Standardisation of river classifications:

Framework method for calibrating different biological survey results against ecological quality classifications to be developed for the Water Framework Directive

Contract No: EVK1-CT 2001-00089
Deliverable N1, due 31/05/04, entitled:

Comparison of (cost) effectiveness between various macroinvertebrate field and laboratory protocols

Compiled by Hanneke Vlek
Partner no 4 (Alterra, The Netherlands)

A project under the 5th Framework Programme
Energy, Environment and Sustainable Development
Key Action 1: Sustainable Management and Quality of Water

Table of contents

Preface

This deliverable is one of the outputs of the STAR project, which was funded by the European Commision, 5th Framework Program, Energy, Environment and Sustainable Development, Contract No. EVK1-CT2001-00089. The Water Framework Directive (WFD) provides a framework for the ecological assessment of surface waters and the development of catchment management plans to guarantee sustainable management of surface waters.

The wide variety of assessment methods for streams and rivers in Europe, some of which developed during the EU project AQEM, provide great opportunities for developing effective biological monitoring and assessment protocols. However, the development of many new approaches may not be compatible with an important objective of the WFD: to gain comparability of stream assessment results all over Europe. To gain comparable results assessment must be preformed in a standardised way.

Assessment results are influenced by the methods used to collect and process biological samples, therefore this deliverable focuses on issues relates to the standardisation of sampling and sample processing of macroinvertebrates. In standardising field and laboratory protocols for macroinvertebrate sampling it is important to know what should be standardised and which choices yield optimal results both in terms of cost efficiency and metric/assessment results.

The objective of this project was to study the results of different macroinvertebrate field and laboratory methods particularly in terms of errors, precision, and effectiveness in relation to the assessment/management objective. To investigate the cost effectiveness of different approaches to the collection and processing of macroinvertebrate samples involving varying levels of resource intensity.

The following partners contributed to this deliverable:

- ALTERRA Green World Research, Centre for Ecosystem Studies, Team Freshwater Ecosystems, Wageningen, the Netherlands (Hanneke Vlek)
- BOKU - University of Natural Resources and Applied Life Sciences, Department of Water, Atmosphere & Environment, Institute for Hydrobiology & Water Management, Working Group on Benthic Ecology and Ecological Status Assessment (Otto Moog, Patrick Leitner and Philipp Wenzl)
- Comenius University of Bratislava, Faculty of Science, Department of Ecology, Bratislava, Slovakia (Il'ja Krno)
- Institute of Zoology, Slovak Academy of Sciences, Bratislava, Slovakia (Ferdinand Sporka)
- Masaryk University, Faculty of Science, Department of Zoology & Ecology, Division of Hydrobiology, Brno, Czeck Rebublic (Karel Brabec)

1 Introduction

The wide variety of assessment methods for streams and rivers in Europe provide great opportunities for developing effective biological monitoring and assessment protocols for all types of stressors and geographic regions. However, the variety of organism groups and types that need to be considered and the multiplicity of established methodologies in current usage present significant problems for consistent inter-state interpretation and allocation of the Ecological Status. Given these potential problems, inter-calibration of protocols and standardisation of interpretation and allocation of the Ecological Status is crucial to the implementation of the WFD. The STAR project aims to solve these problems by addressing a series of key issues. One of these key issues is the standardisation of field and laboratory protocols for macroinvertebrate samples.

Since the beginning of this century a wide variety of methods for the biological assessment of streams has been developed. Macroinvertebrates are commonly used in bioassessment of lakes and streams (Hawkes, 1979; Hellawell, 1986; Bailey *et al.*, 2001). The collection and processing of macroinvertebrate samples is a process that consists of several steps (Figure 1.1). Each step in the sample processing chain stands for choices that has to be made, e.g. "Do we sample all habitats or not?" and "Do we identify to genus or species level?" Depending on the choices made the actual structure and condition of the macroinvertebrate community may be misinterpreted (Diamond *et al.*, 1996). The choices made will influence the final result, the taxa list. Since assessment is based on this taxa list, assessment results (Ecological Status) can vary based on the choices made during sampling and sample processing.

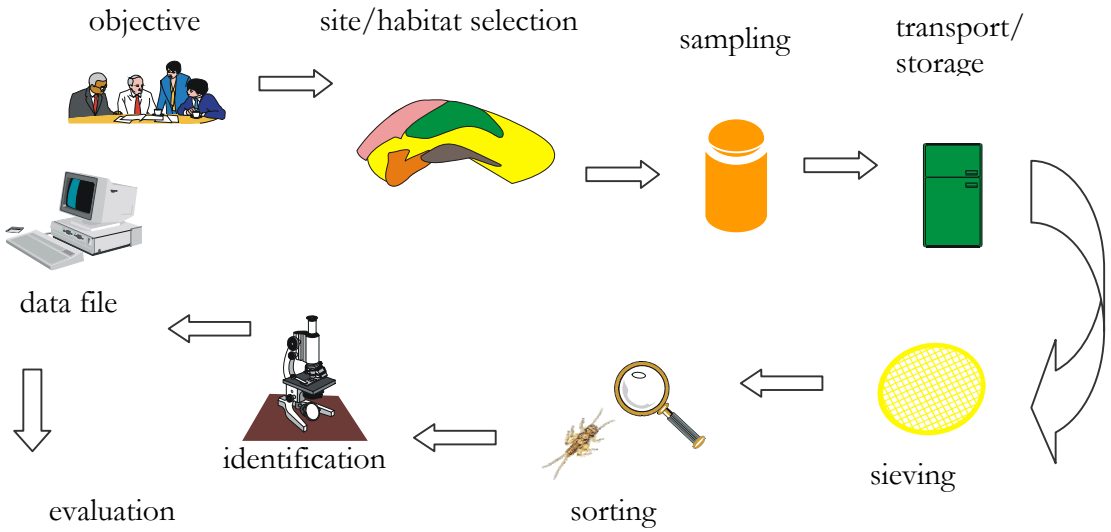


Figure 1.1. Overview of the different steps forming the sample processing chain.

Over the years many different field and laboratory protocols regarding macroinvertebrates have been developed in Europe and the United States, e.g. RIVPACS, PERLA, Rapid Bioassessment Protocols. These protocols are often based on

subjective arguments like tradition or convenience, rather than on optimising the data obtained per unit effort (Norris *et al.*, 1993; Carter & Resh, 2001). Standardisation of protocols within Europe will not be easy due to these different traditions within the different member states. Collecting objective information on the consequences of choices made during sampling and sample processing may facilitate the process of standardisation. In standardising field and laboratory protocols for macroinvertebrate sampling it is important to know what should be standardised and which choices yield optimal results both in terms of cost efficiency and metric/assessment results.

The objective of this project was to study the results of different macroinvertebrate field and laboratory methods particularly in terms of errors, precision, and effectiveness in relation to the assessment/management objective. To investigate the cost effectiveness of different approaches to the collection and processing of macroinvertebrate samples involving varying levels of resource intensity.

Since it was not possible to study all aspects of sampling and sample processing a selection was made. The research project was divided into several smaller studies, described in six different chapters. The first chapter describes the effects of seasonal variation on bioassessment. In the second chapter the effectiveness of samples taken with a 500 µm and 1000 µm mesh size net is compared. In the third chapter the (cost) effectiveness of samples that were preserved directly after sampling was compared with the (cost) effectiveness of samples that were not preserved. In the fourth chapter the (cost) effectiveness of five different protocols (ALTERRA, EBEOSWA, RIVPACS, STAR and STARp) for macroinvertebrate sampling and sample processing is compared. The fifth chapter deals with the (cost) effectiveness of subsampling procedures (only a part of the sample is sorted).

The sampling method applied in all studies consisted of quantitative sampling of multiple habitats with hand-nets, surber samplers or shovels. The focus was on this sampling technique because it was applied in the EU project AQEM and during this project assessment systems were developed that complied with the demands of the Water Framework Directive.

2 Seasonal variation

2.1 Introduction

One of the objectives of the Water Framework Directive is to standardise bioassessment of surface waters. Water managers prefer cost efficient methods, e.g. sampling only once a year for the purpose of surveillance monitoring. A higher level of standardisation would be reached when samples from the same area would be collected in the same time period. This raises the question which time period would be most suited for taking samples. In many European countries there is some general understanding about the preferred time period for sampling. In a lot of cases, however, scientific background is lacking.

The aim of this study was (1) to examine the variation in macroinvertebrate community composition between months (2) to assess the effect of this variation on metrics and assessment results and (3) to determine whether a preferred time period for sampling can be indicated for lowland streams in the Netherlands and lower mountainous streams in the Slovakia.

2.2 Methods

2.2.1 Study site and data collection

The Netherlands

All samples were collected in the Heelsumse beek, a small sand bottomed lowland (<200 m above sea level) stream in the eastern part of the country. The study site consisted of a relative uniform 100 m section of stream. The 100 m section was divided into two 50 m stretches. Two replicate samples were taken every other month, alternate from the two stretches, for one year. In total 12 composite samples were taken in the first week of six different months (May, July, September, November, January and March). For the collection of the samples a hand-net (25 cm wide with a 500 μm mesh) was used. Samples were collected by pushing the net through the upper part (2-5 cm) of the substratum. Prior to sampling habitat coverage was estimated for the complete 100 m section. A sample was taken from each habitat (subsample) with coverage of more than 5%. The sampled area of each habitat was the same at all sampling occasions. All samples were collected by the same operator. The subsamples were stored separately in buckets, transported to the laboratory and stored in a refrigerator, where they were oxygenated. All subsamples were kept separate during sample processing. The subsamples were sieved using a 1000 and 250 μm sieve. The samples were sorted by eye. Sorting was performed by a group of three to five people. Identification was done by the same two people for all samples. Organisms were identified to the lowest taxonomic level possible (species level for almost all groups).

Slovakia

All samples were collected in the Stupavsky potok, a small calcareous lower mountainous (200-500 m above sea level) stream in the Carpathian. The study site consisted of a relative uniform 100 m section of stream. The 100 m section was divided into two 50 m stretches. Two replicate samples were taken every other month, alternate from the two

stretches, for one year. In total 12 composite samples were taken at the last week of six different months (April, June, August, October, December, and February). The samples from December was actually taken on the 8th of January, but is referred to as the sample from December. For the collection of the samples a hand-net (25 cm wide with a 500 μm mesh) was used. Prior to sampling habitat coverage was estimated for the complete 100 m section. A sample was taken from each habitat (subsample) with coverage of more than 5%. The sampled area of each habitat was the same at all sampling occasions. All samples were collected by the same operator. The subsamples were preserved in 4% formaldehyde. All subsamples were kept separate during sample processing. The subsamples were sieved using a 1000 and 500 μm sieve. The samples were sorted by eye. Sorting was performed by a group of three people. Identification was done by the same two people for all samples. Organisms were identified to the lowest taxonomic level possible (species level for almost all groups). All identifications were preformed by the same specialist for each major organism group. For a more elaborate description of the used protocols for sampling and sample processing we refer to the AQEM sampling protocol and revised AQEM/STAR sorting protocol described by STAR Consortium (2003), in this paper the protocol is referred to as the STAR protocol. The samples from April form an exception to the above description. In April the subsamples form different habitats were not kept separate during sample processing, the sample was however fully sorted.

2.2.2 Data analysis

A correspondence analysis (CA) was performed separately for both sites to gain insight in community composition variation between months. Species data were log 2 (x+1) transformed before analysis.

CA was followed by evaluation of metrics commonly used in Europe (Table 2.2). The metrics were selected from an extensive list given by Hering *et al.* (2004). Apart from metrics selected form the list by Hering *et al.* (2004) the number of taxa and the number of individuals for each major macroinvertebrate group (e.g. Diptera, Ephemeroptera, Plecoptera) was also evaluated. Some major groups were only present in a few samples and in low abundances. These groups were excluded from evaluation, because it is difficult to normalize through transformation in case many zero values occur (Metzling *et al.*, 2003).

Prior to analysis the number of individuals per taxon was standardised to a total sample area of 1.25 m^2 for each composite sample based on habitat coverage and sampled area. Metric values were calculated with the AQEM River Assessment Program (AQEMrap version 2.3). With the same program the final ecological quality classes were calculated for the Heelsumse beek, characterising the samples as being from a small Dutch lowland stream. The multimetric index used to calculate the final ecological quality class is a revised version of the multimetric index described by Vlek *et al.* (2004). The multimetric index consists of 11 metrics which are indicated in table 2.2. The ecological quality class for the Stupavsky potok could not be calculated because no suited assessment system was available.

One-way analysis of variance (ANOVA) was used to look for significant differences between months (α=0.05). Assumptions for normality and homogeneity of variance could not be tested in a reliable way due to the low number of samples. For this reason it would have been more appropriated to perform a non-parametric test. However, a non-parametric test would never be able to detect significant differences between protocols based on two replicates. Therefore it was decided to use an ANOVA

and to transform metric values based on experiences in other studies. Abundance metrics were ln(x+1) transformed, taxa counts were not transformed (Kerans *et al.*, 1992) and proportions were ln(x+1)-ln(y+1) transformed. Biotic index data (e.g. Sapobic Index, BMWP, ASPT) were not transformed (Norris & Georges, 1993). Metrics like XENO (%), SHRED (%) and littoral (%) are not simple proportional metrics. The values for these metrics also depend on the strength with which a species prefers a certain category (AQEM consortium, 2002). The decision was made not to transform values of these metrics, since no references were found on a suited transformation.

The coefficient of variation (CV) was calculated for the 11 metrics incorporated in the Dutch multimetric index based on the 12 samples form the Heelsumse beek. For the same 11 metrics the CV was calculated based on the 35 samples from the AQEM project post-classified as indicating good ecological quality, that were used to calibrate the Dutch multimetric index. By comparing the two CVs information is gained on how much variation was taken into account in developing the multimetric index. Appendix 1 gives an overview of mean values and standard deviations per metric for the Heelsumse beek and the Stupavsky potok. The appendix is meant to inform people on the variation they can expect on an annual basis so they can take this into account when developing an assessment system.

2.3 Results

2.3.1 General

The Netherlands

In total 136 taxa were collected during this study. One replicate contained on average 37% of the total number of taxa found during this study. In most months the total number of taxa occurring in both replicates from one month varied between 51% and 61%. The macroinvertebrate community of the Heelsumse beek consisted for a large part of Diptera taxa (Figure 2.1). Seasonal changes were not marked by major changes in the number of taxa for the different organism groups (Figure 2.1).There was no significant difference between months for the total number of taxa ($p=0.846$).

Similar to the total number of taxa there was also no significant difference between months for the total number of individuals ($p=0.356$). However, the number of individuals per major macroinvertebrate group did vary between months (Figure 2.2). During all months the Crustacea formed a large proportion of the community (varying between 40 and 63%). While in September the Plecoptera made up a large proportion of the community (32%) next to the Crustacea, in May it were the Diptera (48%), in November the Oligochaeta, and in July the Hydracarina (23%) (Figure 2.2). In January and March there was no dominant organism group apart from the Crustacea.

Based on multivariate analysis the samples could by divided into three groups: (1) September samples (2) July and May samples and (3) November, January and March samples (Figure 2.3).

The samples from the Heelsumse beek were characterised by a higher CV than the samples from the AQEM project for the metrics ALPHA-MESO (%) and hyporhithral (%). The differences between CVs for the metrics littoral (%), PSA (%), GSI and OLIGO (%) were not very large (Table 2.1).

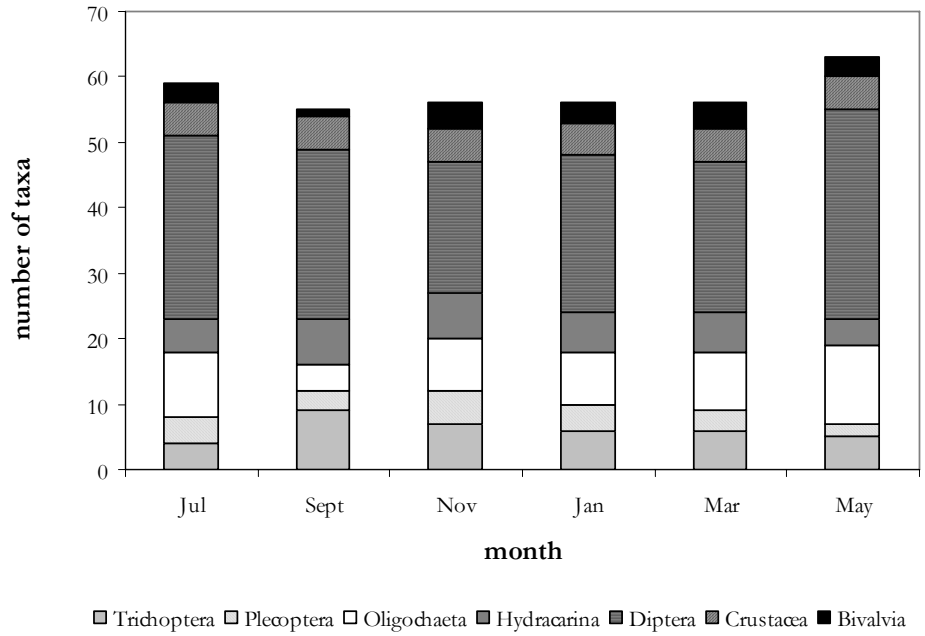


Figure 2.1. Bi-monthly variation in faunal composition for the Heelsumse beek. Number of taxa based on the sum of both replicates. Only groups that formed more than 5% of the total community were included in the graph.

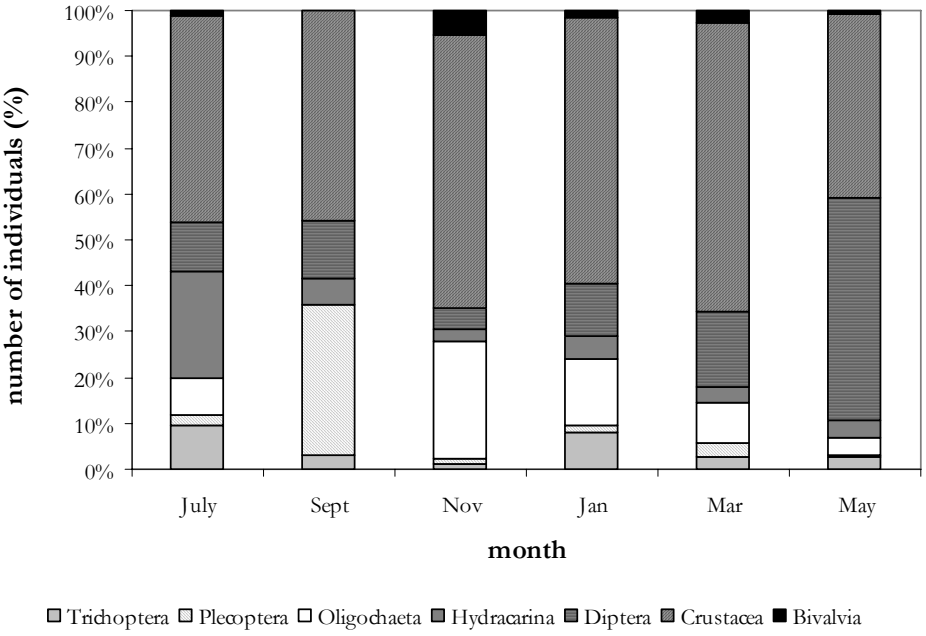


Figure 2.2. Bi-monthly variation in faunal composition for the Heelsumse beek. Percentage of individuals based on the average of both replicates. Only groups that formed more than 5% of the total community were included in the graph.

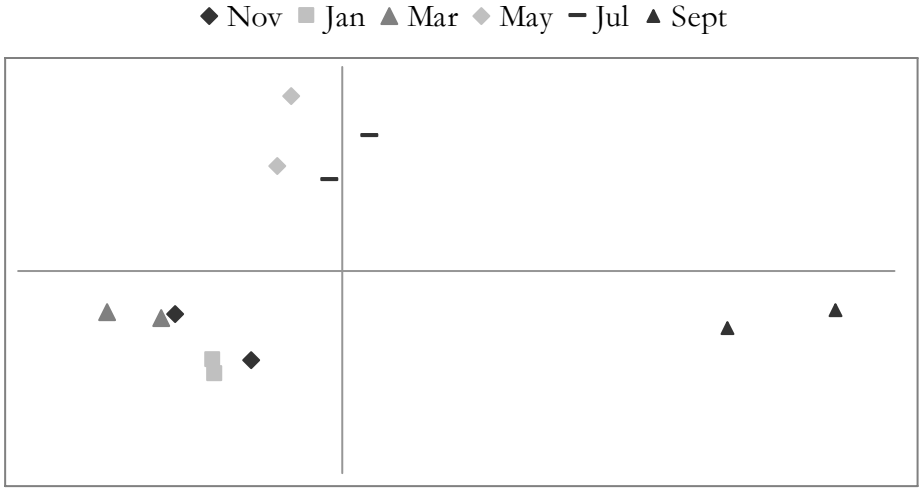


Figure 2.3. CA ordination diagram of axis 1 and 2 for the Heelsumse beek.

Table 2.1.Coefficient of variation (CV) for the samples from the Heelsumse beek and for the samples from the AQEM project post-classified as good and used for calibration of the Dutch multimetric index.

acronym	CV Heelsumse beek samples	CV AQEM samples
ALPHA-MESO (%)	32%	19%
ASPT	6%	12%
EPT/DIP-taxa	29%	55%
EPT-taxa (%)	23%	39%
GSI	7%	12%
hyporhithral (%)	26%	21%
littoral (%)	57%	64%
OL+DIP-taxa (%)	12%	23%
OLIGO (%)	29%	35%
AKA+LIT+PSA (%)	14%	20%
PSA (%)	41%	43%

Slovakia

In total 219 taxa were collected during this study. One replicate contained on average 42% of the total number of taxa found during this study. The total number of taxa occurring in both replicates from one month varied between 56% and 70%.

The macroinvertebrate community of the Stupavsky potok consisted for a large part of Diptera taxa (Figure 2.4). Seasonal changes were not marked by major changes in the number of taxa for the different organism groups (Figure 2.4).There was no significant difference between months for the total number of taxa ($p=0.170$).

Similar to the total number of taxa there was also no significant difference between months for the total number of individuals ($p=0.057$). However, the number of individuals per major organism group did vary between months (Figure 2.5). During most months (except from April and February) the Crustacea formed the largest proportion of the community (varying between 25 and 57%), followed by the Diptera (varying between 15 and 38%). Instead of the Crustacea (less than 1%) the Diptera made

up a large proportion of the community in February (38%). In April the Trichoptera made up 31% of the community, while Diptera (9%) and Crustacea (10%) formed a relative small part of the community (Figure 2.5).

Based on multivariate analysis the samples could by divided into three groups: (1) April samples (2) June and August samples (3) October, December and February samples (Figure 2.6).

