Master Thesis

Analysing mesocosm experiments: Principal Response Curves vs. Multivariate Generalized Linear Models

Eduard Szöcs

Student number: 207 220 191

11.3.2014

Supervisor:

Prof. Dr. Ralf Schäfer

Erklärung über selbstständige Anfertigung der Masterarbeit

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen - benutzt habe und die Arbeit von mir vorher nicht in einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD-ROM).

Landau, 11.3.2014

_____

Eduard Szöcs

Abstract

Mesocosm experiments in ecotoxicology lead to complex multivariate datasets, which are challenging to analyse and interpret, because multiple species are sampled on many dates from many mesocosms. Such data is usually analysed using the multivariate method 'Principal Response Curves'. Recently, the multivariate extension of generalised linear models (GLM) were introduced as a tool to analyse ecological community data. Moreover, data aggregation techniques that can be analysed with univariate statistics have been proposed. The aim of this study is to compare the performance of these new tools.

We compiled datasets of mesocosm experiments and re-analysed them. Multivariate GLM and univariate endpoints were compared to PRC regarding their power and potential to identify affected taxa. In addition, we analysed the inter-replicate variability encountered in the studies.

Inter-replicate variability could be mostly explained by the total number of species detected during the experiments and the number of zero counts in the samples. Multivariate GLM performed equally well as Principal Response Curves and gave better indication of species responding to treatments. Univariate data aggregation methods performed poorer compared to Principal Response Curves.

Multivariate community data, which is generated during mesocosm experiments, needs to be analysed using multivariate methods. Multivariate GLM are a legitimate alternative to the widely used Principal Response Curves.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Mesocosm experiments

Freshwater model ecosystems (hereafter mesocosms) are a valuable tool for ecological risk assessment (ERA) since they allow assessing contaminant effects on community structure and functioning (Leeuwen and Vermeire, 2007). Mesocosms are at an intermediate level between laboratory and field studies, regarding ecological relevance and experimental control (Newman and Clements, 2008) and usually applied if lower-tier studies indicate potential hazards (de Jong et al., 2008).

Lentic and lotic mesocosm are used in ecotoxicology. Their aim is to simulate natural conditions and realistic exposure regimes for the communities (Brock et al., 2011). Communities in mesocosms are monitored multiple times before and after pesticide exposure and therefore mesocosm experiments can show causality between treatments and effects.

Such experiments generate large and complex data sets - multiple species are sampled on many dates from many mesocosms (van den Brink, 2013). Analysing, summarising and interpreting such data is challenging and special statistical techniques are needed to assess the effects on the community level.

The standard tool to analyse mesocosm data is a multivariate technique termed Principal Response Curves (van den Brink and ter Braak, 1999). Recently generalised linear models and their multivariate extension were proposed as a tool to analyse ecological community data (Warton et al., 2011). To date, however, this new method has only been applied to mesocosm data in one study (Cañedo-Argüelles et al., 2014). Moreover, data aggregation techniques have been suggested to analyse such data (Beketov et al., 2013; Sanchez-Bayo and Goka, 2012).

1.2 Tools to analyse mesocosm experiments

We will demonstrate these statistical methods using data from the mesocosm experiment of van den Brink et al. (1996):
Twelve experimental ditches were used in this experiment: Four ditches served as control and the remaining eight were treated in duplicates once with the insecticide chlorpyrifos at doses of 0.1, 0.9, 6 and 44 µg / L. Invertebrates were sampled 15 times from week -4 pre-treatment through week 56 post-treatment. A total of 189 taxa were identified. This resulted in a table with 180 rows (12 * 15 samples) and 189 columns.
For illustration purposes we will use only data from week -4 till week 24 and only a subset of eight representative taxa which showed different responses (Figure 1.1):

Negative effects with recovery *Caenis horaria, Cloeon dipterum and Chaoborus obscuripes*

Negative persisting effects *Gammarus pulex*

Positive effects Oligochaeta and *Bithynia tentaculata*

No effects Agrypnia complex

Rare species Libellulidae

Variation in this experiment can be explained by three sources: the time course during the experiment, the chlorpyrifos treatment and the random variation between replicates. When analysing such an experiment, researchers are mainly interested in the effect of the treatment and how this effect changes over time (the treatment x time interaction). Mesocosms will show a temporal change during the experiment, however, the effect purely related to time is of minor interest.

1.2.1 Principal Response Curves

Principal Response Curves (PRC) (van den Brink and ter Braak, 1998 1999) is the most widely method to analyse mesocosm experiments. It belongs to the multivariate methods of constrained ordination and is a special form of Redundancy Analysis (RDA),

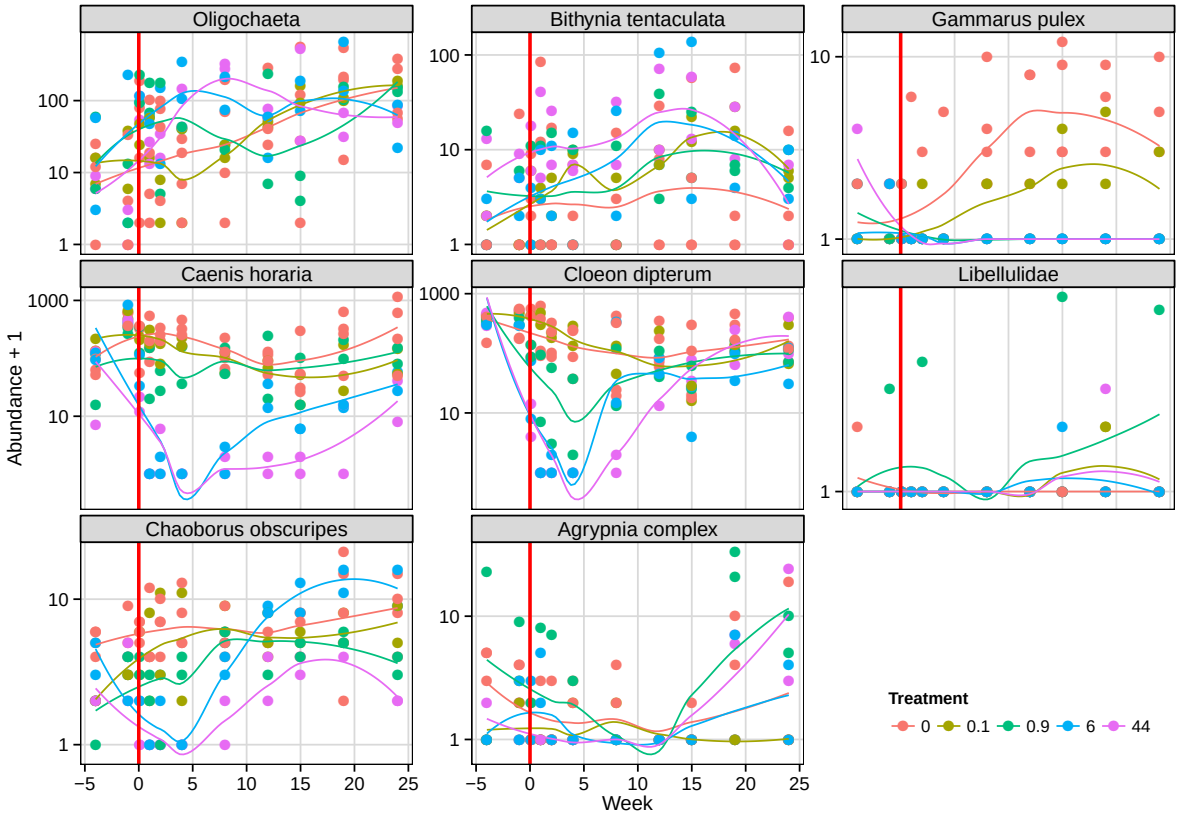

Figure 1.1: **Abundances (on a log-scale) of eight species during a mesocosm study with chlorpyrifos (van den Brink et al., 1996).** Lines represent a LOESS-smooth to show the main patterns in the data. The red vertical line indicates the date of chlorpyrifos application.

the multivariate extension of linear regression (Legendre and Legendre, 2012). As the temporal changes are of minor interest, these are 'partialled out', i.e. the effect of time is fitted to the data and only the residuals are used in further analysis (partial RDA). This removes the effect of time from the response. A RDA model is then fitted using treatment and its interaction with time as predictors.

Results of this ordination are usually presented in a diagram, where the first axis of the resulting diagram displays the maximum variation explained by the treatment and treatment x time interaction. RDA preserves the euclidean distance between samples, therefore species abundances should be transformed before analysis to avoid dominance of species with high abundances.

From such an ordination, one can obtain different information (Figure 1.2). The three most common in ecotoxicology are:

1. **Deviations of species composition from control:** Displayed as the mean difference of site scores between treatment and control on first axis during time. This plot shows the sampling date on the x axis and deviations in species composition from the control on the y axis (Figure 1.2, upper right).

2. **Species responsible for the observed differences:** Species scores on the first axis indicate the contribution species to the observed pattern (Figure 1.2, middle right).

3. **Test of significance** The first PRC axis can be tested, whether it displays a statistically significant amount of variation using permutations (Legendre et al., 2011)(Figure 1.2, lower right).

Species scores should be interpreted with caution and a low species weight does not necessarily translate to a small response (van den Brink and ter Braak, 1999). If a taxon shows a strong response, but is different from the global pattern, this will result in a low species score. In the example (Figure 1.1) *G. pulex* shows a persistent effect without recovery, but has a low species weight.
If the first axis is significant and shows a treatment-time interaction (non-parallel lines), the nature of this interaction can be further explored with separate RDAs for every sampling event. This allows also to determine community recovery, which is indicated by the failure of permutation tests to detect a treatment effect.
Given that mesocosms are sampled multiple times during the experiment, samples are not independent which should be considered when testing the significance. This is achieved by restricting the possible permutations: repeated observations of a mesocosm are kept together and only whole time series are permuted.

1.2.2 Multivariate Generalized Linear Models

Multivariate Generalized Linear Models (hereafter mvGLM) are the multivariate extension of Generalized Linear Models (GLM). As demonstrated recently, these may have superior properties compared to classical multivariate techniques like RDA (Warton et al., 2011). GLMs are a valuable tool for modelling data that show inherently a not

normally distributed response variable. Other distributions, than the normal distribution, should be used to model fractions (between 0 and 1), counts (positive integers) or presence/absence data (0 or 1).
MvGLM treat multivariate response as the sum of individual univariate responses. Separate GLMs are fitted to each species in the dataset. Two commonly used distributions for count data are the poisson and the negative binomial distribution. Model assumptions can be checked using residual plots. Combining these models using the sum-of-likelihood-ratios statistic allows testing for a significant community response (Warton, 2011; Warton et al., 2011). Univariate responses are directly available because a GLM is fitted to each species. Species contributions to the community response can be derived from the deviance of the univariate GLMs (Figure 1.3).
Different models could be fitted and compared: The model

$$y \sim time + treatment + time\ x\ treatment$$

incorporates a treatment-time interaction, therefore allows the treatment effect to vary over time. This model can be compared to the model without interaction

$$y \sim time + treatment$$

to test for statistical significance of the interaction. Comparing it to a model including only time as predictor,

$$y \sim time$$

allows to test for any treatment effect.
If there is a statistically significant interaction between treatment and time, this can be scrutinised by separate analyses per date. MvGLM assess the significance analogous to PRC via permutations, therefore restricted permutations should also be used.

1.2.3 Univariate endpoints

Univariate endpoints aggregate the multivariate response to a single response variable, which can be analysed using univariate techniques. Sanchez-Bayo and Goka (2012) studied four community endpoints: total abundance, species richness, a diversity index (Shannon-Wiener) and a similarity index (percentage difference, i.e. Bray-Curtis

coefficient). Beketov et al. (2013) showed in a simulation study that using the total abundance as response variable had a better statistical power then using RDA, especially if inter-replicate variation was high.

