INTRODUCTION

The industrialized world we live in today uses large amounts of chemicals, e.g., pesticides, that may enter aquatic ecosystems and produce undesirable side effects on their biological and functional properties. At an appropriate stage of the assessment procedure, microcosm and mesocosm experiments may be carried out with these chemicals. Normally, these experiments result in large data sets comprising information about temporal changes in the structure and function of control and treated microcosms or mesocosms. A generally recognized disadvantage of these experiments is that they require great effort in sampling and measuring of endpoints, particularly for the identification of the biological populations, whereas data for only a few taxa or other endpoints appear to suffice for standard univariate statistical analysis [1].

Multivariate statistical analysis may be used to describe the effects of chemical stress at the community level. Van den Brink et al. [2] proposed a multivariate method based on redundancy analysis (RDA), which is akin to the well-known method of principal components analysis (PCA) [3,4]. The advantage of this and related multivariate methods over univariate methods is that they use and summarize all information on the investigated populations simultaneously, and in doing so they evaluate the effects of contaminants at the community level. A serious drawback of multivariate methods is that the results are more abstract than those of univariate methods. In particular, results are often reported as ordination diagrams or biplots, which are unfamiliar to most ecotoxicologists and legislators. Time-dependent multivariate responses often result in diagrams that are too cluttered to allow easy interpretation of the changes in treatment effects over time. In this paper we introduce a novel multivariate method, called the principal response curve method (PRC), that overcomes these problems.

RDA

Redundancy analysis (RDA) can be considered to be a constrained form of principal components analysis (PCA) [3,4]. We therefore first introduce PCA and then derive RDA from it. With ecological data, multivariate methods are commonly used to find those factors that best explain the differences in species composition between samples. Of these, PCA, a type of factor analysis, is the most commonly used [3,4]. Principal components analysis uses linear models similar to the linear model underlying regression analysis. A difference with regression analysis is that the explanatory variables are not measured (manifest) but latent. The values of the latent variables for the various samples are called sample scores, and the regression coefficients of the linear model are called species weights. The rank 1 model (the model with one latent variable) of PCA can be written as

\[ y_{ik} = \bar{y}_k + b_k x_i + \epsilon_{ik} \]  

in which

- \( y_{ik} \) is the abundance of species \( k \) in sample \( i \) \((k = 1, \ldots, m; i = 1, \ldots, n) \);
- \( \bar{y}_k \) is the mean of \( y_{ik} \) for species \( k \) across all samples;
- \( x_i \) is the sample score of sample \( i \) on the latent variable (principal component);
- \( b_k \) is the species weight, i.e., the regression coefficient for species \( k \) with respect to the sample scores (\( x_i \)); and
- \( \epsilon_{ik} \) is the error term with mean zero and variance \( \sigma^2_{ik} \).
Some remarks on this model are in order. Because organisms multiply or die, count data are naturally modeled by proportional changes, i.e., by a multiplicative model. We analyze the counts after taking the logarithm so as to turn the multiplicative model for the counts into a linear model. Because the logarithm of 0 is undefined, the counts are increased by the value 1 before the natural logarithm is taken. The term "abundance" refers to the so-transformed count. Finally, note that the term $b_k x_i$ in Equation (1) expresses the deviation of species $k$ in sample $i$ from the mean $\bar{y}_k$. In terms of the original counts, $1 - \exp(b_k x_i)$ expresses the proportional change in count relative to the geometric mean count, $\exp(\bar{y}_k)$.

The sample scores and species weights can be displayed in an ordination diagram as the first, horizontal axis (see Fig. 1). Together the scores and weights explain a particular fraction of the total variance of the data set. A second latent variable can be extracted from the remaining variance, yielding a rank 2 model. The second set of sample scores and species weights is displayed as the second, vertical axis of the ordination diagram. After extracting more and more latent variables, PCA eventually accounts for all the variance of a data set.

If, like in RDA, explanatory variables are manifest, i.e., fixed a priori, the total variance can be partitioned by multivariate regression analysis into explained variance and residual variance. Redundancy analysis, in contrast to PCA, extracts information from the explained variance only; the axes of RDA (e.g., Fig. 1) represent a percentage of that variance. More formally, RDA can be defined in two equivalent ways. Redundancy analysis is a PCA that is applied to the fitted species data. Redundancy analysis is a PCA in which the sample scores are constrained to be linear combinations of the explanatory variables. Both definitions ensure that the ordination diagram displays only those differences between the samples that can be explained by the explanatory variables at hand. In this paper, the models are fitted by unweighted least-squares, disregarding possible differences in the residual variances ($\sigma^2$). More theoretical background information and technical details may be found in the references [3,5–8].
**The use of RDA in the 1990 chlorpyrifos study**