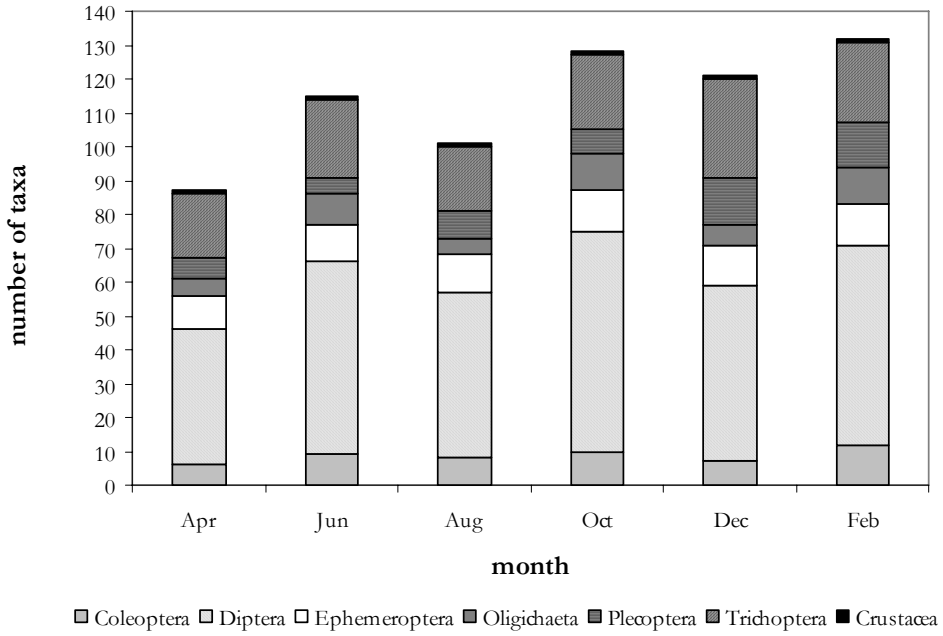


Figure 2.4. Bi-monthly variation in faunal composition for the Stupavsky potok. Number of taxa based on the sum of both replicates. Only groups that formed more than 5% of the total community were included in the graph.

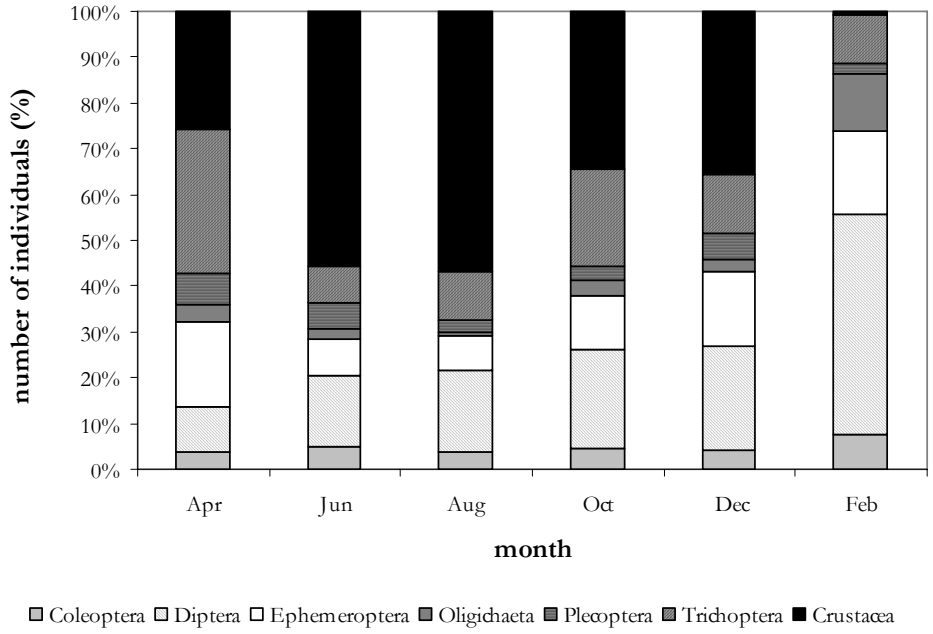


Figure 2.5. Bi-monthly variation in faunal composition for the Stupavsky potok. Percentage of individuals based on the average of both replicates. Only groups that formed more than 5% of the total community were included in the graph.

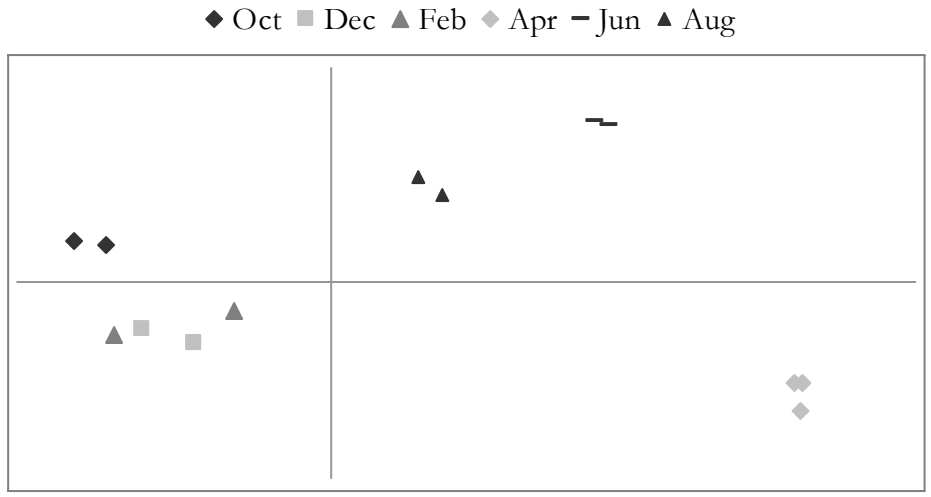


Figure 2.6. CA ordination diagram of axis 1 and 2 for the Stupavsky potok.

The Netherlands-Slovakia

Apart from geology of the catchment and the altitude at which the streams are located the Stupavsky potok and the Heelsumse beek especially differ in habitat composition. The substrate in the Stupavsky potok consists for a large part of gravel, stones and cobbles, while the Heelsumse beek is characterised by submerged macrophytes in summer and the rest of the year the substrate mainly consists of sand.

Comparing the macroinvertebrate community from the Stupavsky potok and the Heelsumse beek taxa richness was higher in the Stupavsky potok (219) than in the Heelsumse beek (136). The total number of individuals on average collected from the Stupavsky potok (4473) was also higher than the total number of individuals on average collected from the Heelsumse beek (3442). Both macroinvertebrate communities consisted for a large part of Diptera taxa and the Crustacea formed a large part of the community in most months. Only seven species were found in both streams. In total 31 taxa were found in both streams.

The multivariate analysis results are similar for the Stupavsky potok and the Heelsumse beek. Samples in Slovakia were taken at the end of the month compared to the beginning of the month in the Netherlands, so what seems to be a difference of one month is never more than two weeks. In this light the groups formed in Slovakia and the Netherlands are very comparable. The biggest difference is that in the Heelsumse beek September/August seems to differ most from all other months and in the Stupavsky potok this was May/April.

2.3.2 Metrics and assessment

The Netherlands

Eighteen out of 62 metrics showed significant ($p<0.05$) differences between months (Table 2.2). Of the 18 metrics showing significant differences between months, 14 were quantitative metrics. To discover what was causing these differences the taxa lists were studied in more detail. It turned out that for six out of nine metrics (metrics directly

related to a major macroinvertebrate group were not considered) the significant differences between months could be explained by the contribution of just one (or two) of the taxa mentioned in table 2.3. Table 2.3 shows that one taxon can make up a large proportion of the total number of individuals in a sample in one month and be (almost) absent in another month. Depending on the metric these shifts in abundance can strongly influence metric results; an example is the metric PSA (%). The values for this metric in September differed from the values in all other months. *Nemurella pictetii* and *Proasellus coxalis* made up 58% of the total number of individuals in September, while in other months they did not reach more than 5% of the total number of individuals (Table 2.3). Both taxa have a strong preference for stones and gravel, so automatically values for the proportion of individuals preferring a sandy substrate will be significantly lower in September.

Between which months significant differences occur strongly depends on the metric. Eleven out of the 18 metrics showed significant differences between September and at least four other months (Table 2.4).

Table 2.2.Summary of ANOVA results for comparison of months per metric on the Heelsumse beek and Stupavsky potok (a=0.05). Metrics incorporated in the Dutch AQEM assessment system are indicated in grey.

acronym	**metric description**	**season F**	
		the Netherlands	**Slovakia**
ABUN	Abundance	1.36	4.12
NTAX	Number of taxa	0.38	2.29
NGEN	Number of genera	0.39	1.9
NFAM	Number of families	0.54	1.7
ZSI	Saprobic Index (Zelinka & Marvan)	4.26	0.86
ALPHA-MESO (%)	Proportion of individuals with a preference for alpha-meso saprobic circumstances	4.81*	16.37*
OLIGO (%)	Proportion of individuals with a preferences for oligo saprobic circumstances	2.61	1.2
XENO (%)	Proportion of individuals with a preferences for xeno saprobic circumstances	0.54	1.76
GFI D03	German Fauna Index D03	4.98*	3.98
GFI D04	German Fauna Index D04	1.91	1.3
BMWP	Biological Monitoring Working Party	2.8	2.81
ASPT	Average Score per Taxon	0.38	3.88
GSI	German Saprobic Index (new version)	2.29	1.73
CSI	Czeck Saprobic Index	3.33	7.25*
MTS	Mayfly Total Score	-	8.86*
HAI	Acid Index (Hendrikson & Medin)	2	2.8
LIFE	LIFE-index	5.37*	0.1
NSTA	Number of sensitive taxa (Austria)	0.4	3.36
DIM	Diversity Index (Margalef)	0.78	2.71
DIS	Diversity Index (Shannon & Wiener)	0.63	39.76*
RP (%)	Proportion of individuals with a preference for zones with moderate to high current (rheophil)	1.86	12.46*
PEL (%)	Proportion of individuals with a preference for muddy substrates	1.3	1.14
PSA (%)	Proportion of individuals with a preference for sandy substrates	12.78*	8.70*
AKA+LIT+PSA (%)	Proportion of individuals with a preferences for gravel, littoral and sand	5.79*	2.64
littoral (%)	Proportion of individuals with a preference for the littoral	3.71	18.29*

acronym	metric description	season F	
		the Netherlands	Slovakia
metarhithral (%)	Proportion of individuals with a preference for the lower-trout region	2.22	27.63*
hyporhithral (%)	Proportion of individuals with a preference for the greyling region	3.81	3.96
epirhithral (%)	Proportion of individuals with a preference for the upper-trout region	5.32*	0.72
IBR	Index of Biocoenotic Region	18.14*	1.56
GAT/COL (%)	Proportion of gatherers to collectors (individuals)	2.61	4.03
SHRED (%)	Proportion of shredders (individuals)	1.17	12.32*
PASF (%)	Proportion of passive filter feeders (individuals)	8.24*	10.65*
GRA+SCRA (%)	Proportion of grazers and scrapers (individuals)	2.36	20.18*
RETI	Rhithron Feeding Type Index	5.85*	3.71
EPT-taxa	Number of Ephemeroptera, Plecoptera and Trichoptera taxa	3.02	4.2
EPT/DIP-taxa	Proportion of EPT-taxa to Diptera taxa	4.59*	3.54
OL+DIP-taxa (%)	Proportion of Oligochaeta and Diptera taxa	3.44	1.83
OL-taxa	Number of Oligochaeta taxa	1.6	3.66
TRIC (%)	Proportion of Trichoptera individuals	5.22*	1.18
PLEC (%)	Proportion of Plecoptera individuals	1.7	6.69*
EPT-taxa (%)	Proportion of EPT-taxa	3.83	3.54
OL	Number of Oligochaeta individuals	5.20*	1.96
CRUS-taxa	Number of Crustacea taxa	1	-
CRUS	Number of Crustacea individuals	0.4	16.46*
EPHE-taxa	Number of Ephemeroptera taxa	1.8	1.55
EPHE	Number of Ephemeroptera individuals	5.47*	5.47*
TUR-taxa	Number of Turbellaria taxa	4.37*	1
TUR	Number of Turbellaria individuals	2.15	1.33
TRIC-taxa	Number of Trichoptera taxa	5.28*	2.87
TRIC	Number of Trichoptera individuals	16.73*	10.24*
COL-taxa	Number of Coleoptera taxa	0.29	2.96
COL	Number of Coleoptera individuals	0.19	10.60*
DIP-taxa	Number of Diptera taxa	1.71	1.4
DIP	Number of Diptera individuals	3.15	6.58*
HYD-taxa	Number of Hydrachnidia taxa	1	-
HYD	Number of Hydrachnidia individuals	12.47*	-
GAS-taxa	Number of Gastropoda taxa	0.6	-
GAS	Number of Gastropoda individuals	0.69	-
HIRU-taxa	Number of Hirudinea taxa	0.55	-
HIRU	Number of Hirudinea individuals	0.34	-
PLEC-taxa	Number of Plecoptera taxa	1.14	5.53*
PLEC	Number of Plecoptera individuals	80.08*	6.78*

*$p<0.05$

Table 2.3. Overview of taxa that incidentally show high abundances. Percentage of individuals based on the average of both replicates.

month	number of individuals (%)			
	Simulium sp	*Nemurella pictetii*	*Proasellus coxalis*	*Limnephilus lunatus*
July	1	2	1	9
September	6	34	24	0
November	0	1	4	0
January	1	1	1	6
March	1	0	5	1
May	11	0	1	2

Table 2.4. Overview between which months metric values significantly differ ($p<0.05$) for the Heelsumse beek, based on the Least Significant Difference (LSD, a=0.05).

acronym	**significant differences between**
ALPHA-MESO (%)	Sept-other
GFI D03	Sept/Nov-Jan/May
LIFE	
PSA (%)	Sept-other
AKA+LIT+PSA (%)	May-Jul/Sept
epirhithral (%)	Sept-other
IBR	Sept-other
PASF (%)	May/Sept-other
RETI	Nov-other
EPT/DIP-taxa	Sept/Nov-May/Jul
TRIC (%)	Sept-Mar/May/Jul Nov/Jan-Jul
OL	Sept-other Nov-May
EPHE	Sept-other
TUR-taxa	Sept-other
TRIC-taxa	Sept-Jan/Mar/May/Jul Nov-Jul
TRIC	Sept-Mar/May Jul-Mar/May/Sept/Nov Jan-Mar/May/Nov
HYD	Jul-other Sept-Jan/Mar/May/Nov
PLEC	Sept-other Mar/Jul-Jan/May/Nov May-Jan/Nov

In November, January and March the samples indicated that the Heelsumse beek was of good ecological quality. In May and July one of the two replicates indicated poor ecological quality. In September one replicate indicated bad ecological quality class and the other moderate ecological quality class. The indications of poor quality class in May and July were caused by slightly lower values for the metric EPT-taxa (%). When this metric would not have been considered the samples would have indicated good quality. The indication of bad (instead of good) ecological quality class in September was caused by relatively very low values for the metric ALPHA-MESO (%).The indication of moderate (instead of good) ecological quality class in September was caused by relatively low values for the metric hyporhithral (%).

Slovakia

Nineteen out of 62 metrics showed significant ($p<0.05$) differences between months (Table 2.2). Between which months significant differences occur strongly depends on the metric. Of the 19 metrics showing significant differences between months, 16 were quantitative metrics. Six out of the 19 metrics showed significant differences between April and at least four other months (Table 2.5).

Table 2.5.Overview between which months metric values significantly differ ($p<0.05$) for the Stupavsky potok, based on the Least Significant Difference (LSD, a=0.05).

acronym	**significant differences between**
ALPHA-MESO (%)	Apr-other
CSI	Apr-Dec/Oct Aug-Oct Jan-Feb
MTS	Dec-Apr/Jun/Aug/Oct Feb-Apr
DIS	Jun/Aug-other
RP (%)	Feb/Aug-other
PSA (%)	Apr-other (except Dec) Oct/Feb-Dec
littoral (%)	Apr-other Jun/Aug-Dec
metarhithral (%)	Apr-other
SHRED (%)	Jun/Aug-other Feb-Dec
PASF (%)	Aug/Dec-other
GRA+SCRA (%)	Apr-other Aug-other (except (jun)
PLEC (%)	Jun/Oct/Dec-Feb/Aug Oct-Apr
CRUS	Jun/Aug/Oct-Feb/Apr Oct-Dec
EPHE	Oct/Feb-Jun/Aug Oct-Apr
TRIC	Oct-other
COL	Oct/Feb-Apr/Aug/Dec Jun-Apr
DIP	Feb-Apr/Jun/Aug/Dec Apr-Jun/Aug/Dec
PLEC-taxa	Dec/Oct-Apr/Jun Dec-Aug/Oct
PLEC	Jun/Oct/Dec-Feb/Aug Oct-Aug

The Netherlands-Slovakia

The following metrics showed significant differences between months in the Stupavsky potok and the Heelsumse beek: ALPHA-MESO (%), PSA (%), PASF (%), EPHE, TRIC, PLEC (Table 2.2). The metrics ALPHA-MESO (%) and PSA(%), showed differences between September and all other months in the Heelsumse beek and between April and all other months in the Stupavsky potok (Table 2.4 and 2.5). Although the month differed between the two streams April and September were in both streams the month that differed most from all other months based on multivariate analysis results.

This pattern could not be detected for the other metrics that showed significant differences between months in both streams.

2.4 Discussion

Significant differences in metric values between months were found. We didn't test whether sorting efficiency between people could have influenced these results. The assumption was made that inter-sorter differences were not significant, as proven by Armitage *et al.* (1995).

Multivariate analysis and also some metric results indicated that macroinvertebrate community composition in the Stupavsky potok in April differed from all other months. Sample processing of the April samples differed from sample processing of the other samples, what might have caused differences. However, this is not likely since the only difference in sample processing compared to samples collected during other months was that the habitats were not kept separately during sorting.

Armitage *et al.* (1995) detected significant differences in the total number of individuals, taxa richness BMWP values and ASPT values between seasons for some mesohabitats. In this study no differences between months could be detected for these metrics, probably because (meso)habitats were not analysed separately. The majority of metrics showing significant differences between months were quantitative metrics. So, when using quantitative metrics in assessment one should realize that the moment of sampling can have a strong influence on the results. For the individual metrics differences between months strongly depend on the metric under evaluation, this makes it difficult to give a general recommendation on the preferred time period for sampling. An option is to select a preferred time period for each individual metric. For metrics directly related to the number of taxa or the number of individuals the preferred time period for sampling would be the month in which values are highest. However, for metrics like PSA (%) and RP (%) a higher value automatically mean a better value. For metrics were the optimal time period is not directly related to the highest metric value the best solution would be to always sample in the same month or months or to take into account seasonal variation in setting class boundaries.

The other reason it is difficult to make general recommendations is that some metrics showed differences between months in both streams, but other metrics showed differences between months in only one of the two streams. For metrics that showed differences between months in both streams, the months that differed depended on the stream. Based on these results it seems that the recommended time period for sampling not only depends on the metric, but also on the streamtype (both streams belong to a different streamtype) or maybe even the stream.

The influence of incidentally high abundant taxa can be stronger for some metrics than others. Many of the metrics evaluated in this study work with indicator values, like PSA (%), littoral (%) et cetera. In many cases indicator values for taxa were unknown. For example, the microhabitat indicator values were unknown for 76% of the taxa. When indicator values of most taxa in a sample are unknown, the influence of taxa with indicator value (and high abundance) will become even stronger. This problem can be solved by increasing autecological knowledge.

In many cases not an individual metric, but a multimetric index is used for assessment. Unfortunately, no assessment system was available to assess the Stupavsky potok. The metrics that caused assessment results other than good were not always the metrics that showed significant differences between seasons. This was also observed in comparing different protocols (paragraph 5.3.2.). The metric ALPHA-MESO (%) did

show differences in values between September and all other months. EPT-taxa (%) and hyporhithral (%) values happened to fall near a breakpoint in scoring criteria (Fore *et al.*, 2001). The metrics PSA (%), AKA+LIT+PSA (%) and EPT/IP-taxa showed differences between months and are part of the Dutch assessment system, however, these metrics didn't have a major influence (other quality class) on assessment results. In short it is very difficult to predict the effects of differences in individual metrics on the final assessment. The only way to decided on the most preferred time period for sampling with the goal of determining the ecological quality status of a stream is to directly compare assessment results between months. Based on expert-judgement the Heelsumse beek the ecological status can not be considered bad or poor. The majority of the samples (75%) indicated that the Heelsumse beek was of good ecological quality. Months where sampling resulted in classifications other than good therefore were judged not suited for sampling. To assess whether the results of this study are applicable on a wider scale in the Netherlands and Slovakia more research has to be done.

During this study it was discovered that in some cases variation between months and replicates in the Heelsumse beek was higher than the variation taken into account in the development of the Dutch multimetric index. This is exceptional since the samples used for the development the Dutch multimetric index didn't only include variation between replicates and months, but also spatial variation (22 different sites). The variation between months for the AQEM dataset clearly underestimated the total variation, probably due to the fact that only samples form two different months were available. These findings indicate that monthly/seasonal variation may be higher than people expect and that it is very important to really know variation between months in developing a robust assessment system.

2.5 Recommendations

The time period preferred for sampling depends on the system or metrics used for assessment and the streamtype. In general, it is not recommended to sample in September, May or July in Dutch lowland streams, because of a deviation in community composition and metric results in these months. When using the Dutch AQEM assessment system it is also advised not to sample in May, July or September. However, when the metric EPT-taxa (%) in the assessment system would be replaced by another more robust metric sampling can take place in all months except September. To conclude, more autecological knowledge is required to improve future biological assessment.

3 Hand-net mesh size

3.1 Introduction

There are several EU member states that have a national standard for sampling and sample processing. These standard protocols for sampling and sample processing include instructions on the mesh size of the sample device. The standard protocol in the UK, RIVPACS, requires the use of a 1000 µm mesh size net. Many national standards in Europe, however, require the use of a 500 µm mesh size sample device, e.g. EBEOSWA (the Netherlands), IBGN (France) and PERLA (Czech).

According to Reish (1959), McIntyre (1961), Lewis & Stoner (1981), Eleftheriou & Holme (1984) and Rees (1984) the mesh size used in macrobenthic studies will influence the interpretation of community structure. However, these studies were all undertaken in a marine environment. To our knowledge no studies have been published on the efficiency of sample devices with a mesh size larger than 600 µm in streams. The objective of this study was (1) to compare efficiency between samples collected with a 1000 µm mesh net and samples collected with a 500 µm mesh net (2) to assess whether differences in net mesh sizes can lead to significant differences in metric and assessment results.

3.2 Methods

3.2.1 Study site and data collection

For this study, we used data collected from two different streams in Slovakia. Samples were taken from the Bystrica and the Pokútsky potok, two silicious mountain streams in the West Carpathian. Both streams were characterised by a catchment area between 10 and 100 km^2 and were located between 200 and 500 m above sea level. The substrate in both streams consisted mainly of cobbles and coarse blocks (mesolithal and macrolithal) with variable percentages of gravel and sand (Table 3.1).

Both streams were sampled in June 2003. In each stream a uniform 100 m stretch of the stream was selected for sampling. Prior to sampling habitat coverage was estimated. A complete sample consisted of 20 subsamples of 25 cm taken from all habitats with coverage of at least 5%. The 20 subsamples were distributed according to their share of coverage. A hand-net was used for sampling. After sampling all subsamples were collected in one bucket and preserved in 4% formaldehyde. In each stream six composite samples were collected, three replicate samples with a 500 µm mesh net and three replicate samples with a 1000 µm mesh net. The buckets were transported to the laboratory. At the laboratory the samples were stored until sorting. The composite samples were washed through a 1000 and 500 µm mesh size sieve before sorting. The total amount of sampled material was reduced by taking a subsample. The subsample contained at least 1/6 of the total amount of sampled material and 700 individuals. In case 1/6 of the sample contained less than 700 individuals the subsample was increased in size. All individuals were removed from the subsample. The taxa list composed after identification of the sorted individuals was extrapolated to the whole sample.