We compared these three statistical methods to analyse multivariate mesocosm data. Our aim was to evaluate these tools with respect to test power and identification of responsive species.

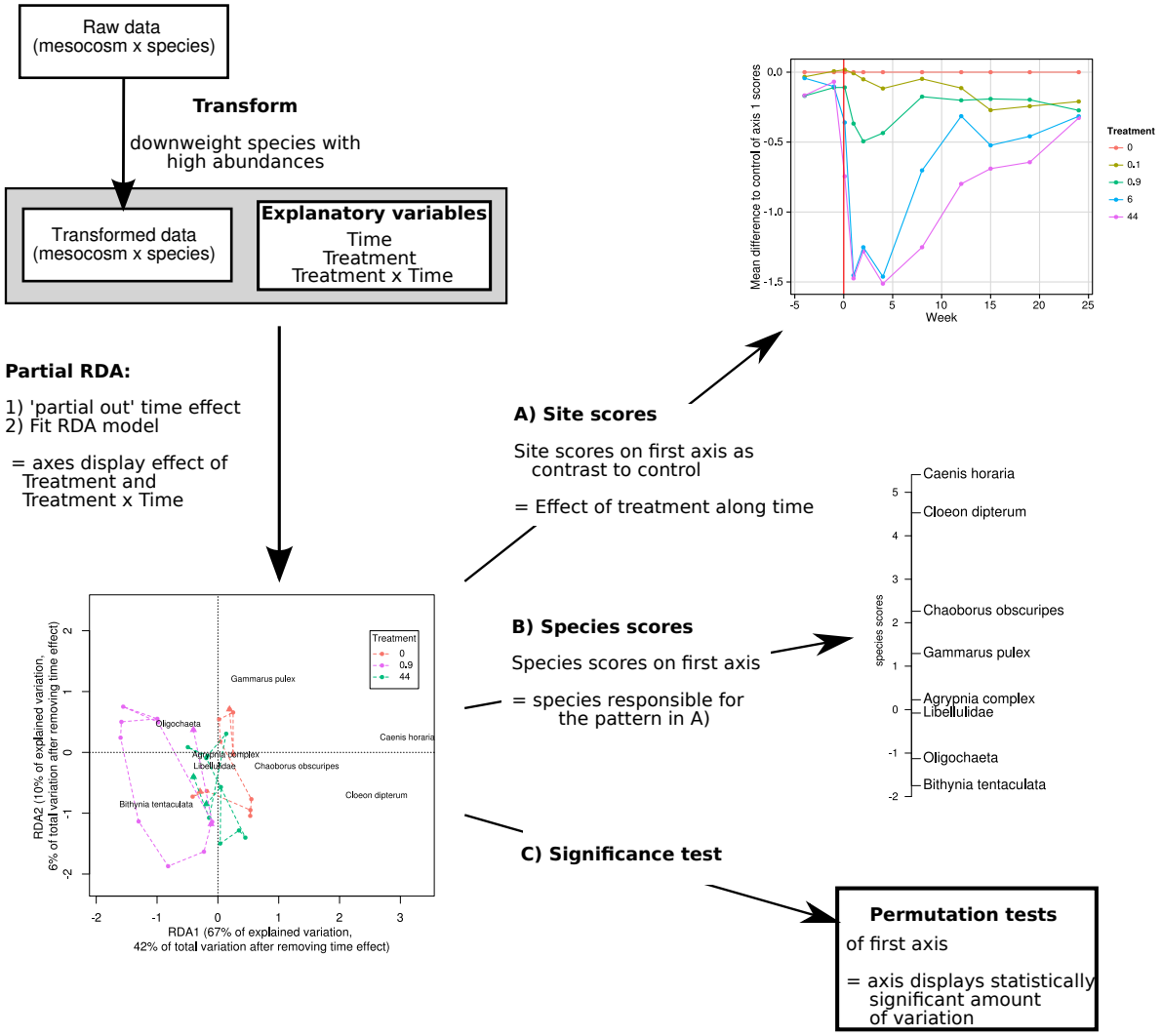

Figure 1.2: **Principal Response Curves as a special form of partial RDA.** Species data should be transformed to avoid overly influence of abundant taxa. Fit a partial RDA to abundance data, partialling out the time effect and using treatment and the treatment x time interaction as explanatory variables. The first axis displays the highest fraction of variation that can be explained by the explanatory variables. The mean difference of site scores between treatment and control on first axis displays the deviation of communities from control. Species scores indicate, which species are responsible for this pattern. Finally, the significance of the first axis can be tested.

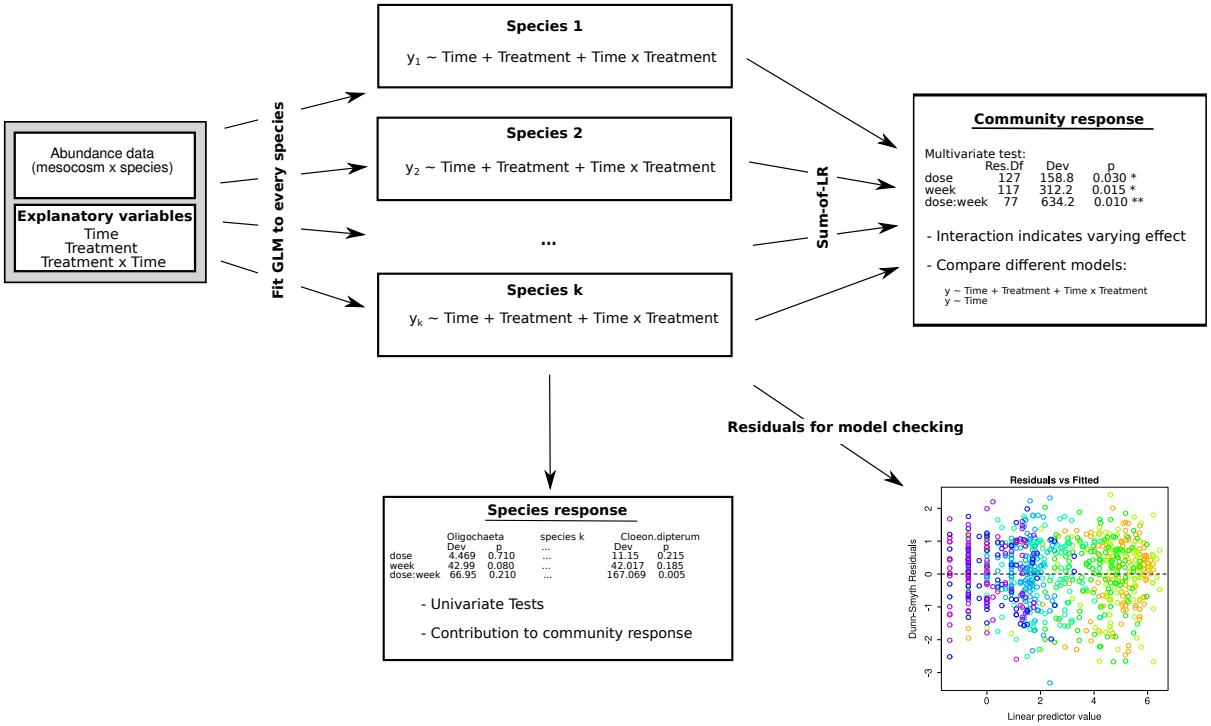

Figure 1.3: **Overview Multivariate Generalized Linear Models.** Separate GLMs are fitted to each species. These provide individual species responses. Using the sum of likelihood-ratios from these models allows testing the community response. The deviance of each taxon indicates how much it contributes to the community response. Model assumptions can be checked using residual plots.

2 Methods

2.1 Data sets

We compiled 11 datasets (published and unpublished, see Table 2.1) of mesocosm experiments. These datasets covered a wide range of experimental designs: Pond and stream systems, single and repeated applications as well as different sample sizes and durations. Only macroinvertebrate communities were considered in this study. Each of these studies was analysed using PRC, mvGLM and univariate responses.

2.2 Inter-replicate variability

We measured inter-replicate variability as variation in community structure. This can be expressed as dispersion in multivariate space and quantified by using any distance measure between observations (Anderson, 2006). We measured the multivariate dispersion ($\hat{\sigma}^2$) based on euclidean distance using log-transformed abundances (see next section), as this is also the used distance metric in PRC. For the euclidean distance the multivariate dispersion is equal to the sum of species variances (Anderson et al., 2011). Dispersions of replicated mesocosms were calculated for every treatment and at each sampling event.

We used linear mixed effect modelling (Bolker et al., 2009) to identify factors determining inter-replicate variability. The study characteristics total number of taxa, duration of the experiment and the within replicate characteristics proportion of zero counts, number of replicates, as well as treatment (binary variable: control or treatment) were used as predictors of variability. Study and treatment within study were used as random intercepts. R^2 values were calculated using the approach of Nakagawa and Schielzeth (2013).

Table 2.1: **Overview and source of the analysed mesocosm datasets.**

Study ID	Mesocosm type	Chemical	Application	No. taxa	No. treatments [a]	No. sampling events	$n_{control}$	$n_{treatment}$	Reference
1	indoor microcosm	Carbendazim	repeated	86	6	7	2	2	(Cuppen et al., 2000)
2	outdoor pond	-[b]	single	27	5	12	4	2	pers. comm. Ebke, P.
3	outdoor pond	mix	repeated	69	3	32	4	3	(Auber et al., 2011)[c]
4	outdoor pond	mix	repeated	69	3	32	4	3	(Auber et al., 2011)[c]
5	outdoor streams	mix	repeated	46	3	5	5	2	(Ippolito et al., 2012)
6	outdoor streams	Thiacloprid	repeated	35	4	9	10/4[d]	2/4[d]	(Liess and Beketov, 2011)
7	outdoor ditches	Chlorpyrifos	single	188	5	15	4	2	(van den Brink et al., 1996)
8	outdoor pond	Deltamethrin	single	40	4	21	5	4/3	(Caquet et al., 2007)
9	outdoor ditches	Lambda-Cyhalothrin	repeated	75	6	4	2	2	(Roessink et al., 2005)
10	outdoor pond	Thiram	repeated	55	3	14	3	2	(Bayona et al., 2013) [c]
11	outdoor pond	Hydrocarbon	repeated	60	5	14	3	2	(Bayona et al., 2013) [c]

[a] incl. control

[b] proprietary data

[c] split into two separate analyses, due to experimental design

[d] changed during experiment

2.3 Principal Response Curves

Prior to performing PRC abundances were $ln(Ax + 1)$ transformed. The term Ax was chosen to be equal 2 for the lowest abundance value (x) greater than zero (van den Brink et al., 2000). Time and treatment were used as categorical explanatory variables. The significance of the PRC and the overall treatment effect (main effect and interaction) were tested using the first eigenvalue (van den Brink and ter Braak, 1998). To scrutinise the interaction, separate RDAs using treatment as predictor were applied to each sampling event.

Significance was tested using 1000 restricted permutations. For PRC whole time series from mesocosms were permuted, accounting for the repeated measure design (van den Brink and ter Braak, 1999).

2.4 Multivariate Generalized Linear Models

MvGLM were fitted using the negative binomial distribution, because it is more flexible than the poisson distribution and appropriate for count data (O'Hara and Kotze, 2010). The overall treatment effect was assessed by comparing two nested models. The model

$$y \sim time + treatment + time\ x\ treatment$$

was compared to the model

$$y \sim time$$

where time and treatment were categorical variables. The difference between these two models was tested using a Likelihood-Ratio test based on permutations. For each sampling event the treatment effect was also assessed separately.

