model) or maximizes allocation success (in the case of group model), subject to the constraint that the percentage of variability in the original dissimilarity matrix described by the first m axes be at least 60%. This generally results in a choice of m that is not too large, but nevertheless is large enough to encapsulate the majority of the information inherent in the dissimilarity matrix.

In the case of groups, the adequacy of the model, even after the choice of m has been made, can be seen to a large extent by the overall allocation success achieved – is it substantially better than what could have been achieved with a random set of points for this number of groups (e.g. 50% for 2 groups, 33.3% for 3 groups, 25% for 4 groups and so on)? On the other hand, it is still not clear for the gradient model, even after the choice of m has been made (by minimizing the residual sum of squares)

whether or not the model is substantially better than what would have been obtained with random data.

The original description of the canonical analysis of principal coordinates (CAP) included the possibility of performing a permutation test (see Anderson and Willis 2003 and Anderson and Robinson 2003 for details). This test should provide us with a measure of the probability of observing a test statistic (i.e. a canonical correlation) that is equal to or larger than the one we observed with our dataset. It does this by generating a number of "new" datasets in which the null hypothesis of "no relationship" with the pollution ranking is true, by simply re-ordering the original data, while keeping the rankings fixed. If the pairing of specific sites with particular rankings was not important (i.e. if there was no relationship), then it would not matter what order the sites took by reference to the rankings. For each re-ordering the test statistic is re-calculated and our original observed value for the test statistic is then compared to the distribution of values it takes under permutation. If our value is "extreme" by reference to this distribution, then we have a basis for claiming that the original ordering of the data was not something random ordering, but rather signifies a relationship between the fauna and the pollution ranking. The P-value is a measure of how extreme our observation is: the smaller the P-value, the smaller the probability that we would observe such an extreme value by chance.

The problem with the original permutation procedure is that it did not take into account the selection of m PCO axes on the basis of minimizing the residual sum of squares. This would be expected to bias the P-value, making it appear smaller than it really should be. Consider the following logic: it is the ordering of the original observations giving rise to the dissimilarity matrix and the complete set of PCO axes that are exchangeable under the null hypothesis, *not* the observations *after* having selected an appropriate value for m. Thus, the process of an appropriate selection for the value of m, using the criterion of minimizing the residual sum of squares, should be done for each permuted data set. Those permuted datasets which obtain the same choice for m as the original dataset should then be the ones that provide the distribution against which our original value should be compared. Although this new proposed permutation technique increases the computing time considerably, it is nevertheless the one that takes into account the selection procedure and is therefore producing the correct null distribution of the test statistic. The two permutation test approaches are outlined schematically in Figure 2.

2.2   Analysis and re-analysis of the gradient model

2.2.1   Reduced tidal creek data set

The gradient model was analysed with the reduced tidal creek dataset using the original CAP procedure and associated permutation test, with 4999 permutations. The distribution of the test statistic under permutation (i.e., the squared canonical correlation) is shown in Fig. 3a. Clearly the observed value of 0.823 is in the right-hand tail of this distribution, yielding a P-value of $P = 0.0136$. However, what of the influence of the selection procedure? Has it biased this result? We re-analysed the data using the newly proposed permutation procedure, outlined above and in Fig. 2. In

essence, a new value of m was selected for each randomization of the data and the canonical analysis ensued. Out of the 4999 permutations done, 232 of the permuted datasets selected a value of m equal to 11, which was the value chosen by the original ordering of data. These were therefore comparable for purposes of the test. The distribution of these 232 test statistics from permuted datasets is shown in Fig. 3b. Clearly, this distribution is shifted to the right compared to the one obtained using the original permutation method (the mean for the top distribution in Figure 3 is 0.52, while for the bottom distribution the mean is 0.65). Thus, the selection of m using a chosen criterion (such as minimizing the residual sum of squares by reference to the hypothesis) has an effect on P-values and therefore on type I error. The P-value associated with the significance of the gradient model of tidal creek data has therefore increased as a consequence of this new permutation method which takes into account the bias inherent in the CAP "selection of m" procedure.