*Invertebrate data set.* The invertebrate data set used as an example in this paper was described in detail by Van Wijngaarden et al. [1] and Van den Brink et al. [2]. This data set was obtained from an experiment in outdoor experimental ditches. Twelve mesocosms were allocated at random to treatments; four served as controls, and the remaining eight were treated once with the insecticide chlorpyrifos, applied as Dursban® 4E, with nominal dose levels of 0.1, 0.9, 6, and 44 µg/L in two mesocosms each. The example data set comprises that of the invertebrates, which is a combination of macro-invertebrate and zooplankton data sets. Sampling was done 11 times, from week –4 pretreatment through week 24 posttreatment, giving a total of 132 samples (12 mesocosms × 11 sampling dates) in the statistical analyses. A total of 189 different taxa were identified and counted in these samples. The responses and recovery of the invertebrate community after chlorpyrifos treatment were analyzed in time using RDA, which was performed using the CANOCO computer program, version 3.14 [9]. The input data for RDA consisted of the 132 × 189 matrix of species data and a 132 × 55 matrix of explanatory variables. These 55 explanatory variables are dummy variables that indicate to which combination of the factors treatment (five levels) and sampling week (11 levels) each sample belonged. The 5 × 11 dummy explanatory variables ensure, by way of the linear constraints in RDA, that the mesocosms that received the same treatment (the replicates) also receive identical sample scores at each week. In an equivalent manner, this RDA can be obtained from a PCA in which the data values are replaced by the treatment means. The RDA thus allowed us to focus on the variance in the invertebrate data set that can be attributed to time, treatment, and their interaction. Within CANOCO, scaling 1 (Euclidean distances) was used because of the presence of the dummy explanatory variables [6]. As in the previous section, the counts were natural log transformed before analysis. Apart from this, the default options were chosen.

The RDA summarized the effects on the invertebrate community in time while still showing the effects at the species level (Fig. 1). In Figure 1, samples (open circles) with nearly identical species composition lie close together in the diagram, whereas samples with very different species composition lie far apart. The following rules are necessary to properly interpret the figure. If an imaginary line is drawn through a species point (solid circles) and the origin of the plot, the relative abundances of this species in all samples can be derived by projecting the sample points onto this imaginary line. The samples projecting on the ‘species line’ far away from the origin, but on the same side of the origin as the species point, contain relatively high numbers of this species. The greater the distance between the projection of a sample and the origin, the more abundant the species will be in this sample. If a sample point projects on the other side of the origin, compared to the species point, numbers of the species in this sample will be relatively low. Using these rules, we infer from the diagram that the species Cloeon dipterum is relatively abundant in all control samples and (almost) absent in the samples for weeks 1, 2, and 4 at the highest dose level. In general, Crustacea and Insecta showed a rapid concentration-dependent decrease in numbers after insecticide application (direct effects). An increase in Gastropoda (e.g., Bithynia tentaculata) and Oligochaeta was found, suggesting indirect effects [1,2].

In the CANOCO computer program, RDA is accompanied by Monte Carlo permutation tests to assess the statistical significance of the effects of the explanatory variables on the species composition of the samples [2,10]. In the 1990 chlorpyrifos study, Monte Carlo permutation tests were performed per sampling date, using the natural log-transformed nominal dose as the explanatory variable (see Van den Brink et al. [2] for rationale). This allowed the significance of the treatment regime to be tested per sampling date. The results of the permutation tests indicated that the treatment regime had a significant effect \( p < 0.05 \) on the invertebrate community in all posttreatment weeks up to week 24. At 24 weeks posttreatment, no effect could be demonstrated, suggesting recovery of the invertebrate community at all treatment levels.

Besides the overall significance of the treatment regime, we also wanted to know which treatments differed significantly from the control so that we could infer the no-observed-effect Concentration (NOEC) at the community level. This could not be done by testing every treatment against the control because there were too few permutation possibilities; testing two treated mesocosms against four controls only provides 15 permutation possibilities \( (6!/2!4!) \) and a corresponding lowest possible permutation-based \( p \) value of 0.07 \((1/15)\) [11]. In the corresponding univariate case, however, the Williams test [12] can be applied to calculate a NOEC. The Williams test can be applied to the multivariate data set if the data set is reduced to a single variable. The first principal component of a PCA suits this purpose. The NOEC calculations were therefore performed by applying the Williams test to the sample scores of the first principal component of each sampling date in turn [2]. These analyses indicated a NOEC\_community of 0.1 µg/L for the invertebrates [2].