Significance was tested using 1000 restricted permutations. Whole time series from mesocosms were permuted, taking the repeated measures design into account (van den Brink and ter Braak, 1999). Univariate test statistics were calculated from single species GLMs.

2.5 Univariate endpoints

Three community endpoints were calculated for each sampling event and treatment:

Total abundance (N) The sum of all individuals. Additionally, the sum of log-transformed abundances (N (log)) was used, as this down-weights highly abundant species.

Species richness (SR) The number of species.

Diversity index (H') Expressed as Shannon-Wiener index.

The overall treatment effect was assessed for these endpoints by analysis of variance (ANOVA).
Performance of mvGLM and univariate endpoints were compared to PRC a) whether the conclusions drawn based on statistical significance are in agreement and b) species responses.

2.6 Software

All computations were performed using R (version 3.0.2 on Linux, 64-bit (R Core Team, 2013)). Linear mixed effect models were fitted using the LME4 package (Bates et al., 2014). Principal Response Curves were calculated using the VEGAN package (Oksanen et al., 2013), Multivariate Generalized Linear Models using the MVABUND package (Wang et al., 2012) and restricted permutations were created using the PERMUTE package (Simpson, 2013). The data cannot be provided because it is partly proprietary. However, a reproducible tutorial how to analyse mesocosm data with R can be found in appendix B. This tutorial uses the data of van den Brink et al. (1996) and is similar to the analyses performed in this study.

3 Results

Most studies showed comparable inter-replicate variability (median $\hat{\sigma}$ = 5.08, median absolute deviation $\hat{\sigma}$ = 1.19). Study no. 7 showed highest (median $\hat{\sigma}$ = 14.19) and study no. 2 median $\hat{\sigma}$ = 2.89) lowest variability (Figure 3.1). $\hat{\sigma}$ decreased as more zero counts were recorded in the samples, and increased slightly during the time course of the experiments. However, the most important factor determining variability was the total number of species recorded in the studies, leading to an increase in variability. We found no relationship between variability and the number of replicates as well as treatment (Table 3.1 and Appendix A).

MvGLM performed equally well as PRC. Analysing the overall pattern by comparing full with reduced models resulted in matching conclusions on 10 of 11 studies (Figure 3.2, A). For study no. 5 PRC showed a statistically significant result, whereas mvGLM did not. Except for this study, p-values obtained from mvGLM were equal or lower compared to PRC.

MvGLM showed high species deviances, where the PRC assigned low species scores (Figure 3.3). We found comparable outcomes for mvGLM and RDA when applied per sampling event (Figure 3.2, B).

Univariate data aggregation methods did not perform as well as PRC (Figure 3.4). Compared with the outcome of PRC, false negative rates (FNR) for univariate data aggregation methods were high (mean FNR = 48.6 $\pm$2.9%). For total abundance, down-weighting abundant species by using the logarithm improved performance slightly (FNR = 46.5 %). We found that species richness (FNR = 45.9 %) performed best of all four investigated univariate methods.

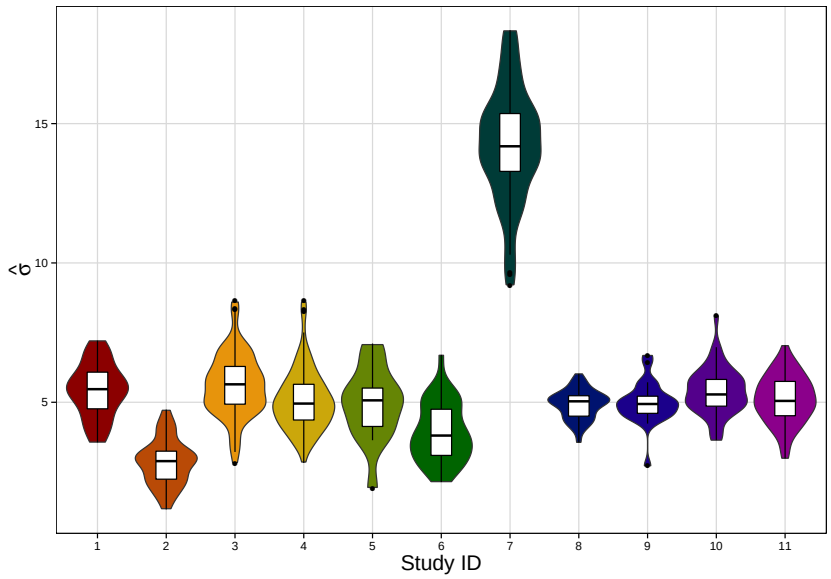

Figure 3.1: **Multivariate inter-replicate variability ($\hat{\sigma}$) of the studies.** Data pooled from treatment and sampling event per study. Study ID refers to table 2.1.

Table 3.1: **Mixed-effects modelling of the effects of study properties on inter-replicate variability.** Predictors have been scaled to zero mean and unit variance. All parameters were estimated using restricted maximum likelihood.

Model parameters			Model statistics	
Fixed effects	b	[95% CI]	$R^2_{GLMM(m)}$	80.3%
Intercept	**5.908**	[5.263, 6.552]	$R^2_{GLMM(c)}$	92.2%
% zero counts	**-0.670**	[-0.798, -0.535]	AIC	1845.6
Total no. taxa	**3.213**	[2.614, 3.811]	BIC	1885.7
Day	**0.172**	[0.081, 0.263]	N	640
No. replicates	0.004	[-0.163, 0.179]	N_{Study}	11
Treatment	-0.118	[-0.596, 0.360]	$N_{Treatment}$	47
Random effects	Variance			
Treatment in Study	0.30			
Study	0.81			
Residual	0.86			

Abbreviations: CI, confidence interval; $R^2_{GLMM(m)}$, marginal R^2_{GLMM}; $R^2_{GLMM(c)}$, conditional R^2_{GLMM}; AIC, Akaike Information Criterion; BIC, Bayesian information criterion.

Bold values indicate parameters with zero outside the confidence interval.

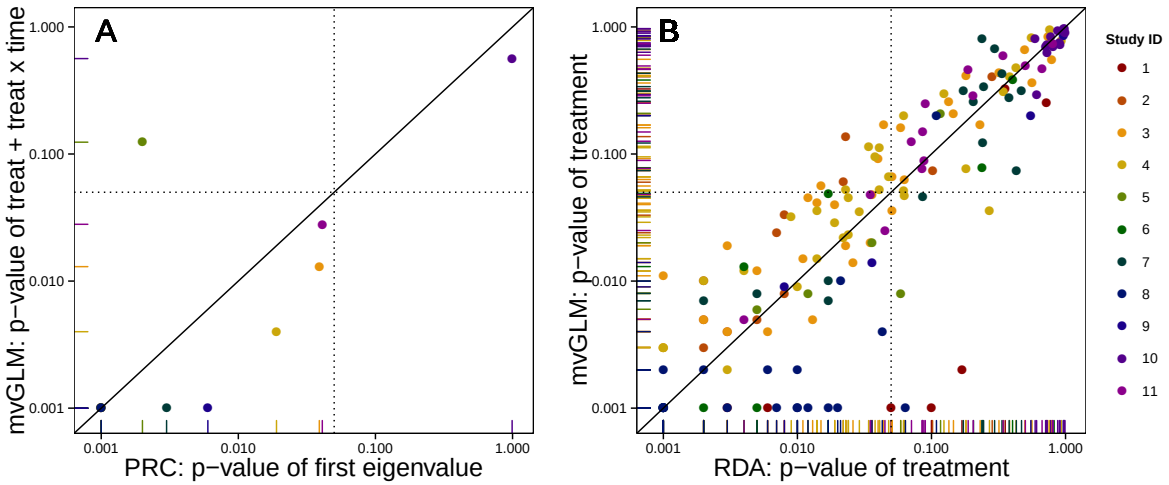

Figure 3.2: **p-values analysing the 11 datasets with Principal Response Curves and Multivariate Generalized Linear Models.** A) p-value of PRC vs. mvGLM assessing overall treatment effect, B) p-values of treatment effect per week. Black line represents the 1:1 line, dashed lines show an alpha-level of 0.05. Study ID refers to table 2.1.

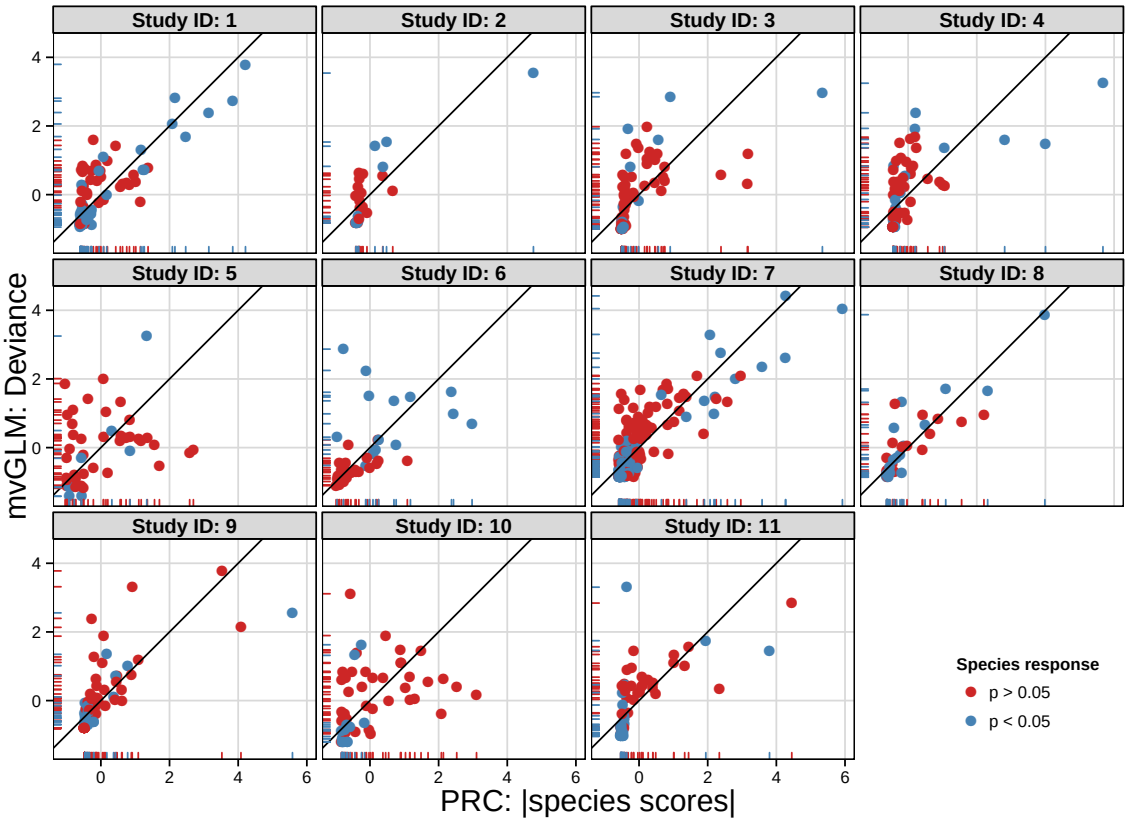

Figure 3.3: **Absolute PRC species scores compared to GLM deviances.** Both represent a measure of species response to treatment. Each point represents a taxon and both measures represent the contribution of this taxon to the community response. Deviances and species scores were standardized within study. Colour indicates whether the univariate GLMs showed a statistically significant treatment effect. Reported single species p-values are unadjusted for multiple testing. Black line represents the 1:1 line. Study ID refers to table 2.1.

