Detection of Human-Induced Stress in Streams

Comparison of Bioassessment Approaches using Macroinvertebrates

Joakim Dahl
Department of Environmental Assessment
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2004

Acta Universitatis Agriculturae Sueciae
Silvestria 332

ISSN 1401-6230
ISBN 91-576-6716-0

Tryck: SLU Service/Repro, Uppsala 2004

Abstract

Dahl, J. 2004. *Detection of human-induced stress in streams. Comparison of bioassessment approaches using macroinvertebrates.* Doctoral dissertation.
ISSN 1401-6230, ISBN 91-576-6716-0.

Running water habitats are among the most precious, yet most threatened, ecosystems on earth. Hence, there is a need for reliable methods for detecting the effects of pollution on this valuable commodity. This thesis examined a number of different approaches commonly used in bioassessment of stream ecosystems for their ability to detect ecological change. In particular, focus was on assessing the utility of single metric, multimetric and multivariate approaches using macroinvertebrate communities, with organic pollution and acidity as stress gradients. For detecting the effects of organic pollution, two metrics (the DJ index and the CA scores) representing two different approaches (multimetric and the multivariate) were found to be reliable tools for detecting the effects of stress in streams of southern Sweden. These methods were sensitive (high coefficients of variation) and had high precision (low error) to the stress gradient. Notwithstanding, single metric approaches might also be used when the knowledge of multimetric or multivariate methods is not available. The saprobic index was one such metric that showed promise as a tool for detecting human-induced change in streams. However, many of the single metrics tested had low precision (high error) and hence should be used with caution. Regarding the effects of acid stress, one index (out of ten) was highly correlated between spring and autumn, and was thus deemed to be a reliable metric for use with autumn samples to assess even spring conditions. This multimetric index was developed for use in southern Sweden and is included in Swedish EPAs Environmental Quality Criteria. A 'new' weighted averaging method was examined and found to be useful for assessing the acidity of streams. However, more research is needed to develop the method as well as a user-friendly platform. Both multimetric and multivariate (the weighted averaging model) methods were shown to be reliable tools for assessing the effects of acid stress on stream macroinvertebrate assemblages. Moreover, the results of this thesis support the use of multimetric and multivariate approaches in bioassessment. If possible, a combination of these methods is recommended in assessing the effects of ecological change. Single metric indices could be used, however a deeper ecological understanding is needed to evaluate the inherent errors associated with these metrics. For the future, user-friendlier methods should be developed so that multimetric and multivariate approaches can be more widely used by water managers.

Keywords: benthic macroinvertebrates, streams, single metric, multimetric, multivariate, organic pollution, acid stress

Author's address: Joakim Dahl, Department of Environmental Assessment, SLU, Box 7050, SE-750 07 Uppsala, Sweden. E-mail: Joakim.Dahl@ma.slu.se

'In an age when man has forgotten his origins and is blind even to his most essential needs for survival, water along with other resources has become the victim of his indifference'

- Rachel Carson

Contents

Appendix

Papers I – V

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. Dahl, J., Johnson, R.K. & Sandin, L. 2004. Detection of organic pollution in southern Sweden using benthic macroinvertebrates. *Hydrobiologia 516*, 161-172.

II. Dahl, J. & Johnson, R.K. 2004. A multimetric macroinvertebrate index for detecting organic pollution of streams in southern Sweden. *Archives für Hydrobiologie 160 (4)*, 487-513.

III. Dahl, J. & Johnson, R.K. Detection of ecological change in stream macroinvertebrate assemblages using single metric, multimetric or multivariate approaches. (Manuscript).

IV. Sandin, L., Dahl, J. & Johnson, R.K. 2004. Assessing acid stress in Swedish boreal and alpine streams using benthic macroinvertebrates. *Hydrobiologia 516*, 129-148.

V. Dahl, J. & Johnson, R.K. Assessing the acidity of Swedish streams using benthic macroinvertebrates and weighted averaging (WA) regression and calibration. (Manuscript).

Introduction

Background

Human-generated pressures act singly or in concert in deleteriously affecting the biological integrity of freshwater ecosystems. Running water habitats, in particular, is one of the most important resources on earth, but also among the most threatened (*e.g.* Malmqvist & Rundle, 2002). Across Europe humans have altered the integrity of streams and rivers through changes in hydromorphology (*e.g.* channelisation), acidification, and inputs of nutrients, organic matter, and toxic compounds. To address the effects of these and other human-induced pressures on the ecological integrity of running waters there has been a strong impetus to improve monitoring programmes and assessment techniques. Indeed, in Europe the area of freshwater science is experiencing a rapid expansion since the recent ratification of the European Water Framework Directive (European Commission, 2000). In contrast to earlier legislation pertaining to aquatic ecosystems, the European Water Framework Directive is probably the most significant piece of ordinance to be assembled in the interests of preserving and restoring the biodiversity of inland waters, wetlands and coastal areas. Moreover, whilst previous statutes focused on curbing emissions and monitoring using chemical indicators, the Water Framework Directive focuses on catchment planning and management, viewing aquatic ecosystems not as isolated entities but holistically as larger interconnected ecosystems. A key feature of the Directive is its focus on detecting ecological change (*i.e.* degradation and recovery) and determining what human-generated pressures (or stressors) are acting as drivers of change using a suite of biotic indicators such as fish, benthic invertebrates, and aquatic flora. These new guidelines in water quality assessment have generated a strong demand for the development of new, or adaptation of existing, assessment systems.

In the last few decades, focus in monitoring programmes has shifted from the sole use of chemical metrics to the combined use of both biological and chemical metrics. Ideally, selection of the most 'appropriate' metric should include consideration of the goals of the monitoring programme as well as knowledge of the levels of uncertainty associated with the metric(s) selected. Biological metrics have gained in favour since they are often considered to integrate changes over relatively long time scales. For example, although a chemical metric will provide information on the quality of the water at the moment the sample is taken, a biological metric may incorporate a quality measure during the life span of the organism selected (*e.g.* from weeks to years). Another argument for the use of organisms in biomonitoring is that species generally occur in a characteristic and limited range of habitats within their geographic range, and tend to be most abundant around their particular environmental optimum (ter Braak & Prentice, 1988). Using this relatively well-established axiom, a number of biotic metrics have been constructed in the last two decades to evaluate the structural and functional integrity of surface waters using macroinvertebrates (Johnson, 1995).

The single most popular biological approach for assessing streams and rivers is the use of benthic macroinvertebrates as indicators of pollution (*e.g.* Knoben *et al.,* 1995). Benthic macroinvertebrates refers to organisms that inhabit the bottom substrates of freshwater habitats, for at least part of their life cycle (Rosenberg & Resh, 1993), and are those retained by mesh sizes ≥ 200 to 500 µm (Slack *et al.,* 1973; Weber, 1973; Wiederholm, 1980; Suess, 1982). The idea that benthic invertebrates reflect the quality of their environment is not new. The Greek philosopher Aristotle taught his students, for example, that worms were born of slime (Time Magazine; 11-Oct-1993, as cited in Johnson, 1995). Furthermore, as early as the late 19th century benthic invertebrates have been used to indicate pollution of freshwater ecosystems (Kolkwitz & Marsson, 1902). There are many reasons for the widespread proliferation of the use of benthic invertebrates in biomonitoring. Metcalfe (1989) summarised five reasons for why macroinvertebrates are commonly selected: (1) they are differentially sensitive to pollutants of various types, and react to them quickly; *i.e.* macroinvertebrate communities are capable of a graded response to a broad spectrum of stressors; (2) they are ubiquitous, abundant and relatively easy to collect and their identification and enumeration is not as tedious and difficult as that for microorganisms; (3) they are relatively sedentary, and are therefore representative of local conditions; (4) they often have life spans long enough to provide a record of environmental quality; and (5) they are a very heterogeneous group, consisting of representatives of several phyla. Given these many positive attributes there is a high probability that human-induced change will result also in a change in benthic community composition. It is therefore not surprising that the use of benthic macroinvertebrates is widespread and constitutes the basis for most aquatic biomonitoring programmes currently in use (e.g. Whitton, 1979; Wiederholm, 1980; Sladecek *et al.,* 1982; Metcalfe, 1989; Rosenberg & Resh, 1993).

Different assessment approaches

More knowledge is needed on how organisms respond to human-generated as well as natural environmental changes to better understand how to protect and restore the structure and function of streams and rivers. The most common bioassessment methods range from relatively simple algorithms or biotic indices, to combinations of multiple indices (a.k.a. multimetric approaches), or relatively complex, multivariate approaches for pattern recognition and prediction. All three approaches are commonly used in monitoring and assessment studies to detect ecological change (Johnson *et al.,* 1993), but only a few studies have been conducted that compare the performance of different assessment methods (*e.g.* Reynoldson *et al.,* 1997; Hawkins, submitted).

As mentioned, Europe has a long history of using benthic macroinvertebrates in bioassessment programmes. The Saprobien system, which focuses on organic pollution and the associated decrease in dissolved oxygen, was first developed by Kolkwitz & Marsson (1902). Since its introduction, this approach has undergone several, often simplifying, modifications and is now used in a number of Eastern European countries as well as in Germany (*e.g.* Friedrich, 1990) and in Austria (*e.g.* Zelinka & Marvan, 1961; Moog, 1995). Arguing that the Saprobic index was taxonomically too demanding, researchers in the UK developed more simplified biotic metrics, such as the Trent Biotic Index (TBI) (Woodiwiss, 1964), which rely on family- and/or genus-level resolution. Metcalfe (1989) showed that many of the biotic metrics commonly used in European biomonitoring programmes stem from these two (*i.e.* Saprobic and TBI) slightly different approaches. For example, recognising that one of the shortcomings of the use of TBI is that macroinvertebrate abundance is ignored, several modifications of this index (*e.g.* the Chandler Biotic Score developed for use in Scotland (Chandler, 1970), the Belgian Biotic Index (BBI) (De Pauw & Vanhooren, 1983) and the Extended Biotic Index (EBI or IBE) developed for use in Italy (Ghetti, 1997) now include the relative abundance of indicator taxa. In Denmark, a modified TBI, with a diversity measure included, is commonly used (*i.e.* the Danish Stream Fauna Index, DSFI) (Skriver *et al.,* 2000). In the UK, the TBI was later modified to the Biological Monitoring Working Party (BMWP) and the Average Score Per Taxon (ASPT) (Armitage *et al.,* 1983) for use in national river surveys (*e.g.* ISO, 1979). Even modifications of these latter two biotic indices are used in a number of European countries (*e.g.* a Spanish BMWP has been developed for use in Spain (Alba-Tercedor & Sanchez-Ortega, 1988)).

Recently there has been a tendency towards developing more complex bioassessment methods in Europe through the amalgamation of the information from single metrics into a multimetric score (*e.g.* Hering *et al.,* 2004) and prediction (Wright 1995). The multimetric index concept was first developed by Karr (1981) as an Index of Biotic Integrity (IBI) to assess stream quality using fish assemblages. Since Karr's (1981) paper, multimetric indices have been developed for a number of purposes and using a number of organism groups (*e.g.* fish, macroinvertebrates and periphyton), and in the last decade the development and use of multimetric approaches has increased markedly (Resh *et al.,* 2000). Although the use of the multimetric approach in biological monitoring is widespread in the USA, in Europe single metrics are still more commonly used (Metcalfe, 1989; De Pauw *et al.,* 1992; Knoben *et al.,* 1995). However, as the need for reliable methods for detecting the effects of pollution increases (*e.g.* as problems become more widespread and complex), more complex assessment techniques may be warranted. Barbour *et al.*, (1999) suggest that combinations of multiple measures should minimise the weaknesses of individual metrics resulting in more robust metrics. Recently, a EU-funded project proposed a number of multimetrics for use across Europe (AQEM-consortium, 2002).

In addition to the use of single and multimetric approaches, more complex, multivariate approaches are increasingly being used in bioassessment. Reynoldson *et al.* (1995) proposed simply the use of ordination of community composition in bioassessment, where sites lying outside a pre-selected range of reference sites (*e.g.* the 95% confidence ellipse) are deemed to differ in composition. Prediction of taxon occurrence is also increasing in bioassessment, and a number of different approaches are currently used for predicting the taxa occurrence: generalised linear models (*e.g.* Nicholls, 1989), logistic regression (Agresti, 1990), Bayesian models (*e.g.* Brzeziecki *et al.,* 1995), partial least squares regression (Wold, 1982), β-functions (Austin *et al.,* 1994), and taxon-specific models (Bio *et al.,* 1998). The use of the approach first developed in the UK (RIVPACS or River InVertebrate Prediction And Classification System; Wright *et al.,* 1984) is commonly used for predicting community composition of macroinvertebrate communities in streams (*e.g.* Wright, 1995; Simpson & Norris, 2000; Hawkins *et al.,* 2000) and lakes (Johnson, 2003).

Effects of organic pollution

Eutrophication and organic pollution of freshwater ecosystems is presently considered a pan European problem ‘of major concern’ (Stanner & Bordeau, 1995). Both of these impacts may, through nutrient enrichment and siltation, cause habitat degradation and subsequently loss of biodiversity. The major source of pollutant discharge to rivers is organic matter derived from diverse human activities. The decomposition and breakdown of the organic matter is mediated by microorganisms and takes place mainly at the surface of the sediment and vegetation in smaller rivers, and in the water column in larger rivers. As the process requires the consumption of oxygen, severe organic pollution may lead to rapid deoxygenation of the river water and hence to the disappearance of fish and aquatic invertebrates. The habitat then becomes uniform, consisting of only few robust species that are able to tolerate the low oxygen concentration. The effects of both direct and indirect sources of organic pollution are hereafter referred to as ‘organic pollution’ in this thesis.