*Physicochemical data set.* The physicochemical data set resulting from the above experiment has been described in detail by Kersting and Van den Brink [13]. In the mesocosms, continuous oxygen, pH, and temperature measurements were made to determine the ecosystem metabolism. In addition, alkalinity and conductivity were determined weekly. The dissolved oxygen (DO)–pH–alkalinity–conductivity syndrome was analyzed with RDA and indicated a temporary effect of chlorpyrifos (Fig. 2). After being standardized to zero mean and unit variance, in this analysis the parameters of the syndrome played the part of the species. The RDA diagram was very cluttered, however, so the interpretation was not as easy as that of Figure 1. The analysis indicated higher pH and, though less pronounced, higher oxygen concentration, as well as lower alkalinity and conductivity for the highest chlorpyrifos concentration, compared to the control treatment for the period of week 4 to roughly week 10 posttreatment (Fig. 2). These differences can be explained from the drop in gross primary production and oxygen consumption of the whole system in relation to the treatment. Oxygen consumption showed the largest decrease. These effects were explained as a decrease in the decomposition rate after the arthropods had been killed [13].

**PROPOSED NEW ANALYSIS: PRINCIPAL RESPONSE CURVES**

*Introduction*  
Redundancy analysis often results in very cluttered biplots, as can be seen in Figure 2. The problem is that differences between treatments and controls do not stand out and, because of the jagged trajectories, time is not expressed in a single
where $y_d$ is the abundance of species $k$ in replicate mesocosm $j$ of treatment $d$ at time $t$, $\bar{y}_{djk}$ is the mean abundance of species $k$ in week $t$ in the control ($d = 0$), and $e_{djk}$ is an error term with mean zero and variance $\sigma^2_{e}$. Note that $c_{ao} = 0$ for every $t$, because by definition $T_{0k} = 0$ for every $t$ and $k$. The least-squares estimates of the coefficients $c_{ao}$ can be found by partial RDA, which is also known as reduced-rank regression with concomitant regressors [7,8]. In CANOCO, these additional regressors are called covariables. There are two differences compared to the RDA of the previous section: (1) a 132×11 matrix of dummy variables is specified, indicating the sampling week from which each sample comes, and (2) the 11 explanatory dummy variables that indicate the control treatments are deleted from the analysis to ensure that the treatment effects are expressed as deviations from the control. As required above, $c_{ao} = 0$ for all $t$ (reference coding). The desired estimates of the coefficients $c_{ao}$ are the canonical coefficients of the partial RDA so specified. For further mathematical and computational details and an illustration on a small artificial data set we refer to Appendices 1 and 2.

When the coefficients $c_{ao}$ are plotted against the sampling date ($t$), curves are obtained, one for each treatment, that can be interpreted as the principal response curves of the community (see Fig. 3 for example). The accompanying species weights ($b_t$) allow an interpretation at the species level: the weight $b_t$ is the multiple by which the principal curves must be multiplied to obtain the fitted response curves of species $k$, because $T_{dh} = b_tC_d$. The higher the weight, the more the actual response pattern of the species is likely to follow the pattern in the PRC. Taxa with a high negative weight are inferred to show the opposite pattern, whereas taxa with near zero weight either show no response or a response that is unrelated to the pattern shown by the PRC. Note that, in terms of the original counts, $1 - \exp(b_t \times c_{ao})$ expresses the proportional change of species $k$ in treatment $d$ and week $t$ relative to its count in the control. More formally, the fitted value for the count in mesocosms with treatment $d$ at week $t$ is $\exp(b_t \times c_{ao})$ times the geometric mean count in the controls.

The significance of the PRC diagram can be tested by performing a Monte Carlo permutation of the mesocosms, i.e., by permuting whole time series, in the partial RDA from which PRC is obtained, using an $F$-type test statistic based on the eigenvalue of the component. The null hypothesis of this test is that $T_{dh} = 0$ for all $t$, $d$, and $k$. The second axis of the RDA can be used to generate a second PRC diagram (see Appendix...
1). The significance of the second PRC can be tested by adding the sample scores of the first axis to the covariables.

**Analyzed data sets**

In order to make a direct comparison between the results of RDA and PRC, the invertebrate data set, the RDA results of which were displayed in Figure 1, was also analyzed with PRC.

The physicochemical data set (Fig. 2) was also analyzed with PRC, but the data set used for PRC was larger than that for the RDA displayed in Figure 2. Kersting and Van den Brink [13] deleted some sampling dates from those available to avoid overcluttering the RDA biplot. For the same reason they also left out the 0.1 and 0.9 µg/L treatments. In this paper the whole data set, containing all available data from week -4 through week 24, was analyzed.

Monte Carlo permutation tests and NOEC_{community} calculations were also performed for each data set.