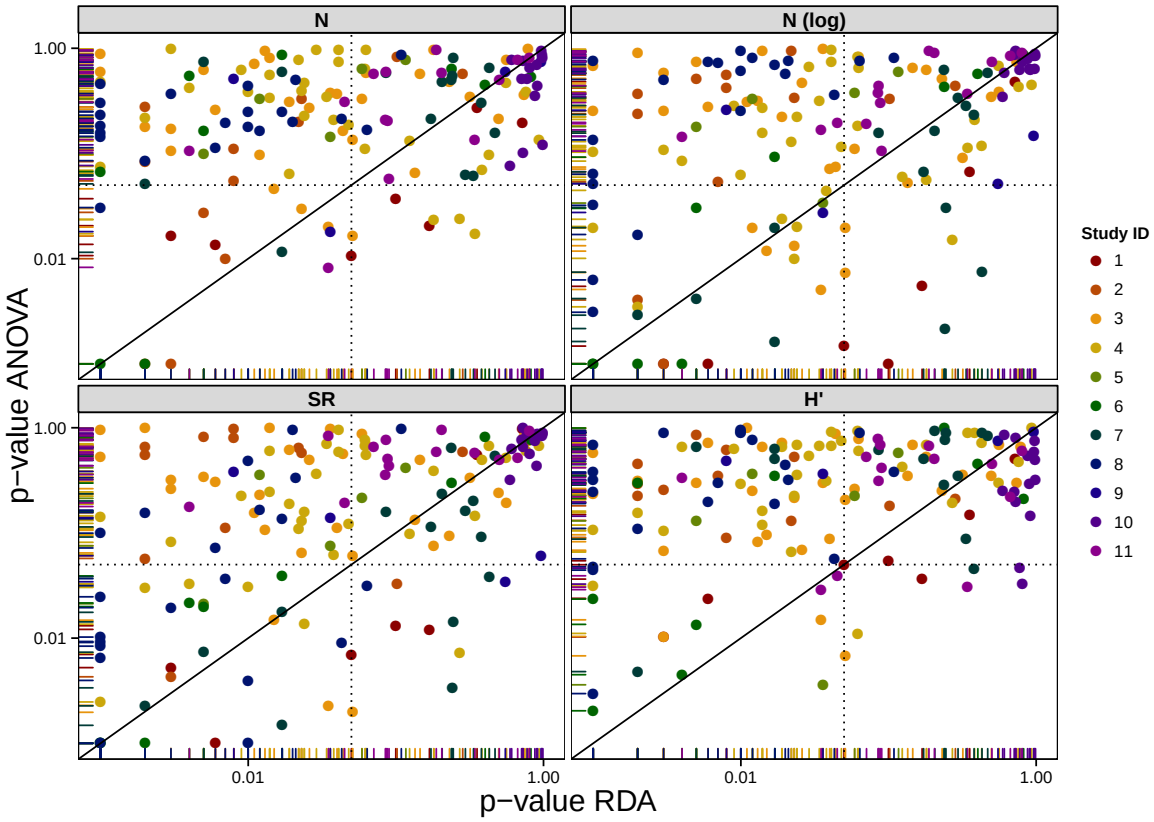

Figure 3.4: **Performance of univariate endpoints compared to PRC**. p-values of treatment effects per sampling. Black line represents the 1:1 line, dashed lines indicate an alpha-level of 0.05. p-values for ANOVA have been truncated to 0.001 for graphical representation. Study ID refers to table 2.1.

4 Discussion

4.1 Inter-replicate variability

The total number of species encountered in the studies was the main determinant of community variation. This could be either explained by an increased community complexity: The removal of a species may lead to stabilisation or destabilisation of the community, depending on the function of the species within the community (Ives et al., 2000). Therefore, a stochastic loss of species may lead to increased variability the more complex the communities are. Or, by the used dispersion measure: For euclidean distances $\hat{\sigma}$ increases as more species are added to the community.
Variability decreased as more zero counts were encountered in the samples. This could be attributed to the underlying measure of multivariate dispersion: The euclidean distance takes the joint absence of a species from two observations into account. For example, samples having no species in common, may be more similar to each other than to other samples sharing species. This phenomenon is also widely known as the *double-zero problem* (Legendre and Legendre, 2012). Therefore, replicates with many zero counts tend to be more similar and show a reduced variability.
During the experiments variability tended to increase slightly. Similarly, Sanderson et al. (2009) found an initial increase, however, they used univariate population level statistics. Knauer et al. (2005) used multivariate dispersion measures to analyse inter-replicate variability of zooplankton over a period of three years. They found an increase of variability over time only during one year and higher between year than within year variability. Similary, our results suggest that this observed increase is negligible compared to temporal variation (Figure A.3).

4.2 Multivariate Generalized Linear Models

We found that mvGLMs showed a similar performance as PRCs. Using mesocosm data, both methods resulted in the same conclusions drawn from statistical analysis. Except for study no. 5 mvGLM resulted in lower p-values than PRC, which is an indication of higher statistical power. The experimental setup in study no. 5 was profoundly different from the setups of all other considered studies: No pre-application data was available, as samples were taken immediately after treatments (Ippolito et al., 2012). Mesocosm studies usually show a characteristic pattern: Diverging communities directly after treatment followed by possible recovery during the time course of the experiment. This study lacked this pattern, with constant treatment effect over time (no significant treatment x time interaction, p = 0.375). In such a case, including the interaction into the model is not sensible, but for comparability we fitted to every dataset the same model. Omitting the interaction term leads to a statistical significant treatment effect on the communities (p = 0.012).

Warton et al. (2011) showed that mvGLMs have better power properties than RDA when error variance of affected taxa is low. This is in line with our findings: The observed error variances of the most affected taxa per study varied between 0.06 and 1.23. This is within the range where Warton et al. (2011) reported better performance of mvGLM compared to RDA.

MvGLM showed high single species deviances, where the PRC assigned low species scores. Therefore, mvGLM is able to identify responding taxa that would not be missed by PRC. This is the advantage of modelling single species responses: Univariate statistics are directly obtained, whereas PRC can only identify species that follow the global pattern. However, it is recommended that a PRC analysis should be followed by an univariate analysis of single taxa to ensure that all important species responses have been captured (de Jong et al., 2008; van den Brink, 2013).

The results of PRC can be displayed in an easy-to-interpret diagram (see Figure 1.2), whereas mvGLM are lacking such a representation. Nevertheless, a similar graphical representation can be extracted from the results. Species can be sorted by their deviance and the sum-of-Likelihood-Ratios statistic from per-week tests can be plotted along a time axis (Figure 4.1). MvGLM is computationally much more demanding than

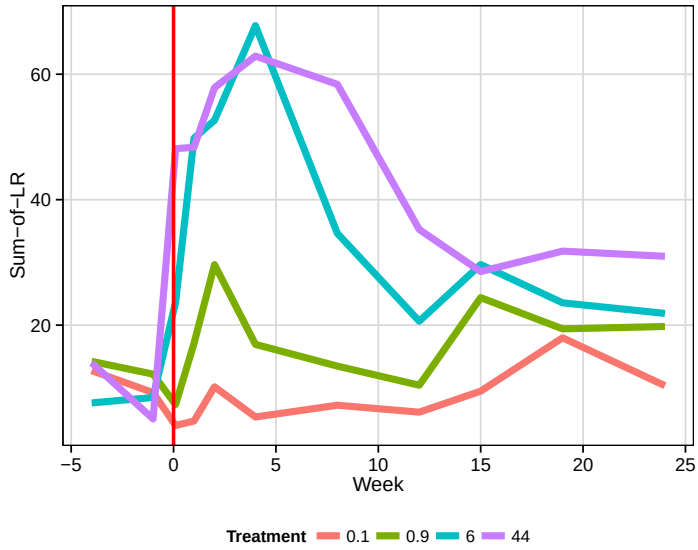

Figure 4.1: **Sum-of-Likelihood ratios through time**. An example plot how results of mvGLM could be represented in an ecotoxicological context. Compare with the result of PRC (Figure 1.2).

PRC. The computations for a medium sized study (e.g. study no. 10, 55 taxa and 14 sampling events) took 4 seconds for PRC, whereas mvGLM took 18 Minutes (on a Linux machine with 64bit, 4GB RAM and 2.27 GHz).

RDA and PRC are both based on the euclidean distance between samples. As previously outlined, this distance measure is sensitive to double zeros in the data. Therefore, other distance measures that do not take the joint absence of species from samples into account (*Asymmetric coefficients*) are frequently used to analyse field studies. The mesocosm studies considered here showed partly zero counts as high as those observed in field studies (Figure A.1): For example, the data presented in (Szöcs et al., 2012) showed a mean proportion of zero counts of 78% and (Szöcs et al., 2014) 83%. Especially when zero counts are high, asymmetric distance measures may result in a better performance. A PRC-like analysis, using any distance measure, can be performed by distance-based RDA (db-RDA) (Legendre and Anderson, 1999; McArdle and Anderson, 2001). In fact, dbRDA gives identical results as RDA when used with euclidean distances. Asymmetric distance measures, like the percentage difference coefficient, have already been used to analyse mesocosm experiments, however, only as univariate endpoint (e.g. van den Brink and ter Braak (1998), Sanchez-Bayo and Goka (2012)). Alternatively, special transformations of community data allows using

RDA with data containing many zeros (Legendre and Gallagher, 2001). This method is known as transformation-based RDA (tb-RDA). To our knowledge both methods have not yet been applied to mesocosm experiments in ecotoxicology. However, they may have a better performance when zero counts are high and therefore, their suitability to analyse mesocosm experiments should be further explored.

4.3 Univariate endpoints

All studied univariate endpoints performed considerably poorer than PRC. Beketov et al. (2013) showed in their simulation study that univariate t-tests on total abundance have better power properties than RDA. This is not supported by our results. The cause of this discrepant conclusion could be the very simplified simulation they used, which did not capture the complexity that is inherent to mesocosm studies. For example, in their simulation all species showed similar decrease in abundances. However, some species may benefit from treatment because of biotic interactions or show no effect. This is likely to occur in mesocosms and would reduce the power of the total abundance endpoint, but would be captured by PRC.

Liess and Beketov (2011) adapted a trait-based indicator for mesocosm application ($SPEAR_{mesocosm}$). The relative (log-transformed) abundance of sensitive species is calculated based on a binary species classification system and can be analysed using univariate statistical techniques. Notwithstanding that aggregation of sensitive taxa may be more powerful, this could be not tested here, as $SPEAR_{mesocosm}$ was developed for lotic systems. However, a similar indicator for lentic systems is currently under development (personal communication; Roucaute, M.; INRA). In the past, the use of this method has been subject to debate (Liess and Beketov, 2012; van den Brink and ter Braak, 2012).

Data aggregation comes along with a loss of information, since all the complexity of the response is pooled into a single variable. For instance, the species response is an essential information for an ecological and mechanistic interpretation of results. Univariate endpoints lose this information because species responses cannot be directly

derived. Multivariate techniques on the other hand, as the two methods examined here, preserve most of the information in the data. They enable to draw conclusions on species responses, with mvGLM resulting to a better overview of responsive species.

4.4 Implications for ecotoxicology

Data sets frequently encountered in community ecotoxicology are inherently not normally distributed, like proportions (bounded between 0 and 1), abundances (positive integer data) or biomass (strictly positive) (Wang and Riffel, 2011). GLMs offer a powerful and flexible parametric framework to analyse such non-normal data. However, it is rarely used in ecotoxicology. For example GLMs are not discussed in the guideline on statistical analysis of ecotoxicological data (OECD, 2006). Instead researchers often try to transform data to achieve normality and variance homogeneity or use non-parametric tests (Wang and Riffel, 2011). We advocate the use of GLMs in such cases.
A lot of data is generated for environmental risk assessment, but only a small fraction is made publicly available (Schäfer et al., 2013). All 11 raw data sets studied here were provided to the authors only upon personal request. Requests to industry remained unsuccessful. We argue to *not let data go waste* (sensu Poisot et al. (2013)) and make data publicly available. Ecotoxicology will undoubtedly benefit from open data: For example, it enhances transparency and by combining data sets from different experiments patterns may emerge that might be overlooked otherwise.
When dealing with community data, statistical methods that take the multivariate nature of the data into account are advisable. Data aggregation methods are, despite their simplicity, an inappropriate tool to analyse mesocosm data. Multivariate Generalized Linear Models are an alternative to the widely used Principal Response Curves, yielding to similar results and a better indication of responding taxa (see Table 4.1 for a comparison).