The samples were sorted by eye by a group of three people. Organisms were identified to the lowest taxonomic level possible (species level for almost all groups). All identifications were preformed by the same specialist for each major organismn group.

For a more elaborate description of the used protocols for sampling and sample processing we refer to the AQEM sampling protocol and revised AQEM/STAR sorting protocol described by STAR Consortium (2003), in this paper the protocol is referred to as the STAR protocol.

Table 3.1. Habitat coverage in the Bystrica and the Pokútsky potok in June 2003.

stream	habitat	coverage (%)
Bystrica	macrolithal	35
	mesolithal	45
	microlithal	15
	submerged macrophytes	5
Pokútsky potok	macrolithal	70
	mesolithal	10
	microlithal	10
	submerged macrophytes	5
	psammal	5

3.2.2 Data analysis

Metrics commonly used in Europe were selected for evaluation (Table 3.2). The metrics were selected from an extensive list given by Hering *et al.* (2004). Apart from metrics selected form the list by Hering *et al.* (2004) the number of taxa and the number of individuals for each major macroinvertebrate group (e.g. Diptera, Ephemeroptera, Plecoptera) was also evaluated. Some major groups were only present in a few samples and in low abundances, only metrics with values higher than zero in four out of six samples for one of the streams were used for analysis.

Prior to analysis the number of individuals per taxon was standardised to a total sample area of 1.25 m^2 for each composite sample. Metric values were calculated with the AQEM River Assessment Program (AQEMrap version 2.3). The ecological quality class for the Bystrica and the Pokútsky potok was not be calculated because no suited assessment system was available.

To test for significant differences in metric values between the two protocols (500 and 1000 μm mesh net) a one-way ANOVA was used (α=0.05) with the two streams as blocks. The block design made it possible to test whether differences between protocols were consistent among streams. Assumptions of normality and homogeneity of variance could not be tested in a reliable way due to the low number of samples. For this reason it would have been more appropriate to perform a non-parametric test. However, the non-parametric test that gives the possibility to include a block design (Friedman ANOVA by ranks) doesn't allow for repetition of observations. During this study three replicates per stream were collected and this information would be lost in using a Friedman ANOVA by ranks. Taking all the above into account it was decided to perform an ANOVA and to transform metric values based on experiences in other studies. Abundance metrics were ln(x+1) transformed, taxa counts were not transformed (Kerans *et al.*, 1992) and proportions were ln(x+1)-ln(y+1) transformed. Biotic index data (e.g. Sapobic Index, BMWP, ASPT) were not transformed (Norris & Georges, 1993). Metrics like XENO (%), SHRED (%) and littoral (%) are not simple proportional metrics. The values for these metrics also depend on the strength with which a species prefers a certain category (AQEM consortium, 2002). The decision was made not to transform values of these metrics, since no references were found on a suited transformation.

3.3 Results

3.3.1 General

In total 147 taxa were collected from the two streams. Hundred eighteen taxa were collected from the Bystrica and 108 taxa from the Pokútsky potok. Macroinvertebrate community composition differed between streams (Figure 3.1). Diptera dominated the community in the Bystrica. (70%). The macroinvertebrate community in the Pokútsky potok also consisted for a large part of Diptera (27%), but next to the Diptera the Crustacea (21%) and Trichoptera (20%) also formed a large proportion of the community.

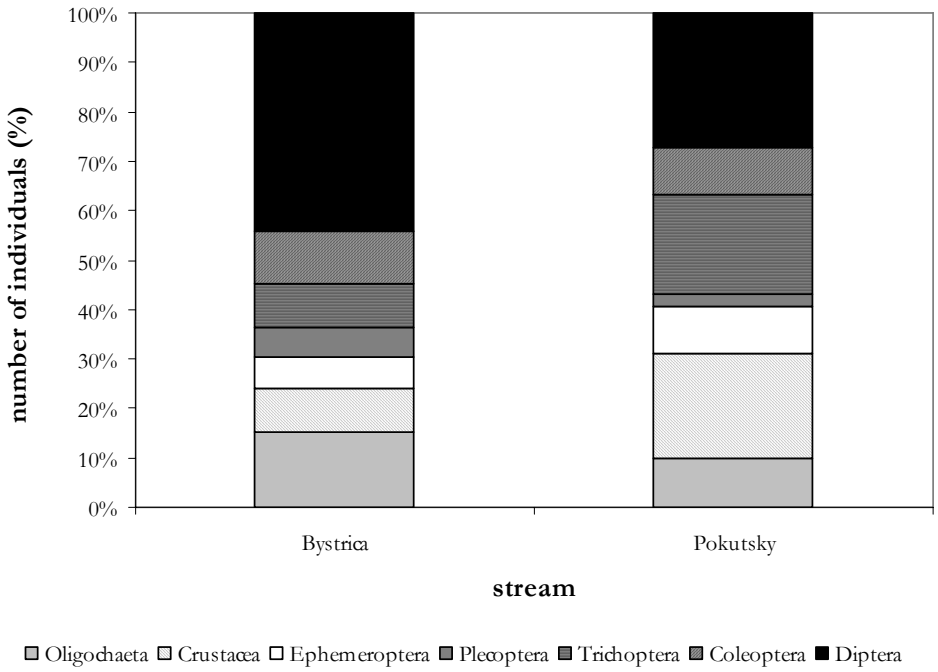


Figure 3.1. Macorinvertebrate community composition per stream. Only groups that formed more than 5% of the total community were included in the graph. Mean percentages of individuals are based on the six samples per stream.

3.3.2 Metrics and assessment

In total nine of the 56 metrics showed significant ($p<0.05$) differences between samples collected with a 500 µm mesh size net and samples collected with a 1000 µm mesh size net (Table 3.2). Higher numbers of individuals were collected from samples taken with a 500 µm mesh size net for all major organism groups (Figure 3.2), except for the Turbellaria and the Oligochaeta. NTAX and RP(%) values were also higher for samples taken with a 500 µm mesh size net (Figure 3.3).

Table 3.2. Summary of ANOVA results for comparison of protocols (n=12 12, a=0.05) and overview of replicate means (coefficient of variation) per metric for the Bystrica and the Pokútsky potok. Metrics incorporated in the Dutch AQEM assessment system are indicated in grey.

acronym	metric description	mesh size F	mean (coefficient of variation)			
			Bystrica		Pokútsky	
			1000 µm	500 µm	1000 µm	500 µm
ABUN	Abundance	5.25*	2526.6 (61)	4038.8 (35.8)	1672.2 (55)	3153.6 (20.6)
NTAX	Number of taxa	5.10*	53 (15)	67.7 (3.1)	59.7 (5.1)	60.3 (3.8)
NGEN	Number of genera	1.85	42 (21.2)	54.7 (7.4)	49 (6.1)	46.7 (5.4)
NFAM	Number of families	0.52	26.3 (13.3)	31 (8.5)	32 (8.3)	30 (5.8)
ZSI	Saprobic Index (Zelinka & Marvan)	1.24	1.8 (3.5)	1.8 (4.7)	1.7 (5)	1.7 (2.8)
ALPHA-MESO (%)	Proportion of individuals with a preference for alpha-meso saprobic circumstances	0.17	16.6 (5.8)	15.8 (4.4)	15.2 (11.2)	15.4 (7.6)
OLIGO (%)	Proportion of individuals with a preferences for oligo saprobic circumstances	2.34	32.3 (7.4)	33.6 (10)	33.8 (6.7)	36.8 (5.2)
XENO (%)	Proportion of individuals with a preferences for xeno saprobic circumstances	0.07	4.3 (35.8)	3.7 (56.5)	5.1 (34.3)	6.1 (8.9)
GFI D03	German Fauna Index D03	1.31	1.3 (10.7)	1.3 (9.3)	0.7 (28.8)	0.9 (7.4)
GFI D04	German Fauna Index D04	0.42	1.1 (5.9)	1.2 (4.7)	1.3 (4.2)	1.2 (5.8)
BMWP	Biological Monitoring Working Party	0.49	119 (22)	139.3 (21.8)	166.7 (5.6)	163 (2.2)
ASPT	Average Score per Taxon	0	6.6 (7)	6.7 (2.2)	7.2 (1.7)	7.1 (2.2)
GSI	German Saprobic Index (new version)	0	1.4 (0.9)	1.3 (4.3)	1.4 (1.8)	1.4 (2.2)
CSI	Czeck Saprobic Index	0.97	1.6 (8.3)	1.5 (12.7)	1.2 (12.4)	1.1 (9.1)
MTS	Mayfly Total Score	0.47	17.3 (43.7)	19.7 (21.2)	27.3 (7.6)	28.7 (14.1)
HAI	Acid Index (Hendrikson & Medin)	8.1	8.3 (6.9)	9.3 (6.2)	9 (0)	10 (10)
LIFE	LIFE-index	0.37	6.9 (2.2)	6.8 (1.6)	6.9 (0.4)	6.9 (1.7)
NSTA	Number of sensitive taxa (Austria)	0.3	10.7 (19.5)	13.3 (15.6)	13.3 (15.6)	12 (8.3)
DIM	Diversity Index (Margalef)	0.21	6.8 (21.6)	8.1 (8.1)	8.1 (11)	7.4 (4.4)
DIS	Diversity Index (Shannon & Wiener)	0	3.1 (6.5)	3.2 (4.2)	3.2 (4.5)	3.1 (1.9)
RP (%)	Proportion of individuals with a preference for zones with moderate to high current (rheophil)	9.31*	69.4 (10.3)	76.8 (8.2)	67.3 (10.5)	83 (6.2)
PEL (%)	Proportion of individuals with a preference for muddy substrates	0.72	4.3 (25.4)	4.2 (12.7)	3 (36.1)	2.2 (32.6)

acronym	metric description	mesh size F	mean (coefficient of variation)			
			Bystrica		Pokútsky	
			1000 µm	500 µm	1000 µm	500 µm
PSA (%)	Proportion of individuals with a preference for sandy substrates	2.33	9.7 (11.6)	9.6 (29)	10.6 (14.2)	14.1 (6.7)
AKA+LIT+PSA (%)	Proportion of individuals with a preference for gravel, littoral and sand	0.92	71.2 (6)	68.5 (7.9)	75 (3.1)	73.5 (4)
littoral (%)	Proportion of individuals with a preference for the littoral	1.66	3.3 (46.2)	3.3 (67)	3.1 (14)	5.3 (13.4)
metarhithral (%)	Proportion of individuals with a preference for the lower-trout region	1.46	22.1 (6.4)	23.8 (2.7)	26.4 (12.9)	27.7 (8.8)
hyporhithral (%)	Proportion of individuals with a preference for the greyling region	0.21	19.7 (10.6)	20.4 (12.8)	18.9 (5.7)	19.1 (3.7)
epirhithral (%)	Proportion of individuals with a preference for the upper-trout region	0.2	19.3 (7.7)	18.9 (5.8)	20 (5)	19.7 (10.8)
IBR	Index of Biocoenotic Region	2.32	4.3 (2.7)	4.3 (3.6)	4 (3.7)	4.2 (2.4)
GAT/COL (%)	Proportion of gatherers to collectors (individuals)	0.25	24.1 (17.3)	24 (9.4)	20.9 (4.8)	19 (28.2)
SHRED (%)	Proportion of shredders (individuals)	0.19	11.1 (37.7)	8.6 (52.8)	19.7 (31.5)	24.6 (11.2)
PASF (%)	Proportion of passive filter feeders (individuals)	0.44	4.7 (13.3)	4.8 (43.4)	6.3 (1.1)	8.2 (59.4)
GRA+SCRA (%)	Proportion of grazers and scrapers (individuals)	0.78	39.8 (15.8)	38.6 (27.2)	36 (18.7)	30.2 (8.5)
RETI	Rhithron Feeding Type Index	0.89	0.6 (6.3)	0.6 (0.4)	0.7 (1.1)	0.7 (2.9)
EPT-taxa	Number of Ephemeroptera, Plecoptera and Trichoptera taxa	0.16	23.3 (26.2)	27.7 (17.8)	31 (6.5)	28.7 (4)
EPT/DIP-taxa	Proportion of EPT-taxa to Diptera taxa	4.86	1.7 (27.6)	1.2 (25.5)	2.1 (18.3)	1.7 (17.9)
OL+DIP-taxa (%)	Proportion of Oligochaeta and Diptera taxa	2.41	37.9 (18.6)	43.9 (12.1)	34 (9.2)	35.8 (10.5)
OL-taxa	Number of Oligochaeta taxa	0.03	6 (28.9)	7 (14.3)	5.3 (10.8)	4 (66.1)
TRIC (%)	Proportion of Trichoptera individuals	0.32	9.3 (42.9)	8.5 (76.3)	20.8 (38.2)	19.1 (14.4)
PLEC (%)	Proportion of Plecoptera individuals	0.81	4.5 (43.7)	7 (39.2)	2.4 (59.7)	2.3 (76.3)
EPT-taxa (%)	Proportion of EPT-taxa	0.57	43.7 (15.4)	40.8 (16.4)	52 (6.7)	47.6 (6.7)
OL	Number of Oligochaeta individuals	0.27	369.1 (58.4)	548.8 (41.6)	259.7 (82.7)	239 (73.8)
CRUS-taxa	Number of Crustacea taxa	0	1.3 (43.3)	1.3 (43.3)	2 (0)	2 (0)
CRUS	Number of Crustacea individuals	8.04*	216.6 (24.3)	273.3 (20.7)	261 (45.5)	797 (38.9)

acronym	metric description	mesh size F	mean (coefficient of variation)			
			Bystrica		Pokútsky	
			1000 µm	500 µm	1000 µm	500 µm
EPHE-taxa	Number of Ephemeroptera taxa	0.65	5.7 (36.7)	7 (24.7)	9.3 (27)	9.7 (6)
EPHE	Number of Ephemeroptera individuals	10.43*	129.1 (32.7)	272.5 (35)	124.3 (70.7)	336.8 (59.2)
TUR-taxa	Number of Turbellaria taxa	0	0.3 (173.2)	0.3 (173.2)	1 (0)	1 (0)
TUR	Number of Turbellaria individuals	0.28	4 (173.2)	1.3 (173.2)	14.8 (70.5)	35.5 (46.4)
TRIC-taxa	Number of Trichoptera taxa	0	12.3 (32.8)	14.3 (21.3)	18.3 (3.1)	16.3 (9.4)
TRIC	Number of Trichoptera individuals	10.21*	194.3 (14.3)	285.3 (34.3)	299.8 (32)	601.6 (26.6)
COL-taxa	Number of Coleoptera taxa	1.58	7.7 (30.1)	8.7 (26.6)	5.3 (10.8)	7 (28.6)
COL	Number of Coleoptera individuals	4.13*	253.1 (72.3)	457.8 (30.9)	163.6 (83.1)	325.8 (17.5)
DIP-taxa	Number of Diptera taxa	8.00*	14.3 (31.5)	22.7 (9.2)	15 (13.3)	17.7 (18.2)
DIP	Number of Diptera individuals	1.24	1264 (81.6)	1913 (67.4)	517.3 (65.1)	751.7 (14.7)
PLEC-taxa	Number of Plecoptera taxa	0.09	5.3 (10.8)	6.3 (24.1)	3.3 (17.3)	2.7 (21.7)
PLEC	Number of Plecoptera individuals	7.73*	96.3 (21.6)	287 (53.9)	31.9 (16.4)	66.3 (61.2)

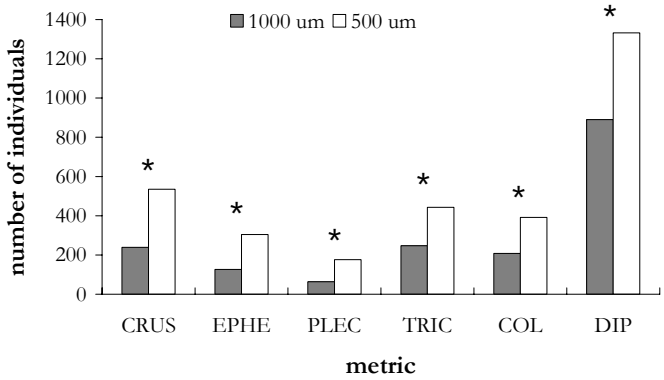


Figure 3.2. Mean metric values for 500 and 1000 µm mesh size samples from the Bystrica and the Pokútsky potok. Asterisks indicate significant differences between means (One-way ANOVA with blocks, Table 3.2)

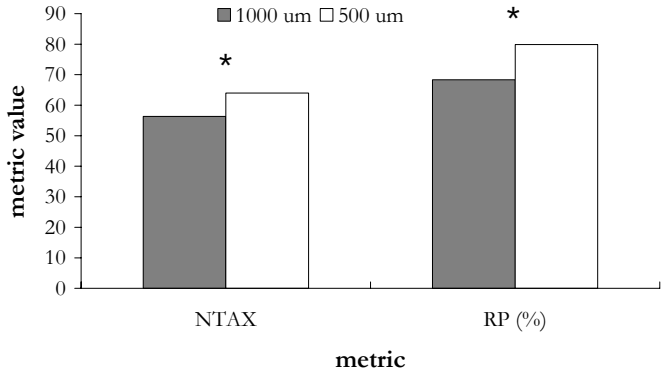


Figure 3.3. Mean metric values for 500 and 1000 µm mesh size samples from the Bystrica and the Pokútsky potok. Asterisks indicate significant differences between means (One-way ANOVA with blocks, Table 3.2)

The variation in metric values depends on the metric and stream (Table 3.2). In several cases variation was higher using a 1000 um mesh size net in the Bystrica, while in the Pokútsky potok variation was higher using the 500 um mesh size net (Table 3.2), or the other way around.

3.4 Discussion

Significant differences in metric values between the two protocols were detected. We didn't test whether sorting efficiency between people could have influenced the results. The assumption was made that inter-sorter differences were not significant, as proven by Armitage *et al.* (1995).

Significant differences in metric values between protocols were primarily found for abundances measures. Nets with a 500 µm mesh size showed higher efficiency in removing individuals from the stream than nets with a 1000 µm mesh size. These results confirm the assumption that small individuals are missed using larger mesh sizes (Ferraro *et al*, 1989). The differences in abundances between protocols were not reflected in other metrics, except for RP (%), this is an indication that the relative abundance of the major organism groups doesn't differ between samples taken with a 500 or 1000 um mesh size net.

The efficiency for collecting taxa (per major organism group) from a stream didn't seem to depend on the mesh size of the net. Significant differences were detected for the total number of taxa between protocols, but these differences were very small for the Potusky potok. The difference between protocols for the average total number of taxa was 0.6 in the Pokutsky potok.

Since variation in metric values depended on the metric and the stream no general statements can be made on the variation of one protocol compared to the other. Unfortunately, time required for sorting and identification was not recorded during this study. The cost efficiency of STAR samples taken with a 500 um mesh size net and STAR samples taken with a 1000 um mesh size net could, therefore, not be compared.

3.5 Recommendations

Care should be taken with extrapolation of the results from this study to streams with very different substrate and macroinvertebrate community composition. Based on the results form this study it seems that in most cases sampling with a 1000 μm mesh size net will result in comparable estimates of metric values as when sampling with a 500 μm mesh size net. However, when absolute abundances will be used for assessment a 500 μm mesh size net will result in more accurate estimates.

4 Comparison of preserved and unpreserved macorinvertebrate samples

4.1 Introduction

In case no preservative is added directly after taking a macroinvertebrate samples, this means living organisms will have to be collected from the sample (a process to which we will refer as live sorting). Live sorting is commonly applied in Australia for the rapid biological assessment (RBA) of rivers (Metzling *et al.*, 2003), either for set periods (Chessman & Robinson, 1987) or until a fixed number of specimens is collected (Chessmann, 1995). Apart form Australia live sorting is frequently applied in Southern European countries (Alba-Tercedor & Ortega, 1988; Buffagni *et al.*, 2002), Germany (Braukmann, 2000) and the United States. In all these cases, however, live sorting is preformed in the field. In this study we will focus on sorting in the laboratory.

At the moment the choice to use a preservative is based on the preferences of researches involved in a study. These preferences, however, are related to experiences and feelings and not based on evidence gained through research. People in favour of using preservatives often mention the following disadvantages of not using preservatives: (1) specimens may be eaten by others before sorting is completed (2) specimens may disintegrate before sorting is completed (3) removing fast moving taxa like *Gammarus sp.* from a sample may be time consuming (3) samples have to be sorted as soon as possible (within 5 days) after collection, which makes it impossible to collect a large number of samples at the same time. People not in favour of using preservative often mention the following disadvantages of using preservatives: (1) the health risk of formaldehyde (2) it is easier to spot living than dead organisms due to movement and (3) preservatives can make it harder to identify certain taxa. In table 4.1 the (dis)advantages related to the identification of the two most commonly used preservatives (formaldehyde and ethanol) are mentioned.

Table 4.1. Advantages and disadvantages of using formaldehyde or ethanol as a preservative (pers. com. Van den Hoek & Higler, 2003).

Formaldehyde	**Ethanol**
Animals stiffen (body musculature) and are difficult to prepare in horizontal position (chironomids, oligochaetes): vulnerable body parts like antennae, throracic horns etc. can break off. Many characteristics for identification may be missing and identification takes more time.	Animals are flexible and therefore better to prepare (i.g. oligochaetes and chironomids).
Animals rapidly dry during preparation, with result that only a restricted number can be placed on an object glass (chironomids), which takes more time. Probably, the concentration of formaldehyde influences the stiffness of the musulature.	Animals don't dry rapidly, and therefore larger series can be placed on the object glass (chironomids).
Fixation preserves colour pattern (Ephemeroptera, Hirudinea) helps tracheal gills to adhere (Ephemeroptera).	Animals loose colour patterns earlier and loose gills (esp. Ephemeroptera). Probably, the duration time in the pails plays a role.
The valve-surface of Bivalvia becomes brittle and detaches from the shell, resulting in the loss of ribbles on the surface (identific characteristics). The valves cannot be taken apart, so the cardial teeth stay invisible.	Animals are often damaged because of long handling time and conservation (esp. Ephemeroptera, and also Chironomidae)
Claws and antennae of Tanypodinae and Tanytarsini	

(Chironomidae) are retracted in formaldehyde and characteristics are invisible.	
The epidermus of Hydracarina, Chironomidae, Trichoptera and Diptera is strongly wrinkeled.	
The legs and palps of Hydracarina are folded over the ventral side. Characteristics of the epimeres are difficult to judge and positioning of the animal is impossible.	
Flatworms are hardly to identify in samples with formaldehyde	
Many species of Micropsectra and of Orthocladiinae cannot be identified to species level	

The question is whether the disadvantage and advantages of using preservatives that have been describe in this paragraph will significantly influence the efficiency of sample processing. The aim of this study was (1) to compare (cost) efficiency between preserved and unpreserved samples (2) to determine whether preserved and unpreserved samples show significant differences in metric and assessment results.