**RESULTS**

**Invertebrate data set**

The diagram of the first PRC of the invertebrate data set (Fig. 3) shows small variations in the pretreatment period and larger concentration-dependent deviations from the control after the chlorpyrifos application. The higher the dose, the larger the deviations. Taxa indicated with a positive species weight in Figure 3 are expected to decrease in abundance, relative to the controls, after treatment with the higher doses, whereas
taxa with negative weights are expected to increase. In particular, in Figure 3 Caenis horaria has the highest high positive weight and is thus inferred to decrease most strongly in abundance after the chlorpyrifos treatment. In contrast, Bithynia tentaculata has a small negative weight, indicating a small increase in abundance due to the chlorpyrifos treatment. Note that the taxa with weights between −0.5 and +0.5 are not shown because they are likely to show either a weak response or a response that is unrelated to that shown in Figure 3. Eighty percent of the taxa with a species score of 0.5 or higher were arthropods (Fig. 3). None of the taxa with a negative weight belonged to Phylum Arthropoda.

For a quantitative evaluation of PRC, the quotient \(\text{exp}(b_k \times c_p)\) can be calculated for each species \(k\) at treatment \(d\) and sampling date \(t\). For example, the species weight of \(C.\) horaria is 4.75 (Fig. 3), so the fitted reduction for the highest treatment level in week 1 is \(\text{exp}(-1.4 \times 4.75) = 0.0013\). The PRC analysis thus predicts that after the treatment \(C.\) horaria is reduced in the highest treatment to approximately 0.13% of its geometric mean count in the control. This is in good accordance with its actual response (see Van den Brink et al. [2]).

Table 1 shows that 22% of total variance can be attributed to time and treatment regime for the two analyzed data sets. The remaining fraction of the variance is residual. The table also indicates which fraction of the variance explained by the treatment regime is captured by the first and second principal response curves (PRC).

<table>
<thead>
<tr>
<th>Data set</th>
<th>Time</th>
<th>Treatment regime</th>
<th>First PRC</th>
<th>Second PRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invertebrates Syndrome</td>
<td>22</td>
<td>34</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>23</td>
<td>83</td>
<td>10</td>
</tr>
</tbody>
</table>

**Discussion**

Comparison between RDA and the principal response curves

Mesocosm experiments result in large data sets on the effects of pollutants on biological communities and how they change in time. The principal response curve (PRC) method that we propose for analysis of such data sets distills the complexity of time-dependent, community-level effects of pollutants to a graphic form that is easy to comprehend. Moreover, the responses of individual species to the treatments can be inferred from the PRC curves.

The PRC method has distinct advantages over the RDA method from which it is derived. These advantages can be appreciated by comparing Figures 3 and 4 (PRC) with Figures 1 and 2 (RDA). Although the RDA diagrams show the changes in species composition between the treatments in time, they are difficult to interpret in terms of how the treatment effects develop in time and how they depend on the dose level. An important reason for this is that the trajectory for the control cosms in time may appear quite chaotic, as is illustrated in Figure 2. As a consequence, the treatment effects that are of most interest, i.e., the deviations of the treatments from the control, do not stand out. The PRC method overcomes these difficulties of interpretation by representing the time trajectory for the controls as a horizontal line (Figs. 3 and 4). This is achieved by taking the control treatment as the reference to which the other treatments are compared and by defining “time” as the horizontal axis of the diagram. Notice that treatment effects displayed in Figures 1 and 2 could in principle be reexpressed in this new graphical form. The PRC method does better than this by accounting a priori for the development of the biological communities in the control cosms over time. This is why PRC is based on a partial RDA with, as covariables, indicator variables for the sampling dates. As a result, the PRC gives immediately discernible information on how the treatment effects develop over time (Figs. 3 and 4). With the help of species weights, the PRC can be used to derive...
Fig. 4. Principal response curves (PRC) with species weights for the dissolved oxygen (DO)–pH–alkalinity–conductivity syndrome data set, indicating the effects of a single application of the insecticide chlorpyrifos. In this analysis the parameters of the syndrome played the role of the species. For percentages of variance accounted for and the significance level see Tables 1 and 2.

Inferences about the response of individual species to the stressor.

Sometimes RDA and PRC represent essentially the same treatment effects. This happens when the time trajectory of the controls is approximately a straight line in the RDA diagram. For example, Figure 1 is quite similar to Figure 3 with a 45° rotation. The taxa with a relatively high weight in the PRC are also discriminating in the RDA (Figs. 1 and 3). Even in this case, however, the PRC diagram is easier to explain to nonexperts than the RDA diagram.

We close this section with a cautionary note. A small species weight of a taxon in the PRC cannot be translated automatically into a low susceptibility of the taxa to the stressor. For instance, Gammarus pulex has a relatively low species weight in the PRC diagram (Fig. 3), but it also happens to be the taxon that, as far as we know, is most susceptible to chlorpyrifos [14]. The reason for this discrepancy is that G. pulex has a very specific response pattern (see Van den Brink et al. [2]) that differs from that shown in Figure 3. PRC shares this limitation with RDA and other multivariate techniques that search for global patterns in community data.