Table 4.1: **Comparison of Principal Response Curves (PRC), Multivariate Generalized Linear Models (mvGLM) and univariate endpoints.**

	PRC	mvGLM	univariate endpoints
Power	✓	✓	✗
Species Responses	(✓)	✓	✗
Graphical Display of results	✓	(✓)[b]	✓
Computation time	✓	✗	✓
Flexibilty	(✓)[a]	✓	(✓)

[a] if specified as partial RDA.

[b] but see Figure 4.1.

5 Acknowledgements

We are thankful to all authors and institutions that were involved creating and providing their data to realise this study.

References

Anderson, M. J. (2006). Distance based tests for homogeneity of multivariate dispersions. *Biometrics*, 62(1):245–253.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C., and Swenson, N. G. (2011). Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters*, 14(1):19–28.

Auber, A., Roucaute, M., Togola, A., and Caquet, T. (2011). Structural and functional effects of conventional and low pesticide input crop-protection programs on benthic macroinvertebrate communities in outdoor pond mesocosms. *Ecotoxicology*, 20(8):2042–55.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. R package version 1.1-0.

Bayona, Y., Roucaute, A., Roucaute, M., Gorzerino, C., Cailleaud, K., Lagadic, L., Bassères, A., and Caquet, T. (2013). Secondary production of freshwater zooplankton communities exposed to a fungicide and to a petroleum distillate in outdoor pond mesocosms. *Environmental Toxicology and Chemistry*. Advance online publication. DOI: 10.1002/etc.2495.

Beketov, M. A., Kattwinkel, M., and Liess, M. (2013). Statistics matter: data aggregation improves identification of community-level effects compared to a commonly used multivariate method. *Ecotoxicology*, 22(10):1–10.

Bolker, B., Brooks, M., Clark, C., Geange, S., Poulsen, J., Stevens, M., and White, J. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135.

Brock, T., Arts, G., ten Hulscher, T., de Jong, F., Luttik, R., Roex, E., Smit, C., and van Vliet, P. (2011). Aquatic effect assessment for plant protection products; dutch proposal that addresses the requirements of the plant protection product regulation and water framework directive. Alterra Report 2235, Alterra, Wageningen.

Caquet, C., Hanson, M., Roucaute, M., Graham, D., and Lagadi, L. (2007). Influence of isolation on the recovery of pond mesocosms from the application of an insecticide. II benthic macroinvertebrate responses. *Environmental Toxicology and Chemistry*, 26(6):1280–1290.

Cañedo-Argüelles, M., Bundschuh, M., Gutiérrez-Cánovas, C., Kefford, B. J., Prat, N., Trobajo, R., and Schäfer, R. B. (2014). Effects of repeated salt pulses on ecosystem structure and functions in a stream mesocosm. *Science of The Total Environment*, 476-477:634–642.

Cuppen, J., Van den Brink, P., Camps, E., Uil, K., and Brock, T. (2000). Impact of the fungicide carbendazim in freshwater microcosms. i. water quality, breakdown of particulate organic matter and responses of macroinvertebrates. *Aquatic Toxicology*, 48(2-3):233–250.

de Jong, F. M. W., Brock, T. C. M., Foekema, E. M., and Leeuwangh, P. (2008). Guidance for summarizing and evaluating aquatic micro- and mesocosm studies. RIVM Report 601506009/2008, RIVM.

Ippolito, A., Carolli, M., Varolo, E., Villa, S., and Vighi, M. (2012). Evaluating pesticide effects on freshwater invertebrate communities in alpine environment: a model ecosystem experiment. *Ecotoxicology*, 21(7):2051–2067.

Ives, A. R., Klug, J. L., and Gross, K. (2000). Stability and species richness in complex communities. *Ecology Letters*, 3(5):399–411.

Knauer, K., Maise, S., Thoma, G., Hommen, U., and Gonzalez-Valero, J. (2005). Long-term variability of zooplankton populations in aquatic mesocosms. *Environmental Toxicology and Chemistry*, 24(5):1182–1189.

Leeuwen, C. J. v. and Vermeire, T. G. (2007). *Risk Assessment of Chemicals: An Introduction*. Springer Netherlands.

Legendre, P. and Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69(1):1–24.

Legendre, P. and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2):271–280.

Legendre, P. and Legendre, L. (2012). *Numerical ecology*. Elsevier, Amsterdam; Boston.

Legendre, P., Oksanen, J., and ter Braak, C. J. F. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, 2(3):269–277.

Liess, M. and Beketov, M. (2011). Traits and stress: keys to identify community effects of low levels of toxicants in test systems. *Ecotoxicology*, 20(6):1328–1340.

Liess, M. and Beketov, M. A. (2012). Rebuttal related to “Traits and stress: Keys to identify community effects of low levels of toxicants in test systems” by liess and beketov (2011). *Ecotoxicology*, 21(2):300–303.

McArdle, B. H. and Anderson, M. J. (2001). Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, 82(1):290–297.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142.

Newman, M. C. and Clements, W. H. (2008). *Ecotoxicology : A Comprehensive Treatment*. CRC Press.

OECD (2006). *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application*. Number 54 in Series on Testing and Assessment. OECD, Paris.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2013). vegan: Community ecology package. R package version 2.1-40.

O'Hara, R. B. and Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, 1(2):118–122.

Poisot, T. E., Mounce, R., and Gravel, D. (2013). Moving toward a sustainable ecological science: don't let data go to waste! *Ideas in Ecology and Evolution*, 6(2).

R Core Team (2013). R: A language and environment for statistical computing. ISBN 3-900051-07-0.

Roessink, I., Arts, G. H. P., Belgers, J. D. M., Bransen, F., Maund, S. J., and Brock, T. C. M. (2005). Effects of lambda-cyhalothrin in two ditch microcosm systems of different trophic status. *Environmental Toxicology And Chemistry*, 24(7):1684–1696.

Sanchez-Bayo, F. and Goka, K. (2012). Evaluation of suitable endpoints for assessing the impacts of toxicants at the community level. *Ecotoxicology*, 21(3):667–80.

Sanderson, H., Laird, B., Brain, R., Wilson, C. J., and Solomon, K. R. (2009). Detectability of fifteen aquatic micro/mesocosms. *Ecotoxicology*, 18(7):838–845.

Schäfer, R. B., Bundschuh, M., Focks, A., and von der Ohe, P. C. (2013). Letter to the editor. *Environmental Toxicology And Chemistry*, 32(4):734–735.

Simpson, G. L. (2013). permute: Functions for generating restricted permutations of data. R package version 0.8-0.

Szöcs, E., Coring, E., Bäthe, J., and Schäfer, R. B. (2014). Effects of anthropogenic salinization on biological traits and community composition of stream macroinvertebrates. *Science of The Total Environment*, 468–469:943–949.

Szöcs, E., Kefford, B. J., and Schäfer, R. B. (2012). Is there an interaction of the effects of salinity and pesticides on the community structure of macroinvertebrates? *Science of the Total Environment*, 437(1):121–126.

van den Brink, P. J. (2013). Assessing aquatic population and community-level risks of pesticides. *Environmental Toxicology and Chemistry*, 32(5):972–973.

van den Brink, P. J., Hattink, J., Brock, T. C. M., Bransen, F., and van Donk, E. (2000). Impact of the fungicide carbendazim in freshwater microcosms. II. zooplankton, primary producers and final conclusions. *Aquatic Toxicology*, 48(2-3):251–264.

van den Brink, P. J. and ter Braak, C. J. F. (1998). Multivariate analysis of stress in experimental ecosystems by principal response curves and similarity analysis. *Aquatic Ecology*, 32(2):163–178.

van den Brink, P. J. and ter Braak, C. J. F. (1999). Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry*, 18(2):138–148.

van den Brink, P. J. and ter Braak, C. J. F. (2012). Response to “traits and stress: keys to identify community effects of low levels of toxicants in test systems” by liess and beketov (2011). *Ecotoxicology*, 21(2):297–299.

van den Brink, P. J., van Wijngaarden, R., Lucassen, W., Brock, T., and Leeuwangh, P. (1996). Effects of the insecticide dursban 4E (active ingredient chlorpyrifos) in outdoor experimental ditches: II. invertebrate community responses and recovery. *Environmental Toxicology and Chemistry*, 15(7):1143–1153.

Wang, M. and Riffel, M. (2011). Making the right conclusions based on wrong results and small sample sizes: interpretation of statistical tests in ecotoxicology. *Ecotoxicology and environmental safety*, 74(4):684–92.

Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). mvabund- an r package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, 3(3):471–474.

Warton, D. (2011). Regularized sandwich estimators for analysis of High–Dimensional data using generalized estimating equations. *Biometrics*, 67(1):116–123.

Warton, D. I., Wright, S. T., and Wang, Y. (2011). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3(1):89–101.

A Additional material

A.1 Supplemental Figures to Table 3.1

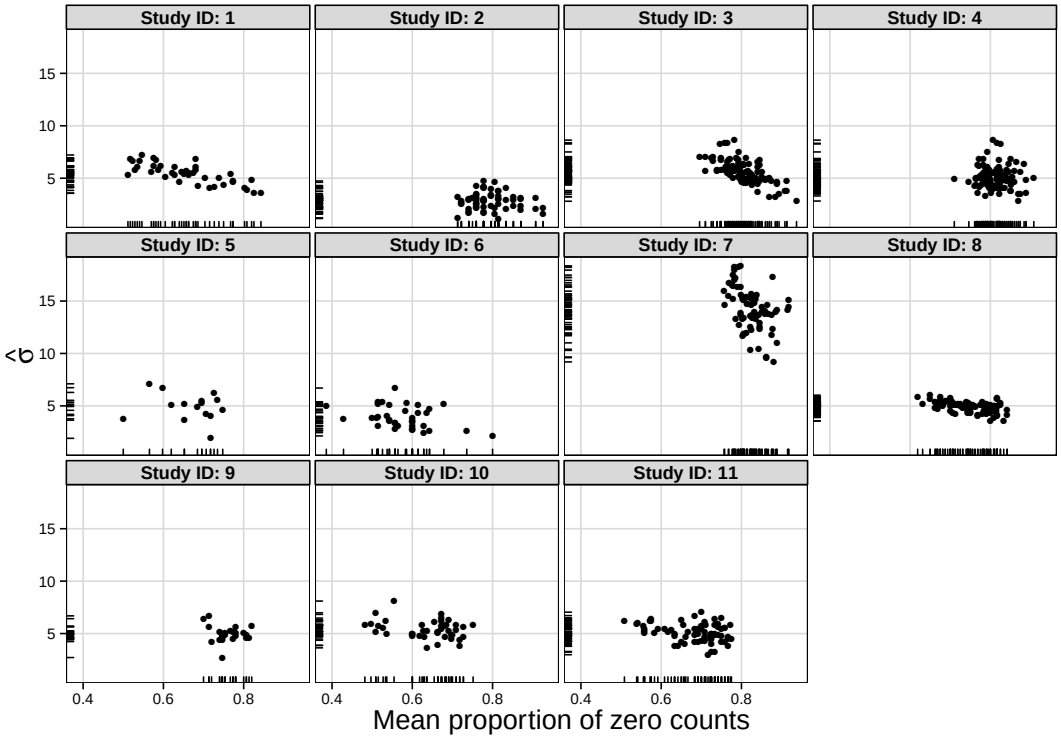

Figure A.1: Inter-replicate variability depending on the proportion of zero counts in the samples. Each point represents a sampling event and treatment.