The response of macroinvertebrates to organic pollution is well documented (Hynes 1960; Hellawell, 1978; Hellawell, 1986; Mason, 1996), and many European countries have developed biotic metrics using macroinvertebrates for monitoring these effects. In Sweden, present-day bioassessment programmes commonly use single metrics such as taxon richness and diversity for determining ecological change due to human-generated pressures. Two metrics are presently recommended for detecting the effects of organic pollution in southern Sweden, namely the Danish Stream Fauna Index (Skriver *et al.,* 2000) and the Average Score Per Taxon (Armitage *et al.,* 1983) (Johnson, 1998). More recently, two multimetric indices have been developed and proposed for use in southern Sweden: the AQEM Type S05 index (AQEM-consortium, 2002) and the DJ index (Dahl & Johnson, 2004).

Effects of acid stress

The anthropogenic acidification of lakes and streams is one of the biggest environmental problems in Sweden today (Miljödepartementet, 1996). In the last two decades considerable focus has been placed on quantifying and mitigating (by liming) the effects of acidification on aquatic ecosystems in the southwest parts of the country. The potential problem of acidification in northern Sweden was not recognised until the beginning of the 1990´s, when Ahlström & Isaksson (1990) reported that large-scale biological damage, mainly in streams, was caused by acid deposition. Slow weathering of the bedrock increases the susceptibility of streams in northern Sweden to acid deposition. Much of the area has low buffering capacity, and consequently episodic surface water acidification, caused by elusion of SO^{2-}_4 and NO^-_3 from the snow pack at melting and the short residence time of the water in the soil (Bishop *et al.,* 2000) that occurs during spring flood events. Due to episodic nature of acidic rain- or melt-water to runoff, a large number of measurements at short time intervals are needed to chemically detect the acidic levels of streams.

A simple biological model using few measurements to infer acidity from biota might therefore be a more cost-effective approach than measures of water chemistry. Lowered pH and/or increased metal concentrations of stream water are particularly important factors associated with changes in benthic macroinvertebrate community structure of running waters (*e.g.* Townsend *et al.,* 1983; Raddum & Fjellheim, 1984; Herrman *et al.,* 1993; Larsen *et al.,* 1996). Hildrew *et al.,* (1984) found that the pool of suitable species was limited at more acidic sites compared to pH-circumneutral ones and that the available food resources were lower in acid streams. The rather straightforward relationship between acid conditions and the presence/absence of acid-sensitive/tolerant benthic macroinvertebrate species has therefore been used to assess the effects of acid stress on stream ecosystems (*e.g.* Henrikson & Medin, 1986; Raddum *et al.,* 1988; Bæken & Aanes, 1990; Degerman *et al.,* 1994).

Objectives

The overall aim of this thesis is to examine how different approaches commonly used in bioassessment of stream ecosystems differ in their ability to detect ecological change. In particular, the thesis focuses of the use of single indicator, multimetric and multivariate approaches and organic/nutrient enrichment and acidification stress gradients.

In paper **I** the objective is to determine which single metrics are the most appropriate for detecting the effects of organic pollution of streams in southern Sweden using benthic macroinvertebrates. A comparison of the 'best' single metrics with two multimetric indices is also made.

The objectives of paper **II** are to develop a multimetric index using benthic macroinvertebrates, and to test if this approach is more robust than currently used methods for detecting the effects of organic pollution on benthic communities.

An examination of which approach (single metric, multimetric, or multivariate) should be preferred in assessing the effects of organic pollution of streams is done in paper **III**.

Paper **IV** focuses on acidity, and the objectives of this paper are: (i) to evaluate the ability of selected indices to detect the effects of acid stress on stream ecosystems in northern Sweden, (ii) to compare benthic macroinvertebrate assemblage composition and index values between spring and autumn, and (iii) to test the typology (system A) as suggested by the EC Water Framework Directive for partitioning macroinvertebrate variance among stream types.

Finally, in paper **V** the objective is to explore if weighted averaging (WA) regression and calibration can better our understanding of the effects of acidification in Swedish stream ecosystems. Simply put, do WA methods perform better than commonly used acidity indices?

Material and methods

Four datasets were used in this thesis:

(1) Data from the AQEM-project with 15 sites in southern Sweden (paper **I**, **II** and **III**);
(2) Data from the AQEM-project with 60 sites in northern Sweden (paper **IV**);
(3) Data from the 1995 national stream survey with 281 sites in southern Sweden (paper **II**);
(4) Data from the 2000 national stream survey with 233 sites in northern Sweden, and 154 sites in southern Sweden (paper **V**).

Datasets 1 and 2

Dataset 1 was used for different purposes in paper **I**, **II**, and **III**, and dataset 2 was used for examining the strength of different acidity indices in paper **IV** (Fig. 1). Dataset 1 consists of samples collected from 15 stream sites, from one stream type in ecoregion 14 (Illies, 1978) in southern Sweden. Ecoregion 14 (hereafter referred to as the mixed forest ecoregion) consists of mixed forest and agricultural landscapes. Dataset 2 consists of samples collected from 60 stream sites from four stream types situated in ecoregions 20 and 22 (Illies, 1978) in northern Sweden (*i.e.* the arctic/alpine region and the northern boreal forest region). These stream sites were sampled within the EU-financed project AQEM (*The Development and Testing of an Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates*) (Hering *et al.,* 2004). Criteria as defined by the EC Water Framework Directive (European Commission, 2000) were used in a first categorisation of stream types. According to System A classification, sites were classified by: ecoregions according to Illies (1978), size classes based on catchment area, geology, and altitude classes.

The streams of dataset 1 all had siliceous geology, and altitudes ranged from 15 to 200 m above sea level (a.s.l.). A pre-classification of catchment areas ranging from 100 to 1000 km^2 was desired, but post-classification showed that stream sites were situated in catchments from 32 to 1005 km^2. All but one site were situated in catchments less than 500 km^2. Stream sites of dataset 2 also had siliceous geology, but altitudes ranged from 15 to more than 800 m a.s.l.

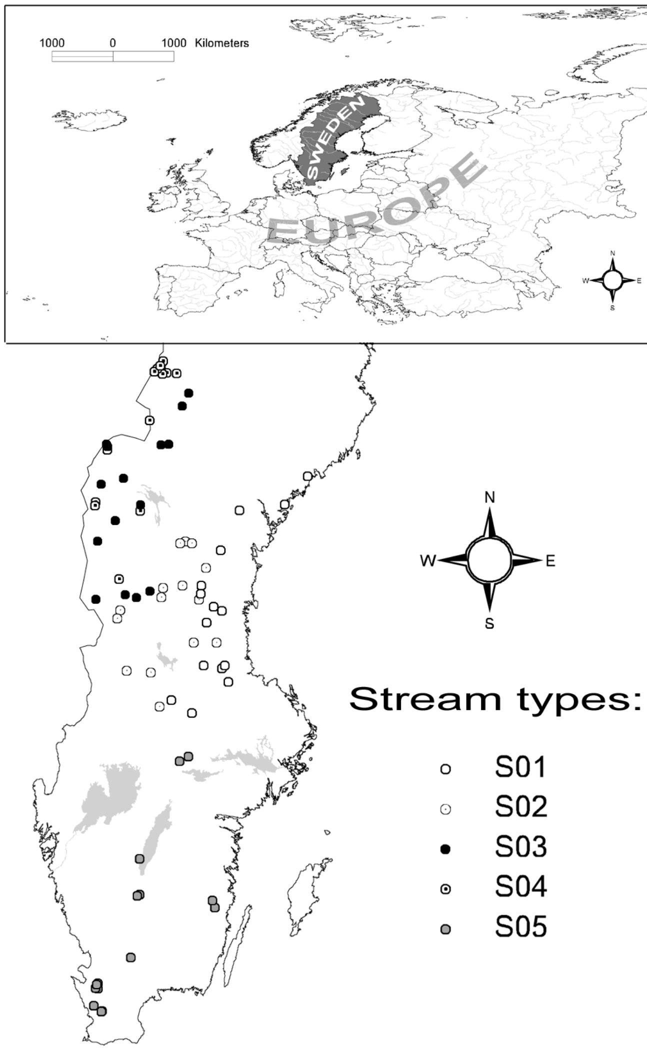

Fig. 1. Location of the 15 streams in southern Sweden (S05) used in dataset 1, and 60 streams in northern Sweden (S01-S04) used in dataset 2. Streams in both datasets were sampled in the spring and autumn of 2000.

The sampling sites of dataset 1 were selected *a priori* to represent an organic-pollution gradient within the stream type S05, whilst the sampling sites of dataset 2 were selected to represent an acidity gradient within the stream types S01, S02, S03, and S04. Fifteen streams were sampled within each stream type. Using biological and chemical data from earlier studies, 10 of the 15 sites were pre-classified into four ecological quality classes: good, moderate, poor, and bad. The five remaining sites were considered as having no or minimal human-generated impacts (Hering *et al.,* 2004) and hence were pre-classified as reference sites. The criteria used in the selection of these reference sites were partly taken from Hughes (1994) and Wiederholm & Johnson (1997), and more explicitly specified in Hering *et al.* (2004). Since these preliminary classifications were based on existing data of varying quality it was deemed necessary to reclassify the sites after sampling to achieve a proper organic pollution gradient. This reclassification was based on stream characteristics (*e.g.* hydromorphology) and water chemistry.

The streams were sampled twice in 2000; in spring (May-July) and in autumn (September-October). All sites were sampled for benthic macroinvertebrates using a multi-habitat sampling procedure (Hering *et al.,* 2004). In brief, the microhabitat composition (*i.e.* bottom substratum types) within a 25 to 50 m stream stretch was first estimated and microhabitats that constituted > 5% of the total area were sampled for benthic macroinvertebrates. Twenty replicate samples were distributed according to habitat type and coverage. For example, if the microhabitats encountered consisted of 25% psammal (*i.e.* sand), 40% mesolithal (*i.e.* pebbles and stones) and 35% macrolithal (*i.e.* coarse blocks), then sampling effort was allocated as 5, 8 and 7 sampling units, respectively. For each replicate an area of 25 x 25 cm was sampled by kick-sampling using a 500 μm handnet. All replicate samples of a site and sampling date were pooled into one sample. Each sample was preserved in 70% ethanol and later completely sorted in the lab. All macroinvertebrate samples were sorted and taxonomically identified according to quality control and assurance protocols given in Wilander *et al.,* (2003). Taxonomic identification was done to the lowest taxonomic unit possible, usually to species or species groups, with the exception of oligochaetes and chironomids. A thorough description of the taxa included and the errors involved in sample processing (sorting, counting and identification) is given in Wilander *et al.,* (1998). In addition to collecting macroinvertebrates, 0.5-l samples for water chemistry were taken and analysed according to international (ISO) or European (EN) standards when available (Wilander *et al.,* 1998). Water depth and average current velocity (at the level of '0.6 x water depth') were also measured for each site and sampling unit. When possible, 130 physico-chemical and other abiotic parameters were investigated for the sampling sites. Some data were taken from maps or already existing data sources, while other data were recorded directly in the field. Site-related information consisted of general data on the site and its catchment, data on stream morphology and hydrology, site description and human impacts on stream morphology. Sample-related information consisted of stream morphology and hydrology at the sampling-site, human impacts on stream morphology, hydrology and floodplain, pollution at sampling site, as well as microhabitat composition, hydraulics, and chemistry.

The major requirement for comparing 'good' samples with 'bad' samples in paper **III** is a dataset with a clear organic pollution gradient. Two groups were created from the dataset using only non-biological data; group A with low or no organic pollution and group B with high organic pollution. To refine the gradient and create two clearly separated groups some samples had to be excluded. First, the site-samples were divided into five chemical classes of total phosphorus (P) concentration according to a standardised procedure (SEPA, 1999). Class one and two were grouped together (group A) to represent sites with low total P concentration (< 25 µg l^{-1}) and class four and five were grouped together (group B) to represent sites with high total P concentration (> 50 µg l^{-1}). Samples in class three (25-50 µg l^{-1}) were excluded from the calculations to assure a clear difference in concentration between the two groups. A principal component analysis (PCA) was done, using parameters indicative of organic pollution, to extract a more accurate gradient of organic pollution. The parameters included in the PCA were Total-P, PO_4-P, NO_2-NO_3-N, NH_4-N, SO_4, BOD_5, conductivity, and percentage land use of the catchment area classified as cropland and pasture. The first axis explained 62% of the variance by the original nine variables and represented a gradient of organic pollution. Hence, the scores along this axis were used to represent a organic pollution gradient. Finally, these sample scores were regressed against the phosphorus grouping (group A and B), to determine if any site-sample from group A overlapped with site-samples from group B or vice versa. In this step one sample from group A was found to overlap with group B and was therefore excluded from further analyses. After this procedure a dataset of 21 samples remained, and was used for comparing the 'good' site-samples (group A; n = 9 samples) and 'bad' site-samples (group B; n = 12 samples) in paper **III**.

Dataset 3

Dataset 3, used in the calibration of the multimetric index constructed in paper **II**, was taken from the Swedish national stream survey performed in the autumn of 1995 (n = 694 streams) (Wilander *et al.,* 1998). A number of factors indicate that the 1995 stream survey is a robust dataset for establishing a multimetric index. The streams included in the database were randomly selected; hence this dataset represents an unbiased selection of the stream population. A number of considerations were also taken to reduce natural and/or operator-induced variability. Macroinvertebrates were sampled from mainly single habitats (riffle regions) in autumn to minimise spatial (habitat) and temporal (seasonal) effects. Moreover, to reduce sampling variability, samples were collected using standardised kick-sampling. Univariate and multivariate analyses of sampling methods have shown little differences among four of the Nordic countries (Sweden, Denmark, Finland and Norway) using standardised kick-sampling (Johnson *et al.,* 2001). Despite being a qualitative/semi-quanitative sampling procedure, a recent evaluation showed that even single kick-samples allowed the reliable detection of differences in assemblages between sampling sites in upland streams (Bradley & Ormerod, 2002). Five kick-samples were taken from each site (a 10 m longitudinal stretch x stream width) with a handnet (0.5 mm mesh size) according to the Swedish and European standards (SS-EN-27828, 1994). Each sample consisted of a 60 second disturbance of the bottom substratum along a one meter transect. The five samples were pooled in the field and stored in 70% ethanol until analysed. Similar to the samples of datasets 1 and 2, all macroinvertebrates were sorted and taxonomically identified according to quality control and assurance protocols. Taxonomic identification was done to the lowest taxonomic unit possible, usually to species or species groups, with the exception of oligochaetes and chironomids. A thorough description of the taxa included and the errors involved in sample processing (sorting, counting and identification) is given in Wilander *et al.,* (1998). In addition to collecting macroinvertebrates, samples for water chemistry were taken and analysed according to international (ISO) or European (EN) standards (Wilander *et al.,* 1998).