4.2 Methods

4.2.1 Study site and data collection

For this study, we used data collected from three different sites in the Netherlands. Samples were taken from the Springendalse beek, the Tongerensche beek and the Swalm. All streams are characterised by a catchment area smaller than 100 km^2 and an elevation level below 200 m above sea level. The Springendalse beek is a small stream in the eastern part of the Netherlands. The sample site was located upstream (average width 1.2 m) and characterised by a great slope, intermediate stream velocity (v=30 cm/sec) and a small natural profile. The substrate consisted mainly of sand (Table 4.2). The second site was located in the middle course of the Tongerensche beek (average width 3.5 m) a small stream in the central part of the Netherlands. The sample site was stream is characterised by a small slope and low stream velocity (v < 30 cm/sec). The substrate consisted mainly of mud and vegetation (Table 4.2). The third site was located in the middle course of the Swalm (average width 7.5 m). The Swalm is a small stream in the southern part of the Netherlands. The stream is characterised by a great slope, high stream velocity (v>30 cm/sec) and a natural profile. The substrate consisted mainly of gravel and sand (Table 4.2).

The Springendalse beek was sampled in September 2002, the Tongerensche beek in June 2003 and the Swalm in April 2003. In each stream a uniform 100 m stretch of the stream was selected for sampling. At each site six replicate samples were collected. For the collection of the samples a hand-net (25cm wide with a 500 μm mesh) was used. The samples were taken by pushing the net through the upper part (2-5 cm) of the substratum. Each sample consisted of subsamples from different habitats. Each habitat with more than 5% coverage was sampled over a distance that ensured collection of most species present at the habitat (expert-judgement). Before sampling habitat coverage at the site was estimated (Table 4.2). The collected subsamples were stored separately in buckets. Three subsamples of each habitat were preserved in 4% formaldehyde. The buckets were transported to the laboratory. At the laboratory the subsamples without formaldehyde were stored in the fridge, where they were oxygenated, until sorting. All subsamples were kept separate during sample processing. The subsamples were sieved using a 1000 and 250 μm sieve. The samples were sorted by eye by a varying group of

three to five people. Organisms were identified to the lowest taxonomic level possible (species level for almost all groups). All identifications were preformed by the same two qualified persons. The time necessary to sort and identify all specimens in each subsample was recorded.

Table 4.2. Habitat coverage and sampled length of each habitat for the three streams sampled in this study.

stream	habitat	sampled length (m)	coverage (%)
Tongerensche beek	mud	0.5	50
	sand	0.5	20
	submerged vegetation	0.5	30
Swalm	mud/detritus	0.25	5
	gravel	0.75	75
	sand	0.75	20
Springendalse beek	gravel	0.5	5
	sand	0.5	95
	submerged vegetation	0.25	5

4.3 Data analysis

Metrics commonly used in Europe were selected for evaluation (Table 4.3). The metrics were selected from an extensive list given by Hering *et al.* (2004). Apart from metrics selected form the list by Hering *et al.* (2004) the number of taxa and the number of individuals for each major macroinvertebrate group (e.g. Diptera, Ephemeroptera, Plecoptera) was also evaluated. Some major groups were only present in a few samples and in low abundances, only metrics with values higher than zero in 17 out of 18 samples were used for analysis.

Prior to analysis the number of individuals per taxon was standardised to a total sample area of 1.25 m^2 for each composite sample based on habitat coverage and sampled area. Metric values were calculated with the AQEM River Assessment Program (AQEMrap version 2.3). With the same program the final ecological quality classes were calculated for each of the three streams, characterising the samples as being from small Dutch lowland streams. The multimetric index used to calculate the final ecological quality class is a revised version of the multimetric index described by Vlek *et al.* (2004). The multimetric index consists of 11 metrics which are indicated in table 3.3.

We tested for differences in metric values between preserved and unpreserved samples with an exact non-parametric permutation test (α=0.05) (Lehman, 1975). A non-parametric test was used because many metrics were not normally distributed and some could not readily be normalized by transformation (Vinson & Hawkins, 1996; Metzling *et al.*, 2003). Also the low number of samples made it difficult to examine heterogeneity of variance and normality assumptions. An exact non-parametric permutation test was used because this made it possible to use a block design. This block design made it possible to test whether differences between preserved and unpreserved samples were consistent among streams. The Friedman ANOVA by ranks also gives the possibility to include a block design, but it doesn't allow for repetition of observations. During this study three replicates per stream were collected and this information would be lost in using a Friedman ANOVA by ranks.

To test for differences in the recorded time for sorting and identification a one-way ANOVA per stream was preformed (α=0.05). To test if differences were consistent among streams an ANOVA with blocks (streams) was applied (α=0.05). Data on recorded time were ln(x) transformed prior to analysis, according to Growns *et al.* (1997).

Differences between replicate samples in the number of individuals could confound results, because the time required

for sorting and identification will stongly depend on the number individuals in a sample. Therefore, the recorded time was corrected for the number of specimens in a sample. With the corrected times a one-way ANOVA with blocks (streams) was applied (α=0.05) to test whether differences between streams were consistent. Data on recorded time were ln(x) transformed prior to analysis, according to Growns *et al.* (1997).

4.4 Results

4.4.1 General

In total 297 taxa were collected from the three streams. Seventy four taxa were collected from the Springendalse beek, 171 taxa from the Swalm and 182 taxa from the Tongerensche beek. Macroinvertebrate community composition highly differed between streams (Figure 4.1). Crustacea formed a large proportion of the community in the Springendalse beek (70%), while in the Swalm Oligocheata made up a large proportion of the community (58%). In the Tongerensche beek Diptera were present in high numbers (73%).

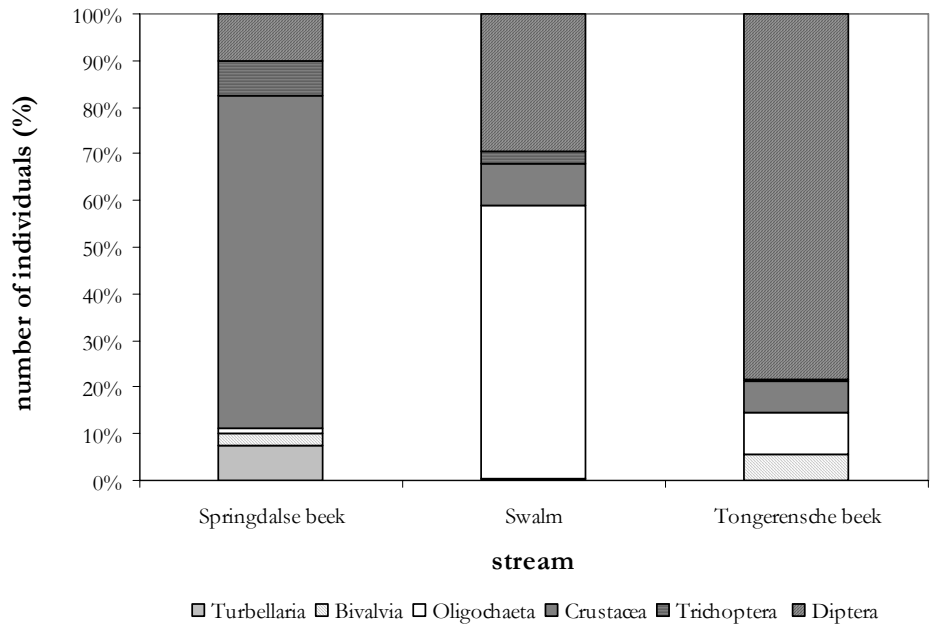


Figure 4.1.Macorinvertebrate community composition per stream. Only groups that formed more than 5% of the total community were included in the graph. Mean percentages of individuals are based on the six replicate samples (preserved and not-preserved) per stream.

4.4.2 Metrics and assessment

In total nine of the 58 metrics showed significant ($p<0.05$) differences between preserved and unpreserved samples (Table 4.3). Metric values for BMWP, RETI, SHRED (%), EPHE and TRIC-taxa were consistently higher in preserved than in unpreserved samples (Figure 4.2). Metric values for AKA+LIT+PSA (%), metahithral (%), HYD-taxa and HYD were significantly higher in samples that were not preserved (Figure 4.2).

Table 4.3. Summary of permutation test results for comparison of preserved and unpreserved samples on three streams (n=18, a=0.05). Metrics incorporated in the Dutch AQEM assessment system are indicated in grey.

acronym	metric description	p-value
ABUN	Abundance	0.622
NTAX	Number of taxa	0.506
NGEN	Number of genera	0.149
NFAM	Number of families	0.941
ZSI	Saprobic Index (Zelinka & Marvan)	0.132
ALPHA-MESO (%)	Proportion of individuals with a preference for alpha-meso saprobic circumstances	0.321
OLIGO (%)	Proportion of individuals with a preferences for oligo saprobic circumstances	1.000
XENO (%)	Proportion of individuals with a preferences for xeno saprobic circumstances	0.077
GFI D03	German Fauna Index D03	0.358
GFI D04	German Fauna Index D04	0.622
BMWP	Biological Monitoring Working Party	*0.046*
ASPT	Average Score per Taxon	0.321
GSI	German Saprobic Index (new version)	0.212
CSI	Czeck Saprobic Index	0.622
DSFI	Danish Stream Fauna Index	
MTS	Mayfly Total Score	1.000
HAI	Acid Index (Hendrikson & Medin)	0.251
LIFE	LIFE-index	0.321
NSTA	Number of sensitive taxa (Austria)	0.700
DIM	Diversity Index (Margalef)	0.321
DIS	Diversity Index (Shannon & Wiener)	0.321
RP (%)	Proportion of individuals with a preference for zones with moderate to high current (rheophil)	0.806
PEL (%)	Proportion of individuals with a preference for muddy substrates	0.132
PSA (%)	Proportion of individuals with a preference for sandy substrates	0.132
AKA+LIT+PSA (%)	Proportion of individuals with a preferences for gravel, littoral and sand	*0.041*
littoral (%)	Proportion of individuals with a preference for the littoral	0.321
metarhithral (%)	Proportion of individuals with a preference for the lower-trout region	*0.020*
hyporhithral (%)	Proportion of individuals with a preference for the greyling region	0.622
epirhithral (%)	Proportion of individuals with a preference for the upper-trout region	1.000
IBR	Index of Biocoenotic Region	0.459
GAT/COL (%)	Proportion of gatherers to collectors (individuals)	0.132
SHRED (%)	Proportion of shredders (individuals)	*0.041*
PASF (%)	Proportion of passive filter feeders (individuals)	0.077
GRA+SCRA (%)	Proportion of grazers and scrapers (individuals)	0.212
RETI	Rhithron Feeding Type Index	*0.041*

acronym	metric description	p-value
EPT-taxa	Number of Ephemeroptera, Plecoptera and Trichoptera taxa	0.076
EPT/DIP-taxa	Proportion of EPT-taxa to Diptera taxa	0.321
OL+DIP-taxa (%)	Proportion of Oligochaeta and Diptera taxa	0.459
OL-taxa	Number of Oligochaeta taxa	0.412
TRIC (%)	Proportion of Trichoptera individuals	1.000
PLEC (%)	Proportion of Plecoptera individuals	0.600
EPT-taxa (%)	Proportion of EPT-taxa	0.459
BIVAL-taxa	Number of Bivalvia taxa	1.000
BIVAL	Number of Bivalvia individuals	0.459
OL	Number of Oligochaeta individuals	0.077
CRUS-taxa	Number of Crustacea taxa	1.000
CRUS	Number of Crustacea individuals	0.321
EPHE-taxa	Number of Ephemeroptera taxa	1.000
EPHE	Number of Ephemeroptera individuals	*0.007*
PLEC-taxa	Number of Plecoptera taxa	0.600
PLEC	Number of Plecoptera individuals	0.600
TRIC-taxa	Number of Trichoptera taxa	*0.046*
TRIC	Number of Trichoptera individuals	0.679
COL-taxa	Number of Coleoptera taxa	0.655
COL	Number of Coleoptera individuals	0.662
DIP-taxa	Number of Diptera taxa	0.578
DIP	Number of Diptera individuals	0.806
HYD-taxa	Number of Hydrachnidia taxa	*0.013*
HYD	Number of Hydrachnidia individuals	*0.001*

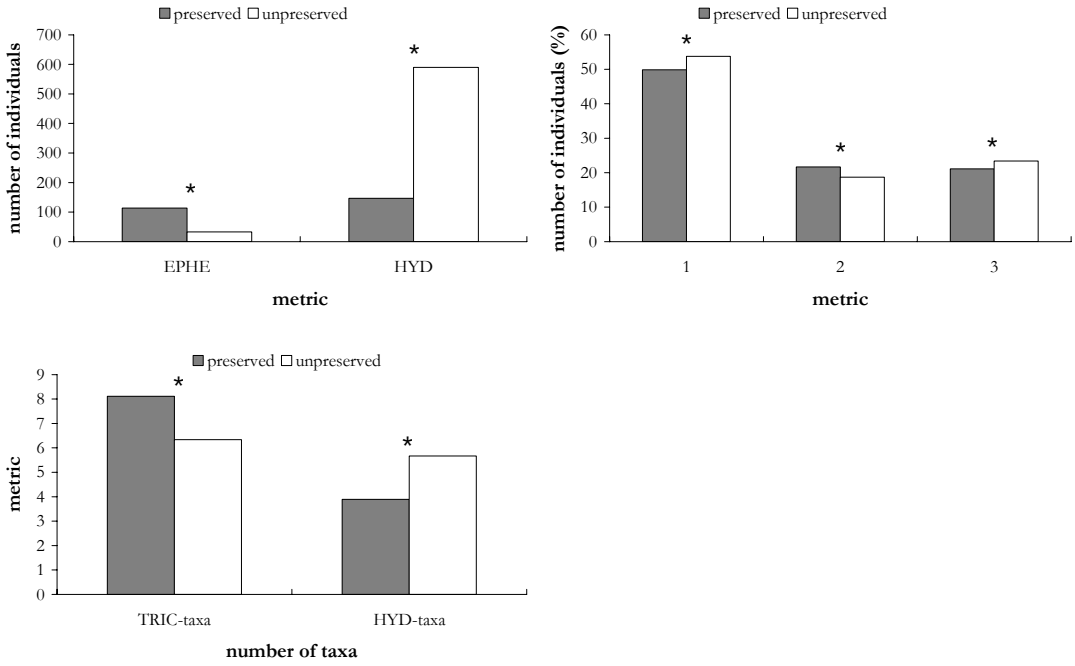


Figure 4.2. Mean metric values for preserved and unpreserved samples from the Springendalse beek, Swalm and Tongerensche beek. (1) AKA+LIT+PSA (%) (2) SHRED (%) (3)metarhithral (%). Asterisks indicate significant differences between means (Exact non-parametric permutation test, Table 3.3).

The assessment results for the preserved and unpreserved samples from the Tongerensche beek didn't differ; all samples indicated poor ecological quality (Table 4.4). The same was true for the Springendalse beek where for preserved as unpreserved one

sample indicated high ecological quality and the two other samples indicated good ecological quality (Table 4.4). The only difference between assessment results for preserved and unpreserved samples was found in the Swalm, where one preserved sample indicated good ecological quality and all other samples indicated poor ecological quality (Table 4.4).

Table 4.4. Assessment results per stream for the preserved and unpreserved samples.

	ecological quality class	
	preserved samples	unpreserved samples
Tongerensche beek	poor	poor
	poor	poor
	poor	poor
Springendalse beek	high	high
	good	good
	good	good
Swalm	good	poor
	poor	poor
	poor	poor

4.4.3 Cost efficiency

The costs of sample processing (sorting and identification) depend on the time required for sample processing. Since the costs of a person hour are variable only the time required for sample processing for preserved and unpreserved samples was compared. No significant difference in the total time required for sample processing was detected between preserved and unpreserved samples (Table 4.5). Comparing the time required for sorting and identification separately also no significant differences between preserved and unpreserved samples were detected (Table 4.5). In comparing preserved and unpreserved samples per stream also no significant differences were found in time required for sorting and identification (Table 4.6). After correction of the recorded time for the number of individuals per sample the results were unchanged (Table 4.5).

Table 4.5. Summary of ANOVA result for comparison of recorded time and recorded time corrected for the number of individuals (corr) in the sample for sorting and identification of preserved and unpreserved samples from three streams.

	protocol F	protocol F (corr)
sorting time	0.02*	0.25*
identification time	2.69*	1.11*
total time recorded for sample processing	2.53*	-

*no significant difference (ANOVA with blocks, α=0.05, n=18)

Table 4.6. Mean time (and standard deviations) recorded for sorting and identification of preserved and unpreserved macroinvertebrate samples from three different streams.

	mean sorting time (min)		mean identification time (min)	
	preserved	unpreserved	preserved	unpreserved
Springendalse beek	502 (122)*	421 (116)	474 (68)*	411 (85)
Swalm	867 (224)*	920 (61)	5564 (2027)*	2978 (741)
Tongerensche beek	1800 (459)*	1885 (248)	3800 (248)*	4318 (742)

*no significant difference (ANOVA, α=0.05, n=6)

4.5 Discussion

Significant differences in metric values between preserved and unpreserved samples were detected. We didn't test whether sorting efficiency between people could have influenced these results. The assumption was made that inter-sorter differences were not significant, as proven by Armitage *et al.* (1995).

The higher number of Ephemeroptera individuals and Trichoptera taxa collected from the preserved samples indicates higher efficiency of sample processing for the preserved samples. It possible that sample efficiency is higher for the preserved samples because Ephemeroptera individuals disintegrate during transportation, storage and sorting. This suggestion is supported by the fact that during sorting often only parts instead of complete Ephemeroptera specimens were found. For the Trichoptera taxa nothing was observed during sorting that could explain the higher number of taxa in the preserved samples. The number of Hydrachnidia individuals and taxa indicated higher efficiency of sample processing for the unpreserved samples. This was expected since Hydrachnidia are very small and therefore hard to detect in preserved samples where they are not moving. The metrics BMWP, RETI, SHRED (%) and AKA+LIT+PSA (%) can not be directly related to sample efficiency, because a higher value doesn't automatically mean higher efficiency. It can only be stated that for these metrics differences were found between preserved and unpreserved samples. The results indicate that sample processing efficiency of preserved and unpreserved samples depends on the organism group and that there is no general rule that efficiency of sample processing increases with the preservation of samples. Some organism groups were not included in the analyses, because they were not present in a large number of the samples. However, these organism groups may show significant differences between protocols in streams where they are more abundant.

AKA+LIT+PSA (%), one of the metrics that showed significant differences between values of preserved and unpreserved samples, was also incorporated in the Dutch assessment system. In only one out of nine occasions a difference in the assessment result between preserved and unpreserved samples was found. The difference in assessment result for the Springendalse beek, where one preserved sample indicated good instead of poor ecological quality class, was the result of higher values for the metric EPT-taxa (%). This corresponds to the findings in chapter 2 where the metric EPT-taxa (%) also caused differences in assessment results between seasons. This might seem strange because the values for EPT-taxa (%) didn't show significant differences between protocols, while the values for the metric AKA+LIT+PSA (%) did show differences between protocols. The fact that differences between metric values occur doesn't necessarily mean this will also create differences in assessment results. Nor does the fact that no differences occur between metric values necessarily means there will be no differences in assessment results. These findings were also presented by Lorenz *et al.* (2004) and Fore *et al.* (2001). Differences in assessment results are created when metric values happen to fall near a break point in the scoring criteria (Fore *et al.*, 2001) and the values for EPT-taxa (%) happened to fall near this break point.

No significant differences were found between the preserved and unpreserved samples in the time required for sorting. This can mean two things: (1) the advantages of moving organism that are easy to spot is cancelled out by the disadvantage that it is harder to catch moving organisms or (2) the variation between replicates is so high that the differences do to preservation are not visible (3) other sources of variation might have confounded the results. If the second is true this means that the exact spot you choose for sampling is far more important than the decision to preserve a sample or not. This doesn't only apply to the time required for sample processing, but also to the

metrics that didn't show significant differences between protocols. It is also likely that other sources of variation have confounded the results. Concerning the time required for identification the results may have been confounded by the fact that two people were used for identification. One person may work faster than the other. Another thing that might have confounded the times recorded for identification is that samples from two streams contained species that the analysts were not very familiar with. This means that times required for identification might have gone down during the identification process as the analysts got to know the species better.

Finally it should be mentioned that there are a few restrictions to this study. First, this study was only preformed in three Dutch streams. Extrapolation of the results to streams with a very different macroinvertebrate community composition is not recommended. Second, formaldehyde was used to preserve the samples. The question is whether the use of ethanol as a preservative would have given the same results. Third, the sorting of the preserved and unpreserved samples took place in the laboratory. The results of sorting unpreserved samples in the laboratory can not be compared to results of sorting the samples in the field. Especially since the circumstances for sorting in the field are in some cases far from optimal (Carter & Resh, 2001; Rawer-Jost, 2001).

4.6 Recommendations

Overall the effects of differences between replicate samples seem to be larger than the effects due to differences in protocol (preserved versus unpreserved samples). This statement is true both for the metric values and the costs of sample processing. This makes the question whether to use a preservative or not almost irrelevant.

In some cases a significant difference between protocols was detected and the choice to use a preservative becomes relevant. In streams with Ephemeroptera or Trichoptera the preservation of samples is preferred. In streams that contain Hydrachnidia it is not recommended to preserve samples. Problems arise when both groups are present in a stream and choices have to be made. The decision made should always depend on the system/metric(s) used for assessment. In this case the assessment result did not depend on the protocol used. For the assessment of Dutch streams the choice to preserve samples or not can be based on individual preferences.

5 Comparison of four protocols for macroinvertebrate sampling and sample processing

5.1 Introduction

Much research has been focused on the comparison of quantitative sampling protocols and on the comparison between quantitative and qualitative sampling protocols. In most cases where the efficiency of collecting species is a major criterion, the pond-net is considered to perform at least as well as any other method (Macan, 1957, 1958, 1977; Morgan & Egglishaw, 1965; Armitage *et al.*, 1975, Mackey *et al.*, 1984; Kerans *et al.*, 1992). Nowadays pond-net and kick net sampling are widely accepted sampling methods. Protocols associated with pond-net and kick-net sampling often do not only include guidelines on sampling but also include guidelines on sample processing. The RIVPACS protocol is a very well known example of a protocol that includes guidelines on sampling as well as on sample processing. Since protocols associated with qualitative sampling are often used for monitoring purposes, the cost efficiency of such protocols is a big issue. A general problem with many sampling methods, used for generating macroinvertebrate data for stream assessment purposes, is the high number of individuals collected (Vinson & Hawkins, 1996; Carter & Resh 2001; Lorenz *et a*l., 2004). Different protocols try in different ways to reduce the time required for sample processing. The most time consuming aspects of sample processing are sorting and identification. Large cost reductions can be achieved by reducing time spent on sorting and identification.

Since the use of different protocols can lead to differences in assessment results (Hering *et al.*, 2004) the implementation of the EU Water Framework Directive (WFD) requires the development of standardized methods for sampling and sample processing of macroinvertebrates. Important questions related to the standardization of protocols are: "In which cases is standardization necessary?" and "If standardization is necessary, which protocols yield optimal results both in terms of cost efficiency and metric/ assessment results? To answer these questions information is needed on different protocols, therefore this study aimed (1) to compare (cost) efficiency between five different protocols for sampling and sample processing; ALTERRA, EBEOSWA, RIVPACS, STAR and STARp (2) to determine whether differences between protocols can lead to significant differences in metric and assessment results.