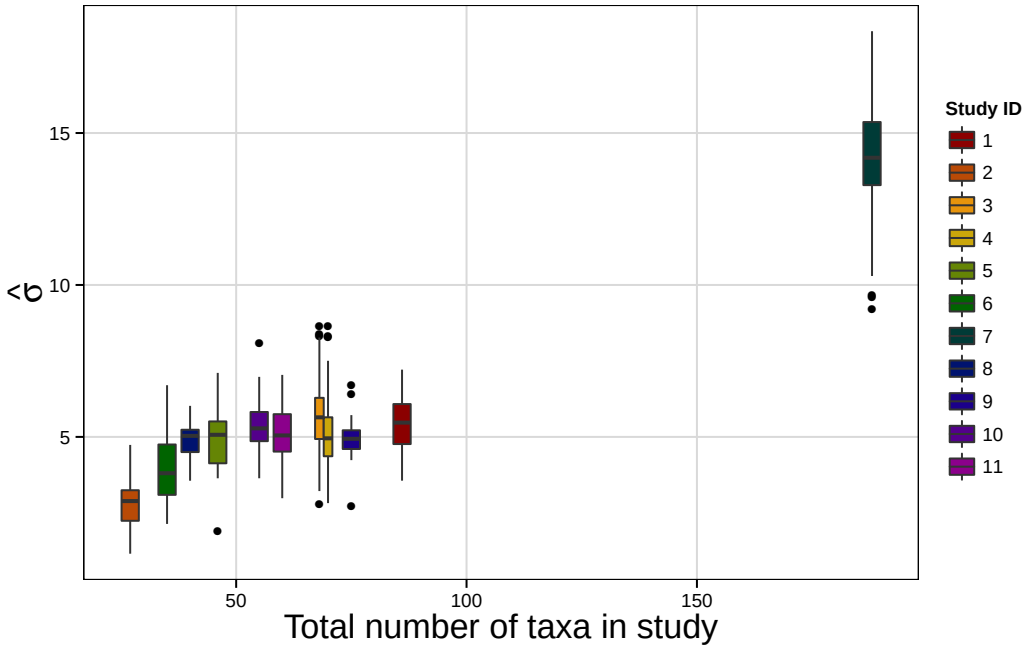

Figure A.2: Inter-replicate variability of the studies sorted by the total number of taxa in the studies.

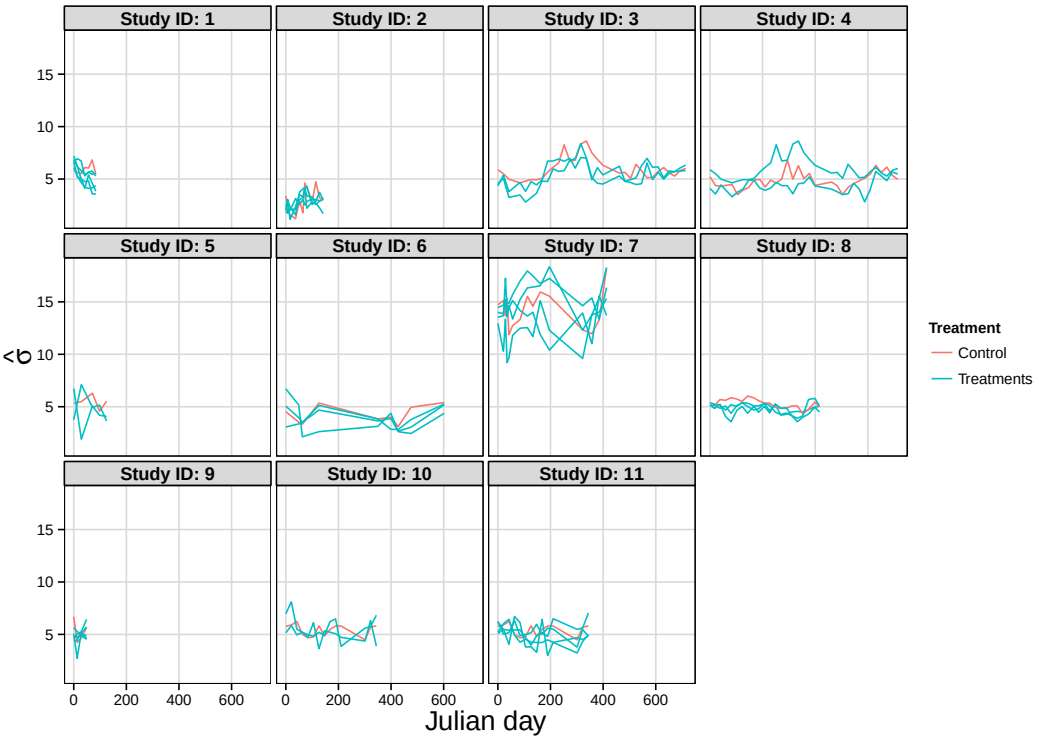

Figure A.3: Inter-replicate variability over study period. Each line represents a treatment.

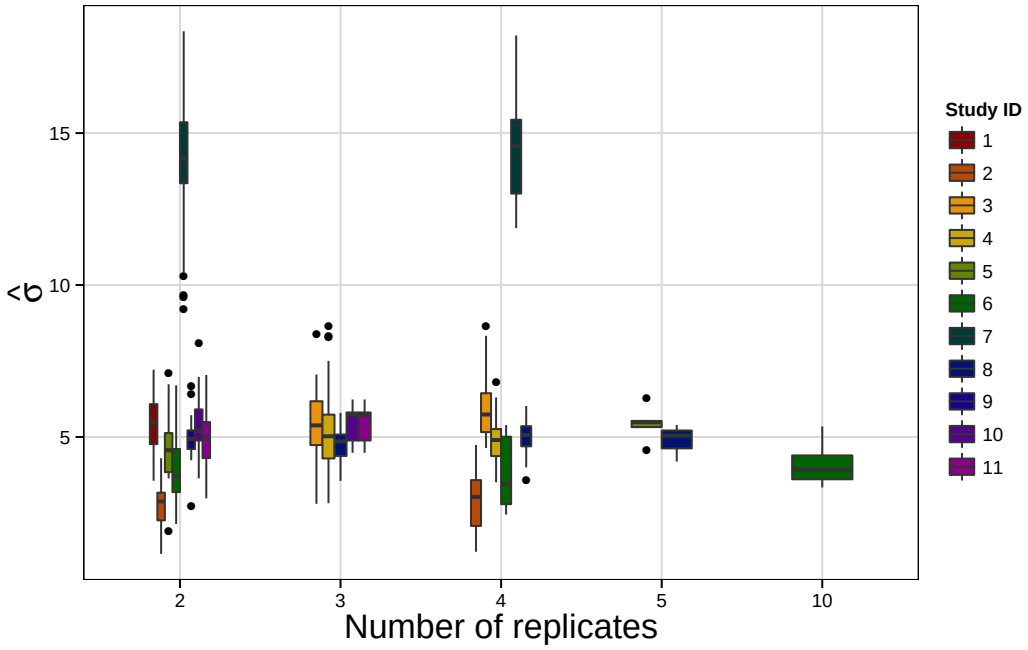

Figure A.4: Inter-replicate variability of the studies sorted by the number of replicates for each treatment.

B Tutorial: Analysing mesocosm data with R

B.1 Why use R to analyse ecotoxicological data?

There are many reasons to use R instead of other statistical software:

- It's free.
- It's platform independent (runs on Windows, Mac, Linux).
- It has enormous functionality (over 5000 packages) - some of them not yet available in other software.

However, the two main reasons for ecotoxicologists to use R are:

It's open-source. Everyone can take a look at the source code to see the exact computation. This is not the case with commercial software. This also enables to spot bugs.

It facilitates reproducible research. If the raw data is available and code is distributed everyone can reproduce and check the results.

In the next sections we will show you how mesocosm experiments can be analysed using R. Moreover, we show the general procedure of the individual analyses of this study.
Two packages need to be installed before running these examples: `vegan` and `mvabund`, both are available on official package repository (The Comprehensive R Archive Network (CRAN), `cran.r-project.org`).

B.2 Principal Response Curves (PRC)

B.2.1 Example data

We will analyse the pyrifos data set of van den Brink and ter Braak (1999) which is shipped with the `vegan` package. Twelve experimental ditches were used in this experiment: Four ditches served as control and the remaining eight were treated in duplicates once with in the insecticide chlorpyrifos at doses of 0.1, 0.9, 6 and 44 µg / L. Invertebrates were sampled 11 times from week -4 pre-treatment through week 24 post-treatment. A total of 178 taxa were identified, this resulted in a table of 132 rows (11 * 12 samples) and 178 columns (taxa).

```
require(vegan)
data(pyrifos)
head(pyrifos[, c(1:8)])

##        Simve Daplo Cerpu Alogu Aloco Alore Aloaf Copsp
## w.4.c1 3.951     0     0     0     0     0     0 2.773
## w.4.c2 2.303     0     0     0     0     0     0 2.079
## w.4.c3 4.595     0     0     0     0     0     0 3.761
## w.4.c4 2.398     0     0     0     0     0     0 3.296
## w.4.c5 4.025     0     0     0     0     0     0 3.466
## w.4.c6 2.303     0     0     0     0     0     0 2.197
```

Rows correspond to samples and columns are the species (with abbreviated names) - a *species x sites* matrix as commonly used in community ecology. The column names code treatment and time, but we will create a separate data.frame with variables providing information on the experimental ditch, sampling time and treatment:

```
ditch <- gl(12, 1, length = 132)
week <- gl(11, 12, labels = c(-4, -1, 0.1, 1, 2, 4, 8, 12,
    15, 19, 24))
dose <- factor(rep(c(0.1, 0, 0, 0.9, 0, 44, 6, 0.1, 44,
    0.9, 0, 6), 11))
```

```
pyrifos_env <- data.frame(ditch, week, dose)
```

B.2.2 Analysis of overall treatment effect

To calculate PRCs we can use the `vegan` package and the `prc()` function therein. It takes the abundance data (it is already log-transformed) as response and the treatment and time as explaining variables:

```
pyrifos_prc <- prc(response = pyrifos, treatment = dose,
    time = week)
pyrifos_prc_sum <- summary(pyrifos_prc, scaling = 1)
```

A plot can be produced using `plot()`, see Figure B.1. Note that only species with scores greater or smaller than 1 are displayed to avoid cluttering.

```
plot(pyrifos_prc, select = abs(pyrifos_prc_sum$sp) > 1,
    scaling = 1)
```

The plot (Figure B.1) shows on the x axis the time and on the y-axis the difference from the control treatments. The farther apart from the x-axis the more different are the communities compared to the control group.
We clearly see a treatment related effect: After application at week 0 treated communities change treatment-dependent. However, at the end of the experiment the treated and the control communities become similar again, which indicates a *recovery*.
On the right-hand side of this plot we see the species names and their scores. The more extreme the scores the more this species contributed to the observed pattern. However, you cannot directly infer from these species scores which species are more susceptible. For example *Gammarus pulex* (gammapule) has a relatively low score, although its response pattern (Figure B.2) shows a strong response, but without recovery. PRC displays global pattern in the community, but the pattern of *G. pulex* is different from most other species, therefore it has a lower species score.
We can also look at the numerical output for this plot using the summary method (Only a shortened output is given here.):