For calibration of the multimetric index in paper **II** only sites situated in southern Sweden (n = 356 streams) and deemed not to be affected by acidification and liming activities were used. Consequently, 75 acidified (exceedance of sulphur critical load > 0) or limed sites were removed from the dataset, resulting in a calibration dataset of 281 stream sites (Fig. 2a).

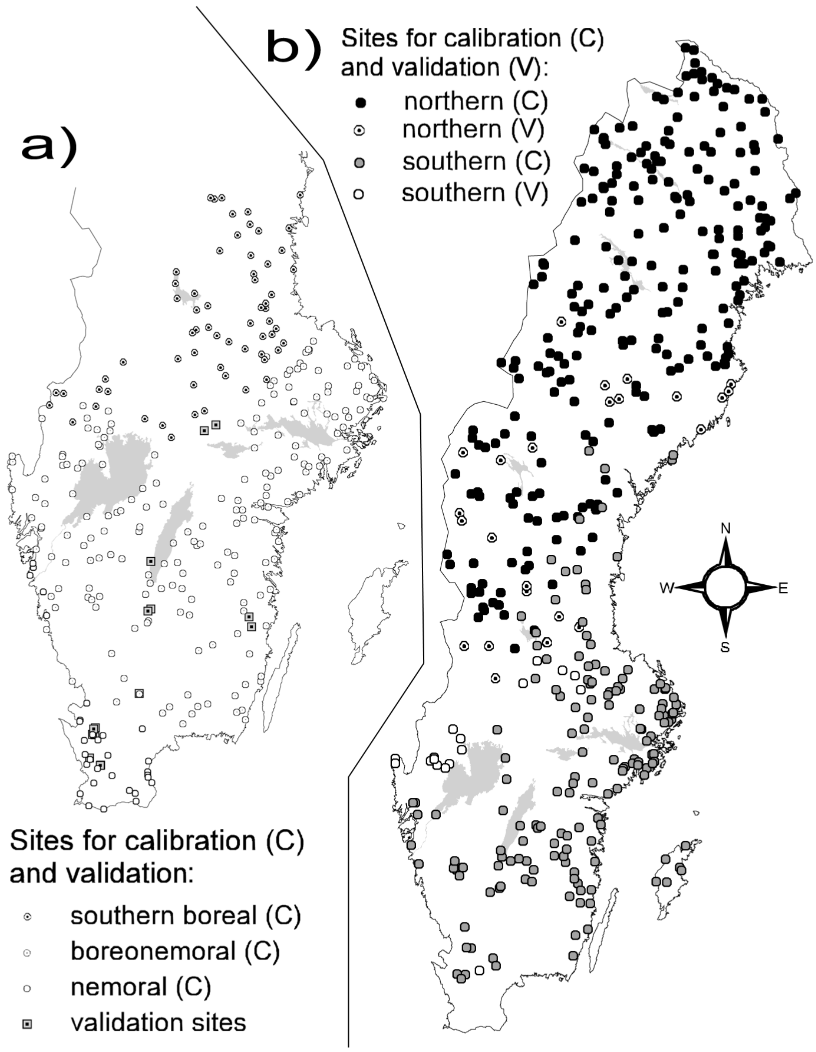

Fig. 2.a) Location of 281 streams in the southern boreal, boreonemoral, and nemoral ecoregions of Sweden sampled in the 1995 national stream survey and used in dataset 3. Validation sites used in paper II are also shown. *b)* Location of 258 streams in northern Sweden and 171 streams in southern sampled in the 2000 national stream survey and used in dataset 4 for calibration and validation.

Dataset 4

Dataset 4 was used for both training and testing of the weighted averaging (WA) methods in paper **V**. This dataset was taken from the Swedish national stream survey in 2000 (n = 708 streams) (Wilander *et al.,* 2003). As with the 1995 national stream survey, a number of factors indicate that the 2000 stream survey is a robust dataset for establishing a multimetric index. Similar to the national stream survey in 1995, the streams included in this database were randomly selected; hence this dataset represents an unbiased selection of the stream population. The sampling procedure, the sorting, the taxonomical identification, and other analyses are identical to those done for the national survey in 1995.

In order to fit the objectives of paper **V** some adjustments had to be done on this dataset. For the training and testing datasets, non-eutrophic (< 20% arable land in the catchment area), and non-limed streams were retained in the dataset. Due to effects of climatic differences between the northern and southern parts of the country on macroinvertebrate community composition (Sandin & Johnson, 2000), the sites were divided into two datasets. This resulted in two datasets of 258 stream sites in northern Sweden and 171 stream sites in southern Sweden (Fig. 2b). Twenty-five streams were randomly selected from the northern dataset and 17 from the southern as the validation datasets, which resulted in 233 stream sites in the northern and 154 stream sites in southern datasets to be used as training datasets. For the third modelling approach (see below) the pH-circumneutral sites (*i.e.* pH = 6.5-7.5) were removed and the remaining 53 sites in northern Sweden and 47 sites in southern Sweden were used for both calibration and validation (using jack-knifing).

Analyses

Single metric indices (paper I)

Eighty-four single metrics from six attribute groups were calculated from the compiled taxa data (Table 1): richness, composition, tolerance, feeding, habit, and habitat measures. Five steps were taken for determining which metrics were the most appropriate for detecting organic pollution of streams in southern Sweden using benthic macroinvertebrates: (step 1) one-way ANOVA of spring versus autumn data (both chemical and taxonomical) was done to determine if spring and autumn data could be considered as independent of season, and hence provide unique information; (step 2) principal component analysis (PCA) and clustering (TWINSPAN) of species data, were used to visually examine the possible differences in species composition between spring and autumn; (step 3) multiple stepwise regression of cluster groupings versus chemical, physical, substratum, and land use variables was used to compile the most important parameters for explaining the variation in species composition; (step 4) correlation of metrics versus variables indicating the effects of organic pollution (phosphorus concentration, conductivity, and percentage cropland) was done to select metrics that respond to the organic pollution gradient; and (step 5) comparisons between the single metrics and two multimetric methods were made to examine if

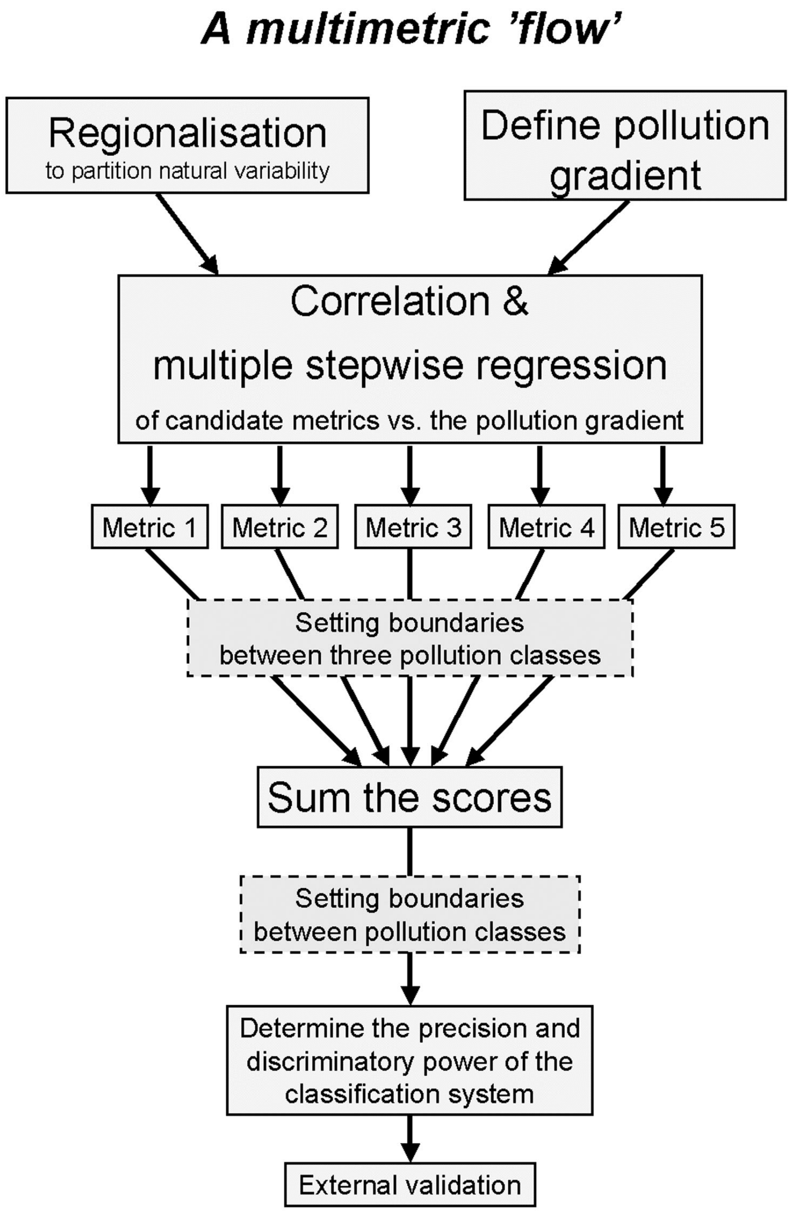

Fig. 3. Flowchart of the steps taken in calibration and validation of the DJ index, a multimetric index constructed to detect the effects of organic pollution in streams of southern Sweden.

multimetric approaches were more robust than single metric approaches for detecting the effects of organic pollution.

Multimetric index development (paper II)

Eighty-four macroinvertebrate candidate metrics were calculated and divided into six attribute groups; richness, composition, tolerance, feeding, habit, and habitat measures (Table 1). Metrics were promoted from candidacy (to inclusion in a final multimetric index) if they demonstrated a significant correlation with organic pollution. Six steps were taken in calibration and validation of the multimetric index for organic enrichment (Fig. 3): (step 1) regionalisation by four regional categories was used to partition natural variability; (step 2) an organic pollution gradient was defined using principal component analysis (PCA) on chemical variables; (step 3) testing of candidate metrics sensitivity to the organic pollution gradient was done using correlation (Spearman's rho-test) and multiple stepwise regression; (step 4) class boundaries between pollution classes were set using percentile distributions; (step 5) determination of precision and discriminatory power of the classification system was done by comparing how well the pollution classes were able to discriminate human-generated effects using ANOVAs and Tukey-Kramer HSD (honestly significant difference) tests; and (step 6) an external validation by correlating (Spearman's rho) the final index with measures of disturbance made at the reach and the catchment scale was done with dataset 1.

Comparison of different approaches (paper III)

Single metric (DSFI, BBI, IBE, Saprobic index, BMWP, and ASPT), multimetric (the DJ index and the AQEM type S05 index), and multivariate (CA, DCA, and the BEAST model) approaches were compared in this study, to examine performance of the different methods at detecting the effects of organic pollution. The scores of the first CA- and DCA-axes correlated with measures of acid conditions. However, the scores of the second CA- and DCA-axes were used, since these axes explained the main part of the effects of organic pollution. Also the BEAST model was included in this study, but only for examining the error frequencies. The BEAST model was constructed from a CA of species x site-sample matrix, using the 75% probability ellipse around the 'good' group A-samples as a limit between type I (*i.e.* false postive) and type II (*i.e.* false negative) errors.

The indicator methods were regressed against a complex organic pollution gradient derived from a principal component analysis (PCA) of parameters associated with the effects of organic pollution to determine the strength of the relationship between the various methods and the stressor gradient. The performance of the different approaches was assessed using sensitivity, which was estimated as the coefficient of variation (r^2) and p-values and precision, which was estimated as the prediction errors (RMSE). Secondly, the sensitivity of the different methods to stress was assessed by testing for differences between a group of presumably unimpaired (*i.e.* group A) and presumably impaired (*i.e.* group B) samples. ANOVA and boxplots of single, multimetric and multivariate (CA and DCA) scores versus the phosphorus grouping of site-samples (group A and B) were done to extract the r^2- and p-values of the different indicator methods. The variance associated with the presumably unperturbed sites was also used as an indication of the precision of the difference methods. In the third step, the potential errors associated with the classification of impairment were assessed using the same ANOVA approach described above. Here the interest was in a comparison of the frequency of type II (or false negative) errors associated with the different methods. A type I error occurs when a 'good' sample is incorrectly classified as 'bad', and a type II error occurs when a 'bad' sample is incorrectly classified as 'good'. Here type I error is defined as 'group A'-samples below the 25-percentile, and the type II-error is defined as 'group B'-samples above the same limit used for type I errors. Since by definition the type I error frequency was set at 25%, it is the type II error frequency that is of primary interest for comparing the precision of the different assessment methods. Finally, the different methods were correlated (Spearman's rho) with measures of disturbance made at the reach and the catchment scale (*i.e.* conductivity, percentage silt of the bottom substrata, and percentage cropland of the catchment area) that could be associated with organic pollution of streams, and these findings used to better interpret the results.

Assessing acid stress (paper IV)

Seven indices, developed to assess the effects of acid stress on running water ecosystems in Sweden and Norway, were evaluated in paper **IV**: namely, Norwegian acidity index I (N I) (Raddum *et al.,* 1988), Norwegian acidity index II (N II) (Bæken & Aanes, 1990), Norwegian acidity index III (N III) (Bæken & Kjellberg, 1999), Swedish acidity index I (S I) (Henrikson & Medin, 1986), Swedish acidity index II (S II) (Degerman *et al.,* 1994), and the Swedish acidity index III (S III) (Lingdell & Engblom, 2002). With the exception of S I (Henrikson & Medin, 1986), the five remaining indices are based on an extensive taxa-list, where each taxon is classified according to its tolerance or sensitivity to acid stress. Also three multimetric acidity indices developed for the AQEM project (AQEM-consortium, 2002) were included in the evaluation.