There were 9 permuted datasets that achieved a squared canonical correlation greater than or equal to the value of 0.823 observed for our original data, thus the new P-value is 10/233 = 0.0429, which is still statistically significant. Therefore, although there is bias inherent in the CAP procedure due to the selection of an appropriate subset of PCO axes to use for the analysis, it is clear that the gradient model is still a significant and useful model of ecosystem health for the tidal creek data.

2.2.2 Harbour data set

The canonical analyses of the harbour dataset for the group and the gradient models are shown in Figure 4. Once again, the model for the gradient ($\delta^2 = 0.632$, Fig. 4b) appears more convincing than the model using 5 separate groups (Fig. 4a). The allocation success results for the harbour dataset when analysed using the group model are shown in Table 2. Although perhaps better than for a set of random points (for which we would expect a success of only about 20% in the case of 5 groups), the overall allocation success of only 55% still does not instill one with great confidence in the group model. Quite a few of the "mis-classified" observations are, nevertheless, only out by one. For example, those that belonged in group 2 but were misclassified were generally allocated to either group 1 or group 3, but never to group 5. This suggests that a gradient model may well be more appropriate.

Table 2. Results of "leave-one-out" classification of observations into each of 5 groups for the harbour data set.

		Classified into groups…						
		1	2	3	4	5	Total	%correct
Original	1	23	9	0	0	0	32	71.9%
groups…	2	13	11	1	2	0	27	40.7%
	3	1	2	4	0	0	7	57.1%
	4	1	1	2	1	1	6	16.7%
	5	0	1	0	0	3	4	75.0%
							Overall	55.3%

The gradient model for the harbour dataset was therefore assessed using the new permutation procedure. The differences between the distributions of the test statistic under permutation using the new versus the old procedure are shown in Fig. 6. Clearly, once again, the permutation distribution using the new method was "shifted to the right" by comparison to the old method, given the bias inherent in the

"selection of m" procedure. Out of 4999 permutations done, only XX of these chose a value of $m = 15$, equivalent to that obtained by the original data (using the criterion of minimizing the residual sum of squares with the additional constraint that the amount of variability in the original data retained had to exceed 60%). Out of these, XX values of the test statistic exceeded the value of $\delta^2 = 0.632$, which was obtained using the original data. This yields a P-value of XX/XX = XX. Thus, the gradient model is still useful and robust for the harbour data set as well, even given the presence of selection bias.

3 Recommendations

The analysis of the reduced tidal creek dataset using 5 separate groups was no longer useful for the allocation of new sites to pollution groupings. The use of the group model for the harbour dataset was also deemed to have little overall success in allocation to specific individual groups. However, the models of the sites in terms of a single rank pollution gradient were still quite useful and significant for each of the datasets, even after biases inherent in the CAP methodology were corrected using a new permutation method. We recommend that, in the absence of further data, the method of assessment of the "health" of a new site be amended so that it be based only on the gradient relationship. The value of new observations along this canonical axis (Fig. 2b for tidal creeks and Fig. 4b for harbours) may be tracked for individual sites through time as a measure of ecosystem health. When a greater number of observations from more sites and locations are obtained from across the region, this model may be re-run and modified in light of the new data. The assessment of the utility of any future model (i.e. the calculation of a P-value for the canonical correlation) must take into account the inherent selection bias of the canonical method, which is achieved by the new permutation technique described here.

4 References

Anderson, M.J. 2003. CAP: a FORTRAN computer program for canonical analysis of principal coordinates. Department of Statistics, University of Auckland, New Zealand.

Anderson, M.J., Hewitt, J. and Thrush, S. 2002. Using a multivariate statistical model to define community health. Report prepared for Auckland Regional Council, NIWA client report HAM2002-048, NIWA, Hamilton.

Anderson, M.J. and Robinson, J. 2003. Generalised discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics* 45(3): 301-318.

Anderson, M.J. and Willis, T.J. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84: 511-525.