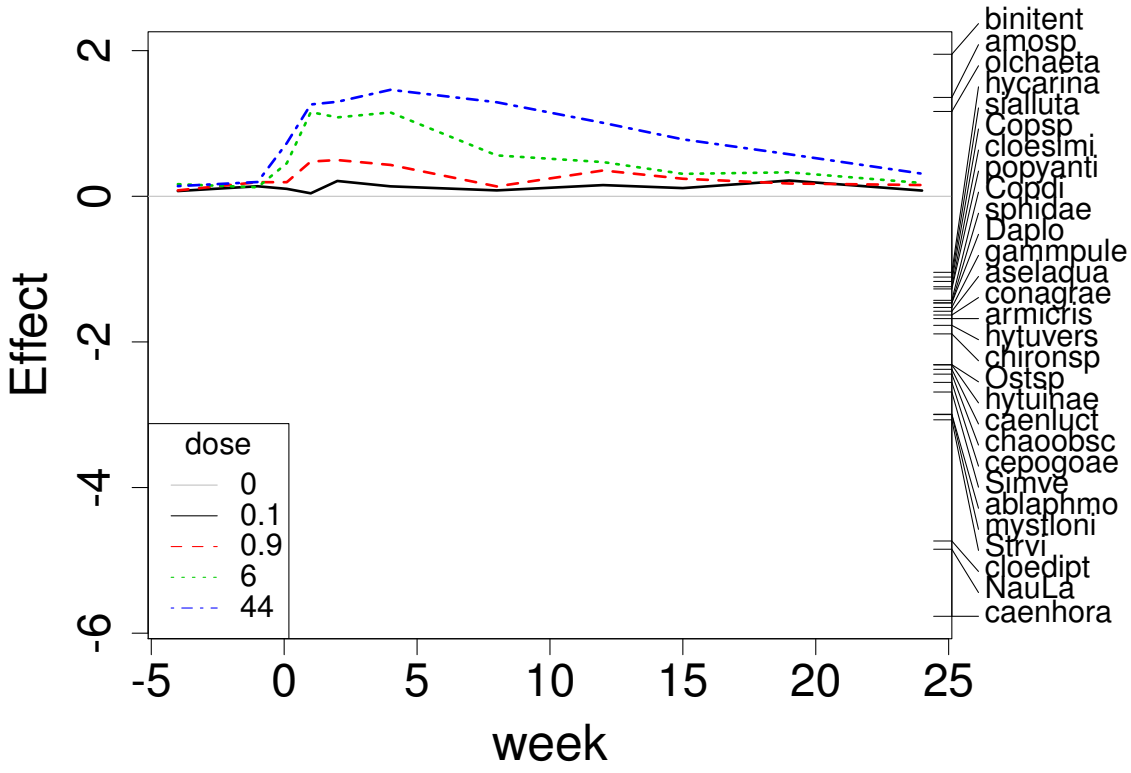

Figure B.1: Principal response curves (PRC) with species weights for the pyrifos data set indicating effects of the insecticide on the invertebrate community. Only species with scores greater or smaller than 1 are displayed to avoid cluttering of the plot. Only abbreviated species names are available in this dataset.

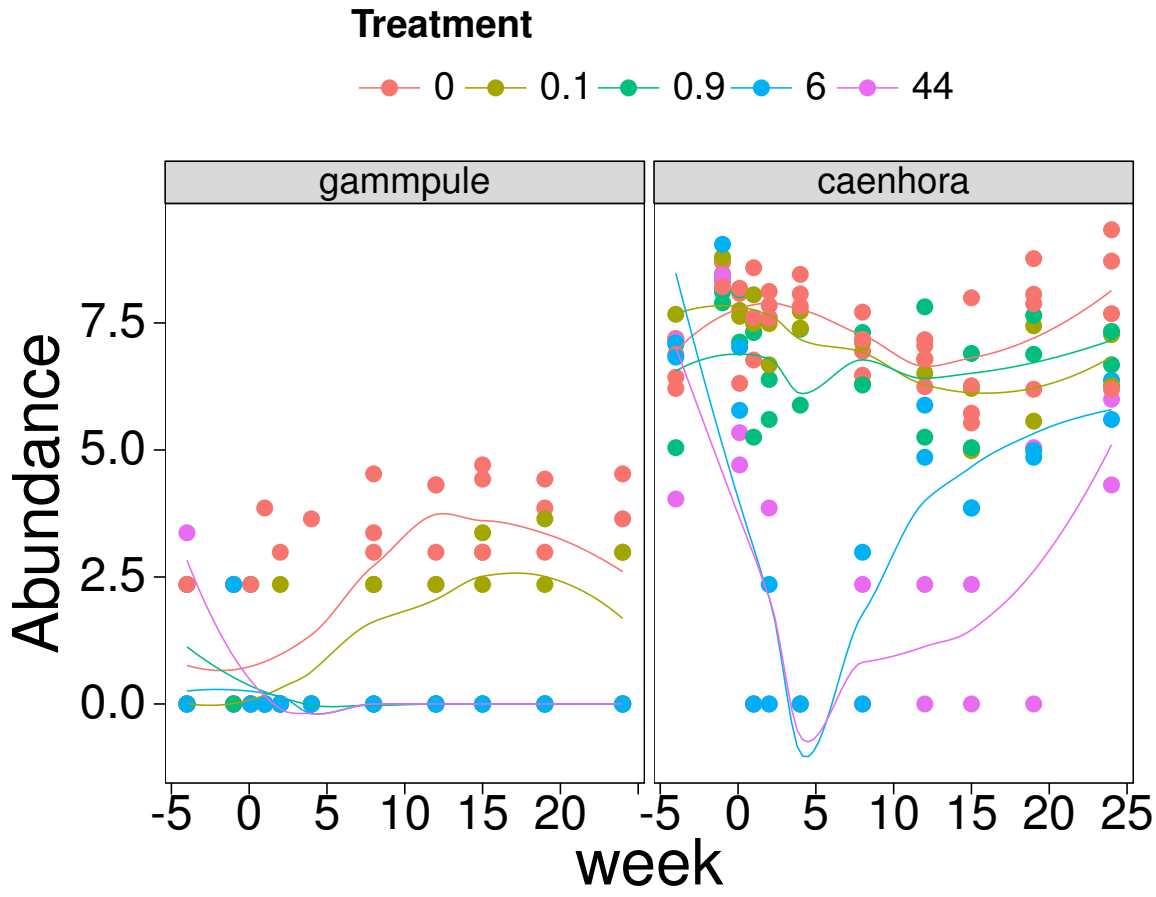

Figure B.2: Responses of *G. pulex* and *C. horaria* to chlorpyrifos.

```
pyrifos_prc_sum
```

```
##
## Call:
## prc(response = pyrifos, treatment = dose, time = week)
## Species scores:
##    Simve    Ostsp    NauLa    Strvi binitent caenhora
##   -2.688   -2.312   -4.847   -3.070    1.951   -5.768
## caenluct cloedipt hytuinae ablaphmo cepogoae chaoobsc
##   -2.376   -4.734   -2.316   -2.993   -2.555   -2.442
## mystloni
##   -2.998
##
## Coefficients for dose + week:dose interaction
## which are contrasts to dose 0
## rows are dose, columns are week
##          -4     -1    0.1       1      2      4
## 0.1 0.07218 0.1375 0.1020 0.04068 0.2101 0.1364
## 0.9 0.08106 0.1935 0.1936 0.47699 0.4977 0.4306
## 6   0.16616 0.1232 0.4539 1.15638 1.0835 1.1511
## 44  0.13979 0.1958 0.7308 1.26088 1.2978 1.4627
```

This summary returns the numerical species and sites scores.

The output of `prc()` gives more detailed information about the RDA model:

```
pyrifos_prc
```

```
## Call: prc(response = pyrifos, treatment = dose,
## time = week)
##
##               Inertia Proportion Rank
## Total         289.000      1.000
```

```
## Conditional    63.300      0.219   10
## Constrained    96.700      0.335   44
## Unconstrained 129.000      0.446   77
## Inertia is variance
##
## Eigenvalues for constrained axes:
##  RDA1  RDA2  RDA3  RDA4  RDA5  RDA6  RDA7  RDA8
## 25.28  8.30  6.04  4.77  4.15  3.86  3.59  3.33
##
## Eigenvalues for unconstrained axes:
##   PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
## 17.16  9.19  7.58  6.06  5.73  4.84  4.52  4.10
## (Showed only 8 of all 77 unconstrained eigenvalues)
```

We see that 21.9% of the variance can be attributed to time (Conditional, this is removed in partial RDA), 33.5% can be explained by the treatment regime (Constrained) and 44.6% of residual variance (Unconstrained), which cannot be explained by time and treatment.
The first RDA axis has an eigenvalue of 25.3. If we divide this eigenvalue by the sum of all eigenvalues, we obtain the proportion of explained variance which is displayed on the first axis.

```
pyrifos_prc$CCA$eig[1]/sum(pyrifos_prc$CCA$eig) * 100
```

```
##  RDA1
## 26.15
```

Since a PRC is related to RDA, we can also use the `summary()` function for rda-objects as a convenient way to access this information:

```
vegan:::summary.cca(pyrifos_prc, scaling = 1, display = "sites")
```

The output is not displayed because it is humongous.

The significance of the PRC diagram can be tested via permutations. However, observations from an experimental ditch are not independent, since the same ditch was measured repeatedly during the experiment. We have to take this into account: each ditch represents a time-series. Permuting the whole series of one ditch, keeping the temporal order, take this into account.
To setup such a permutation scheme we can use the `permute` package, which is automatically loaded with vegan:

```
control = how(plots = Plots(strata = ditch, type = "free"),
    within = Within(type = "none"), nperm = 199)
```

This sets the permutation scheme:

plots Permute mesocosms, without any restrictions.

within Within observations from one mesocosm there will be no permutations (keeping the time-series together).

nperm We request 199 permutations.

Note that we requested for demonstration purpose only 199 permutations here, so the best achievable p-value is 1/200 = 0.005. Usually 1000 or more permutations should be used (giving a minimal p-value of 1/1000 = 0.001). Permutation tests for PRC can be performed using the `anova()` function of vegan:

```
set.seed(1234)
anova(pyrifos_prc, permutations = control, first = TRUE)

## Permutation test for rda under reduced model
## Plots: ditch, plot permutation: free
## Permutation: none
## Number of permutations: 199
##
## Model: prc(response = pyrifos, treatment = dose, time = week)
##          Df Variance    F Pr(>F)
```

```
## RDA1      1     25.3 15.1   0.01 **
## Residual 77    129.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This runs a permutation test for the first eigenvalue of our model. We see that our first axis explains a statistically significant proportion of variation. The minimum p-value that we could obtain is 0.01 (=1/no. of permutations).
One can also test the significance of each axis separately Legendre et al. (2011) - the results for the first axis are identical, but testing only the first eigenvalue is computationally more efficient since we are generally not interested in all axes.

```
anova(pyrifos_prc, permutations = control, by = "axis")
```

or the terms in the model:

```
anova(pyrifos_prc, permutations = control, by = "terms")

## Permutation test for rda under reduced model
## Terms added sequentially (first to last)
## Plots: ditch, plot permutation: free
## Permutation: none
## Number of permutations: 199
##
## Model: rda(formula = pyrifos ~ dose * week + Condition(week))
##           Df Variance    F Pr(>F)
## dose       4     30.7 4.58  0.045 *
## week       0      0.0
## dose:week 40     66.0 0.99  0.050 *
## Residual  77    129.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This shows that the interaction term in the model is statistically significant. Note that no output for week is generated - this is because with the specified permutation scheme the significance of a week effect cannot be assessed.