The similarity between spring and autumn chemical measurements was evaluated using standardised Principal Component Analysis (PCA) of chemical variables. Detrended Correspondence Analysis (DCA) (Hill & Gauch, 1980) was used to examine the variability in the macroinvertebrate data, and Canonical Correspondence Analysis (CCA) (ter Braak, 1987) was run for spring and autumn separately, to determine gradients and examine the relationships between macroinvertebrate indices and water chemistry. The down-weighting option of rare species was invoked in DCA and CCA. To further evaluate the performance of the different acidity indices used in Sweden and Norway, correlations between index values measured in spring and autumn were compared using simple linear regression, paired *t*-tests, and Student's *t* statistics.

Table 1. *Macroinvertebrate attributes and their predicted response to increasing organic pollution. References for the metrics can be found in the AQEM* CONSORTIUM *(2002).*

Macroinvertebrate attribute	Pred. resp. to incr. org. pollution	Macroinvertebrate attribute	Pred. resp. to incr. org. pollution
Richness measures		Feeding measures	
Abundance (ind. m^{-2})	Increase	RETI (Rhitron Feeding Type Index)	Increase
Total number of individuals	Increase		
Total number.taxa	Decrease	*Feeding types (number of individuals):*	
Number of EPT individuals	Decrease	Active filter feeders	Variable
Number of EPT taxa	Decrease	Detrivores	Decrease
Shannon-Wiener Diversity Index	Decrease	Grazers	Decrease
Simpson Dominance Index	Decrease	Parasites	Variable
Evenness	Decrease	Passive filter feeders	Decrease
		Predators	Decrease
Composition measures		Shredders	Decrease
Order (% of community):		Xylophagous taxa	Decrease
Bivalvia (%)	Increase		
Coleoptera (%)	Decrease	*Feeding types (% of community):*	
Crustacea (%)	Increase	Active filter feeders (%)	Increase
Diptera (%)	Increase	Gatherers/Collectors (%)	Decrease
Ephemeroptera (%)	Decrease	Grazer and scrapers (%)	Decrease
Gastropoda (%)	Increase	Miners (%)	Increase
Heteroptera (%)	Increase	Parasites (%)	Increase
Hirudinea (%)	Increase	Passive filter feeders (%)	Variable
Lepidoptera (%)	Decrease	Predators (%)	Variable
Megaloptera (%)	Variable	Shredders (%)	Increase
Nematoda (%)	Decrease	Xylophagoustaxa (%)	Decrease
Nematomorpha (%)	Decrease		
Odonata (%)	Decrease	Habit measures	
Planipennia (%)	Variable	*Locomotion type (% of community):*	
Plecoptera (%)	Decrease	Burrowing/boring (%)	Variable
Trichoptera (%)	Decrease	Sprawling/walking (%)	Increase
Turbellaria (%)	Increase	Swiming/skating (%)	Variable
		Swimming/diving (%)	Variable
EPT taxa (%)	Decrease	Semisessil (%)	Variable
		Habitat measures	
		Microhabitat preference (% of community preferring a certain microhabitat):	
Tolerance measures		Type Akal (%)	Variable
DSFI (Danish Stream Fauna Index)	Decrease	Type Argyllal (%)	Variable
ASPT (Average Score Per Taxon)	Decrease	Type Lithal (%)	Decrease
BBI (Belgian Biotic Index)	Decrease	Type POM (particular organic matter) (%)	Increase
IBE (Extended Biotic Index)	Decrease	Type Pelal (%)	Increase
BMWP (Biological Monitoring Working Party)	Decrease	Type Phythal (%)	Decrease
		Type Psammal (%)	Increase
		Current preference (% of community preferring a certain current):	
		Type limnobiont (%)	Decrease
Saprobic indices:		Type limnophil (%)	Increase
Dutch Saprobic Index	Increase	Type limno- to rheophil (%)	Increase
German Saprobic Index	Increase	Type rheophil (%)	Decrease
Saprobic Index (Zelinka & Marvan)	Increase	Type rheo- to limnophil (%)	Decrease
		Type rheobiont (%)	Variable
		Type indifferent (%)	Increase
		Zonation (% of community preferring a certain zone):	
Saprobic Valence (% of community):		crenal (%)	Increase
xeno saprobic (%)	Variable	hypocrenal (%)	Decrease
oligo saprobic (%)	Decrease	epirhithral (%)	Variable
beta-meso saprobic (%)	Variable	metarhithral (%)	Variable
alpha-meso saprobic (%)	Increase	hyporhithral (%)	Variable
poly saprobic (%)	Increase	epipotamal (%)	Variable
		metapotamal (%)	Increase
		hypopotamal (%)	Increase

Inference of stream pH (paper V)

Weighted averaging (WA) with and without tolerance down-weighting and with classic and inverse deshrinking was used to determine which modeling approach gave the best fit or predictive power. Tolerance down-weighting gives taxa with a narrow pH tolerance or amplitude a greater weight than taxa with a narrow pH tolerance in WA calibration. Furthermore, the WA method takes the averages twice, and therefore the estimated *x*-values are shrunken. The amount of shrinking can be estimated from training data by regressing either x_2 on x_1 (classic deshrinking) or x_1 on x_2 (inverse deshrinking). For the analyses weighted averaging regression predictions were compared with the predictions obtained using four biotic indices that are commonly used in Swedish stream bioassessment programs. Acidity index I (Henrikson & Medin, 1986), a form of a multimetric index, consists of a combination of four scored categories. The first category consists of the presence/absence of *Gammarus* and the acid sensitive species of the orders Ephemeroptera, Plecoptera, and Trichoptera. The second category consists of the presence/absence of groups of acid-sensitive taxa (leeches, the coleopteran family Elmidae and molluscs). The third category is the ratio of *Baetis* to Plecoptera, and the fourth is simply a score for taxa richness by a pre-determined taxa list. The three other biotic metrics tested here are mathematically and ecologically simpler than acidity index I. Acidity indices II (Raddum *et al.,* 1988), III (Lingdell *et al.,* 1990) and IV (Bæken & Aanes, 1990) are simply based on presence/absence and tolerance of selected indicator taxa to pH. In other words, a site is scored as unimpaired if, for example, one individual from a sensitive class of species is recorded as present.

Three different measures of an impairment gradient were used to compare the performance of the WA modelling approaches. In the first model, pH alone was used to define an acidity gradient. In the second model, an acidity gradient was extracted from a principal component analysis (PCA) of selected metrics indicative of acidity; namely, pH, calcium concentration (Ca; µg l^{-1}), acid neutralizing capacity (ANC; meq l^{-1}), alkalinity (ALK; meq l^{-1}), and CBALK (meq l^{-1}). The acidity level is generally expressed with pH, a measure for the content of hydrogen ions (H^+). Alkalinity (ALK), the buffering capacity of a pure carbonate bearing water, is defined according to Hemond (1990), and another measure of this is ANC, which also includes the buffering capacity of organic anions (Stumm & Morgan, 1981). A third measure is CBALK (Hemond, 1990; Köhler *et al.,* 2000), which only needs ALK and Total Organic Carbon (TOC) for the calculation of buffering capacity. Ions of calcium are correlated with buffering capacity, since these ions accompany an equivalent amount of ALK. Finally, in a third model pH-circumneutral sites (pH = 6.5-7.5) were removed from the dataset and, as in model 2, an acidity gradient was extracted by PCA on pH, Ca, ANC, ALK, and CBALK. Removal of pH-circumneutral sites in this third model was done to isolate a more homogeneous acidity gradient. For all modeling approaches r^2- and RMSE-values were used to assess model performance. The number of sites was low in model three, hence jack-knife, cross-validation tests were used to test the performance of the WA methods.

The computer programme WACALIB version 3.3 (Line *et al.,* 1994) was used for inferring stream pH from macroinvertebrate assemblages in streams. Bootstrapping was used to estimate the sample-specific errors of the predictions. Another programme, CALIBRATE (Juggins & ter Braak, 1993), was used to calculate the errors (using a jack-knife, cross-validation method) of both calibration and validation data.

Results and discussion

The focus of this thesis is on evaluating the strengths and weaknesses of different approaches that are commonly used in bioassessment of stream integrity using macroinvertebrates. Single metric indices used for assessing organic pollution of streams are evaluated (paper **I**), as well as multimetric indices and multivariate methods (paper **III**). Particular focus is placed on appraising the performance of the multimetric approach, through construction of a new index for assessing the organic pollution of streams in southern Sweden (paper **II**). The acidity of streams in northern Sweden is an area of concern and hence reliable assessment methods are needed to detect changes in ecological quality. The use of single and multimetric indices of acidity is therefore evaluated (paper **IV**). A relatively new approach using weighted averaging is also tested and compared to more traditional measures of stream quality (paper **V**).

Single metric indices (paper I)

ANOVA of chemical variables showed that most of the chemical factors have a similar distribution both in spring and autumn. ANOVA of higher taxonomic level data (*e.g.* family) showed no difference between spring and autumn, and PCA of taxa data showed that only three sites differed in species composition between spring and autumn. This gives a good base for evaluating the relation between chemical gradients and biological assessment systems.

Comparison of the 84 single metrics (see Table 1) showed that seven metrics in particular, namely, the Saprobic Index (Zelinka & Marvan), the German Saprobic Index, percentage hypopotamal preferences, percentage pelal preferences, the BBI, the ASPT, and the DSFI were good indicators for assessing the effects of organic pollution on streams in southern Sweden. Ranking these seven ‘best’ single metrics according to the cumulative r^2-values showed that BBI was ranked first (combined r^2-value score = 4.71), followed by ASPT (4.54) and the two saprobien indices (4.16 and 3.89) (Table 2). A common denominator for these metrics is that they all are tolerance measures. Both BBI and ASPT are biotic indices, which are taxonomically simpler than the more demanding saprobien indices. However, the saprobien indices are well-known indices for detecting organic pollution, and they are commonly used in several European countries. A saprobic index adapted to Swedish conditions would probably perform better than, for example, the German saprobic index and could be a useful contribution to Swedish stream quality assessment.

Table 2. *Comparison of the 'best' single metrics and two multimetric indices; their responses (using linear regression) to total phosphorus concentration, conductivity, and percentage cropland. Coefficients of variation (r^2-values) are shown. *$p < 0.05$; **$p < 0.025$*

	Total phosphorus concentration			Conductivity			Percentage cropland			
Metric	Spring	Autumn	Spring and autumn combined	Spring	Autumn	Spring and autumn combined	Spring	Autumn	Spring and autumn combined	Combined r^2-value score
BBI	0.47**	0.55**	0.44**	0.33*	0.74**	0.48**	0.32*	0.85**	0.53**	4.71
ASPT	0.58**	0.41**	0.40**	0.40**	0.62**	0.48**	0.41**	0.70**	0.54**	4.54
Saprobic Index (Zelinka & Marvan)	0.49**	0.46**	0.47**	0.48**	0.54**	0.46**	0.42**	0.43**	0.41**	4.16
German Saprobic Index	0.45**	0.46**	0.35**	0.38**	0.61**	0.45**	0.29*	0.53**	0.37**	3.89
hypopotamal preferences (%)	0.54**	0.29*	0.38**	0.39**	0.43**	0.25**	0.27*	0.49**	0.31**	3.35
DSFI	0.47**	0.27*	0.31**	0.34**	0.43**	0.36**	0.30*	0.30*	0.30**	3.08
pelal preferences (%)	0.45**	0.52**	0.27**	0.29*	0.33*	0.29**	0.28*	0.35**	0.26**	3.04
DJ index	0.67**	0.52**	0.55**	0.54**	0.70**	0.58**	0.54**	0.66**	0.59**	5.35
AQEM Type S05 index	0.54**	0.51**	0.45**	0.32*	0.65**	0.45**	0.36**	0.69**	0.51**	4.48

Comparison of the utility of single metrics versus multimetric indices for detecting the effects of organic pollution gives credence to the use of multimetric indices in bioassessment. In particular, the DJ index was generally more strongly correlated with the effects of organic pollution than the single metrics tested here.

The DJ index had a combined r^2-value score of 5.35, and the other multimetric index, the AQEM Type S05 index (AQEM-consortium, 2002), had a slightly lower combined r^2-value score (4.48), but performed well in comparison with the other metrics (Table 2). Single metric approaches, especially saprobien indices, have a long history of use in Europe, and have been shown to be useful in detecting the effects of a number of human-generated disturbances. Stream ecosystems are often affected by a multitude of human-generated pressures, which may singly or in concert affect the integrity of stream ecosystems. Moreover, human-generated stress may even exacerbate the effects of natural stress. Hence, whereas single metrics are often aimed at detecting a specific type of degradation, multimetric indices are often considered to give a broader perspective of the disturbance and thus might perform better in situations where more than one stressor is prevalent (*e.g.* Barbour *et al.,* 1999).

Multimetric index development (paper II)

Principal component analysis (PCA) on chemical variables, land use in the catchment area, and stream bottom substrata was used to isolate an organic pollution gradient. The first PC-axis explained 30% of the total variance, and was related to organic pollution. Hence, the scores along this first PC axis were used to represent an organic pollution gradient.

After defining a organic pollution gradient, 84 single metrics were calculated and tested for possible inclusion in the final multimetric index. Five metrics correlating with the effects of organic pollution remained and were included in the final multimetric index; (i) number of EPT taxa, representing richness measures; (ii) percent Crustacea and (iii) percent EPT taxa, representing composition measures; (iv) ASPT and (v) Saprobic index, representing tolerance measures. After each metric had been divided into three classes, the sum of the five metrics was divided into both three and five pollution classes (Fig. 4). The validation of these final classes showed that the final multimetric index developed here is a promising tool for detecting the effects of organic pollution on stream ecosystems. When measures of taxon richness, composition and tolerance were incorporated, the index was found to significantly correlate with human disturbance, in particular with variables indicative of organic pollution.

Encouragingly, the DJ index was also correlated with a number of single environmental variables indicative of organic pollution stress. For example, the index was correlated with both habitat- (silt substratum) and catchment-level predictors (*e.g.* conductivity and percentage of the catchment classified as urban and arable). An external validation showed that the DJ index was also highly correlated with total phosphorus concentration (TP) both in spring ($r^2 = 0.50$) and autumn ($r^2 = 0.57$).