Comparison with other techniques used in mesocosm experiments

In this section we attempt to outline the similarities and dissimilarities of multivariate methods used in ecotoxicology, focusing on nonmetric multidimensional scaling (MDS, incorporated in the PRIMER computer program; [15]), nonmetric clustering (NMC, incorporated in the Riffle computer program; [16]), and our own approach of canonical ordination using RDA and PRC (incorporated in the CANOCO computer program; [6,9]).

Multivariate analysis of ecotoxicological experiments consists of (1) testing the statistical significance of treatment effects and (2) estimation of the magnitude of treatment effects. Simultaneous multivariate analysis of many endpoints (many taxa) requires dimension reduction if one wishes to find patterns and display the results. CANOCO and PRIMER use ordination and thus reduce the data set to continuous variables, the axes of their ordination diagrams (e.g., Fig. 1). Riffle uses cluster analysis and thus reduces the data set to a nominal variable (indicating to which cluster a sample belongs). The display of results in reduced dimensions helps the analyst grasp the essential information given by the multivariate experiment. Often, however, this leaves the layman wondering what is actually being displayed. The dimension reduction frequently makes it difficult to explain what is actually estimated by these displays: the translation of the diagram into treatment effects is often nontrivial. This is the main problem we attempted to overcome with PRC.

Up to this point the similarities between the CANOCO and PRIMER approaches are evident. There are considerable differences as well. CANOCO works on the original ‘sample by species’ abundance matrix, whereas the MDS in PRIMER works on a derived ‘sample by sample’ matrix containing similarities or dissimilarities between samples [15,17]. It is therefore difficult, although not impossible [18], to show individual taxa in an MDS diagram. As a consequence, a direct interpretation of treatment effects displayed in the diagram at the species level is not easily obtainable. In addition, because MDS does not use the original species data set, the fractions of the total variance of the data set that are explained by the axes cannot be calculated. These parameters are important for evaluation of the merits of the analysis. An advantage of MDS,
Multivariate analysis of mesocosm experiments

Environmental Toxicology and Chemistry, 18, 1999

Table 3. Matrices used as input files for the principal response curve (PRC) analysis in CANOCO with the notation of Appendix 1

<table>
<thead>
<tr>
<th>Species counts</th>
<th>Explanatory variables</th>
<th>Covariables [w]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>CW0</td>
<td>C W0</td>
</tr>
<tr>
<td></td>
<td>LW0</td>
<td>L W0</td>
</tr>
<tr>
<td></td>
<td>HW0</td>
<td>H W0</td>
</tr>
<tr>
<td></td>
<td>CW1</td>
<td>C W1</td>
</tr>
<tr>
<td></td>
<td>LW1</td>
<td>L W1</td>
</tr>
<tr>
<td></td>
<td>HW1</td>
<td>H W1</td>
</tr>
<tr>
<td></td>
<td>CW2</td>
<td>C W2</td>
</tr>
<tr>
<td></td>
<td>LW2</td>
<td>L W2</td>
</tr>
<tr>
<td></td>
<td>HW2</td>
<td>H W2</td>
</tr>
<tr>
<td></td>
<td>CW3</td>
<td>C W3</td>
</tr>
<tr>
<td></td>
<td>LW3</td>
<td>L W3</td>
</tr>
<tr>
<td></td>
<td>HW3</td>
<td>H W3</td>
</tr>
</tbody>
</table>

Table 4. Output of CANOCO for obtaining principal response curves (PRC): the standardized canonical coefficients ($r_{ck}$), the standard deviations ($s_{ck}$), and the species weights ($b_k$). The total sum of squares was 15.98

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>$r_{ck}$</th>
<th>$s_{ck}$</th>
<th>Species</th>
<th>$b_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S1</td>
<td>1.3484</td>
</tr>
<tr>
<td>LW0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S2</td>
<td>0.0000</td>
</tr>
<tr>
<td>HW0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S3</td>
<td>-1.3484</td>
</tr>
<tr>
<td>CW1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S4</td>
<td>-1.3484</td>
</tr>
<tr>
<td>LW1</td>
<td>0.1805</td>
<td>0.2179</td>
<td>S5</td>
<td>-2.6968</td>
</tr>
<tr>
<td>HW1</td>
<td>0.3970</td>
<td>0.3000</td>
<td>S6</td>
<td>-2.6968</td>
</tr>
<tr>
<td>CW2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S7</td>
<td>0.0000</td>
</tr>
<tr>
<td>LW2</td>
<td>0.1805</td>
<td>0.2179</td>
<td>S8</td>
<td>0.0000</td>
</tr>
<tr>
<td>HW2</td>
<td>0.7716</td>
<td>0.3000</td>
<td>S9</td>
<td>0.0000</td>
</tr>
<tr>
<td>CW3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>S10</td>
<td>0.0000</td>
</tr>
<tr>
<td>LW3</td>
<td>0.0852</td>
<td>0.2179</td>
<td>S11</td>
<td>0.0000</td>
</tr>
<tr>
<td>HW3</td>
<td>0.5686</td>
<td>0.3000</td>
<td>S12</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