5.2 Methods

5.2.1 Study site and data collection

The study site was located in the Swalm a small stream in the southern part of the Netherlands. The catchment area is smaller than 100 km^2. The stream is located below 200 m above sea level. The stream is characterised by a high slope, high stream velocity (v>30 cm/sec) and a natural profile. The substrate consisted mainly of gravel and sand.
A reasonable uniform stretch of several 100 metres was selected for sampling. The site was sampled using four different protocols:
1) **RIVPACS**; the RIVPACS protocol is the national standard in the UK. It consists of three minutes of active kick-sampling using a 1000 μm mesh size net and one minute of searching for animals that were not caught during active sampling. Sampling time per habitat is distributed according to habitat coverage. After sampling the sample is

preserved in 4 % formaldehyde. In this study nets were used with a shape that differed from the official standard.

2) **EBEOSWA**; the EBEOSWA protocol is the national standard in the Netherlands. An area of 1.25 m^2 is sampled jabbing a hand-net (500 μm mesh size) through the upper part of the substratum (2-5 cm). Each habitat is sampled irrespective of its coverage.

3) **STAR**; the STAR protocol is a European protocol developed in the AQEM project and further adjusted in the STAR project. A composite sample consists of 20 subsamples of 25 cm taken from all habitats with coverage of at least 5%. The 20 subsamples are distributed according to their share of coverage. A hand-net (25 cm wide with a 500 μm mesh) is used for sampling. Sampling can be done using different techniques. In this study the samples were collected, by pushing the net through the upper part (2-5 cm) of the substratum. After sampling the samples should be preserved in formaldehyde or ethanol according to the protocol, however, in this study we chose not to preserve the samples.

4) **ALTERRA**; this is not an existing official protocol. Samples were taken by pushing the hand-net (25cm wide with a 500 μm mesh) through the upper part (2-5 cm) of the substratum. Each habitat was sampled over a distance that ensured collection of most species present at the habitat (based on expert-judgement). Only habitats with coverage of more than 5% were sampled. The samples from the different habitats were stored in separate buckets.

Sampling took place every week from the 14th of October to the 2nd of December. On the 4th and 11th of November no samples were collected due to lack of sorting capacity and high water levels. One week RIVPACS, EBEOSWA and ALTERRA samples were taken, the other week STAR samples were taken. This process was repeated until 12 samples were taken in total, 3 'replicates' per protocol. A certified staff member performed RIVPACS sampling. Apart from the RIVPACS sampling all samples were taken by the same person. The 'replicate' samples were not 'real' replicates because they were not all taken during the same. Unfortunately, it was not feasible to collect all samples on the same day due to practical reasons related to sorting (see below).

After sampling the samples were taken to the laboratory. The STAR, EBEOSWA and ALTERRA samples were stored in a refrigerator, where they were oxygenated, until sorting took place. The samples were treated using four different protocols:

1) **RIVPACS**; prior to sorting the samples are washed through a 500 μm sieve. The analyst selects a proportion of the sample which is completely sorted in the laboratory. The sorted fraction depends on the number of individuals expected to be found in the sample. The unsorted rest of the sample is checked for taxa that have not been found in the sorted fraction. Individuals from these taxa are picked from the sample and retained in a separate vial for "extra" taxa. The taxa list composed after identification of the sorted individuals is extrapolated to the whole sample.

2) **EBEOSWA**; the EBEOSWA protocol does not give guidelines on sample processing. In this study the samples were washed through a 1000 and 250 μm sieve. The samples were completely sorted.

3) **STAR**; the STAR samples are washed through a 1000 and 250 μm sieve before sorting. The total amount of sampled material that is to be sorted is reduced by taking a subsample. The subsample has to contain at least 1/6 of the total amount of sampled material and 700 individuals. In case 1/6 of the sample contains less than 700 individuals the subsample is increased in size. All individuals are removed from the subsample. The taxa list composed after identification of the sorted individuals is extrapolated to the whole sample. Unfortunately the protocol was misinterpreted, which resulted in

subsamples containing less than 1/6 of the sampled material (1/9, 1/10 and 1/12). The subsamples did all contain at least 700 individuals.
After sorting the subsample the remaining sample was completely sorted. The completely sorted STAR sample will be referred to as the STARp protocol.
4) **ALTERRA**; the samples from the different habitats were kept separate during sample processing. The samples were washed through a 1000 and 250 μm sieve before sorting. The samples were completely sorted. The taxa list composed after identification was extrapolated to a sample area of 1.25 m^2, based on recorded habitat coverage and the sampled area of each habiat.
For a more elaborate description of the different protocols for sampling and sample processing we refer to STAR Consortium (2003).

Since all samples, except for the RIVPACS samples were not preserved, living organisms had to be picked from the samples. Living organisms tend to disintegrate after a few days; this meant that the samples had to be sorted at least within five days after sampling. Due to limited sorting capacity (three to five persons) it was impossible to collect all 12 samples in one week.

The samples were sorted by eye by a varying group of three to five people, except for the RIVPACS samples. All RIVPACS samples were sorted by one person trained in sorting RIVPACS samples. Organisms were identified to the lowest taxonomic level possible (species level for almost all groups). All identifications were preformed by the same two qualified persons. The time necessary to sort and identify the specimens in each sample was recorded.

5.2.2 Data analysis

Metrics commonly used in Europe were selected for evaluation (Table 5.1). The metrics were selected from an extensive list given by Hering *et al.* (2004). Apart from metrics selected form the list by Hering *et al.* (2004) the number of taxa and the number of individuals for each major macroinvertebrate group (e.g. Diptera, Ephemeroptera, Plecoptera) was also evaluated. Some major groups were only present in a few samples and in low abundances, only metrics with values higher than zero in 75% of the samples were used for analysis.

Metric values were calculated with the AQEM River Assessment Program (AQEMrap version 2.3). With the same program the final ecological quality classes were calculated for each of the three streams, characterising the samples as being from small Dutch lowland streams. The multimetric index used to calculate the final ecological quality class is a revised version of the multimetric index described by Vlek *et al.* (2004). The multimetric index consists of 11 metrics which are indicated in table 5.1.

To test for significant differences in metric values between the different protocols a one-way ANOVA was used (α=0.05). Assumptions of normality and homogeneity of variance could not be tested in a reliable way due to the low number of samples. For this reason it would have been more appropriate to perform a non-parametric test. However, a non-parametric test would never be able to detect significant differences (α=0.05) between protocols based on three replicates for a two–tailed test. Therefore it was decided to use an ANOVA and to transform metric values based on experiences in other studies. Abundance metrics were ln(x+1) transformed, taxa counts were not transformed (Kerans *et al.*, 1992) and proportions were ln(x+1)-ln(y+1) transformed. Biotic index data (e.g. Sapobic Index, BMWP, ASPT) were not transformed (Norris & Georges, 1993). Metrics like XENO (%), SHRED (%) and littoral (%) are not simple proportional metrics. The values for these metrics also depend on the strength with which a species

prefers a certain category (AQEM consortium, 2002). The decision was made not to transform values of these metrics, since no references were found on a suited transformation.

The differences in time required for sample processing applying different protocols was also tested with a one-way ANOVA (α=0.05). Data on recorded time were ln(x) transformed prior to analysis, according to Growns *et al.* (1997).

5.3 Results

5.3.1 General

In total 193 taxa were collected during this study. The macroinvertebrate community of the Swalm consisted for a large part of Diptera and Oligochaeta taxa (Figure 5.1). Considering the number of individuals the Oligochaeta formed a large proportion of the community, followed by the Curstacea and Diptera (Figure 5.2).
One replicate contained on average 33% of the total number of taxa found during this study. A replicate collected according to the ATERRA protocol contained on average 41% of the total number of taxa found during this study, this was 36% for the EBEOSWA protocol, 27% for the RIVPACS protocol, 28% for the STAR protocol and 51% for the STARp protocol.

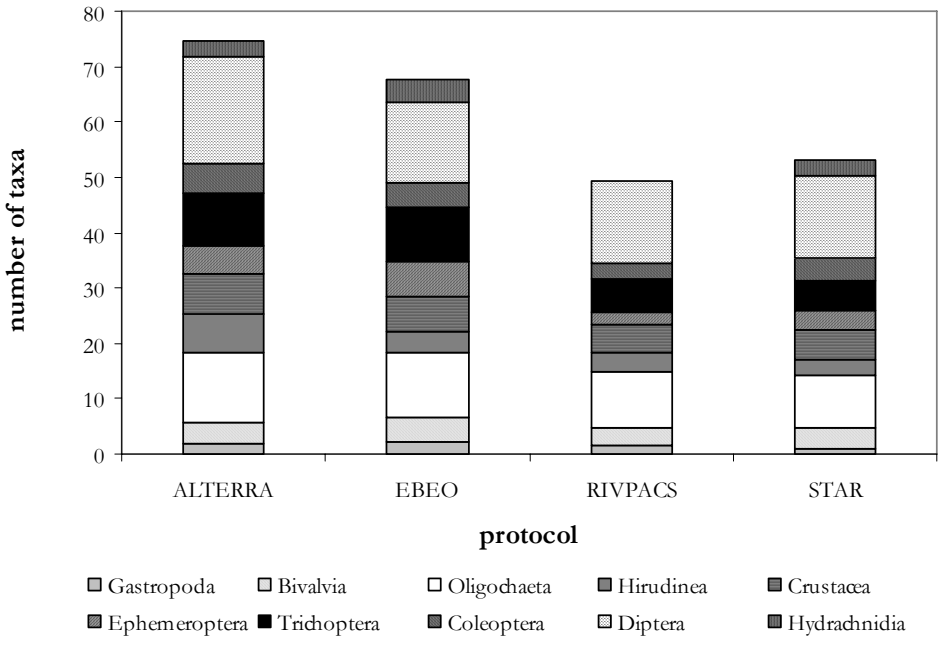


Figure 5.1. Macorinvertebrate community composition of the Swalm per protocol. Number of taxa is based on the average of the three replicate samples per protocol. Only groups that formed more than 5% of the total community were included in the graph.

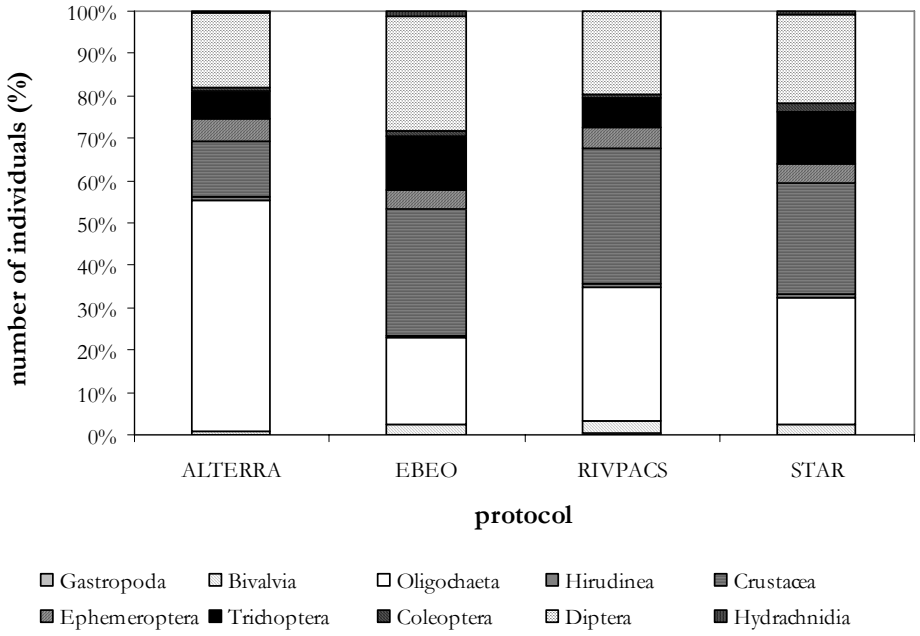


Figure 5.2. Macorinvertebrate community composition of the Swalm per stream. Mean percentages of individuals are based on the three replicate samples per protocol. Only groups that formed more than 5% of the total community were included in the graph.

5.3.2 Metrics and assessment

Significant differences between protocols were detected for 22 out of the 61 metrics tested (Table 5.1). Five of the metrics showing significant differences between protocols were measures of abundance and nine metrics were measures of taxa richness (Table 5.1).

The ALTERRA samples on average contained the highest number of individuals, followed by STARp samples and STAR samples. The STARp, STAR and ALTERRA samples all contained significantly ($p<0.05$) higher numbers of individuals than the EBEOSWA and RIVPACS samples (Table 5.2). The ALTERRA samples also contained the highest number of Oligochaeta, Ephemeroptera and Hirudinea individuals, followed by the STARp and STAR samples. Only in the case of the number of Ephemeroptera individuals the results obtained with the STARp protocol significantly ($p<0.05$) differed from the ALTERRA protocol. In all cases the number of individuals collected with the ALTERRA protocol significantly ($p<0.05$) differed from the number of individuals collected with the EBEOSWA and RIVPACS protocol. In none of the cases the number of individuals collected with the STARp protocol significantly differed from *all* more cost efficient protocols (EBEOSWA, RIVPACS and STAR) (Table 5.2 and 5.4).

The total number of taxa collected with the STARp protocol was significantly ($p<0.05$) higher than the number of taxa collected with any other protocol. The same results were found for BMWP values, DIM values and the number of families and genera (Table 5.2). The number of Crustacea, Trichoptera, Ephemeroptera and Hirudinea taxa collected with STARp and ALTERRA protocol were significantly ($p<0.05$) higher than the number of taxa collected with any other protocol. With the exception that the number of Tricoptera and Ephemeroptera taxa collected with the EBEOSWA protocol didn't significantly ($p<0.05$) differ form the number of taxa collected with the STARp or ALTERRA protocol. The number of Hydrachnididia taxa collected with the RIVPACS

protocol was significantly ($p<0.05$) lower than the number of Hydrachnididia taxa collected with any other protocol. In fact no Hydrachnidia taxa were collected from the RIVPACS samples. The highest number of Ephemeroptera taxa was collected with the EBEOSWA protocol. The number of Ephemeroptera taxa collected with the EBEOSWA protocol significantly ($p<0.05$) differed from the number of Ephemeroptera taxa collected with the STAR and RIVPACS protocol (Table 5.2).

Apart from the 14 measures of abundance and taxa richness that showed significant differences between protocols, there were eight other metrics showing significant differences between protocols. For these metrics no general pattern in the differences between protocols could be detected

Most samples indicated that the Swalm was of good ecological quality (11 out of 15) (Table 5.3). The ALTERRA samples all indicated good ecological quality. The EBEOSWA samples indicated good ecological quality for two samples and high ecological quality for one sample. The metric OL+DIP-taxa (%) showed relative low values for the sample of high ecological quality compared to the other EBEOSWA samples. The coefficient of variation of the OL+DIP-taxa (%) values was high (40%) for the EBEOSWA samples compared to the samples collected according to other protocols (between 4 and 14%). The metric EPT/DIP-taxa (%) showed relative high values for the same sample. The coefficient of variation of the EPT/DIP-taxa values for the EBEOSWA samples was also high (153%) compared to the samples collected according to other protocols (between 10 and 31%). Of the samples collected according to the RIVPACS and STAR protocol two samples indicated good ecological quality and one sample indicated poor ecological quality. In the cases were the assessment resulted in poor ecological quality the values for EPT-taxa (%) were lower than for the samples that indicated good ecological quality. The samples collected according to the STARp protocol indicated good ecological quality in two cases and moderate ecological quality in one case. The metric littoral (%) showed relative high values for the sample of moderate ecological quality compared to the other STAR samples. The coefficient of variation of the hyporhithral (%) values was high (20%) for the STARp samples compared to the samples collected according to other protocols (between 8 and 14%).

Table 5.1. Summary of ANOVA results and overview of replicate means (coefficient of variation) per metric for the ALTERRA, EBEOSWA, RIVPACS, STAR and STARp protocol. Metrics incorporated in the Dutch AQEM assessment system are indicated in grey.

acronym	metric description	protocol F	mean (coefficient of variation)				
			ALTERRA	EBEOSWA	RIVPACS	STAR	STARp
ABUN	Abundance	26.67*	11495 (23%)	2676 (30%)	2818 (14%)	6684 (20%)	6861 (18%)
NTAX	Number of taxa	12.54*	79 (20%)	70 (11%)	53 (14%)	54 (2%)	99 (9%)
NGEN	Number of genera	12.20*	52 (17%)	48 (9%)	42 (15%)	36 (4%)	67 (7%)
NFAM	Number of families	6.22*	33 (12%)	32 (6%)	29 (12%)	24 (8%)	37 (13%)
ZSI	Saprobic Index (Zelinka & Marvan)	1.42	2 (3%)	2 (5%)	2 (5%)	2 (6%)	2 (7%)
ALPHA-MESO (%)	Proportion of individuals with a preference for alpha-meso saprobic circumstances	0.92	24 (7%)	29 (18%)	29 (10%)	28 (13%)	27 (19%)
OLIGO (%)	Proportion of individuals with a preferences for oligo saprobic circumstances	0.79	26 (8%)	22 (24%)	19 (29%)	21 (23%)	21 (30%)
XENO (%)	Proportion of individuals with a preferences for xeno saprobic circumstances	2.04	1 (106%)	2 (45%)	2 (86%)	1 (54%)	1 (62%)
GFI D03	German Fauna Index D03	0.48	0 (158%)	0 (143%)	0 (56%)	0 (27%)	0 (26%)
GFI D04	German Fauna Index D04	1.17	-1 (32%)	-1 (43%)	-1 (35%)	-1 (8%)	-1 (18%)
BMWP	Biological Monitoring Working Party	4.01*	124 (21%)	130 (16%)	103 (17%)	81 (6%)	133 (13%)
ASPT	Average Score per Taxon	3.19	5 (9%)	6 (5%)	5 (4%)	5 (5%)	5 (5%)
GSI	German Saprobic Index (new version)	1.4	2 (6%)	2 (4%)	2 (6%)	2 (2%)	2 (1%)
CSI	Czeck Saprobic Index	3.21	1 (2%)	2 (10%)	2 (11%)	2 (15%)	2 (18%)
MTS	Mayfly Total Score	2.8	8 (43%)	12 (25%)	7 (31%)	7 (25%)	11 (16%)
HAI	Acid Index (Hendrikson & Medin)	1.79	11 (5%)	12 (5%)	10 (6%)	10 (22%)	11 (9%)
LIFE	LIFE-index	1.51	6 (2%)	6 (4%)	6 (3%)	6 (3%)	6 (1%)
NSTA	Number of sensitive taxa (Austria)	4.93*	4 (31%)	5 (11%)	3 (0%)	2 (25%)	5 (35%)
DIM	Diversity Index (Margalef)	12.98*	8 (19%)	9 (10%)	7 (13%)	6 (3%)	11 (8%)
DIS	Diversity Index (Shannon & Wiener)	3.57*	2 (12%)	2 (13%)	3 (9%)	3 (6%)	3 (11%)
RP (%)	Proportion of individuals with a preference for zones with moderate to high current (rheophil)	2.92	30 (32%)	60 (52%)	63 (5%)	37 (18%)	34 (34%)
PEL (%)	Proportion of individuals with a preference for muddy substrates	0.74	12 (9%)	7 (75%)	12 (53%)	11 (24%)	12 (20%)

acronym	metric description	protocol F	mean (coefficient of variation)				
			ALTERRA	EBEOSWA	RIVPACS	STAR	STARp
PSA (%)	Proportion of individuals with a preference for sandy substrates	5.18	37 (6%)	16 (65%)	16 (35%)	26 (22%)	31 (29%)
AKA+LIT+PSA (%%)	Proportion of individuals with a preferences for gravel, littoral and sand	2.01	71 (10%)	70 (18%)	56 (10%)	60 (7%)	64 (10%)
littoral (%)	Proportion of individuals with a preference for the littoral	2.2	12 (95%)	10 (66%)	11 (22%)	24 (21%)	24 (55%)
metarhithral (%)	Proportion of individuals with a preference for the lower-trout region	2.78	18 (15%)	22 (20%)	21 (15%)	16 (11%)	15 (22%)
hyporhithral (%)	Proportion of individuals with a preference for the greyling region	1.14	26 (12%)	26 (8%)	24 (14%)	22 (11%)	22 (20%)
epirhithral (%)	Proportion of individuals with a preference for the upper-trout region	4.58*	3 (37%)	8 (27%)	7 (49%)	4 (24%)	3 (29%)
IBR	Index of Biocoenotic Region	5.39*	6 (8%)	5 (8%)	5 (6%)	6 (3%)	6 (7%)
GAT/COL (%)	Proportion of gatherers to collectors (individuals)	4.28	60 (7%)	29 (55%)	33 (32%)	41 (12%)	47 (23%)
SHRED (%)	Proportion of shredders (individuals)	2.34	11 (60%)	19 (18%)	20 (2%)	20 (24%)	18 (18%)
PASF (%)	Proportion of passive filter feeders (individuals)	1.37	11 (108%)	29 (86%)	18 (70%)	7 (21%)	8 (39%)
GRA+SCRA (%)	Proportion of grazers and scrapers (individuals)	0.35	5 (56%)	7 (54%)	8 (60%)	8 (10%)	7 (18%)
RETI	Rhithron Feeding Type Index	3.51*	0 (45%)	0 (14%)	0 (18%)	0 (15%)	0 (16%)
EPT-taxa	Number of Ephemeroptera, Plecoptera and Trichoptera taxa	10.09*	14 (21%)	16 (9%)	8 (25%)	9 (22%)	16 (7%)
EPT/DIP-taxa (%)	Proportion of EPT-taxa to Diptera taxa	1.13	1 (10%)	7 (153%)	1 (31%)	1 (23%)	1 (26%)
OL+DIP-taxa (%)	Proportion of Oligochaeta and Diptera taxa	0.93	41 (4%)	36 (40%)	48 (14%)	46 (11%)	46 (13%)
OL-taxa	Number of Oligochaeta taxa	2.96	13 (16%)	12 (35%)	10 (31%)	10 (16%)	17 (18%)
TRIC (%)	Proportion of Trichoptera individuals	2.65	6 (98%)	10 (128%)	7 (93%)	12 (20%)	12 (29%)
EPT-taxa (%)	Proportion of EPT-taxa	4.28*	18 (2%)	24 (22%)	16 (18%)	17 (23%)	16 (10%)
BIVAL-taxa	Number of Bivalvia taxa	1.73	4 (42%)	4 (13%)	3 (33%)	4 (31%)	5 (0%)
BIVAL	Number of Bivalvia individuals	0.96	94 (110%)	65 (122%)	77 (73%)	159 (97%)	204 (71%)
OL	Number of Oligochaeta individuals	9.20*	6271 (37%)	544 (65%)	894 (63%)	1991 (8%)	2583 (57%)
CRUS-taxa	Number of Crustacea taxa	13.37*	7 (8%)	6 (9%)	5 (0%)	5 (11%)	7 (8%)
CRUS	Number of Crustacea individuals	3.3	1529 (30%)	800 (15%)	895 (17%)	1775 (40%)	1503 (40%)
EPHE-taxa	Number of Ephemeroptera taxa	6.81*	5 (35%)	6 (18%)	2 (25%)	4 (16%)	5 (11%)

acronym	metric description	protocol F	mean (coefficient of variation)				
			ALTERRA	EBEOSWA	RIVPACS	STAR	STARp
EPHE	Number of Ephemeroptera individuals	4.95*	618 (57%)	120 (61%)	137 (73%)	284 (36%)	180 (27%)
TUR-taxa	Number of Turbellaria taxa	2.6	4 (57%)	2 (25%)	2 (92%)	1 (173%)	3 (17%)
TUR	Number of Turbellaria individuals	0.83	21 (72%)	7 (126%)	15 (88%)	8 (173%)	4 (25%)
TRIC-taxa	Number of Trichoptera taxa	6.38*	9 (22%)	10 (17%)	6 (29%)	5 (29%)	10 (6%)
TRIC	Number of Trichoptera individuals	1.88	726 (115%)	345 (144%)	196 (88%)	833 (39%)	826 (11%)
COL-taxa	Number of Coleoptera taxa	3.46	5 (11%)	4 (35%)	3 (22%)	4 (25%)	6 (33%)
COL	Number of Coleoptera individuals	2.18	127 (136%)	25 (103%)	21 (126%)	124 (66%)	112 (36%)
DIP-taxa	Number of Diptera taxa	2.17	19 (27%)	15 (89%)	15 (7%)	15 (20%)	29 (22%)
DIP	Number of Diptera individuals	1.2	2001 (54%)	730 (91%)	556 (47%)	1402 (31%)	1361 (35%)
HYD-taxa	Number of Hydrachnidia taxa	9.39*	3 (58%)	4 (25%)	-	-	4 (13%)
HYD	Number of Hydrachnidia individuals	8.70*	54 (138%)	28 (124%)	-	-	44 (34%)
GAS-taxa	Number of Gastropoda taxa	1.4	2 (87%)	2 (49%)	2 (35%)	1 (100%)	3 (46%)
GAS	Number of Gastropoda individuals	0.38	4 (104%)	4 (16%)	9 (73%)	9 (95%)	5 (43%)
HIRU-taxa	Number of Hirudinea taxa	7.35*	7 (14%)	4 (42%)	3 (17%)	3 (78%)	7 (14%)
HIRU	Number of Hirudinea individuals	4.10*	50 (42%)	8 (43%)	14 (62%)	44 (88%)	38 (29%)