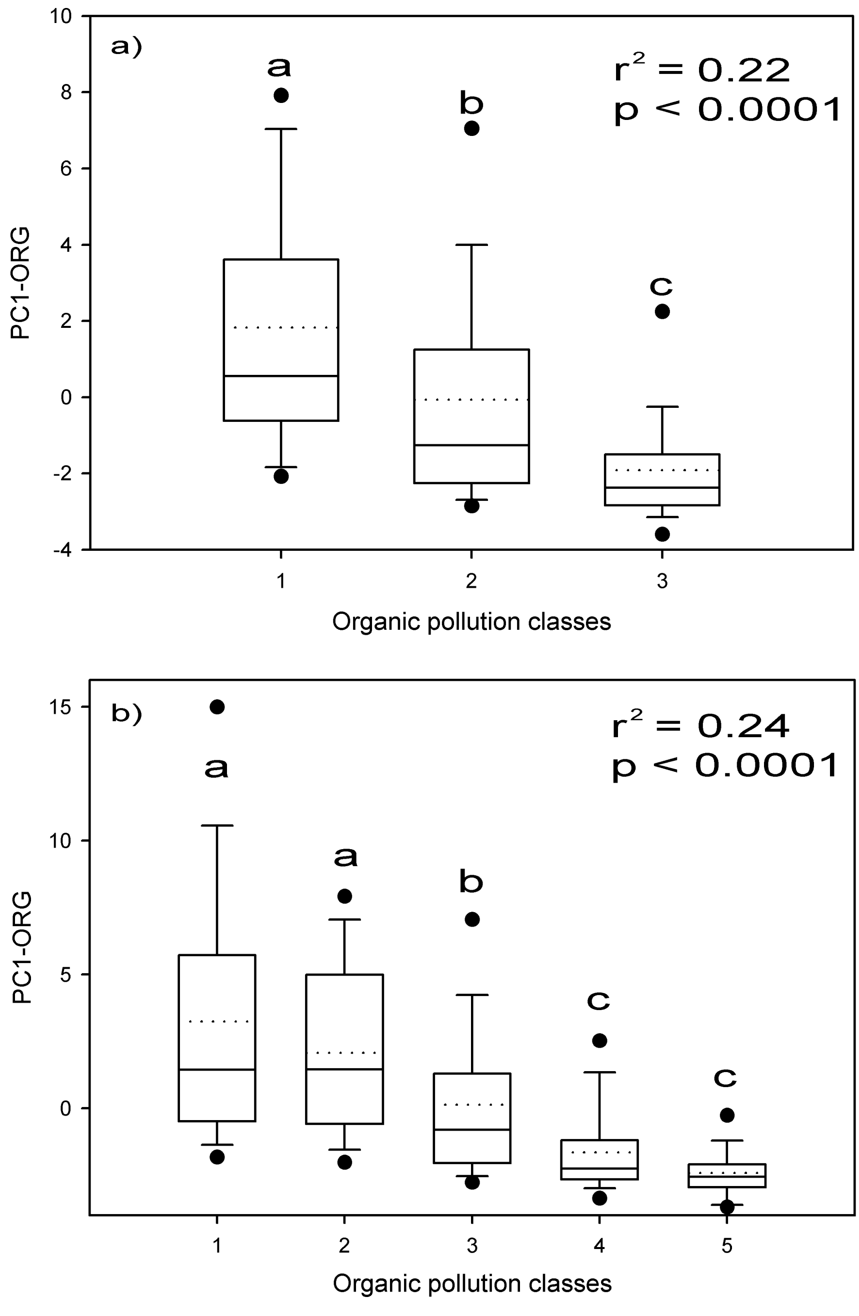

Fig. 4. Boxplots of three (*a*) and five (*b*) pollution classes of the DJ index versus the organic pollution gradient (PC1-ORG) using data from the 1995 national stream survey. Plots of the median (solid line) and mean (dotted line) values as well as 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentiles and the 5^{th} and the 95^{th} percentiles as dots.

In contrast to the prolific use of multimetric approaches in the US, single metrics are at present still more commonly used than either multimetric indices or multivariate analyses in bioassessment programmes using macroinvertebrates in Europe. This disparity is in part due to the early, widespread use of single metrics in Europe. Critics of the multimetric method argue that potentially important ecological information may be lost by aggregating individual measures into a single index (*e.g.* Suter, 1993; Polls, 1994). Advocates, on the other hand, argue that one of the strengths of the multimetric approach is that it incorporates ecological information on how aquatic organisms feed, reproduce, and exploit their habitats (Fore *et al.,* 1996) into assessments of water quality. Further, they suggest that reliance on combinations of multiple measures minimises the weaknesses of individual metrics (*e.g.* Barbour *et al.,* 1999), and that one of the strengths of constructing a multimetric index is the possibility of deleting factors that could cause redundancy or metrics that are not correlated with the pressure or anthropogenic stress to be detected. In the index developed here, for example, information on richness, composition and tolerance were included, while attributes related to feeding, habit and habitat measures were not included as they were either redundant or did not discriminate organic stress. This study also supports the conjecture that the multimetric approach may give a broader perspective of the disturbance, while single metrics are often aimed at detecting only a single type of degradation.

Comparison of different approaches (paper III)

Comparison of the performance of single metric indices, multimetric indices, and multivariate approaches showed that both the multimetric and multivariate methods were better at discriminating impairment than single-metric approaches. In particular, the DJ index and the CA scores seemed to be reliable methods for detecting the effects of organic pollution on macroinvertebrate assemblages in streams in southern Sweden.

Regression of selected metrics against a complex organic pollution gradient (PC1) showed that sensitivity to detect impairment differed with metric, approach and season. Using a spring dataset, eight metrics (of 10) were significantly related to the stressor gradient, compared to six using a autumn dataset and eight using a combined spring and autumn dataset. Using the combined spring and autumn dataset showed that the sensitivity of the various metrics varied generally between that and the sensitivity obtained with the spring and autumn datasets separately. Four of the six single metrics were significantly related to the PC1 organic pollution gradient, whilst all of the multimetric and multivariate approaches were significantly related to the stressor gradient.

Analysis of the various methods to detect impairment by ANOVA showed that sensitivity varied markedly between the two sampling seasons and by metric. ANOVA showed that only four of the 18 tests of single metrics against P concentration were significant ($p < 0.05$). Although coefficients of variation were high for samples collected in spring ($r^2 > 0.33$), only two of the single metrics (ASPT, $r^2 = 0.66$ and BBI, $r^2 = 0.61$) were significantly correlated with P concentration. Both multimetrics and multivariate approaches were slightly better than the single metrics, with r^2-values ranging from 0.58 (DCA scores) to 0.77 (DJ multimetric). Surprisingly, none of the methods tested here showed impairment with the samples collected in autumn ($p < 0.05$). Indeed, metrics that were strongly correlated with P concentration in spring (e.g. ASPT, $r = 0.81$) were poorly correlated with this gradient in autumn (e.g. ASPT, $r = 0.20$). Combining the spring and autumn datasets showed only two single metrics that were significantly related to P concentration (BBI, $r^2 = 0.20$ and Saprobic index, $r^2 = 0.32$), and all but one (DCA score) of the multimetric and multivariate approaches were significantly correlated to P concentration, albeit with lower r^2-values. For example, the multivariate approach with CA scores had the highest explained variability ($r^2 = 0.38$, $p < 0.025$), followed by the DJ index ($r^2 = 0.35$, $p < 0.025$).

Examining error frequencies showed that type II error frequencies differed with metrics and season (Fig. 5). For the spring dataset, none of the multimetric or multivariate metrics showed type II errors, whilst three of the single metrics BMWP, DSFI and Saprobic index had type II error frequencies of 33%. Error frequencies were much higher when the autumn dataset was used (> 33%). For the single metrics, type II error frequencies averaged 63%, and ranged from 33% (Saprobic index) to 89% (IBE). Multimetric and multivariate approaches generally had lower type II error frequencies than the single metrics tested here (mean for multimetric = 56% and mean for multivariate = 44%). Combining the spring and autumn resulted in type II error frequencies generally higher than those observed in spring, but lower than those observed in autumn. For the single metrics, type II error frequencies averaged 53%, ranging from 25% (Saprobic index) to 100% (IBE). The multimetric and multivariate approaches resulted in similar type II error frequencies (mean = 32%, range = 25 to 33%).

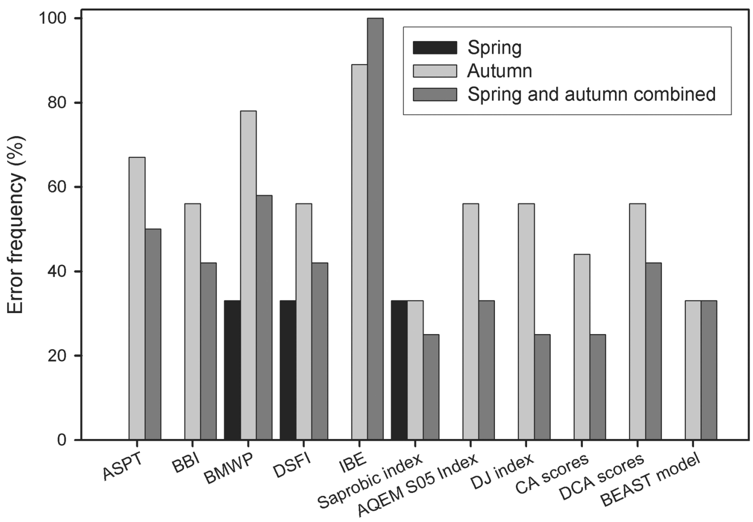

Fig. 5. Frequencies of type II errors ('bad' samples classified as 'good') of eleven assessment methods for detecting the effects of organic pollution using macroinvertebrates.

Table 3. *Correlation (Spearman's rho) matrix of assessment methods and selected parameters associated with organic pollution (percentage silt substratum, conductivity, and percentage cropland in the catchment area). * p < 0.05; ** p < 0.025*

	Percentage silt			Conductivity (log 10)			Percentage cropland		
Method	Spring (n = 8)	Autumn (n = 13)	Spring & autumn combined (n = 21)	Spring (n = 8)	Autumn (n = 13)	Spring & autumn combined (n = 21)	Spring (n = 8)	Autumn (n = 13)	Spring & autumn combined (n = 21)
Single metrics									
ASPT	0.33	0.23	0.25	-0.76*	-0.45	-0.55**	-0.76*	-0.42	-0.53**
BBI	0.18	0.25	0.24	-0.66	-0.59*	-0.61**	-0.68	-0.78**	-0.72**
BMWP	0.65	0.15	0.30	-0.29	-0.29	-0.29	-0.50	-0.39	-0.45*
DSFI	0.26	-0.07	0.03	-0.78**	-0.39	-0.51**	-0.37	-0.28	-0.31
IBE	0.24	-0.21	-0.06	-0.49	-0.01	-0.18	-0.38	-0.05	-0.21
Saprobic index	-0.30	-0.23	-0.26	0.71*	0.63	0.63**	0.50	0.42	0.45*
Multimetrics									
AQEM type S05 index	0.32	0.16	0.22	-0.75*	-0.50	-0.57**	-0.64	-0.48	-0.51**
DJ index	0.26	0.28	0.27	-0.77*	-0.60*	-0.68**	-0.70	-0.54	-0.60**
Multivariate									
CA scores	0.06	-0.07	0.05	0.79**	0.55	0.62**	0.69	0.54	0.59**
DCA scores	0.48	0.33	0.35	0.74	0.43	0.52**	0.44	0.36	0.42

Both ordination (use of CA scores) and multimetrics (use of DJ index) were shown to be robust methods for detecting ecological change in bioassessment. Although disentangling the effects of catchment- versus habitat-level variability on macroinvertebrate assemblage structure is difficult, if not impossible, correlation showed that the metrics studied here were more related (higher Spearman's rho-values) with catchment- than habitat-level descriptors (Table 3). None of the metrics were significantly correlated with the amount of habitat classified as silt substratum, whereas several metrics were significantly correlated with both conductivity and percentage of the catchment classified as cropland (*i.e.* two catchment-level descriptors of land use). Several metrics were significantly correlated with conductivity and/or percent cropland during one season but not the other and *vice versa*. For example, ASPT was significantly correlated with conductivity in spring (rho = -0.76, p < 0.05) but not in autumn (rho = -0.45, p > 0.05), whereas BBI was significantly correlated with conductivity in autumn (rho = -0.59, p < 0.05) but not in spring (rho = -0.66, p > 0.05). However, given the number of site-samples used in these analyses, statistical power is probably low and hence these results should be viewed with caution.

The Saprobic index performed better than other single metric indices compared in this study. However, if only the fourth analysis is considered the BBI performed best. The BBI is constructed for, and based on, benthic macroinvertebrates of Belgian streams, and therefore may not perform as expected with Swedish conditions. Also the Saprobic index should be adapted to Swedish conditions before it is used in Sweden. The DJ index and the CA scores seem to be reliable methods for detecting the effects of organic pollution in southern Sweden. However, if possible, a combination of both multimetric and multivariate methods is preferred in assessing organic pollution effects in streams as the methods are complementary. The multimetric method produces a comparable single 'score' by combining unique ecological information of selected metrics. Multivariate approaches, on the other hand, use all information collected in the taxa list and requires no *a prior* assumptions. Reynoldson *et al.,* (1997) compared multivariate and multimetric approaches and found that precision and accuracy of the multivariate methods tested were consistently higher than for the multimetric assessment. However, they recommended that multimetric approaches should be used in conjunction with multivariate methods.