However, is that it allows the user to choose a specific similarity or dissimilarity index (e.g., the Bray-Curtis index), whereas PRC is restricted to Euclidean distances [9]. For PRC, however, this is less restrictive than it appears because it is still possible to transform the data before the analysis. For the invertebrate data sets from the 1990 chlorpyrifos study, the display of treatment effects provided by the MDS analysis (results not shown) was very similar to that of RDA, presented in Figure 1.

A difference between RDA and PRC, on one hand, and MDS and NMC, on the other, is that MDS and NMC are nonmetric, whereas RDA and PRC are linear and metric. In theory, nonmetric methods are better for dimension reduction and require fewer model assumptions. CANOCO allows the user to adjust the analysis for unwanted effects through covariables and to focus on the treatment effects, thereby counterbalancing its simplistic modeling. The linearity assumption should not be misunderstood: PRC is based on a linear method, but it is well capable of showing nonlinear treatment effects (see, e.g., Figs. 3 and 4).

Multivariate methods are often divided into supervised and nonsupervised methods. Rifflé is nonsupervised; it searches for a clustering of samples on the basis of the endpoints without using the treatment groups. The relationship between the clustering and treatment groups is investigated post hoc in an association analysis by a chi-square test on a contingency table [19,20]. Multidimensional scaling in PRIMER is also nonsupervised. The relationship with the treatments (or other predictor variables) is searched for by overlays on the ordination diagram. Discriminant analysis, as applied by Shaw and Manning [21], as well as RDA, PRC, and related constrained ordination methods, are supervised; the information on the treatment regime is used in the dimension reduction process. Large residual variance does not cause these methods to deviate from their primary aim, providing an optimal demonstration of the treatment effects.

Microcosm and mesocosm experiments are designed to provide information on direct and indirect treatment effects. The fact that supervised methods focus on the treatment effects is therefore an advantage. Note that MDS can be supervised if applied to treatment means of endpoints (or the derived dissimilarities; see Ter Braak [22]).

Unsupervised methods allow statistical tests to be carried out after the dimension reduction, i.e., in reduced space. Examples are the association analysis after Rifflé and the Williams test that calculates a $NOEC_{community}$ on the basis of the
first principal component of PCA. When the same test procedures are applied to results of supervised methods, the tests will become too liberal. For example, application of the Williams test to the first RDA axis would be invalid because the relationship between this axis and the treatment is already maximized in RDA. This problem can be circumvented by using permutation tests, as available in CANOCO and PRIMER. Permutation tests avoid making strong distributional assumptions such as those needed for the normal-theory-based tests in discriminant analysis [21]. We tested the significance of the PRC diagram using the first eigenvalue. Multivariate statistical testing does not necessarily require dimension reduction, however; the tests in the ANOSIM approach in PRIMER, as well as one of the test statistics in CANOCO (the sum of all canonical eigenvalues), work in full-dimensional space.

The suite of permutation methods in CANOCO appears to be more extensive than that in PRIMER. For example, the ANOSIM procedure does not allow designs like before-after-control-impact (BACI) to be analyzed [15], whereas this is incorporated in CANOCO [9].

**Perspective in the statistical analysis of mesocosm experiments**

In this paper, PRC has been used to display how treatment effects change over time. We have proposed separate procedures for statistical testing: an overall permutation test based on the first PRC eigenvalue and permutation and Williams tests for each of the time points. In order to improve integration of estimation and testing procedures, current research is focusing on the generation of confidence bands for the PRC curves using bootstrap methods. The statistical properties of PRC curves can be further improved by parametrizing the curves or by imposing smoothness constraints [23].

In this paper we solved the problems of the zero counts by adding an arbitrary small value to the counts before taking logarithms. This problem can be solved by putting the model into the context of generalized biadditive models [18,24], in which least-squares is replaced by Poisson regression.

The multivariate methods discussed above deal simultaneously with direct and indirect effects of the toxicant at the community level (e.g., the decrease in numbers of *C. horaria* and the increase in *B. tentaculata* at the higher dose levels; see Fig. 3). Although these methods may help to generate hypotheses about direct and indirect effects, they cannot distinguish between them. Testing hypotheses on causal chains requires structural equation modelling or causal modelling as attempted by Johnson et al. [25]. Their approach, however, lacks an explicit time component. Dynamic causal models hold great promise for future research.