Table 5.2. Overview of which protocols significantly differ ($p<0.05$) from each other, based on the Least Significant Difference (LSD).Comparison of replicate means (coefficient of variation) per metric for the ALTERRA, EBEOSWA, RIVPACS, STAR and STARp protocol.

acronym	**significant differences between**	**mean (coefficient of variation%)**				
		ALTERRA	**EBEOSWA**	**RIVPACS**	**STAR**	**STARp**
ABUN	STARp – ALTERRA/EBEOSWA/RIVPACS ALTERRA - EBEOSWA/STAR/RIVPACS STAR - EBEOSWA/RIVPACS	11495 (23%)	2676 (30%)	2818 (14%)	6684 (20%)	6861 (18%)
NTAX	STARp – ALTERRA/EBEOSWA/RIVPACS/STAR ALTERRA - RIVPACS/STAR EBEOSWA – RIVPACS	79 (20%)	70 (11%)	53 (14%)	54 (2%)	99 (9%)
NGEN	STARp – ALTERRA/ EBEOSWA/RIVPACS/STAR ALTERRA/EBEOSWA - STAR	52 (17%)	48 (9%)	42 (15%)	36 (4%)	67 (7%)
NFAM	STARp – ALTERRA/EBEOSWA/RIVPACS/STAR ALTERRA/EBEOSWA - STAR	33 (12%)	32 (6%)	29 (12%)	24 (8%)	37 (13%)
BMWP	STARp - ALTERRA/EBEOSWA/RIVPACS/STAR	124 (21%)	130 (16%)	103 (17%)	81 (6%)	133 (13%)
NSTA	STARp/EBEOSWA - RIVPACS/STAR	4 (31%)	5 (11%)	3 (0%)	2 (25%)	5 (35%)
DIM	STARp – ALTERRA/EBEOSWA/RIVPACS/STAR EBEOSWA - RIVPACS/STAR	8 (19%)	9 (10%)	7 (13%)	6 (3%)	11 (8%)
DIS	RIVPACS – ALTERRA/EBEOSWA STAR - ALTERRA	2 (12%)	2 (13%)	3 (9%)	3 (6%)	3 (11%)
epirhithral (%)	EBEOSWA – ALTERRA/STARp/STAR RIVPACS – ALTERRA/STARp	3 (37%)	8 (27%)	7 (49%)	4 (24%)	3 (29%)
IBR	STARp – EBEOSWA/RIVPACS STAR - EBEOSWA/RIVPACS	6 (8%)	5 (8%)	5 (6%)	6 (3%)	6 (7%)
RETI	STAR - ALTERRA RIVPACS - ALTERRA	0 (45%)	0 (14%)	0 (18%)	0 (15%)	0 (16%)
EPT-taxa	STARp/ALTERRA/EBEOSWA - RIVPACS/STAR	14 (21%)	16 (9%)	8 (25%)	9 (22%)	16 (7%)
EPT-taxa (%)	ALTERRA – STAR EBEOSWA – RIVPACS/STAR	18 (2%)	24 (22%)	16 (18%)	17 (23%)	16 (10%)

acronym	significant differences between	mean (coefficient of variation%)				
		ALTERRA	**EBEOSWA**	**RIVPACS**	**STAR**	**STARp**
OL	ALTERRA - EBEOSWA/RIVPACS/STAR STARp - EBEOSWA	6271 (37%)	544 (65%)	894 (63%)	1991 (8%)	2583 (57%)
CRUS-taxa	STARp/ALTERRA - EBEOSWA/RIVPACS/STAR EBEOSWA - RIVPACS/STAR	7 (8%)	6 (9%)	5 (0%)	5 (11%)	7 (8%)
EPHE-taxa	EBEOSWA - RIVPACS/STAR STARp/ALTERRA - RIVPACS	5 (35%)	6 (18%)	2 (25%)	4 (16%)	5 (11%)
EPHE	ALTERRA – STARp/EBEOSWA/RIVPACS	618 (57%)	120 (61%)	137 (73%)	284 (36%)	180 (27%)
TRIC-taxa	STARp/ALTERRA/EBEOSWA - RIVPACS/STAR	9 (22%)	10 (17%)	6 (29%)	5 (29%)	10 (6%)
HYD-taxa	STARp/ALTERRA/EBEOSWA/STAR - RIVPACS	3 (58%)	4 (25%)	-	3 (22%)	4 (13%)
HYD	STARp/ALTERRA/EBEOSWA/STAR - RIVPACS	54 (138%)	28 (124%)	-	54 (10%)	44 (34%)
HIRU-taxa	STARp/ALTERRA - EBEOSWA/RIVPACS/STAR	7 (14%)	4 (42%)	3 (17%)	3 (78%)	7 (14%)
HIRU	ALTERRA – EBEOSWA/RIVPACS STARp/STAR - EBEOSWA	50 (42%)	8 (43%)	14 (62%)	44 (88%)	38 (29%)

Table 5.3. Assessment results per protocol (ALTERRA, EBEOSWA, RIVPACS, STAR and STARp) for the three replicate samples collected.

ecological quality class	number of samples				
	ALTERRA	EBEOSWA	RIVPACS	STAR	STARp
bad	0	0	0	0	0
poor	0	0	1	1	0
moderate	0	0	0	0	1
good	3	2	2	2	2
high	0	1	0	0	0

5.3.3 Cost efficiency

The costs of sample processing (sorting and identification) depend on the time required for sample processing. Since the costs of a person hour are variable only the time required for sample processing was compared between protocols. Total time required for sample processing significantly ($p<0.05$) differed between protocols, except for the EBEOSWA and ALTERRA protocol (Table 5.4). The comparison of protocols for required sorting time led to the same results. The comparison of protocols for required identification time also gave the same results, except for the fact that no significant differences were detected between the EBEOSWA and STAR protocol (Table 5.4). On average sample processing took 155 hours in case of STARp samples, 48 hours in case of ALTERRA samples, 33 hours in case of EBEOSWA samples, 18 hours in case of STAR sample and 9 hours in case of RIVPACS samples (Figure 5.3).

Table 5.4. Summary of ANOVA results for comparison of time recorded for different aspects of sample processing between five different protocols (ALTERRA, EBEOSWA, RIVPACS, STAR and STARp).

	Protocol F	significant difference between
sorting time	80.20*	STARp-ALTERRA/EBEOSWA/RIVPACS/STAR ALTERRA/EBEOSWA-RIVPACS/STAR STAR-RIVPACS
identification time	25.73*	STARp-ALTERRA/EBEOSWA/RIVPACS/STAR ALTERRA – RIVPACS/STAR EBEOSWA-RIVPACS
total time recorded for sample treatment	51.88*	STARp-ALTERRA/EBEOSWA/RIVPACS/STAR ALTERRA/EBEOSWA-RIVPACS/STAR STAR-RIVPACS

*($p<0.05$)

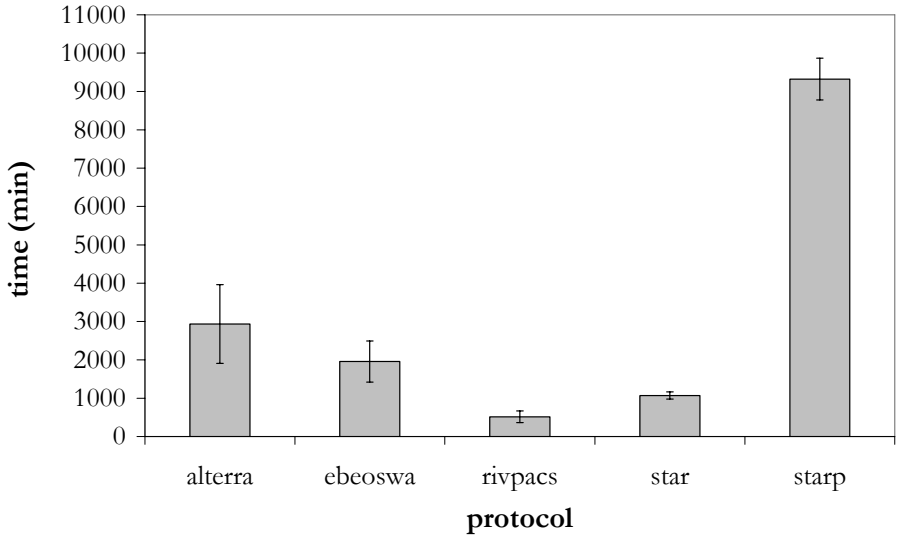


Figure 5.3. Mean time (and standard deviations) recorded for sorting and identification of samples collected form the Swalm according to five different protocols (ALTERRA, EBEOSWA, RIVPACS, STAR and STARp).

5.4 Discussion

Significant differences in metric values between protocols for sampling and sample processing were detected. We didn't test whether sorting efficiency between people could have influenced these results. The assumption was made that inter-sorter differences were not significant, as proven by Armitage *et al.* (1995).

Save for the RIVPACS samples all samples were taken by the same operator. For collecting and sorting the RIVPACS samples a different operator was used experienced in RIVPACS sampling and sample processing. This was done to avoid arguments over the fact that differences may or may not have been detected due to the fact that sampling and sample processing wasn't preformed by an operator experienced in applying the RIVPACS protocol. Armitage *et al.* (1974) and Egglishaw (1964) didn't find significant differences between operators. Clarke *et al.* (2002) stated that inter-operator influences were negligible. Furse *et al.* (1981) and Mackey *et al.* (1984), however, demonstrated significant differences between operators for pond-net samples, therefore it should be kept in mind that significant differences detected between the RIVPACS protocol and any other protocol might have been caused by differences between operators, but it does not seem likely since both operators were adequately trained.

In paragraph 5.2 it was explained that the replicate samples in this study were in fact not 'real' replicates. To make sure that the right choice was made in considering the samples collected according to the same protocol as replicates the ANOVAs on metric values were repeated using the different sampling dates as blocks. By using the sampling dates as blocks the effect of variation between dates could not interfere with the protocol effect. The ANOVA results with blocks differed for only six metrics from the ANOVA results without blocks. In the case of using the sampling dates as blocks RP (%), SHRED (%), GAT/COL (%) , PSA (%) and CRUS values were considered significantly ($p<0.05$) different between protocols. The metric RETI did not significantly differ between protocols. The fact that ANOVA results with and without sample dates as blocks only differed for six metrics was a strong indication the samples were appropriately considered as replicates.

For many metrics evaluated in this study no differences in metric values between protocols were detected. In these cases both the variation in metric values and costs for each protocol should be considered. In general it can't be said that one protocol is more variable than another, variation in metric values for each protocol seems to differ depending on the metric. Only taking costs into consideration the RIVPACS protocol is by far preferred over any other protocol.

The STARp protocol showed higher efficiency in collecting taxa at genus level, family level and collecting the total number of taxa than any other protocol. The problem is that the STARp protocol is very time consuming and will never be adopted as a standard protocol for biological monitoring and assessment. The results presented in this paper will give the (potential) users of the ALTERRA, EBEOSWA, RIVPACS or STAR protocol an idea about the bias in their results. Bias is "….the magnitude and direction of the tendency to measure something other than what was intended" (Green, 1979). The higher efficiency of the STAR protocol confirmed our belief that samples collected according to the STARp protocol give the best picture

of the macroinvertebrate community composition. This belief was held for two reasons. First, in sampling according to the STARp protocol more material is collected than sampling according to any other protocol. STARp samples are collected by sampling a larger area than for the ALTERRA samples. The STARp samples are collected by pushing the net through the upper layer of the substrate. EBEOSWA and RIVPACS samples are collected by jabbing the substrate, which results in the collection of less material. Second, STARp samples are sorted completely, while RIVVPACS and STAR samples are only partly sorted.

The total number of taxa for some major organism groups significantly differed between protocols. In cases of differences between protocols the STARp and ALTERRA showed higher efficiency in collecting taxa than the EBEOSWA, RIVPACS and STAR protocol. Since the ALTERRA protocol is far more cost efficient than the STARp protocol and just as efficient in collecting taxa, the use of the ALTERRA protocol is preferred. The higher efficiency for the ALTERRA and STARp protocol in collecting taxa is most likely due to the fact that these protocols prescribe sorting the whole sample instead of only a fraction of the sample (RIVPACS and STAR). In case of the number of Trichoptera and Ephemeroptera taxa no significant differences were found between EBEOSWA samples and ALTERRA and STARp samples. The EBEOSWA or ALTERRA protocol can be used based on personal preferences, both protocols are far more cost efficient than the STAR protocol. It should be noted that although differences between protocols were detected, these differences might not be considered large enough to be important, depending on the objective of the research. When this is the case choices for a protocol can be based on cost efficiency and variation in metric results.

Metrics expressing the number of taxa collected can easily be related to the efficiency of a protocol. All other metrics can not be directly related to the efficiency of a protocol. For example, in case the collected total number of individuals is much higher for one protocol than the other, this doesn't automatically mean that the protocol removing the most specimens form a sample is also the most efficient. Many protocols use extrapolation (paragraph 5.2) which can lead to underestimation or overestimation of the number of individuals. Considering the STARp protocol as the most optimal protocol all other protocols that do not significantly differ from the STARp protocol can in principle be used for sampling and sample processing. From the point of cost efficiency the EBEOSWA, RIVPACS or STAR protocol are always preferred to the STARp and ALTERRA protocol. For most metrics (other than measures of taxa richness) no differences in values were detected between the STARp protocol and at least one of the more cost efficient protocols (EBEOSWA, RIVPACS or STAR). The only exceptions to this rule were the BMWP and DIM. This means that in most cases a more cost effective protocol (EBEOSWA, RIVPACS or STAR) than the STARp protocol can be chosen for sampling and sample processing, however which protocol depends on the metric. When choosing to use a more cost efficient protocol one should not forget to compare variation in metric values for this protocol with the STARp protocol. Higher variation for the more cost efficient protocol may increase the risk of misclassification or make it difficult to detect changes in status; this depends on the metric and its class boundaries.

No Hydrachnidia were collected from samples treated according to the RIVPACS protocol, resulting in significant differences between the RIVPACS protocol and all other protocols. Hydrachnidia are very small and therefore hard to detect in a sample when they are not moving (paragraph 4.4.2). The fact that RIVPACS samples are preserved prior to sorting and are not sorted completely probably are the reason for the absence of Hydrachnidia in the RIVPACS samples.

In the case of individual metrics the samples collected according to the STARp protocol were considered to best represent the macroinvertebrate community composition of the stream. In case of assessment based on the multimetric index revised by Vlek *et al.* (2004) the ALTERRA protocol was considered most optimal. The ALTERRA protocol was considered most optimal because the multimetric index has been developed based on samples collected according to the ALTERRA protocol. Based on expert-judgement the Swalm was qualified of moderate to good ecological quality. Since 75% of the samples in this study assessed the status of the Swalm as good, good ecological quality was considered to be the correct assessment. The metrics that caused assessment results other than good were not the metrics that showed significant differences between protocols. This was also observed in comparing preserved and unpreserved samples (paragraph 4.4.2). One EBEOSWA sample was incorrectly assessed due to high variation in metric values for two metrics. One RIVPACS and one STAR sample were incorrectly assessed due to the metric EPT-taxa (%). Values for this metric happened to fall near a break point in the scoring criteria (Fore *et al.*, 2001). This was also observed in comparing samples collected during different months (paragraph 2.5) and comparing preserved and unpreserved samples (paragraph 4.). One STARp sample was incorrectly assessed due to the metrics littoral (%) and hyporhithral (%). The metric littoral (%) happened to fall near a breakpoint in scoring criteria and the metric hyporhithral (%) showed high variation in values compared to the ALTERRA samples. In short it is very difficult to predict the effects of differences in individual metrics on the final assessment. The only way to decided on the most appropriate protocol for determining the ecological quality status of a stream is to compare assessment results between protocols.

The samples collected according to the ALTERRA protocol all resulted in the correct assessment of the Swalm, according to the revised method of Vlek *et al.* (2004), contrary to the samples collected according to the other protocols. The STAR and RIVPACS samples gave incorrect assessment results caused by the metric EPT-taxa (%). In other studies described in this paper it was already observed that EPT-taxa (%) values were often the cause of misclassification (paragraph 2.5 and 4.5). When the metric EPT-taxa (%) would be replaced by another more robust metric the RIVPACS protocol and STAR protocol can also be applied for the assessment of Dutch streams.

For this study only one site was sampled this makes it difficult to extrapolate results. Mackey *et al.* (1984) already stated: "…there is reason to doubt that pond-net samples are equally efficient over a wide range of sampling conditions and rivers."

5.5 Recommendations

Based on the findings in this study it is recommended to use the ALTERRA protocol for sampling and sample treatment for the assessment of Dutch streams. However, when the metric EPT-taxa (%) would be replaced by another more robust metric the STAR or RIVPACS might also prove to be appropriate protocols. It can not be ruled out that future research may indicate no significant differences in assessment results for Dutch streams between protocols. Since the Swalm was a stream with relative high species diversity differences between protocols are only likely to be smaller in other streams.

6 Subsampling

6.1 Sorting subsamples versus complete samples

6.1.1 Introduction

The application of the AQEM/STAR multi habitat approach provides a large amount of sample material and organisms. In most cases the sample is too large to sort the sample in total. To reduce the effort for sorting and analysis time subsampling techniques are used. The aim of this study is to obtain quantitative estimates of the difference in macroinvertebrate community composition and derived biotic metrics due to the effects of two subsampling methods described in the RIVPACS and the AQEM/STAR protocol.

6.1.2 Methods

The study focuses on multi habitat samples that have been taken according to the AQEM/STAR protocol. Each sample was treated in two ways: the RIVPACS subsampling technique (RIVPACS subsampling, counting and determination) and the AQEM/STAR subsampling technique (AQEM/STAR subsampling, counting and determination). After the RIVPACS procedure the original sample was restored by replacing the sorted animals and sample content. In a second step the restored sample was treated according to the AQEM/STAR procedure. Finally, the organisms in the rest of the sample were completely removed, counted and identified without any subsampling procedure.

For most groups the taxonomic resolution comprised the species level. Turbellaria, Oligochaeta, Hydrachnidia and Simuliidae were not further determined. Specimen belonging to Pisidium, Baetis, Ecdyonurus, Rhithrogena, Elmis, Esolus, Limnius and Hydraena were identified to the genus level.

The study is based on three (five) multi-habitat-samples from Austrian STAR rivers (Table 6.1).

Table 6.1. List of investigated STAR sites.

site	ecological status classification
Sarmingbach, Wolfsschlucht	1(reference)
Große Isper, Altenmarkt	2 (good)
Stullneggbach, Mainsdorf	3 (moderate)
Stullneggbach, Aichegg	2 (good)
Wildbach, Kramermirtl	1 (reference)

6.1.3 Results

A comparison of metrics' values obtained by RIVPACS subsampling, AQEM/STAR subsampling and totally sorted samples is given in table 6.2. The metrics are combined to metric type groups. With the exception of abundance measures which are based on five samples three samples have been totally sorted for evaluating the subsampling efficiency.

Table 6.2. Comparison of metric values obtained by RIVPACS subsampling, AQEM/STAR subsampling and totally sorted samples.

category	**RIVPACS**	**AQEM/ STAR**	**Total sample**
abundance measures [Ind./m²]			
abundance	1318	1853	1678
	4329	4910	5536
	5757	7054	7118
	1297	1456	1358
	1895	1978	1918
EPT-Abundance	486	590	530
	1093	1483	1334
	2660	3318	2976
	348	400	404
	630	554	584
Chiro.-Abundance	345	420	427
	2200	1838	2438
	759	811	1020
	758	741	678
	514	518	524
richness measures			
taxa richness (species)	72	73	90
	64	61	75
	71	65	82
taxa richness (genus)	54	54	64
	50	50	57
	53	50	59
taxa richness (family)	29	29	32
	30	31	32
	32	32	32
taxa richness (Chiro.)	33	35	48
	25	22	36
	27	21	37
taxa richness without Chiro. (species)	39	38	42
	39	39	39
	44	44	45
taxa richness (species) without post-sorting	72	67	90
	64	51	75

	71	62	82
taxa richness (genus) without post-sorting	54	48	64
	50	41	57
	53	47	59
taxa richness (family) without post-sorting	29	24	32
	30	24	32
	32	29	32
taxa richness (Chiro.-Taxa)	33	35	48
	25	22	36
	27	21	37
taxa richness (EPT-species)	22	20	24
	23	22	23
	24	24	24
taxa richness (EPT-genus)	19	18	21
	17	16	17
	18	18	18
taxa richness (EPT-families)	14	13	14
	15	15	15
	15	15	15
composition measures			
EPT-Taxa % (species)	30,56	27,4	26,67
	35,94	36,07	30,67
	33,8	36,92	29,27
index metrics			
Saprobic Index (Austria)	2,01	2,07	2,04
	2,14	2,14	2,13
	1,76	1,8	1,8
RETI	38,7	40,4	37,2
	35,4	33,6	35,5
	70,1	68,9	68,4
littoral [%]	7,8	8,8	8,2
	9,4	9,3	9,2
	6,7	7,3	7,1
Index of Biocoenotic Region	4,84	4,97	4,84
	4,79	4,81	4,79
	4,22	4,27	4,27
microlithal [%]	28,5	28,9	28,1
	19,3	21,3	21,6
	12,1	14,3	13,4
Shannon-Wiener-Index	2,32	2,44	2,47
	2,62	2,69	2,79
	3	3,03	3,08
Wilhm-Dorris-Index	3,35	3,52	3,56
	3,78	3,88	4,02
	4,33	4,37	4,45
evenness	0,54	0,57	0,54

	0,61	0,65	0,64
	0,7	0,73	0,68

6.1.4 Discussion

Table 6.3 compares the results of RIVPACS subsampling and AQEM/STAR subsampling. The metric values are expressed as mean percentage deviations of subsamples compared to the totally sorted sample. A special focus is given on the importance of post-sorting.