Assessing acid stress (paper IV)

One of the acidity indices evaluated in this study showed strong correlations between spring and autumn sampling, namely the Swedish acidity index I (S I) (Table 4). Samples taken in autumn could therefore be used to evaluate the spring situation when the acid stress (whether natural or anthropogenic) in northern Sweden is most pronounced. This is important since collection of a representative spring sampling is difficult due to varying ice out, spring flood events, and early emergence by insects. Although the S I index might not detect improvement in environmental conditions as fast as the 'simpler' acidity indices (that are simply based on the presence/absence of indicator taxa), the S I index is less prone to vary in time (caused by changes not related to the acid stress) and should therefore be a more reliable tool for evaluating environmental quality in streams in northern Sweden. In total, only eight of the 106 indices tested in this study could differentiate between the sites classified as acid and non-acid using macroinvertebrate data collected either in spring or autumn (*i.e.* the number of taxa, number of EPT taxa, the acidity index S I, BMWP score, BBI, and number of Ephemeroptera, Trichoptera, and Coleoptera taxa). Two of the indices developed for the AQEM project (for stream type S01 and type S03-S04) could also differentiate between the two groups, both for spring and autumn samples.

A total of 184 taxa were found at the 60 stream sites; 132 taxa were recorded in spring and 154 taxa were recorded in autumn. Comparison of both seasons showed that 102 were shared, thus 30 taxa were only found in spring and 52 taxa were only found in autumn. Six taxa were found in at least 50% of the streams in spring (*i.e.* Tanytarsini, Simuliidae, Ortocladiinae, Tanypodinae, Empididae, and *Baetis rhodani* Pictet), whereas four taxa were found in at least 50% of the streams in autumn (*i.e.* Ortocladiinae, Simuliidae, Tanypodinae, and *B. rhodani*). All chemical variables that differed significantly between spring and autumn (10 of 16) had higher values or concentrations in the autumn than in the spring. Canonical correspondence analysis (CCA) of macroinvertebrates and stream physico-chemical descriptors for stream sites situated in the arctic/alpine region (Illies ecoregion 20; stream types S03 and S04), showed a clear temperature/altitudinal gradient using the spring dataset, while latitude together with the percentage of land in the catchment classified as mixed forest came out as the best explanatory variables using the autumn dataset. For streams situated in the northern boreal forest region (Illies ecoregion 22; stream types S01 and S02) two main gradients were found; one related to vegetation and substratum and the other related to stream width and distance to the coast. The apparent temperature/altitudinal gradient noted for stream sites situated in the arctic/alpine region (ecoregion 20) was not obvious for streams situated in the northern boreal forest region (ecoregion 22).

Table 4. *Macroinvertebrate indices that differentiated the nine sites classified as acidic, and the 27 sites classified as non-acidic in spring and autumn, respectively (paired t-test).*

Index	Reference	Spring	Autumn
No. of taxa		< 0.01	< 0.005
No. of EPT taxa		< 0.05	< 0.05
Acid index SI	Henrikson & Medin, 1986	< 0.001	< 0.005
Acid index SII	Degerman *et al.,* 1994	< 0.05	
Acid index SIII	Lingdell & Engblom, 2002	< 0.05	
Acid index NIII	Bækken & Kjellberg, 1999	< 0.001	
BMWP	Armitage *et al.,* 1983	< 0.001	< 0.0001
Saprobic index	Zelinka & Marvan, 1961	< 0.01	
Belgian Biotic index	De Pauw *et al.,* 1992	< 0.001	< 0.0001
% taxa in the hyporithral		< 0.05	
% of rheobiont taxa			< 0.05
% predators			< 0.05
% Plecoptera		< 0.05	
% Coleoptera		< 0.05	
No. of Ephemeroptera taxa		< 0.005	< 0.05
No. of Trichoptera taxa		< 0.05	< 0.001
No. of Coleoptera taxa		< 0.01	< 0.01
Index type S01		< 0.01	< 0.005
Index type S02		< 0.01	
Index type S03 and S04		< 0.01	< 0.005

The 27 sites that were classified as weakly acidic or nearly circumneutral (pH > 6.5) and that had good or very good buffering capacity (alkalinity > 0.10 meq l^{-1}) in both spring and autumn were used to test the typology according to the EC Water Framework Directive. In spring, a total of 116 taxa were collected from these streams. The most homogeneous stream type was the boreal highland (or alpine zone; S04), whereas the other three types showed a more gradual change in macroinvertebrate composition from stream type S01 to S03. Total variability (inertia) in the macroinvertebrate data was 2.890, and the first four axes of the detrended correspondence analysis (DCA) could explain 38.1% of the variability in macroinvertebrate communities among the stream sites. In the autumn, 127 taxa were found at the 27 sites deemed as non-affected by acid stress. Total variability in the data (inertia) was 3.811 and the first four axes of the DCA explained 30.6% of this variability. Hawkins *et al.,* (2000) stated, concerning large-scale features and aquatic biological communities, that 'the variance related to landscape features was not large', and the findings of this study support this conjecture. The findings of this analysis shows that a typology-based classification using only large-scale features is not sufficient for proper stream assessment, whereas a tiered approach, including both large- (*e.g.* catchment area, ecoregions) and local- (*e.g.* substratum composition and stream velocity) scale variables is recommended as a cost-effective strategy to partition natural variability and subsequently detect anthropogenic stress in running water ecosystems.

Inference of stream pH (paper V)

Weighted averaging (WA) approaches were found to be slightly better in predicting stream pH than the four acidity indices studied here. Both WA and the acidity indices performed poorly at the test sites in southern Sweden. This study shows that WA approaches could be useful for assessing acidity of Swedish streams, but the low r^2-values indicate that factors other than pH are also affecting the macroinvertebrate assemblages confounding the pH organism-response relationship.

For the first modelling approach only pH was used to represent an acidity gradient. Comparison of the WA methods showed that WA with down-weighting of tolerance (WA-tol) gave the best r^2-values for the training sets used for both the northern ($r^2 = 0.28$) and southern ($r^2 = 0.28$) parts of the country (Fig. 6a-b). Error estimates (RMSE) were, however, slightly higher for models developed using the southern dataset. Although WA-tol gave higher r^2-values, the predicted RMSE indicated that WA could be a better predictive tool than WA-tol. With training data the WA methods seemed to relate better to pH than the four acidity indices tested here. Using the test datasets, however, the acidity indices gave better results (r^2 = 0.47-0.62) than the WA methods (r^2 = 0.05-0.24).

Principal component analysis (PCA) on the acidity-related variables pH, alkalinity, acid neutralising capacity (ANC), CBALK, and calcium was used to construct a more robust acidity gradient for the second modelling approach. The first PC-axis explained 99% of the variance among the five water chemistry

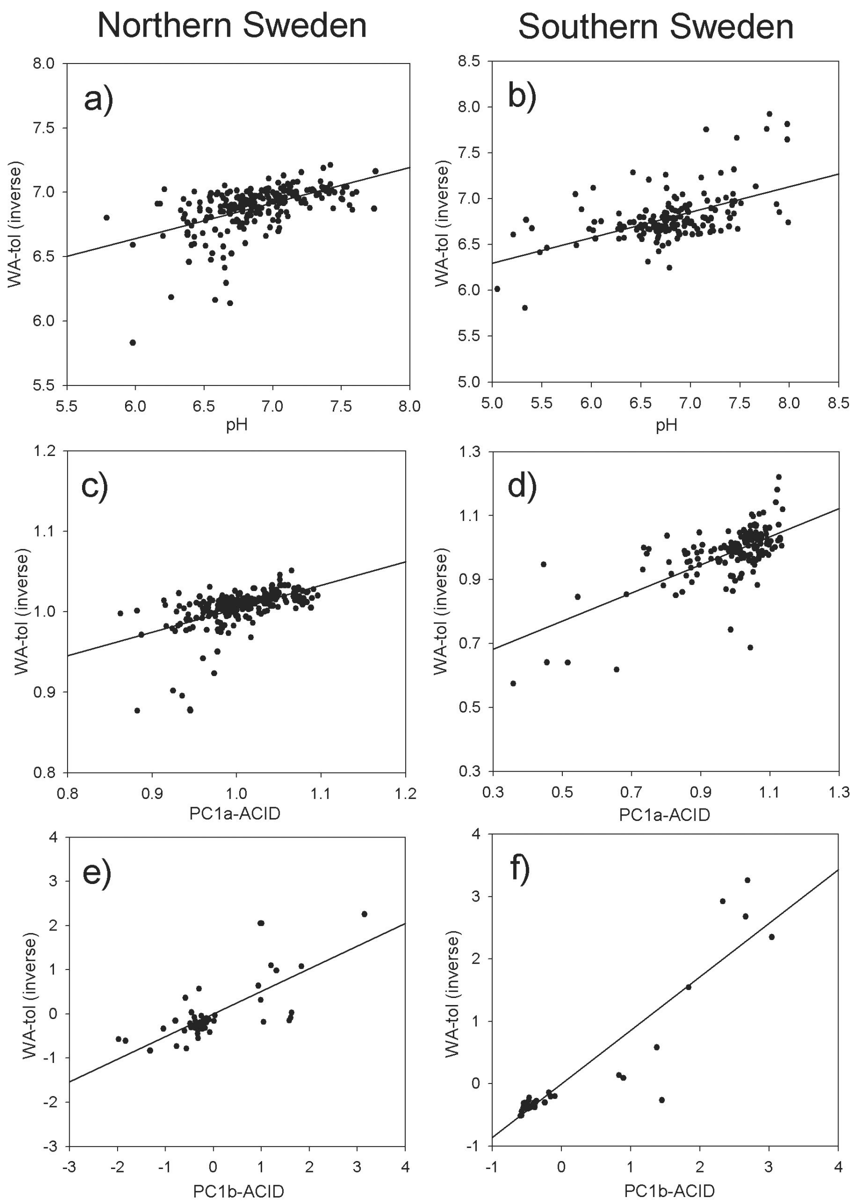

Fig. 6. Examples of the 'best' weighted averaging (WA) models for northern and southern Sweden using modeling approach 1 (*a* and *b*), modeling approach 2 (*c* and *d*), and modeling approach 3 (*e* and *f*).

metrics both for northern and southern Sweden and was used as an acidity gradient. Comparison of the WA methods showed that this new acidity gradient gave only slightly better results (6c-d). WA-tol still showed better r^2-values than WA, but the predicted RMSE was similar for the two methods. Using the northern dataset, the variance explained by WA-tol with this new modelling approach (r^2 = 0.29) was very similar to that from the first modelling approach (*i.e.* r^2 = 0.28). Model performance improved, however, when the southern dataset was used (r^2 = 0.44 versus r^2 = 0.28 with model 1). Error estimates were in both cases somewhat lower using this second acidity gradient. The acidity indices only showed better results than WA methods for test data in northern Sweden, but overall the WA methods showed much lower RMSE than the acidity indices.

For the third modelling approach pH-circumneutral sites (pH = 6.5-7.5) were removed, and a PCA on the acidity-related variables pH, alkalinity, acid neutralising capacity (ANC), CBALK, and calcium was used to establish an acidity gradient. The first PC-axis explained 83% of the variance for northern Sweden and 98% of the variance for southern Sweden and was used as an acidity gradient. The use of this revised acidity gradient showed that WA models were able to better predict pH. The WA-tol had an r^2-value of 0.51 using the northern dataset, and 0.86 using the southern dataset (Fig 6e-f). Cross-validation, with jack-knifing, showed higher r^2-values (for both the northern and southern datasets) than earlier test datasets. Overall, for this modelling approach, the WA methods were better at predicting pH than the four acidity indices, and in general the acidity indices also showed higher RMSE than the WA models.

It may be more reliable to use multimetric or multivariate methods for evaluating the acidity of streams, but in practice a 'simple' method would probably be used more often. The acidity indices tested here are simple to understand and use. However, the WA approach is based on direct weighting of the species recorded, and should therefore be a better method than the acidity indices to assess stream quality. There are also indices using combinations of traditional methods and WA methods. For example, some saprobic indices (*e.g.* Zelinka & Marvan, 1961) use weighted averaging of subjectively assigned indicator values for aquatic organisms in environmental calibration. One drawback of using biotic acidity indices when these are too 'simple' could be that the indices only give a score for the environmental situation at the specific sampling occasion, and if sampling is done for example in the autumn, the lowest pH-levels of the year (normally during the spring flood) could be missed. Generally, organisms are used in bioassessment because they integrate effects over time. The WA methods are 'simple', but could give opportunities to adapt the measurements to the specific situations, meaning that it is possible to collect samples in the autumn without missing the 'spring' information.

This thesis in a wider context (paper I – V)

Single metric indices have a long history in Europe as the 'best' tools for detecting the effects of human activity on running water ecosystems. However, like most areas of science, knowledge evolves and the area of bioassessment has seen a substantial increase in the last two decades. Today, the more 'complex' methods are simpler to use with the proliferation of computers and bioassessment software. Multivariate methods are often reliable and they are being increasingly used in bioasssessment. However, a major drawback is that interpretation and visualisation often demand a deeper ecological understanding. The multimetric approach is also reliable in detecting ecological change, but in contrast to the use of multivariate approaches, the interpretation is more straightforward since the 'output' is a simple 'score'. The potential utility of the multimetric approach in bioassessment of organic polluted streams was shown in paper **I**, **II** and **III**. Also the multivariate approach, with its reliance and stability, is given credit in paper **III**. The index preferred for acid conditions in paper **IV** (the Swedish acidity index I) is a multimetric index, which also gives weight to the multimetric approach. However, if possible, both multimetric and multivariate methods should be used in conjunction, since these methods could work in a complementary way. The multimetric method produces a comparable single 'score' using the unique ecological obtained in selected metrics, while the multivariate method uses all information in the taxa-list and requires no *a prior* assumptions. For among site, region, and country harmonisation of assessment results, it would be advantageous if water managers use the same (or similar) methods, hence user-friendlier software should be developed.