**Acknowledgement**—We are indebted to Margreet Van den Burg for practical assistance; to Margreet Van den Burg, Theo Brock, and Rene Van Wijngaarden for valuable discussions; to the anonymous reviewers and to Theo Brock and Rene Van Wijngaarden for their comments that helped to improve the manuscript; and to Kees Kersting for the use of his data.

**REFERENCES**

15. Clarke KR, Ainsworth M. 1993. A method of linking multivariate analysis and community level effects in outdoor microcosms: Use of principal response curves with smoothness constraints [23] can be obtained by the reduced rank regression specified by Equations A1 and A2 [7,8]. In particular, the \( \{ b_k \} \) and \( \{ c_{ak} \} \) are estimated by the sample weights and canonical coefficients of the first component of this analysis. Notice that these weights and coefficients are determined up to an arbitrary scale value \( \beta \), say, because \( (b_k \beta)^{c_{ak} \beta} = b_k c_{ak} \). For the interpretation of the first PRC the scale is immaterial. To obtain Equation 2 of the main text from Equations A1 and A2, first note that \( y_{ik} \) in (A1) is denoted as \( y_{i,k} \) in equation 2 and that the least-squares estimate of \( a_k \) is denoted by \( y_{i,k} \) in equation 2, then insert equation A2 into equation A1, and finally recall that the \( \{ w_{ik} \} \) and \( \{ z_{ik} \} \) are indicator variables.

**APPENDIX 1**

**The principal response curve (PRC) method**

This appendix shows that least-squares estimates of the coefficients \( \{ c_{ar} \} \) of the PRC model (Eqn. 2) can be obtained by reduced-rank regression with concomitant variables, i.e., by partial redundancy analysis (RDA).

We first introduce some notation. The units (samples) of the analysis, indexed by the subscript \( i \) (\( i = 1, \ldots, n \)), are all combinations of mesocosm and time. Let \( y_{it} \) be the abundance of species \( k (k = 1, \ldots, m) \) in sample \( i \), \( w_i \) be the indicator for sampling date \( t = 1, \ldots, T \), and \( z_{it} \) be the indicator for treatment and sampling date \( d = 0, \ldots, D, t = 1, \ldots, T \). Thus, \( w_i = 1 \) if sample \( i \) is taken at time \( t \) and \( w_i = 0 \) elsewhere, and \( z_{it} = 1 \) if sample \( i \) received treatment dosage \( d \) and was taken at time \( t \). The control treatment is taken to be the 0th treatment level \( (d = 0) \). Now consider the regression model for species \( k \)

\[
y_{it} = \sum_{j=1}^{T} \alpha_{ik} w_i + \sum_{j=1}^{T} \sum_{d=1}^{D} \tau_{ad} z_{it} + \varepsilon_{it} \quad (A1)
\]

Because the \( \{ w_{it} \} \) and \( \{ z_{it} \} \) are indicator variables, this model simply implies that the expected abundance at time \( t \) is \( \alpha_{ik} \) when the sample is from a control mesocosm \( (d = 0) \) and \( \alpha_{ik} + \tau_{ad} \) when the sample received treatment \( d \). As required in the main text, Equation A1 models the abundance of each species as a sum of three terms, namely its level in the control \( (\alpha_i) \), a time-specific treatment effect \( (\tau_{ad}) \), and an error \( (\varepsilon_{it}) \). The essential assumption of the PRC method is that

\[
\tau_{ad} = b_k c_{ar} \quad (A2)
\]

which implies that the \( m \times (DT) \) matrix of regression coefficients \( \{ \tau_{ad} \} \) is of rank 1. By consequence, the least-squares estimates of \( b_k \) and \( c_{ar} \) can be obtained by the reduced rank regression specified by Equations A1 and A2 [7,8]. In particular, the \( \{ b_k \} \) and \( \{ c_{ar} \} \) are estimated by the sample weights and canonical coefficients of the first component of this analysis. Notice that these weights and coefficients are determined up to an arbitrary scale value \( \beta \), say, because \( (b_k \beta)^{c_{ar} \beta} = b_k c_{ar} \). For the interpretation of the first PRC the scale is immaterial. To obtain Equation 2 of the main text from Equations A1 and A2, first note that \( y_{i,k} \) in (A1) is denoted as \( y_{i,k} \) in equation 2 and that the least-squares estimate of \( a_k \) is denoted by \( y_{i,k} \) in equation 2, then insert equation A2 into equation A1, and finally recall that the \( \{ w_{ik} \} \) and \( \{ z_{ik} \} \) are indicator variables.