B.2.3 Effects per week

The PRC indicates that there is a *treatment x time* interaction and we may want to look for effects at individual time-points to explore its nature.
We follow here van den Brink and ter Braak (1999) and use the ln-transformed nominal dose as continuous explanatory variable (regression-like analysis). However, dose can also be as categorical explanatory variable (anova-like analysis).
This code creates the log-transformed continuous dose:

```
dose_c <- log(20 * as.numeric(levels(dose))[dose] + 1)
```

Now we need to compute for every week a RDA and run a permutation test. This could be done using a for-loop:

```
rdas <- NULL
for (i in levels(week)) {
    take_spec <- pyrifos[week == i, ]
    take_dose <- dose_c[week == i]

    rdas[[i]]$rda <- rda(take_spec ~ take_dose)
    rdas[[i]]$anova <- anova(rdas[[i]]$rda, by = "terms",
        permutations = 99)
}
```

This returns a very long list: one list entry per week and each entry itself contains two lists: `rda` (RDA-Model) and `anova` (permutation test). Digging through this manually is to laborious, though, we have to extract only a small part of the information.
We can use `sapply()` to apply a function to every list entry and return results in a vector. For example, to extract the p-values for each week we can use:

```
sapply(rdas, function(x) x$anova[1, 4])

##   -4   -1  0.1    1    2    4    8   12   15   19   24
## 0.42 0.92 0.01 0.01 0.01 0.01 0.01 0.01 0.04 0.04 0.20
```

We can find effects from the application date until week 19.
A R package to simplify these computations with some convenience functions for Principal Response Curves is currently under development (`https://github.com/EDiLD/qetx`).

B.2.4 NOEC

Besides the overall significance of treatment and the effects per week ecotoxicologists are also interested which treatment differ from control. Usually, a no-observed-effect concentration (NOEC) [= the concentration below the lowest significant concentration] is derived from the data. However, the usage of NOEC has been criticised in the past. Testing via permutations fails here, because there are insufficient permutations left (we have only 2 treated and 4 control mesocosm per sampling date). One solution is to break down the multivariate data into a univariate one and use a univariate test.
To break down the community data into one variable one can use Principal Component Analysis (PCA), an unconstrained ordination technique, and take the sites scores on the first axis (which explains) as response variable. Usually a Williams-Test for trend is performed to look whether the treatment effect is statistically significant. In addition, a Dunnett-Test (comparing every treatment to control) could be applied.
Both (as contrast-versions) can be calculated using the `multcomp` package.
As in the previous section calculation are done within every week using a for-loop. For every week we need to:

1. Run a PCA.
2. Extract site scores on first axis.
3. Run a Dunnett/Williams-test on the site scores

The following runs these steps and returns the p-values for the comparisons using Dunnetts test:

```
df <- data.frame(dose = dose, week = week)
require(multcomp)
out <- NULL
for (i in levels(week)) {
    take_spec <- pyrifos[week == i, ]
    pca <- rda(take_spec)  # PCA
    pca_scores <- scores(pca, display = "sites", choices = 1,
        scaling = 1)
    mod_aov <- aov(pca_scores ~ dose, data = df[week ==
        i, ])
    out[[i]] <- summary(glht(mod_aov, linfct = mcp(dose = "Dunnett")))
}
```

Again we use an extractor function to gather the information we want:

```
pvals <- lapply(out, function(x) data.frame(comp = names(x$test$coefficients),
    p = x$test$pvalues))
# shows the p-values of Dunnett-Test on PCA-scores for
# week 1:
pvals[["1"]]

##      comp        p
## 1 0.1 - 0 0.999857
## 2 0.9 - 0 0.060708
## 3   6 - 0 0.002073
## 4  44 - 0 0.001187
```

In week 1 the NOEC would be 0.9 $\mu g/L$, as 6 $\mu g/L$ is the lowest statistically significant treatment different to the control group.

To perform a Williams test, change `mcp(dose = "Dunnett")` to `mcp(dose = "Williams")`. A R package with some convenience functions to simplify these computations is currently under development.

B.3 Multivariate Generalized Linear Models (mvGLM)

B.3.1 Example Data

We will use the same dataset as in the PRC tutorial. The data shipped with the vegan package have already been log-transformed. However, to use multivariate GLMs it is not necessary to transform abundances and therefore we first back-transform the abundances (this step is usually not needed, since raw counts are measured).

```
pyrifos_t <- round((exp(pyrifos) - 1)/10)
```

B.3.2 Analysis of overall treatment effect

Multivariate GLMs can be calculated using the `mvabund` package.
To investigate the overall treatment effect we first build a model with dose, week and their interaction as explaining variables:

```
require(mvabund)
pyrifos_mv <- mvabund(pyrifos_t)
mod <- manyglm(pyrifos_mv ~ dose + week + dose:week, data = pyrifos_env)
```

This sets up the GLMs for every species using a negative binomial error distribution. Other possible distributions are: poisson (for count data), gaussian (linear regression) and binomial (for presence-absence data).
The `plot()` function shows a residual vs. fitted values plot and is used to check if the model assumptions are met (see Figure B.3, there should be no obvious pattern in the residuals).
MvGLMs tests the significance by permutation, as with PRC we need to take the temporal autocorrelation into account. However, we need to construct a *permutation matrix*, holding the permutations generated by the permutation design. This matrix is passed directly to `anova()` instead of passing the permutation design.

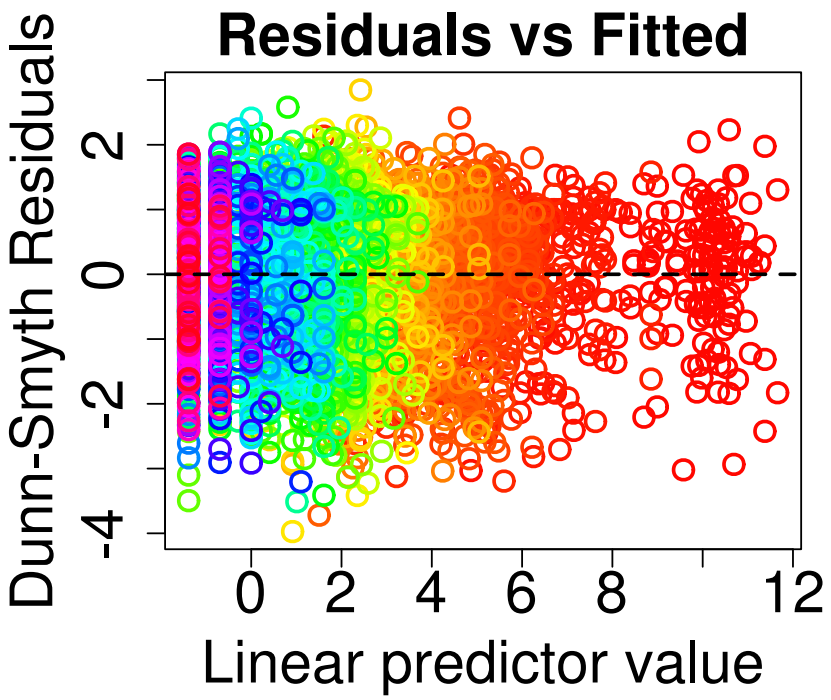

Figure B.3: Residual vs. Fitted plot for the mvGLM model

```
control <- how(within = Within(type = "none"),
    plots = Plots(strata = factor(ditch),
    type = "free"), nperm = 99)

permutations <- shuffleSet(nrow(pyrifos_t), control = control)
```

Significance can be tested using the `anova` function of mvabund:

```
aov_mglm <- anova(mod, bootID = permutations, p.uni = "unadjusted",
    test = "LR", resamp = "perm.resid")
```

Here we specify that

bootID = permutations Use permutations from the specified permutation design.

p.uni = 'unadjusted' Return also univariate test statistics (unadjusted for multiple testing).

test = 'LR' Use the sum-of-likelihood ratio as test statistics.

resamp = 'perm.resid' permute residuals

As output we obtain the result of the multivariate test, as well as the univariate tests:

```
aov_mglm

## Analysis of Deviance Table
##
## Model: manyglm(formula = pyrifos_mv ~ dose + week + dose:week,
## data = pyrifos_env)
##
## Multivariate test:
##             Res.Df Df.diff  Dev Pr(>Dev)
## (Intercept)    131
## dose           127       4 1430     0.13
## week           117      10 4129     0.01 **
## dose:week       77      40 4111     0.01 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Univariate Tests:
##              Simve          Daplo          Cerpu
##                Dev Pr(>Dev)   Dev Pr(>Dev)   Dev
## (Intercept)
## dose           2.7     0.56 34.36     0.07 16.83
## week          36.45     0.42  52.2     0.07 60.82
## dose:week    104.74     0.09 27.52     0.21  5.43

....

## Arguments:
##  Test statistics calculated assuming uncorrelated response
```

```
## (for faster computation).
## P-value calculated using 99 resampling iterations via residual
## permutation (without replacement) resampling (to account for
## correlation in testing.
```

We see that there is a statistically significant interaction between treatment and time for the community, which indicates that the effect of treatment varies over time.
We obtain univariate tests, for example taxon *Simocephalus vetulus* (`Simve`) shows also a statistically significant interaction. Moreover the deviance indicates how much this species contributes to the community response.
Moreover we can compare different models:

```
mod2 <- manyglm(pyrifos_mv ~ week, data = pyrifos_env)
aov_mglm2 <- anova(mod2, mod, bootID = permutations, p.uni = "unadjusted",
    test = "LR", resamp = "perm.resid")

## Using bootID matrix from input.
## Time elapsed: 0 hr 3 min 36 sec


aov_mglm2

## Analysis of Deviance Table
##
## mod2: pyrifos_mv ~ week
## mod: pyrifos_mv ~ dose + week + dose:week
##
## Multivariate test:
##      Res.Df Df.diff  Dev Pr(>Dev)
## mod2    121
## mod      77      44 5998     0.01 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Univariate Tests:
##       Simve          Daplo          Cerpu
##         Dev Pr(>Dev)   Dev Pr(>Dev)   Dev Pr(>Dev)
## mod2
## mod  123.55     0.10 49.95     0.12 31.21     0.35


....


## Arguments:
##  Test statistics calculated assuming uncorrelated response
## (for faster computation).
## P-value calculated using 99 resampling iterations via residual
## permutation (without replacement) resampling (to account for
## correlation in testing.
```

This is analogously to PRC, the output indicates that the two models differ and therefore we can conclude that the treatment effect is substantial (including the interaction with time).

B.3.3 Effects per week

Now that we know that the treatment effect is varying with time, we can investigate this further and look at effects at individual time points.
The concept is like in section B.2.3: Loop through time and for every time point fit and test the model.
Here is a function that does this for us:

```
mv_per_time <- function(response, treatment, time, nperm = NULL) {
    df <- data.frame(treatment = treatment, time = time,
        stringsAsFactors = FALSE)
    out <- NULL
```

```
    for (i in levels(time)) {
        rsp <- mvabund(response[df$time == i, ])
        out[[i]]$mod <- manyglm(rsp ~ treatment, data = df[df$time ==
            i, ])
        out[[i]]$anova <- anova(out[[i]]$mod, nBoot = nperm,
            p.uni = "unadjusted", test = "LR", resamp = "perm.resid",
            show.time = "none")
    }
    return(out)
}
```

```
per_week <- mv_per_time(pyrifos_mv, pyrifos_env$dose, pyrifos_env$week,
    nperm = 99)
```

The output is enormous (for every week a model with multivariate and univariate responses) and is skipped here.
p-values can be extracted for every week:

```
sapply(per_week, function(y) y$anova$table[2, 4])

##     -4     -1    0.1      1      2      4      8
## 0.7576 0.7778 0.1515 0.0101 0.0101 0.0101 0.0404
##     12     15     19     24
## 0.1313 0.2929 0.2525 0.4949
```

Note that this is a different model than in section B.2.3: there we used a log-transformed continuous explaining variable, here we use treatment as factor (dummy variable).

B.3.4 NOEC

Since mvGLM use also permutations to compute a p-value the same problem as with RDA arises: there are insufficient permutations left. Therefore we cannot use mvGLM to derive a NOEC and should use the PCA method (section B.2.4).