Macroinvertebrate community composition might differ between geographic regions (*e.g.* Sandin, 2003), and hence the indices developed for a specific region often should be modified if they are used outside that specific region. In Sweden, the DSFI (developed for Danish streams) and the ASPT (developed for British streams) are recommended for use without any modification (SEPA, 1999). This implies that differences in species pools or taxon-specific differences in sensitivity/tolerance are the same in Sweden, Denmark, and Great Britain. Although this assumption still needs to be tested, the findings of this study lend support to the use of metrics 'borrowed' from other regions. Another example is the BBI, which was constructed for detecting the effects of organic pollution in Belgian streams, but according to paper **I** and **III** can be considered as one of the 'best' metrics for detecting the effects of organic pollution in streams situated in the southern parts of the country. In these cases a deeper ecological understanding of the results is needed to interpret if it is a matter of chance or if it is the truth that makes these indices perform well. A better way would probably be to use a more reliable method (*i.e.* multivariate statistics). A multivariate method also needs an ecological understanding, but often the result is more reliable, which for example is shown in paper **III**.

The sources of pollution change with time, and hence there is a need to develop new and better assessment methods. If a good method for assessing a specific ecological change is lacking, it is sometimes better to construct a new index rather than try to find the 'next best'. Paper **II** develops a new index for detecting the effects of organic pollution on stream macroinvertebrate assemblages in southern Sweden. Paper **V** uses the 'old' weighted averaging method in a 'new' way to assess the acidity of streams. This method is commonly used for algae, but has rarely been used for benthic macroinvertebrates (*e.g.* Hämäläinen & Huttunen, 1996; Larsen *et al.,* 1996). Paper **V** shows that weighted averaging could be a good approach when the prediction of occurrence along single gradients is in focus. In this case, this approach seemed to be at least as good as the commonly used acidity indices.

Conclusions and future perspectives

Papers **I**, **II** and **III** focus on approaches used for assessing the effects of organic pollution on stream macroinvertebrate communities. Paper **I** ranked the seven 'best' single metrics according to the cumulative r^2-values, and showed that BBI was ranked first, followed by ASPT and two saprobien indices. A common denominator for these four metrics is that they all are tolerance measures. The saprobic index was shown in paper **III** as the best of the single metric indices tested. In Sweden, ASPT and DSFI are generally used for assessing the effects of organic pollution. The findings in paper **I** support the continued use of the DSFI index as it was ranked with the seven 'best' indices, but had low statistical power.

Papers **I**, **II** and **III** also examined and compared the utility of single and multimetric indices for detecting the effects of organic pollution. These studies give credibility to the use of multimetric indices in bioassessment. In particular, the DJ index was generally more strongly correlated with the effects of organic pollution than the single metrics tested here. The other multimetric index tested, the AQEM Type S05 index, had a slightly lower statistical power, but performed well in comparison to the other single indices.

A comparison of the performance of single metric, multimetric, and multivariate approaches was done in paper **III**. Both multimetric and multivariate methods were shown to be powerful tools for assessing the effects of organic pollution on stream macroinvertebrate assemblages. In particular, the DJ index and the CA scores seemed to be reliable methods for detecting these effects in southern Sweden. In these comparisons the scores of CA- and DCA-axes represented the multivariate approach. However, the multivariate techniques used today are often sophisticated and require special expertise. Comparison of these techniques with other approaches is often more complicated, since the multivariate methods are often more complex than the single metric or multimetric methods.

Paper **IV** and **V** focus on acid conditions in streams. Ten indices were evaluated regarding their ability to detect acid stress in paper **IV**. Four of the indices are considered as multimetric indices, and one of these multimetric acidity indices, namely the Swedish acidity index I, was strongly correlated between spring and autumn sampling. Hence, samples taken in autumn could therefore be used to evaluate the spring situation when the acid stress is most pronounced. 'Simpler' single metric indices might detect improvement in environmental conditions fast, but the Swedish acidity index I is less prone to vary in time and should therefore be a more reliable tool for evaluating the environmental quality of streams in northern Sweden. These findings give credence to the multimetric approach, while the 'new' weigthed averaging (WA) method examined in paper **V** gave some credence also to the multivariate approach. The weighted averaging method with down-weighting of tolerance performed somewhat better than the four common acidity indices in paper **V**. The weighted averaging model could be a useful tool for assessing acidity of streams, but more research is needed to develop the method.

In summary, this thesis supports the use of multimetric and multivariate approaches in bioassessment. If possible, a combination of multimetric and multivariate methods should be preferred in assessing the effects of ecological change. Single metric indices could be used, but a deeper ecological understanding is needed to evaluate the inherent errors associated with these metrics. This thesis shows, however, that multimetric and multivariate approaches should be considered in future bioassessment.

Further research addressing the following questions should contribute to a better understanding and quality of bioassessment of stream ecosystems:

- Will the single metric DSFI index perform better if it is taxonomically adapted to Swedish conditions?
- Would the development of a 'modern' multimetric acidity index better our understanding of the effects of anthropogenic acidity in streams? Can indicators of brown-water systems be used to determine and normalise for the effects of natural acidity?
- Is the weighted averaging approach an alternative to commonly used indices in assessing the effects of organic pollution? Can weighted averaging regression be used to better our understanding of the effects of liming on stream ecosystems?
- Does a complementary metric approach result in higher sensitivity to detect change in multi-stressor situations?
- Are predictive models the future for stream assessment?

References

Agresti, A. 1990. *Categorical data analysis*. John Wiley & Sons, New York, USA

Ahlström, J. & Isaksson, K.-E. 1990. Försurningsläget i Norrlands inland och fjälltrakter-kalkning. *Report 3781*, *Swedish Environmental Protection Agency*, Stockholm, Sweden.

Alba-Tercedor, J. & Sanchez-Ortega, A. 1988. Un metodo rapido y simple para evaluar la calidad biologica de las aguas corrientes 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 macroinvertebrates, developed for the purpose of the Water Framework Directive.* Version 1.0, February 2002.

Armitage, P.D., Moss, D., Wright, J.F. & Furse, M.T. 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research 17 (3)*, 333-347.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a ß-function. *Journal of Vegetation Science 5*, 215-228.

Bæken, T. & Aanes, K.J. 1990. Use of macroinvertebrates to classify water quality. *Report 2a, The Norwegian Institute for Water Research (NIVA)*, Oslo, Norway.

Bæken, T. & Kjellberg, G. 1999. Klassifisering av surhetsgrad i rennende vann basert på forekomst av bunndyr. Klassifiseringssystem tilpasset humusrike elver og bekker i østlandsområdet. *in* Bæken, T., Kjellberg, G. & Linløkken, A. (eds). *Overåkning av bunndyr i grensekryssende vassdrag i østlandsområdet i forbindelse med kalkningsovervåkning. Samlerapport for undersøkelsene i 1995, 1996 og 1997. DN-notat 2*. Direktoratet for naturforvaltning, Trondheim, Norway.

Barbour, M.T., Gerritsen, J., Snyder, B.D. & Stribling, J.B. 1999. Rapid Bioassessment Protocols For Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish. Second Edition. *EPA Report 841-B-99-002, Office of Water*, Washington, D.C., USA.
Bio, A.M.F., Alkemade, R. & Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *Journal of Vegetation Science 9,* 5-16.
Bishop, K.H., Laudon, H. & Köhler, S. 2000. Separating the natural and anthropogenic components of spring flood pH decline: a method for areas that are not chronically acidified. *Water Resources Research 30,* 1873-1889.
Bradley, B.P. & Ormerod, S.J. 2002. Evaluating the precision of kick-sampling in upland streams for assessments of long-term change. *Archives für Hydrobiologie 155,* 199-221.
Brzeziecki, B., Kienast, F. & Wildi, O. 1995. Modeling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *Journal of Vegetation Science 6,* 257-258.
Chandler, J.R. 1970. A biological approach to water quality management. *Water Pollution Control 69,* 415-421.
Dahl, J. & Johnson, R.K. 2004. A multimetric macroinvertebrate index for detecting organic pollution of streams in southern Sweden. *Archives für Hydrobiologie 160 (4),* 487-513.
De Pauw, N. & Vanhooren, G. 1983. Method for biological quality assessment of watercourses in Belgium. *Hydrobiologia 100,* 153-168.
De Pauw, N., Ghetti, P.F., Manzini, D.P. & Spaggiari, D.R. 1992. Biological assessment methods for running water. Pages 217-248 *in* P. J. Newman, Piavaux, M. A. & Sweeting, R. A. (eds). *River Water Quality. Ecological Assessment and Control.* European Commission.
Degerman, E., Fernholm, B. & Lingdell, P.-E. 1994. Bottenfauna och fisk i sjöar och vattendrag. Utbredning i Sverige. *Report 4345, Swedish Environmental Protection Agency*, Solna, Sweden.
European Commission. 2000. Directive 2000/60/EC of the European Parliament and of the Council - Establishing a framework for Community action in the field of water policy, Brussels, Belgium, 23 October 2000.
Fore, L.S., Karr, J.R. & Wisseman, R.W. 1996. Assessing invertebrate responses to human activities: Evaluating alternative approaches. *Journal of the North American Benthological Society 15 (2),* 212-231.
Friedrich, G. 1990. Eine revision des Saprobiesystems. *Zeitschrift für Wasser und Abwasser Forschung 23,* 141-152.
Ghetti, P.F. 1997. Indicie Biootico Esteso (I.B.E.). I macroinvertebrati nel controllo della qualità degli ambienti di acque correnti. *Provencia Autonoma di Trento,* 222 pp.
Hawkins, C.P., Norris, R.H., Gerritsen, J., Hughes, R.M., Jackson, S.K., Johnson, R.K. & Stevenson, R.J. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society 19 (3),* 541-556.
Hawkins, C.P. Quantifying biological integrity with predictive models: comparisons with three other assessment methods. Submitted manuscript.
Hellawell, J.M. 1978. Biological surveillance of rivers, a biological monitoring handbook. *Water Research Centre*, Medmenham and Stevenage, UK, 331 pp.
Hellawell, J.M. 1986. Biological indicators of Freshwater Pollution and Environmental Management. *Elsevier Applied Science*, London, UK, 546 pp.
Hemond, H.F. 1990. Acid neutralizing capacity, alkalinity, and acid-base status of natural waters containing organic acids. *Environmental Science & Technology (24),* 1486-1489.
Henrikson, L. & Medin, M. 1986. Biologisk bedömning av försurningspåverkan på Lelångens tillflöden och grundområden 1986. *Aquaekologerna*, Rapport till Länsstyrelsen i Älvsborgs län.
Hering, D., Moog, O., Sandin, L. & Verdonschot, P.F.M. 2004. Overview and application of the AQEM assessment system. *Hydrobiologia 516,* 1-20.

Herrman, J., Degerman, E., Gerhardt, A., Johansson, C., Lingdell, P.-E. & Muniz, I.P. 1993. Acid-stress effects on stream biology. *Ambio 22,* 298-307.
Hildrew, A.G., Townsend, C.R. & Francis, J.E. 1984. Community structure in some southern English streams: the influence of species interactions. *Freshwater Biology 14,* 297-310.
Hill, M.O. & Gauch, H.G. 1980. Detrended correspondence analysis, an improved ordination technique. *Vegetatio 42,* 47-58.
Hughes, R.M. 1994. Defining acceptable biological status by comparing with reference conditions. Pages 31-48 *in* Davies, W.S. & Simon, T.P. (eds). *Biological assessment and criteria. Tools for water resource planning and decision making.* Lewis Publishers, Boca Raton, Florida.
Hynes, H.B.N. 1960. *The Biology of Polluted Water.* Liverpool University Press, Liverpool, England.
Hämäläinen, H. & Huttunen, P. 1996. Inferring the minimum pH of streams from macroinvertebrates using weighted averaging regression and calibration. *Freshwater Biology 36,* 697-709.
Illies, J. 1978. *Limnofauna Europaea*. Gustav Fischer Verlag, Stuttgart, Germany.
International Organisation for Standardization (ISO). 1979. Assessment of the biological quality of rivers by a macroinvertebrate 'score'. Draft proposal. ISO/TC 147/SC 5/WG 6 N 5, 18 pp.
Johnson, R.K., Wiederholm, T. & Rosenberg, D.M. 1993. Freshwater biomonitoring using individual organisms, populations and species assemblages of benthic macroinvertebrates. Pages 40-158 *in* Rosenberg, D. M. & Resh, V. H. (eds). *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, New York.
Johnson, R.K. 1995. The indicator concept in freshwater biomonitoring. Thienemann lecture. Pages 11-27 *in* Cranston, P.S. (ed). *Chiromids - from Genes to Ecosystems, Proceedings of the 12th IInternational Symposium in Chiromidae, Canberra, Australia.* CSIRO, Melbourne.
Johnson, R.K. 1998. Benthic macroinvertebrates. Pages 85-166 *in* Wiederholm, T. (ed). *Bedömningsgrunder för miljökvalitet. Sjöar och vattendrag. Bakgrundsrapport 2. Biologiska parametrar.* Report 4921, Swedish Environmental Protection Agency, Uppsala.
Johnson, R.K., Aagard, K., Aanes, K.J., Friberg, N., Gislason, G.M., Lax, H. & Sandin, L. 2001. Macroinvertebrates. Pages 43-52 *in* Skriver, J. (ed). *Biological monitoring in Nordic Rivers and Lakes*. Tema Nord Environment 2001: 513, pp. 43-52.
Johnson, R.K. 2003. Development of a prediction system for lake stony-bottom littoral macroinvertebrate communities. *Archives für Hydrobiologie 158 (4),* 517-540.
Juggins, S. & Terbraak, C.J.F. 1993. CALIBRATE - a program for species-environment calibration by [weighted-averaging] partial least squares regression. Unpublished program.
Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries 6,* 21-27.
Knoben, R.A.E., Roos, C. & van Oirschot, M.C.M. 1995. Biological Assessment Methods for Watercourses. *UN/ECE Task Force on Monitoring and Assessment*, 85 pp.
Kolkwitz, R. & Marsson, M. 1902. Grundsätse für die biologische Beurteiling des Wassers nach seiner Flora und Fauna. *Mitt. Prüfungsanst. Wasserversorg. Abwasserreinig. 1,* 33-72.
Köhler, S., Laudon, H., Wilander, A. & Bishop, K. 2000. Estimating organic acid dissociation in natural surface waters using total alkalinity and TOC source. *Water Research 34,* 1425-1434.
Larsen, J., Birks, H.J.B., Raddum, G.G. & Fjellheim, A. 1996. Quantitative relationships of invertebrates to pH in Norwegian river systems. *Hydrobiologia 328 (1),* 57-74.
Line, J.M., Terbraak, C.J.F. & Birks, H.J.B. 1994. WACALIB version 3.3 - a computer program to reconstruct environmental variables from fossil assemblages by weighted averaging and to derive sample-specific errors of prediction. *Journal of Paleolimnology 10,* 147-152.
Lingdell, P.-E. & Engblom, E. 2002. Bottendjur som indikatorer på kalkningseffekter. *Report 5235, Swedish Environmental Protection Agency*, Stockholm, Sweden.

