ITM-rapport 150

Compilation and analysis of benthic macroinvertebrate data in IKEU-streams

Brendan McKie, Per-Ola Hoffsten och Björn Malmqvist

Institutionen för ekologi, miljö och geovetenskap
Umeå Universitet

Institutionen för tillämpad miljövetenskap

Department of Applied Environmental Science

ITM ENHETEN FÖR AKVATISK
MILJÖKEMI OCH EKOTOXIKOLOGI
PROJEKTLEDARE: CECILIA ANDRÉN

ISSN: 1103- 341X
ISRN: SU-ITM-R-150-SE
Upplaga: 50 ex
Tryckeri: PrintCenter US-AB

Compilation and analysis of benthic macroinvertebrate data in IKEU streams

Brendan McKie
Per-Ola Hoffsten
Björn Malmqvist

Institutionen för ekologi, miljö och geovetenskap
Umeå Universitet
901 87 Umeå

E-mail: brendan.mckie@emg.umu.se,
per-ola.hoffsten@spray.se,
bjorn.malmqvist@emg.umu.se

Table of contents

English Summary

We compiled stream macroinvertebrate and water chemistry data collected as part of the *Integrerad Kalknings Effekt Uppföljning* (IKEU) program over a 12 year period (1994-2005). Analysis of the compiled data addressed two main objectives (i) An assessment of two stream macroinvertebrate sampling methods, M42 and Surber, in relation to the biomonitoring of limed streams, and (ii) an initial assessment of sampling in both autumn and spring. Secondary objectives included (i) an assessment of the impact of liming on stream macroinvertebrate faunas and (ii) an initial assessment of interannual variation in the data.

Most analyses concentrated on the period 1998-2002, for which data for both sampling methods and a complete set of limed and reference streams were available. In order to distinguish between the performance of the Surber and M42 sampling methods, three questions were addressed: (i) which method is best for collecting an assemblage of invertebrate taxa that reflects the acid status of the environment? (ii) Which method better samples a range of acid sensitive taxa? (iii) Which method is better able to distinguish between limed and reference sites? Similar questions were asked in comparing samples collected in autumn and spring. Canonical Correspondence Analysis (CCA) and Mantell's test were used to assess the fit of the multivariate species data to acidity-related environmental data, while the ability of the multivariate data to model environmental data was assessed using Weighted Averaging (WA). Correlations between environmental variables and several acidity indices, calculated from the multivariate species data, were assessed using Kendall's tau correlation. The acid sensitivity of taxa was scored following current definitions used in the formulation of Medins index. The ability of the methods to distinguish liming impacts was investigated using standard hypothesis testing techniques (Analysis of Similarities, Analysis of Variance), and Similarity of Percentages (SIMPER).

In general, data collected using the M42 method was more closely associated with aciditiy-related environmental data, and more effectively sampled acid-sensitive taxa. Discrimination between limed and reference streams was also slightly better for M42. There was little evidence that sampling using two methods added substantial extra information compared with sampling using one method. In contrast, neither spring nor autumn sampling consistently performed better, but spring sampling did appear to add some extra information over that gained from autumn sampling. Overall, there was no marked general divergence in the macroinvertebrate assemblage structure of limed and reference streams, or in the acid status of these streams, as indicated by several acidity indices. However, multivariate analyses indicate the faunas of limed and reference streams are not identical, and the responses of individual streams to liming varied. Limitations in the current set of regularly monitored streams, including the lack of both acid reference sites and control of non-acidity related environmental variation, render a robust assessment of the impact of

liming difficult. The strongest interannual pattern in the data appeared to reflect differences in the intensity of M42 sampling.

Following the work presented here, the following recommendations are made:

Strong recommendations (action is urged on these points):

1) Sample using one method only
2) Sample using the M42 method (but see caveat detailed in the text)
3) Clearly define specifications for M42 sampling and ensure that staff are well-trained in the method, and that specifications are closely followed.
4) Use resources saved from the termination of Surber sampling to expand the breadth of biomonitoring: Expand geographic coverage, and the set of reference streams (consider acid references, and paired reference sites with rigorously defined characteristics).
5) Analyse the current data file more deeply, particularly in relation to the impact of liming and interannual variation in the data.

Additional recommendations (action is suggested on these points):

6) Consider sampling in both autumn and spring, but if only one season can be sampled, autumn is preferred.
7) Consider expanding the IKEU data set by incorporating data from other sources (e.g. from local government authorities)

Svensk Sammanfattning

Vi sammanställde bottenfauna- och vattenkemidata som insamlats under en 12-års period (1994-2000) inom IKEU-programmet. Analyserna av sammanställda data avsåg två huvudmålsättningar: 1) att utvärdera två metoder för provtagning i rinnande vatten av bottenfauna (M42 och Surber), samt 2) en första utvärdering av provtagning utförd både vår och höst. Som delmålsättningar identifierades 1) en utvärdering av kalkningens effekter på bottenfaunan i rinnande vatten, och 2) en första utvärdering av mellanårsvariationen i datamaterialet.