The PRC model also can be obtained from the partial principal components analysis (PCA) model

\[
y_{it} = \sum_{j=1}^{T} \alpha_{ik} w_i + b_k x_i + \varepsilon_{it} \quad (A3)
\]

by imposing the additional constraints (for \( i = 1, \ldots, n \))

\[
x_i = \sum_{r=1}^{T} \sum_{d=1}^{D} c_{ar} z_{it} \quad (A4)
\]

These \( n \) constraints turn the partial PCA model into a partial RDA model. The PRC model is obtained by inserting Equation A4 into Equation A3. Parametric principal response curves or principal response curves with smoothness constraints [23] can be obtained by appropriately parametrizing or constraining equation A4.

A second PRC diagram can be obtained by extending equation A2 to a rank 2 model

\[
\tau_{adk} = b_{k1} c_{ak1} + b_{k2} c_{ak2} \quad (A5)
\]

which yields the PRC model

\[
y_{i,j,k} = y_{i,k} + b_{k1} c_{ak1} + b_{k2} c_{ak2} + \varepsilon_{i,j,k} \quad (A6)
\]
course of treatments in time relative to controls (first PRC),

c_{d1} = second PRC,

b_{k1} = weight of species \( k \) for \( c_{d1} \), and

b_{k2} = weight of species \( k \) for \( c_{d2} \).

The PRC method can be seen as a generalization of van Buuren [26], who used PCA to generate principal curves in an analysis of time-intensity responses in sensory evaluation. The principal curve method developed by Hastie and Stuetzle [27] is, however, unrelated because it addresses nonlinear PCA.

Appendix 2

This appendix gives an example of how to obtain the principal response curves (PRC) by using the computer program CANOCO, Ver. 3 [9,28]. Information on CANOCO can be obtained by internet at www.microcomputerpower.com.

Example data and input for CANOCO

The small example mesocosm experiment consists of three treatments: C (control), L (low dosage), and H (high dosage), labeled \( d = 0, 1, 2 \), respectively. The treatments C and H are replicated, whereas L is not. All mesocosms are sampled four times \( (t = 1, \ldots, 4) \) indicated by W0, W1, W2, and W3 in Table 3 for week 0, 1, 2, and 3. Table 3 shows artificial, noise-free count data for six species (the columns labeled S1–S6) in all \( 5 \times 4 \) combinations of mesocosms and weeks, which are the rows of Table 3. The count data are entered as “species data” in CANOCO. The remaining columns of Table 3 are 12 \( (3 \times 4) \) indicator variables for all combinations of treatment and week, and four indicator variables for week, which are entered in CANOCO as “environmental variables” and “covariables,” respectively. During the analysis in CANOCO, the redundancy analysis (RDA) option is chosen, the species data are natural log(\( x \)) transformed, and the four indicator variables with the letter C in Table 3 that refer to the control treatment are deleted.

Output of CANOCO

The required output items of CANOCO for the formation of the PRC diagram are the species scores, the regression/canonical coefficients for standardized explanatory variables, the standard deviations of treatments, the number of samples, and the total sum of squares in the species data, denoted below by \( b_k, r_{d}, s_{d}, n, \) and TSS, respectively. The output for the example data is listed in Table 4.

Additional calculations

The canonical coefficients \( \{c_{d} \} \) of the PRC diagram are obtained by the formula

\[
c_{d} = \sqrt{(TSS)\ln (r_{d}s_{d})}.
\]

In the example, \( \sqrt{(TSS)\ln} = 4/20 = 0.2 \). This posttransformation is needed because CANOCO standardizes each explanatory variable and the species data. Figure 5 displays the resulting PRC diagram. To obtain inferences about individual responses, note, e.g., that the canonical coefficient for treatment H in week 3 \( (c_{24}) \) is 0.38. So the predicted counts for species 1 and species 4 (with weights \( b_1 = 1.3484 \) and \( b_4 = -1.3484 \)) are \( \exp(0.38 \times 0.3484) = 1.67 \) and \( \exp(0.38 \times -0.3484) = 0.60 \) times the count in the control, respectively. This is in accordance with the counts used as input, namely 130 \( \times 1.67 = 217 \) for species 1 and 70 \( \times 0.60 = 42 \) for species 4 (Table 3). The similarity with Figure 3 could have been enlarged by changing the signs of both \( c_{d} \) and \( b_k \), i.e., by plotting \(-c_{d}\) and \(-b_k\), which would not have affected the above predictions. As Figure 5 shows, the treatment effects of week 0 are zero, which could have been expected, because the observations of week 0 were taken pretreatment. In noisy data, the PRC as specified above tends to show small treatment effects in the pretreatment period (e.g., Figs. 3 and 4). The pretreatment scores in the PRC serve as a yardstick for the size of the random variation in the estimated coefficient \( c_{d} \). Alternatively, the coefficients for week 0 can be forced to be equal to 0 by deleting the explanatory variables, labeled LW0 and HW0 in Table 4, from the analysis.