Lingdell, P.-E., Engblom, E., Norrgrann, O. & Söderberg, H. Assessing acidification in Swedish surface waters with the LD-index. Manuscript.

Malmqvist, B. & Rundle, S. 2002. Threats to running water ecosystems of the world. *Environ. Cons. 29***,** 134-153.

Mason, C.F. 1996. *Biology of freshwater pollution*. Longman, 3rd edition, London.

Metcalfe, J.L. 1989. Biological Water-Quality Assessment of Running Waters Based on Macroinvertebrate Communities - History and Present Status in Europe. *Environmental Pollution 60 (1-2)***,** 101-139.

Miljödepartementet. 1996. Kalkning av sjöar och vattendrag. Betänkande från kalkningsutredningsgruppen. *SOU report 1996:53***,** 144 pp.

Moog, O.E. 1995. Fauna Aquatica Austriaca, Version 1995. *Wasserwirtschaftskataster, Bundesministerium für Land- und Forstwirtschaft, Wien*, ISBN: 3-85 174-004-1.

Nicholls, A.O. 1989. How to make biological models go further with Genaralized Linear Models. *Biological Conservation 50***,** 51-75.

Polls, I. 1994. How people in the regulated community view biological integrity. *Journal of North American Benthological Society 13***,** 598-604.

Raddum, G.G. & Fjellheim, A. 1984. Acidification of early warning organisms in freshwater in western Norway. *Verh. Internat. Verein. Limnol. 23***,** 1973-1980.

Raddum, G.G., Fjellheim, A. & Hesthagæn, T. 1988. Monitoring of acidity by the use of aquatic organisms. *Verh. Internat. Verein. Limnol. 23***,** 2291-2297.

Resh, V.H., Rosenberg, D.M. & Reynoldson, T.B. 2000. Selection of benthic macroinvertebrate metrics for monitoring water quality of the Frazer River, British Colombia: implications for both multimetric approaches and multivariate models. Pages 195-206 *in* Wright, J. F., Sutcliffe, D. W. & Furse, M. T. (eds). *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. Freshwater Biological Association, Ambleside, UK.

Reynoldson, T.B., Bailey, R.C., Day, K.E. & Norris, R.H. 1995. Biological Guidelines for Fresh-Water Sediment Based on Benthic Assessment of Sediment (the Beast) Using a Multivariate Approach for Predicting Biological State. *Australian Journal of Ecology 20 (1)***,** 198-219.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E. & Rosenberg, D.M. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society 16 (4)***,** 833-852.

Rosenberg, D.M. & Resh, V.H. 1993. Introduction to Freshwater Biomonitoring and Benthic Macroinvertebrates. Pages 1-9 *in* Rosenberg, D.M. & Resh, V.H. (eds). *Freshwater biomonitoring and benthic macroinvertebrates.* Chapman & Hall, London and New York.

Sandin, L. & Johnson, R.K. 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. *Journal of the North American Benthological Society 19 (3)***,** 462-474.

Sandin, L. 2003. Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. *Ecography 26***,** 269-282.

Schindler, D.W. 1998. Sustaining aquatic ecosystems in boreal regions. *Conservation Ecology 2 (2), 18.*

SEPA (Swedish Environmental Protection Agency). 1999. Bedömningsgrunder för miljökvalitet. Sjöar och vattendrag. *Report 4913, Naturvårdsverket*, Stockholm.

Simpson, J.C. & Norris, R.H. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125-142 *in* Wright, J.F., Sutcliffe, D.W. & Furse, M.T. (eds). *Assessing the biological quality of freshwaters. RIVPACS and other techniques.* Freshwater Biological Association, Ambleside, England.

Skriver, J., Friberg, N. & Kirkegaard, J. 2000. Biological assessment of running waters in Denmark: Introduction of the Danish Stream Fauna Index (DSFI). *Verh. Internat. Verein. Limnol. 27***,** 1822-1830.

Slack, K.V., Averett, R.C., Greeson, P.E. & Lipscomp, R.G. 1973. Methods for collection and analysis of aquatic biological and microbiological samples. Pages 1-165 *in*

Techniques of Water-Resources Investigations of the United States Geological Survey. *Department of the Inerior, Geological Survey*, Washington, DC.
Sladecek, V., Hawkes, H.A., Alabaster, J.S., Daubner, I., Nötlich, I., Solbé, J.F.d. & Uhlman, D. 1982. Biological examination. *in* Suess, M. J. (ed). *Examination of Water for Pollution Control. A reference handbook. Vol. 3. Biological, Bactereological and Virological Examination.* Pergamon Press, Oxford.
SS-EN-27828. 1994. Vattenundersökningar - Metoder för biologisk provtagning - Riktlinjer för provtagning med handhåv. *European Committee for Standardization*.
Stanner, D. & Bordeau, P. 1995. *Europe´s Environment: The Dobříš Assessment*. European Environment Agency, Luxembourg. 676 pp.
Stumm, W. & Morgan, J.J. 1981. *Aquatic chemistry.* Wiley, New York. 780 pp.
Suess, M.J. (ed). 1982. *Examination of Water for Pollution Control. A Reference Handbook.* Pergamon Press, Oxford.
Suter, G.W. 1993. A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry 12***,** 1533-1539.
ter Braak, C.J.F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio 69***,** 69-77.
ter Braak, C.J.F. & Prentice, I.C. 1988. A theory of gradient analysis. *Advances in ecological research 18***,** 271-317.
Townsend, C.R., Hildrew, A.G. & Francis, J.E. 1983. Community structure in some southern English streams: the influence of physicochemical factors. *Freshwater Biology 13***,** 521-544.
Weber, C.I. 1973. Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents. *EPA report 670/4-73-001, U.S. Environmental Protection Agency*, Cincinnati.
Whitton, B.A. 1979. Algae and higher plants as indicators of river pollution. *in* A. James & L. Evison, (eds). *Biological Indicators of Water Quality*. John Wiley, Chichester.
Wiederholm, T. 1980. Use of benthos in lake monitoring. *Journal of the Water Pollution Control Federation 52***,** 537-547.
Wiederholm, T. & Johnson, R.K. 1997. Monitoring and assessment of lakes and watercourses in Sweden. Pages 317-329 *in* Ottens, J.J., Claessen, F.A.M. *et al.* (eds). *Monitoring Tailor-made II. Information Strategies in Water Management. Proceedings from an international workshop, September 1996, Nunspeet, The Netherlands.*
Wilander, A., Johnson, R.K., Goedkoop, W. & Lundin, L. 1998. Riksinventering 1995. En synoptisk studie av vattenkemi och bottenfauna i svenska sjöar och vattendrag. *Report 4813, Swedish Environmental Protection Agency*, Stockholm.
Wilander, A., Johnson, R.K. & Goedkoop, W. 2003. Riksinventering 2000. En synoptisk studie av vattenkemi och bottenfauna i svenska sjöar och vattendrag. *Report 2003:1, Department of Environmental Assessment, SLU*, Uppsala.
Wold, H. 1982. Soft modeling: the basic design and some extensions. Pages 1-54 *in* Joreskog, K.G. & Wold, H. (eds). *Systems under indirect observations II*, Amsterdam, North-Holland.
Woodiwiss, F. 1964. The biological system of stream classification used by the Trent River Board. *Chemistry and Industry 14***,** 443-447.
Wright, J.F., Moss, D., Armitage, P.D. & Furse, M.T. 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. *Freshwater Biolology 14 (3)***,** 221-256.
Wright, J.F. 1995. Development and Use of a System for Predicting the Macroinvertebrate Fauna in Flowing Waters. *Australian Journal of Ecology 20 (1)***,** 181-197.
Zelinka, M. & Marvan, P. 1961. Zur präzisierung der biologischen klassifikation der reinheit fließender gewässer. *Archives für Hydrobiologie 57***,** 389-407.

Acknowledgements – Tack!

Att skriva en avhandling kan vara något av det största man gör i livet. Ibland kan denna uppgift vara svår och ibland kan den vara lätt. Svårt kan det vara när man tappat motivationen och inte hittar stigen som leder mot målet. Lätt kan det vara när man har människor runt omkring sig som bidrar till motivation och gör stigen lättgången. Sådana människor har jag haft gott om på institutionen för Miljöanalys och dessa vill jag tacka.

*Först av allt vill jag tacka **Richard Johnson**, som lyckats motivera mig även i de svåraste stunder. Du har läst manuskript, som vi sedan diskuterat och skrivit om. Dessutom har du sänt mig till möten och konferenser i både USA och Europa och fått mig att inse att den vetenskapliga världen är mycket trendkänsligare än jag trodde. Nu gäller prediktiva modeller – vad gäller imorgon? AI, WA, eller? Du kommer ofta med nya infallsvinklar, vilket är mycket inspirerande.*

*Denna avhandling skulle nog aldrig ha påbörjats om inte **Leonard Sandin** varit med och ryckt in mig i AQEM-projektet - tack för samarbetet inom det projektet. Jag är också mycket tacksam för att du delat med dig av din kunskap både inom den statistiska och logiska sfären.*

*Jag vill också tacka **Willem Goedkoop**, som har varit en stor tillgång när det gäller alla möjliga typer av frågor och dessutom har du läst igenom och kommenterat flera manuskript.*

***Lars Eriksson**, jag är tacksam för att du hjälpt mig med provtagningar och artbestämningar av bottenfauna. Dessutom är jag glad åt de diskussioner vi haft rörande både ekologiska och humana frågor under provtagningsresor.*

***Oskar Andersson**, vi har under provtagningar åkt många hundra mil i bil och några mil till med helikopter. Det skulle ha varit många ensamma veckor i fält utan dig. Tack för all din hjälp med både provtagning och sortering av bottenfauna.*

***Björn Wiklund**, tack för din hjälp med sortering och artbestämning av bottenfauna.*

***Anders Wilander**, tack för att du har delat med dig av dina kunskaper om sura eller försurade vatten. Sverige skulle nog ha varit surare utan dig.*

*Datasektionen på Miljöanalys är värd ett stort tack. Jag vill särskilt tacka **Jakob Nisell** som varit behjälplig när GIS-kartor inte blivit som man tänkt sig eller när 'avancerade' GIS-beräkningar behövts göras. Jag vill även tacka **Mikael Östlund** för hjälp med grafiska problem. Tack också, **Herman Paz**, som alltid har kunnat hjälpa till när datorn betett sig som en Mac. Ett stort tack också till **Bert Karlsson**, som alltid har kunnat fixa fram data eller fixa ett tekniskt problem när andra vägar varit svåra att nå.*

*Alla de provtagningar som gjorts skulle antagligen inte ha varit möjlig om institutionen saknat det vattenkemiska laboratorium som utför alla vattenanalyser. De data jag har använt mig av härrör främst ifrån resultat som **Mairon Berglund**, **Marie-Louise Berglund**, **Sten Engblom**, **Gun Hölling**, **Lena Lindevall**, **Inger Olsén**, **Bodil Ottosson**, **Gunilla Warnicke**, **Kjell Östling** och **Margaretha Östling** har varit med och tagit fram. Ett stort tack.*

*En av de saker som gjort livet lätt medan man doktorerat är alla ni underbara **doktorander** (gamla som nya) på Miljöanalys som får en att le även en mulen dag. Resan upp till Abisko och Lofoten, spex, kubb, grill, bowling, öl, biljard och en massa annat har varit upplevelser utöver det vanliga. Tack för era glada infall.*

*Ett särskilt tack vill jag rikta till **Sonja Stendera**, som nog är den bästa tänkbara doktorand man kan dela rum med. Med dig kan jag prata om vad som helst. Du har ju dessutom läst och kommenterat sammanfattningen.*

***Ulf Grandin** och **Anneli Widenfalk**, tack för att ni läst och kommenterat manuskript i samband med halvtidsuppföljningen.*

*Jag vill även tacka övriga människor på institutionen som hjälpt till med olika göromål, bl. a. **Hans Eurell**, **Stellan Sjödahl**, **Svante Andersson**, **Ingrid Marelius**, **Tommy Jansson** och **Britta Lidström**.*

*Slutligen ett stort tack också till **Ewa Bringmark**, för att du försöker att få institutionen att må bra trots att ekonomins bistra vindar ibland blåser för kraftigt. Jag hoppas att vinden mojnar och att institutionen fortsätter att visa den starka sida som kännetecknar Sveriges bästa institution.*

Det finns även personer utanför institutionen som har varit till hjälp eller funnits till hands i olika situationer och jag är mycket tacksam för er närvaro.

*I want to thank the **members of the EU-project AQEM**. I hope we will work together again in the future.*

*Tack, **Jan-Erik Åslund** på länsstyrelsen i Jämtland, **Johan Ahlström** på länsstyrelsen i Västerbotten, **Eddie Persson** på länsstyrelsen i Värmland, **Håkan Söderberg** på länsstyrelsen i Västernorrland, samt **Cecilia Holmström** och **Jan Pröjts** på Ekologgruppen i Landskrona, för er hjälp med val av provtagningslokaler i AQEM-projektet. Jag vill också tacka **Anders Dahlén** för att du visade mig vägen till provtagningslokalerna i de Jämtländska fjällen under AQEM-provtagningen.*

*Tack också **Amelie Jarlman** för att du har introducerat mig i kiselalgernas värld. Förhoppningsvis så blir det väl en artikel av det också.*

*Sist, men inte minst, vill jag tacka **Sofie Lücke** för att du har läst och kommenterat sammanfattningen och manuskript. Dessutom har du ju hjälpt till med lite av varje i slutskedet av denna avhandlings tillkomst.*