Table 6.3. Comparison of the mean deviation [%] with standard deviation (SD) of metrics from the subsamples with the metrics obtained from the totally sorted sample

category	**% deviation (underestimation) from total sample (mean of 3 MHS samples)**	
	RIVPACS	**AQEM/STAR**
Abundance*	-13,6% ±10 SD	+1,7% ±8,4 SD
Taxa richness (species)	-16,0% ±3,5 SD	-19,4% ±1,1 SD
Taxa richness (genus)	-12,7% ±2,8 SD	-14,4% ±1,8 SD
Taxa richness (family)	-5,2% ±4,8 SD	-4,2% ±4,8 SD
Taxa richness (Chironomids)	-29,6% ±2,3 SD	-36,4% ±8,4 SD
Taxa richness without Chiro. (species)	-3,1% ±3,7 SD	-3,9% ± 5,0 SD
Taxa richness (species) without post-sorting	-	-27,3% ±4,1 SD
Taxa richness (genus) without post-sorting	-	-24,5% ±3,9 SD
Taxa richness (family) without post-sorting	-	-19,8% ±9 SD
Richness and composition measures**	-8,1% ±12,3 SD	-5,8% ±14,6 SD
Index Metrics	-2,4% ±3,9 SD	-1,3% ±3,7 SD

* mean of 5 samples

** including total-, EPT- and chironomid abundance, No. of total-, EPT- and chironomid species, genus and family and EPT-Taxa %;

The findings can be summarized as follows:

- Subsampling at a high taxonomic resolution (mostly species level) remarkably underestimates the total number of taxa that can be found in the totally sorted sample.
- At species and genus level the RIVPACS subsampling method provides slightly better results.
- At family level the AQEM/STAR subsampling method provides slightly better results.

- Post-sorting clearly reduces the error of missing taxa of the AQEM/STAR subsampling method.
- The error (percentage of not detected taxa) increases with higher taxonomic resolution.
- The underestimation of metrics and indices values is lesser than the underestimation of the total taxa richness measures.
- The errors of "single" metrics' values (based on taxa richness, individuals and composition) are higher than the errors of "combined" metrics based on indices and scores.
- There is only a negligible difference between metrics values based on indices and scores comparing subsamples and totally sorted samples.
- The magnitude of the error seems to be attributed to Chironomid taxa. Considering Chronomids as one taxon clearly reduces the underestimation of taxa to 3 to 4% of total taxa found in fully treated MHS samples.

6.2 Subsample size (Austrian streams)

6.2.1 Introduction

The AQEM/STAR sampling design is based on the multi-habitat-sampling approach. A multi habitat sample (MHS) consists of 20 sampling units taken by a hand net with 500 µm mesh size and a frame width of 25x25 cm. The total sample covers an area of 1.25m^2 of the stream bottom and provides a large amount of material and organisms. If the sample is too large to sort 100%, subsampling techniques are used to reduce time for sorting and analysis.

The AQEM/STAR subsampling-method adapts the approach described by Caton (1991). The sample is thoroughly mixed and the entire sample is distributed evenly over the bottom of a shallow, white subsampling-tray which is divided into 30 identical cells. The subsampling effort is defined by combining a spatial and a fix-count approach: the subsample has to be represented by at least 5 cells and 700 individuals. This means, that if less than 700 specimen are present in the 5 cells, one additional cell after the other is sorted until the number of 700 is reached. The according cell is completely sorted as well.

The aim of this study is to get an opinion of the sub-sampling accuracy for taxa richness associated with the STAR/AQEM sorting protocol.

6.2.2 Methods

To estimate the efficiency of the "5 cells/700 specimen"-rule the multi-habitat-samples of five Austrian STAR sites were sub-sampled with the double effort. The number of taxa of the first 5/700 subsample is compared with the number of taxa of the first and a second set of a 5/700 subsample (together at least 10 cells). For the

above described comparison only "replicate" samples were processed. As additional information the results of the regularly treated "main" samples (5 cells/700 specimen) were included in the analyses. All MHS-"main" and "replicate" samples where taken simultaneously within the same river section.

The investigated sections include five STAR sites. The pre-classified ecological status classes of the sites cover undisturbed reference sites, good sites and a site of moderate quality (Table 6.4).

Table 6.4. List of investigated STAR sites.

Site	Ecological Status Classification
Sarmingbach, Wolfsschlucht	1(reference)
Sarmingbach, Waldhausen	3 (moderate)
Große Isper, Altenmarkt	2 (good)
Stullneggbach, Aichegg	2 (good)
Wildbach, Kramermirtl	1 (reference)

The term taxa refers to a set of 39 higher taxonomic units given in table 6.5.

Table 6.5. List of higher taxonomic units (HTU).

SUPERIOR TAXA	**HTU**	**SUPERIOR TAXA**	**HTU**
Turbellaria	Turbellaria	Ephemeroptera	Baetidae Heptageniidae Caenidae Leptophlebiidae Ephemerellidae Ephemeridae other Ephemeroptera
Nematoda	Nematoda	Odonata	Odonata
Mollusca	Gastropoda Bivalvia	Plecoptera	Predaceous non-predaceous
Oligochaeta	Oligochaeta	Heteroptera	Heteroptera
Polychaeta	Polychaeta	Megaloptera	Megaloptera
Hirudinea	Hirudinea	Coleoptera	Predaceous Non-predaceous Adult predaceous Adult non-predaceous
Hydrachnidia	Hydrachnidia	Trichoptera	Hydropsychidae Polycentropodidae Rhyacophilidae Psychomyidae Philopotamidae Glossosomatidae Goeridae Puppae Other Trichoptera
Isopoda	Isopoda	Diptera	Chironomidae Simuliidae Blephariceridae Other Diptera
Amphipoda	Amphipoda	17 superior taxa	39 HTU

6.2.3 Results

The following tables compare the number of individuals and the number of higher taxonomic units (HTU) in the "main" (subsample: 5 cells/700 specimen) and the first and second set of the "replicate" sample (each subsample: 5 cells/700 specimen).

Table 6.6. Sarmingbach, Wolfsschlucht - number of individuals and number of HTUs in the "main" and the "replicate" sample.

	Main	Replicate 1. Set	Replicate 2. Set
Individuals	703	1511	1499
Number of sorted cells	7	5	5
Average number of individuals per cell	100,42	302,2	299,8
Number of HTU	20	22	+2 (24)

Table 6.7. Sarmingbach, Waldhausen - number of individuals and number of HTU in the "main" and the "replicate" sample.

	Main	Replicate 1. Set	Replicate 2. Set
Individuals	1461	1107	845
Number of sorted cells	5	5	6
Average number of Individuals per cell	292,2	221,4	140,8
Number of HTU	26	24	+1 (25)

Table 6.8. Große Isper, Altenmarkt – number of individuals and number of HTU in the "main" and "replicate" sample.

	Main	Replicate 1. Set	Replicate 2. Set
Individuals	1351	1956	1864
Number of sorted cells	5	5	5
Average number of Individuals per cell	270,2	391,2	372,8
Number of HTU	21	20	+2(22)

Table 6.9. Stullneggbach, Aichegg - number of individuals and number of HTU in the "main" and "replicate" sample.

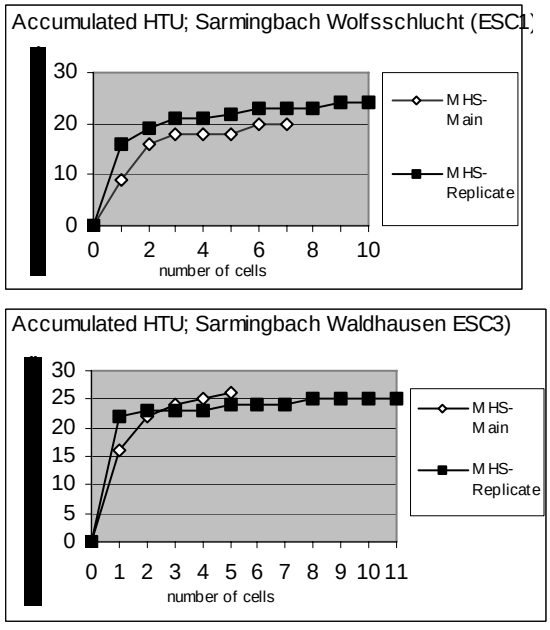
	Main	Replicate 1. Set	Replicate 2. Set
Individuals	1692	1213	870
Number of sorted cells	5	5	5
Average number of Individuals per cell	338,2	242,6	174
Number of HTU	21	22	+0 (22)

Table 6.10. Wildbach, Kramermirtl - number of individuals and number of HTU in the "main" and "replicate" sample.

	Main	Replicate 1. Set	Replicate 2. Set
Individuals	748	780	746
Number of sorted cells	7	7	6
Average number of Individuals per cell	106,8	111,4	124,3
Number of HTU	19	21	+1 (22)

6.2.4 Discussion

The increase of taxa after each newly sorted cell is visualized in figure 6.1. The diagrams cover five STAR sites and contain two kinds of information: the increase of taxa in the "main" samples (5/700 rule) and the increase of taxa in the "replicate" samples (at least 10 cells [2 times 5/700 rule]).



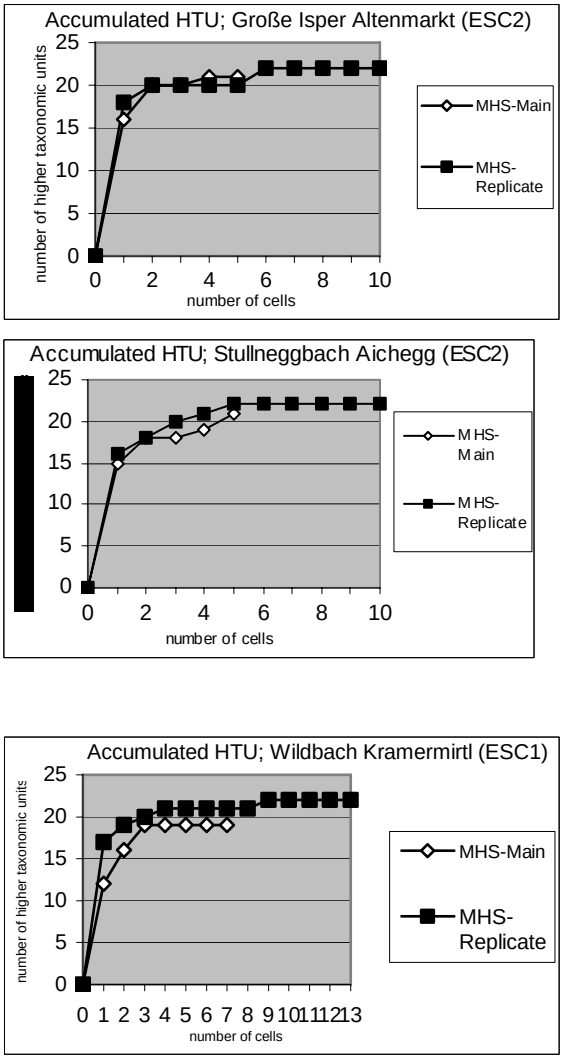


Figure 6.1. Accumulated curves of taxa (higher taxonomic units) in "main" and "replicate" samples from five STAR sites.

Although the shape of the accumulated curves differ among sites the curve-progression of the "main" and "replicate" samples show a similar increase of taxa. The number of taxa increases remarkably within the first cells. After about five cells only a small enhancement of newly recorded taxa can be shown. The number of new taxa found after the fifth sorted cell ranges between 0 and 2. In each case a minimum of 91% of the total taxa of 10 (or more) cells can be found in the first five sorted cells (table 6.11). On average 95% of the taxa are recorded in the first five cells. No influence of the ecological status (ESC) of a site on the accuracy of the subsampling result could be detected.

Table 6.11. Percentage of recorded taxa found in the first five cells ("replicate" samples only).

River / Site	% Taxa
Sarmingbach Wolfsschlucht ; ESC 1	92
Sarmingbach Waldhausen; ESC 3	96
Große Isper Altenmarkt; ESC 2	91
Stullneggbach Aichegg; ESC 2	100
Wildbach Kramermirtl; ESC 1	96

6.2.5 Conclusion

The examples from Austrian STAR rivers demonstrate that subsampling five cells out of 30 seem to be a reasonable effort for estimating the richness of higher taxonomic units. All subsamples where a minimum of 10 cells have been sorted showed that the increase of the number of HTU after 5 sorted cells is less than 10% (with 5% underestimation on average). The number of missing HTU in the first five cells is 2 in the worst case.

6.3 Subsample size (Czeck streams)

6.3.1 Introduction

The subsampling method proposed for the reduction of size of AQEM-type multihabitat samples was tested in terms of biological parameters used for assessment of ecological status. The reduction of effort required for the sorting and identification in macroinvertebrate surveys is needed in order to set their effective design. The estimation of optimal quantity of biological material has to be based on analyses comparing various arrangements of sample partitioning.

Main goals of the case study:

- to test a methodological framework for evaluation of variablitity within samples
- to analyze the effect of sample size on the capture probability of individual taxa, metrics values and final classification
- to evaluate the potential effect of size-dependent variability ofmetrics on the classification

6.3.2 Methods

Samples

We selected 2 sites for the detailed evaluation of subsampling variability where high and good status were expected in pre-classification. Samples taken from these sites allowed comparison of communities of similar total number of individuals, but different in structure and taxa richness. They are summer samples from Luborca and

Nectava streams. While standard subsampling procedure separated only 5 subunits (or more until 700 individuals is reached) we separated and analyzed all 30 subunits of two samples.

Simulation method

Random sampling of subsamples was performed for all possible numbers of subsamples (from 1 to 30). The number of random samples for any given number of subsamples was determined by i) all possible combinations of subsamples (for example, for one subsample there are only 30 possible combinations, i.e. the number of subsamples) or ii) 1000 random samples. Each random sample consists of a given number of subsamples randomly generated from a database of subsamples without repetitions, i.e. every subsample may be in the given random sample only once. The procedure generated a huge set of random samples; a description of each random sample (number of individuals, number of species, indices of diversity and saprobity) was computed and processed in subsequent analysis.

Another type of random sampling used rule of minimum of 5 subunits and 700 and reaching at least 700 individuals. This analyses produced matrix of all metrics supplemented by variation measures which can be used in other workpackages.

6.3.3 Results

Taxa richness

Our results were in agreement with widely accepted principles that taxa richness is related to number of individuals (or more generaly to sample size) (Figure 6.2, 6.3 and 6.4). Although we obtained information about this relationship, their wider application is limited by the fact that community structure and abundance distribution are the most important factors creating specific conditions.

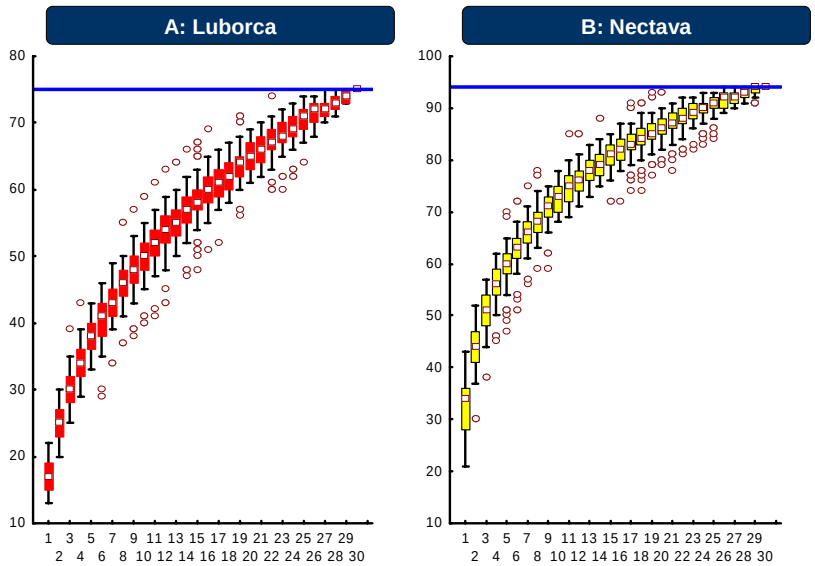


Figure 6.2.Relationship between taxa richness and number of subunits – blue horizontal line indicates total number of taxa in whole sample.

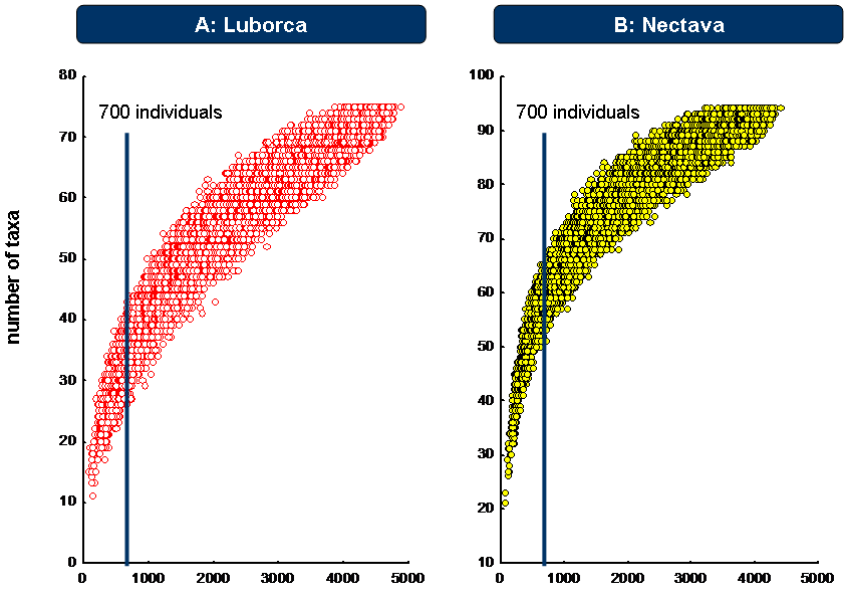


Figure 6.3. Relationship between taxa richness and number of individuals – vertical line indicates 700 individuals.

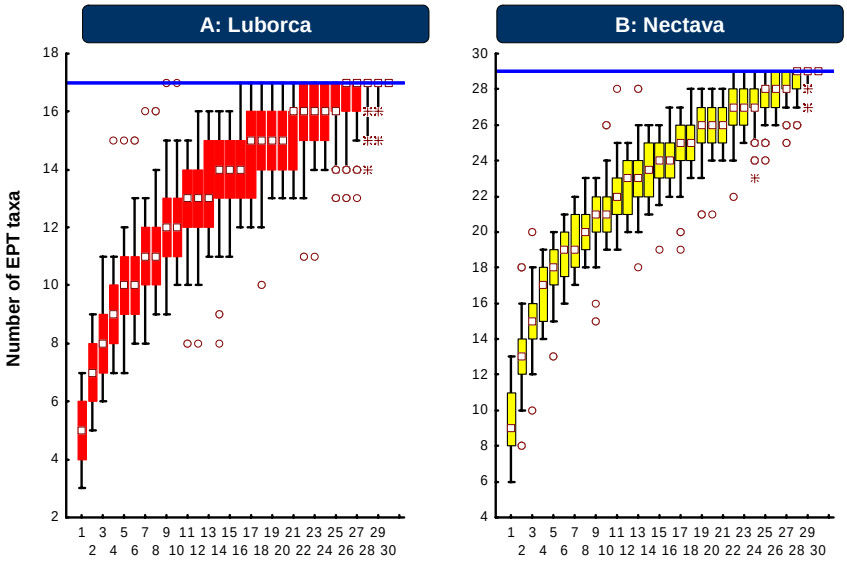


Figure 6.4. Relationship between number of EPT taxa and number of subunits – blue line indicates total number in entire sample.

Relationship between total abundance of individual taxa in entire sample and number of subunits where taxa occured followed similar pattern in both samples (Figure 6.5).

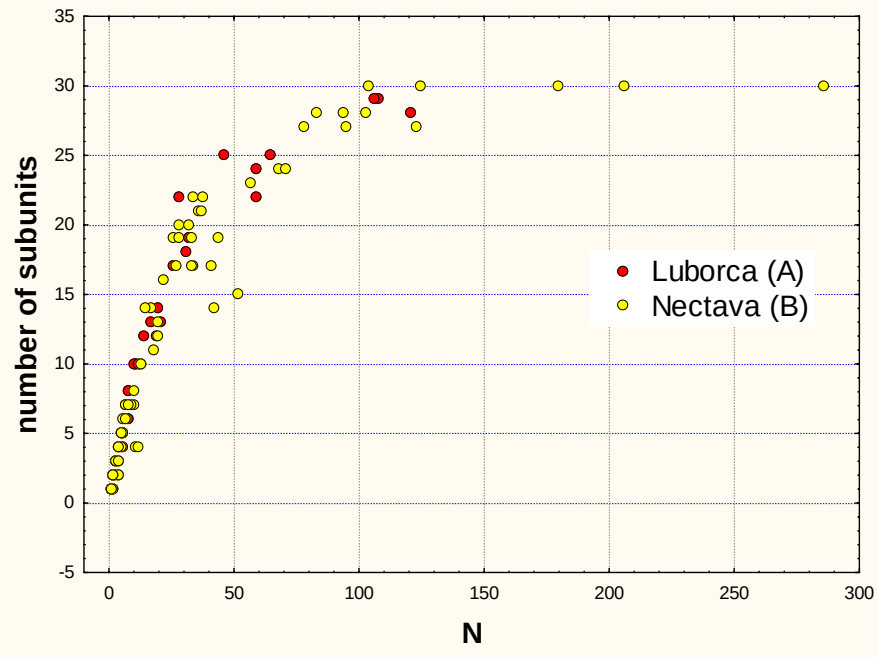


Figure 6.5. Number of subunits where individual taxa (cirles) were recorded related to their total abundance in entire sample (N). Gammarus fossarum present in all 30 subunits and reaching high number of individuals was omitted.

Saprobic index

Saprobic index represents metrics based on the sensitivity of individual taxa to pollution by organic matter. Its values fitted within metric-specific class boundaries of both sites (Luborca – high, Nectava – good) even for the smallest sample size. Medians were very near to the value of the complete sample and the variability exceeded class boundaries in several 1-subunit subsamples (Figure 6.6). The general tendency following increasing sample size is a stable median value and decreasing variability of Saprobic index.

Random simulation of subsamples based on combination of at least 5 subunits and at least 700 individuals resulted in the following mean and standard error values:
Luborca: 0.937 ± 0.001
Nectava: 1.318 ± 0.001

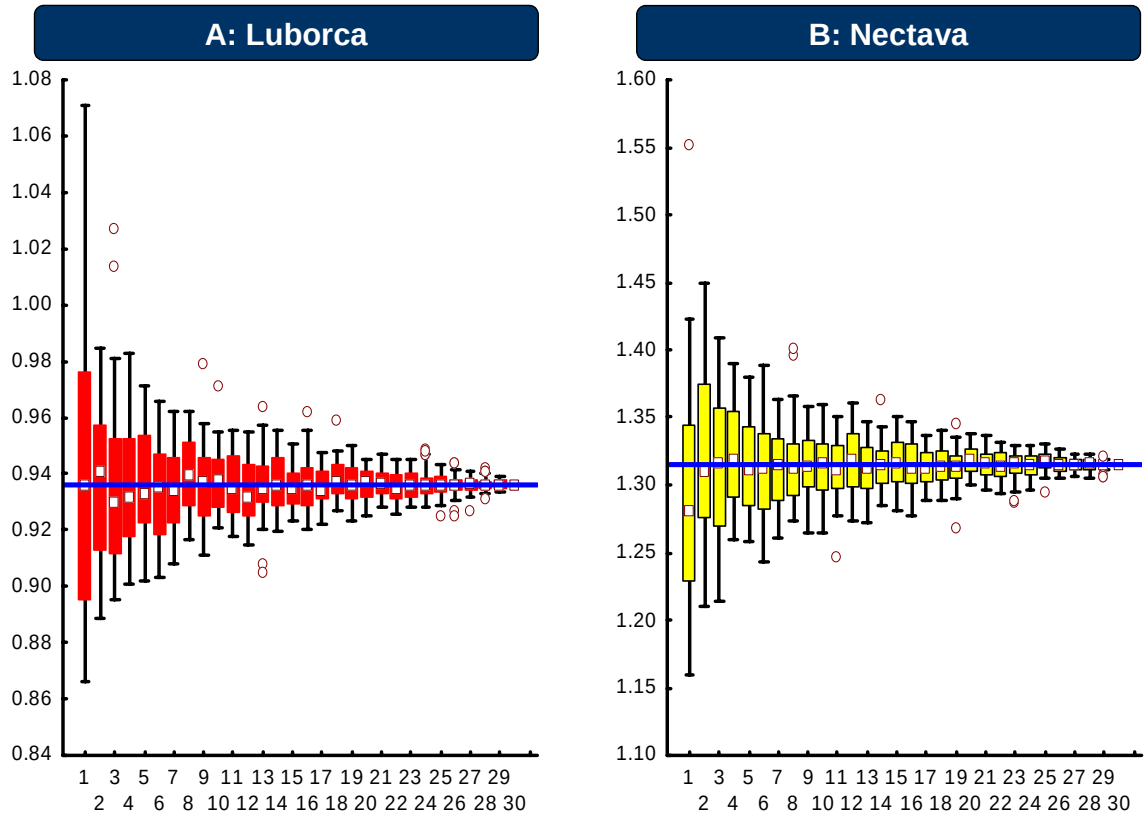


Figure 6.6. The relationship between Saprobic index (Czech version) and number of subunits (value of the whole sample is represented by blue line). The boundary between high and good status was set at the value of 1.0.

Potential effect of sample size to classification results can be demonstrated on share of Oligochaeta individuals incorporated into assessment system of organic pollution in small streams of type C04 (Figure 6.7). Class boundary between high and good status was set at the value 3.54 % so size-dependent variability doesn't effect classification. Oposite situation is valid for number of EPT (Ephemeroptera, Plecoptera and Trichoptera) taxa (Figure 6.7).

Results supplemented findings of sampling effectiveness research carried out in the Netherlands and in Slovakia. The knowledge about linkages between general biological descriptors of sample size (number of individuals, proportion of sample) and metrics used for the assessment increase comparability of samples differing in size.

Functionality of the subsampling gear was also verified in terms of uniform distribution of animals among subunits. Other results will be used for further outputs of the STAR project. Size-dependent variability of all metrics calculated by assessment software can support a decision as to which of the correlated metrics should be incorporated in the assessment system.

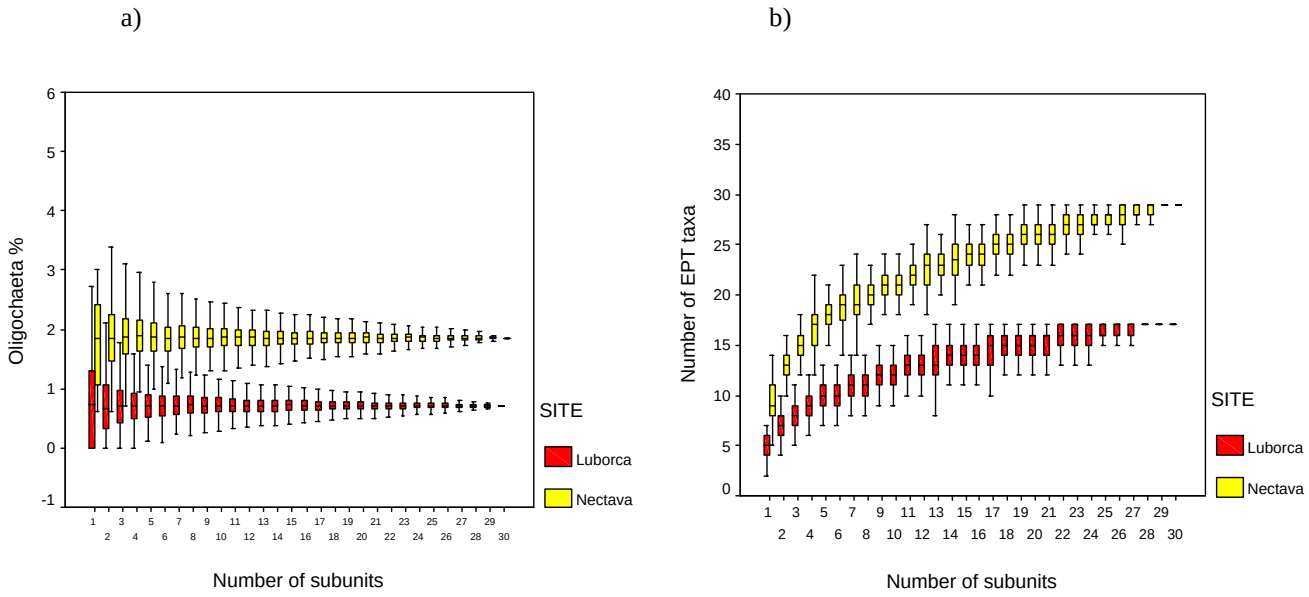


Figure 6.7. Variability of metrics used in newly developed assessment system and their relation to sample size.

Number of individuals

Based on data from real 30 subunits (without simulation) the mean number of individuals per subunit are:
Luborca: 163.2 ± 6.3
Nectava: 147.1 ± 7.8

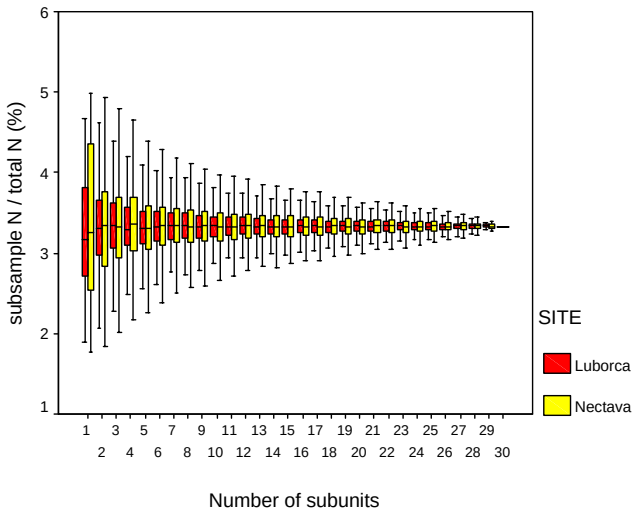
We evaluated relation between number of subunits and 700 individuals for both samples (randomization):

	LUBORCA				
	3	4	5	6	7
< 700	100.00%	90.60%	36.00%	2.10%	0.00%
>= 700	0.00%	9.40%	64.00%	97.90%	100.00%

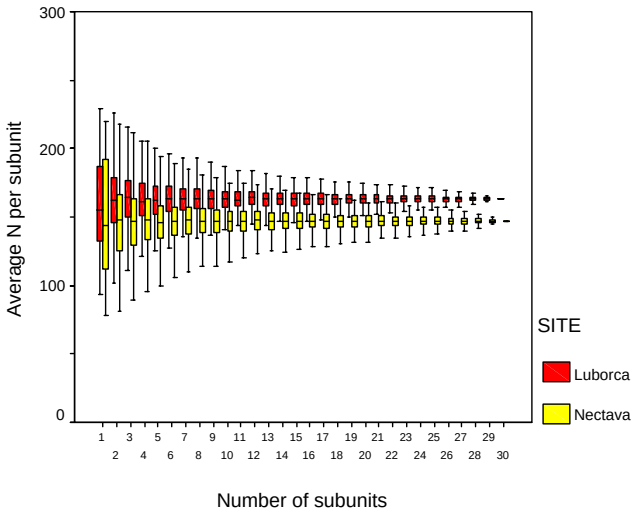
	NECTAVA			
	3	4	5	6
< 700	100.00%	76.00%	5.00%	0.00%
>= 700	0.00%	24.00%	95.00%	100.00%

Abundance distribution in subunits is expressed as the proportion to overall number of individuals (a) and also in absolute values (b). Both parameters are calculated per subunit.

a)



b)



6.3.4 Conclusion

Taxa richness metrics can be used in connection with precisely defined size of standardized sample or conversion formula would be developed. In contrast, another metrics are robust enough to be stable across range of sample size. We strongly recommend processing of more data covering more stream types, community types and different intensity of anthropogenic stressors. Class boundaries developed on the base of samples differing in subsampling procedure should be used with caution to obtained results (e.g. number of EPT taxa).

References

Alba-Tercedor, J. & A. Sanches-Ortega (1988) Un methoda rapido y simple para evaluar la calidadbiologica de las aguas vorrientes basado en el de Hellawell (1978). Limnetica, **4,** 51-56.

AQEM consortium (2002) Manual for the application of the AQEM system. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0, February 2002.

Armitage, P.D., A.M. MacHale, & D.C. Crisp (1974) A survey of stream invertebrates in the Cow Green basin (Upper Teesdale) before inundation. Freshwater Biology, **4**, 369-398.

Bailey, R.C., R.H. Norris and T.B. Reynoldson (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society , **20**(2):, 280-286.

Braukmannn, U. (200) Hydrochemische und biologische merkmale regionaler bachtypen in Baden-Württemberg. Oberirdische Gewässer, Gewässerökologie, 56, 1-501.

Caton, L.W. (1991) Improved sub-sampling methods for the EPA "Rapid Bioassessment" benthic protocols. Bulletin of the North American Benthological Society, **8**(3), 317-319

Carter, J.L. & V.H. Resh (2001) After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. Journal of the North American Benthological Society, **20**(4): 658-682.

Chessman, B.C. (1995) Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family-level identification, and a biotic index. Aus. J. Ecol., **20**, 122-129.

Chessman, B.C. & D.P. Robinson (1987) Some effects of the 1982-83 drought on water quality and macroinvertebrate fauna in the lower La Trobe River, Victoria. Aus. J. Mar. Freshwat. Res., **38**, 289-299.

Clarke, R.T., M.T. Furse, R.J.M. Gunn, J.M. Winder & J.F. Wright (2002) Sampling variation in macroinvertebrate data and implications for river quality indices. Freshwater Biology, **47**, 1735-1751.

Diamond, J.M., M.T. Barbour & J.B. Stribling (1996) Characterizing and comparing bioassessment methods and their results: A perspective. Journal of the North American Benthological Society, **15**(4), 713-727.

Egglishaw, H.J. (1964) The distributional relationship between bottom fauna and plant detritus in streams. Journal of Animal Ecology, **33**, 463-476.

Eleftheriou, A. & N.A. Holme (1984) Macrofauna techniques, p. 140-216. *In* Holme, N.A. & A.D. McIntyre (eds.) Methods for the study of marine benthos. IBP Handbook No.16. 2nd ed. Blackwell Scientific Publications, Oxford.

Ferraro, S.P., F.A. Cole, W.A. DeBen & R.C. Swartz (1989) Power-cost efficiency of eight macrobenthic sampling schemes in Pudget Sound, Washington, USA. Can. J. Fish. Aquat. Sci., **46**, 2157-2165.

Furse, M.T., J.F. Wright, P.D. Armitage & D. Moss (1981) An appraisal of pond-net samples for biological monitoring of lotic macro-invertebrates. Water Research, **15,** 679-689.

Green, R.H. (1979) Sampling design and statistical methods for environmental biologists. John Wiley and Sons, Inc., New York.

Growns, J.E., C.C. Chessman, J.E. Jackson & D.G. Ross (1997) Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. Journal of the North American Benthological Society, **16**(3), 682-693.

Hawkes, H.A., 1979. Invertebrates as indicators of river water quality. *In* James, A. & L. Evison (eds) Biological Indicators of Water Quality. John Wiley, Chichester.

Hellawell, J.M., 1986. Biological indicators of Freshwater Pollution and Environmental Management. Elsevier Applied Science, London.

Hering, D., O. Moog, L. Sandin & P.F.M. Verdonschot (2004) Overview and application of the AQEM assessment system. Hydrobiologia, **516**, 1-20.

Kerans, B.L., J.R. Karr & S.A. Ahlstedt (1992) Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. Journal of the North American Benthological Society, **11**(4), 377-390.

Lehmann, E.L. (1975) Nonparametrics: Statistical Methods Bases on Ranks. Holden-Day, San Fransisco.

Lewis, F.G. III & A.W. Stoner (1981) An examination of methods for sampling macrobenthos in seagrass meadows. Bull. Mar. Sci., **31**, 116-124.

Lorenz, A, L. Kirchner, D. Hering (2004) "Electronic subsampling" of macrobenthic samples: how many individuals are needed for a valid assessment result? Hydrobiologia, **516,** 299-312.

Macan, T.T. (1957) On the station species-curve. *In* Patil, G.P., E.C. Pielou & W.E. Waters (eds.) Statistical Ecology Vol. 3. Many species populations, ecosystems and systems analysis. The Pennsylvania State University Press, University Park, pp.233-235.

Macan, T.T. (1958) Methods of sampling the bottom fauna in stony streams. Mitteilungen der Internationalen Vereinigung fur theoretische und angewandte Limnologie, **8**, 1-21.

Macan , T.T. (1977) The fauna in the vegetation of a moorland fishpond as revealed by different methods of collecting. Hydrobiologia, **55**, 3-15.

Mackey, A.P., D.A. Cooling & A.D. Berrie (1984) An evaluation of sampling strategies for qualitative surveys of macro-invertebrates in rivers, using pond nets. Journal of Applied Ecology, **21**, 515-534.

McIntyre, A.D. (1961) Quantitative differences in the fauna of boreal mud associations. J. Mar. Biol. Assoc. U.K., **41**, 599-616.

Metzling, L., B. Chessman, R. Hardwick & V. Wong (2003) Rapid assessment of rivers using macroinvertebrates: the role of experience, and comparisons with quantative methods. Hydrobiologia, **510**, 39-52.

Morgan, N.C. & H.J. Egglishaw (1965) A survey of the bottom fauna of streams in the Scottish Highlands. Hydrobiologia, **25**, 181-211.

Norris, R.H. & A. Georges (1993) Analysis and interpretation of benthic macroinvertebrate surveys. *In* Rosenberg, D.M. & V.H. Resh (eds.). Freshwater biomonitoring and benthic acroinvertebrates. Chapman & Hall, New York and London, pp. 234-286.

Norris, R.H., E.P. McElravy & V. H. Resh (1993) The Sampling problem. *In* Calow, P. & G.E. Petts (eds.) The rivers handbook. Blackwell, Oxford, England, pp. 282-306.

Rawer-Joost, C. (2001) Eignung und Variabilität von Verfahren zur ökologischen Bewertung von Fließgewässern im Mittelgebirge auf der Basis autökologischer Kenngrössen des Makrozoobenthos. PhD-thesis University of Hohenheim.

Rees, H.L. (1984) A note on mesh size selection and sampling efficiency in benthic studies. Mar. Pollut. Bull., **15**, 225-229.

Reish, D.J. (1959) A discussion of the importance of the screen size in washing quantitative marine bottom samples. Ecology, **40**, 307-309.

STAR Consortium (2003) The AQEM/STAR, IBGN, PERLA, EBEOSWA and RIVPACS protocol to be applied in STAR. Unpublished report, available at http://www.eu-star.at/tops/top_protocols.htm.

Vinson, M.R. & C.P. Hawkins (1996) Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. Journal of the North American Benthological Society, **15**, 392-399.

Vlek, H.E., P.F.M. Verdonschot & R.C. Nijboer (2004) Towards a multimetric index for the assessment of Dutch streams using benthic macroinvertebrates. Hydrobiologia, **516**, 173-189.

Appendix 1 Seasonal variation in metric values for samples from the Heelsumse beek and Stupavsky potok

acronym	metric description	mean (stdev)	
		the Netherlands	Slovakia
ABUN	Abundance	3442.1 (1417)	4473.5 (2442.9)
NTAX	Number of taxa	49.6 (4.3)	91.2 (17.3)
NGEN	Number of genera	37.9 (2.6)	68 (11.4)
NFAM	Number of families	22.5 (2.2)	33.8 (3.2)
ZSI	Saprobic Index (Zelinka & Marvan)	2.2 (0.2)	1.7 (0.2)
ALPHA-MESO (%)	Proportion of individuals with a preference for alpha-meso saprobic circumstances	29.3 (9.4)	16.8 (2)
OLIGO (%)	Proportion of individuals with a preferences for oligo saprobic circumstances	17.2 (5)	38.2 (3.5)
XENO (%)	Proportion of individuals with a preferences for xeno saprobic circumstances	1.6 (1.7)	8.1 (1.1)
GFI D03	German Fauna Index D03	0.9 (0.6)	0.9 (0.2)
GFI D04	German Fauna Index D04	-1.3 (0.3)	1 (0.1)
BMWP	Biological Monitoring Working Party	69.6 (10.4)	168.5 (18.8)
ASPT	Average Score per Taxon	5 (0.3)	7.1 (0.2)
GSI	German Saprobic Index (new version)	2.1 (0.1)	1.5 (0.1)
CSI	Czeck Saprobic Index	2 (0.4)	1.2 (0.4)
MTS	Mayfly Total Score	2.8 (0.5)	25.9 (4.3)
HAI	Acid Index (Hendrikson & Medin)	6.3 (0.5)	9.7 (1.2)
LIFE		5.2 (0.2)	6.7 (0.1)
NSTA	Number of sensitive taxa (Austria)	0.3 (0.5)	13.8 (2.6)
DIM	Diversity Index (Margalef)	6 (0.4)	10.8 (1.6)
DIS	Diversity Index (Shannon & Wiener)	2.3 (0.2)	2.6 (0.4)
RP (%)	Proportion of individuals with a preference for zones with moderate to high current (rheophil)	39.4 (17.2)	73.2 (15.1)
PEL (%)	Proportion of individuals with a preference for muddy substrates	9.6 (4)	3.3 (3.3)
PSA (%)	Proportion of individuals with a preference for sandy substrates	16.8 (6.9)	21.1 (3.3)
AKA+LIT+PSA (%)	Proportion of individuals with preferences for gravel, littoral and sand	51.5 (7.3)	86.8 (3.9)
littoral (%)	Proportion of individuals with a preference for the littoral	8.5 (4.9)	6.2 (1.3)
metarhithral (%)	Proportion of individuals with a preference for the lower-trout region	25.2 (6.7)	24.8 (4.6)
hyporhithral (%)	Proportion of individuals with a preference for the greyling region	27.6 (7.3)	19.2 (1.3)
epirhithral (%)	Proportion of individuals with a preference for the upper-trout region	4.4 (3.2)	23.3 (1.5)
IBR	Index of Biocoenotic Region	4.6 (0.5)	4.2 (0.2)

acronym	metric description	mean (stdev)	
		the Netherlands	Slovakia
GAT/COL (%)	Proportion of gatherers to collectors (individuals)	22 (6.5)	24.3 (7.4)
SHRED (%)	Proportion of shredders (individuals)	32.2 (7.4)	31.2 (9.2)
PASF (%)	Proportion of passive filter feeders (individuals)	2.9 (4.3)	5.9 (3.7)
GRA+SCRA (%)	Proportion of grazers and scrapers (individuals)	8.2 (7.2)	26.2 (9.1)
RETI	Rhithron Feeding Type Index	0.5 (0.1)	0.6 (0.1)
EPT-taxa	Number of Ephemeroptera, Plecoptera and Trichoptera taxa	7.8 (2)	35.1 (7.8)
EPT/DIP-taxa	Proportion of EPT-taxa to Diptera taxa	0.4 (0.1)	0.9 (0.2)
OL+DIP-taxa (%)	Proportion of Oligochaeta and Diptera taxa	51.8 (6.4)	51.7 (4.5)
OL-taxa	Number of Oligochaeta taxa	7.1 (2.5)	5.8 (2.4)
TRIC (%)	Proportion of Trichoptera individuals	4.6 (3.2)	16 (8.7)
PLEC (%)	Proportion of Plecoptera individuals	7.2 (13.6)	4.5 (1.9)
EPT-taxa (%)	Proportion of EPT-taxa	15.8 (3.6)	38.5 (4.8)
OL	Number of Oligochaeta individuals	319.8 (349.3)	183.9 (288.7)
CRUS-taxa	Number of Crustacea taxa	4.9 (0.3)	1 (0)
CRUS	Number of Crustacea individuals	1751.8 (960.4)	1651.6 (872.8)
EPHE-taxa	Number of Ephemeroptera taxa	0.5 (0.7)	9.8 (1.1)
EPHE	Number of Ephemeroptera individuals	5.4 (10.8)	561.2 (313.3)
TUR-taxa	Number of Turbellaria taxa	0.8 (1.2)	0.9 (0.3)
TUR	Number of Turbellaria individuals	3.8 (9.1)	16.6 (14.2)
TRIC-taxa	Number of Trichoptera taxa	4.5 (1.6)	18.1 (3.7)
TRIC	Number of Trichoptera individuals	152.8 (115.6)	676.8 (534.8)
COL-taxa	Number of Coleoptera taxa	0.7 (0.9)	6.8 (1.9)
COL	Number of Coleoptera individuals	1.6 (2.6)	210.5 (120.9)
DIP-taxa	Number of Diptera taxa	18.7 (3.1)	41.4 (8.9)
DIP	Number of Diptera individuals	528.2 (423.8)	1001.6 (964.3)
HYD-taxa	Number of Hydrachnidia taxa	4.7 (1.2)	-
HYD	Number of Hydrachnidia individuals	262.7 (292.6)	-
GAS-taxa	Number of Gastropoda taxa	0.3 (0.5)	-
GAS	Number of Gastropoda individuals	0.9 (2.1)	-
HIRU-taxa	Number of Hirudinea taxa	1.6 (1)	-
HIRU	Number of Hirudinea individuals	4.7 (5.3)	-
PLEC-taxa	Number of Plecoptera taxa	2.8 (1.1)	7.2 (3.5)
PLEC	Number of Plecoptera individuals	344 (697.5)	171.1 (65.5)