Analyserna koncentrerades huvudsakligen till perioden 1998-2002 för vilken fullständiga data beträffande de två metoderna samt för både kalkade och referensvattendrag fanns tillgängliga. För att kunna skilja ut hur väl Surber- resp. M42- metoderna fungerade undersöktes tre frågeställningar: 1) Vilken metod fungerar bäst för att samla in sådana taxa som speglar miljöns försurningsstatus? 2) Vilken metod fungerar bäst för att samla störst antal försurningskänsliga taxa? 3) Vilken metod särskiljer kalkade och referensvattendrag bäst? Kanonisk korrespondensanalys (CCA) och Mantel-test användes för att anpassa multivariata artdata till surhetsrelaterade miljödata, medan 'weighted averaging' (WA) användes för att undersöka förmågan att med hjälp av multivariata artdata modellera miljödata. För korrelationssamband mellan miljövariabler och ett flertal surhetsindex, beräknade från multivariata artdata, användes Kendalls Tau. Olika taxas surhetskänslighet bedömdes enligt Medins surhetsindex. De olika metodernas förmåga att urskilja effekter av kalkningen bedömdes med hjälp av standardmetoder för hypotestestning (likhetsanalys, variansanalys) och 'procentlikhetsanalys' (SIMPER).

I de flesta fall låg data som insamlats med M42-metoden närmare surhetsrelaterade miljödata och dessutom fungerade denna metod bättre för insamling av surhetskänsliga arter. Separation av kalkade vattendrag och referensvattendrag var också något lättare med M42-metoden. Knappast något talade för att en påtagligt ökad mängd information erhölls till följd av att man använde två i stället för en provtagningsmetod. Däremot fungerade varken endera vår- eller höstprovtagning konsekvent bättre än den andra även om vårprovtagning tycktes ge något mer information än höstprovtagning. På det hela taget fanns ingen tydlig generell skillnad mellan bottenfaunans sammansättning i kalkade bäckar respektive i referensbäckarna, och inte heller med avseende på surhetstillståndet så som detta bestämts genom tillämpningen av flera försurningsindex. De multivariata analyserna antydde dock att faunan i dessa vattendragstyper skilde sig åt och att responsen på kalkning varierade individuellt mellan olika bäckar. Begränsningar med avseende på den nuvarande uppsättningen regelbundet undersökta bäckar, inkl. bristen både på sura referensvattendrag och på icke-försurningsrelaterad omvärldsvariation, försvårar en robust uppskattning av kalkningens effekter. Det starkaste mellanårsmönstret i datamaterialet tycktes avspegla skillnader i intensiteten av hur M42-metoden tillämpats.

Baserat på föreliggande arbete kan följande rekommendationer göras:

Primära rekomendationer:

1. Begränsa provtagningen av bottenfaunan till en metod.
2. Använd M42-metoden (men det finns en hake – se texten).
3. Klargör specifikationerna för M42-metoden, och ge personalen ingående träning i metodiken.
4. Använd de resurser som sparas genom att Surber-provtagning upphör till att utöka den biologiska kontrollen: Utvidga den geografiska täckningsgraden och uppsättningen referenslokaler (överväg att inkludera sura referensbäckar samt definiera referensvattendragens egenskaper mera rigoröst).
5. Utför fördjupade analyser av den nu upprättade databasen, särskilt med avseende på kalkningseffekter och mellanårsvariation hos data.

Ytterligare rekommendationer:

6. Överväg fortsatt provtagning av bottenfaunan både vår och höst. Om bara en årstid kan provtas är hösten att föredra.
7. Överväg en utvidgning av IKEU-datamaterialet genom inkorporering av data från andra källor (exempelvis från länsstyrelserna).

Objectives

The current analysis of IKEU data has three main objectives:

1) The compilation of disparate IKEU stream macroinvertebrate species and physico-chemical data, collected by different agencies over many years
2) An assessment of two stream macroinvertebrate sampling methods, M42 and Surber, in relation to the biomonitoring of limed streams
3) An initial assessment of sampling at two different times each year, autumn and spring, in relation to the biomonitoring of limed streams

Secondary objectives of the analyses include:

1) An initial assessment of the impact of liming, with reference to a set of unlimed circumneutral streams
2) An initial investigation of inter-annual patterns in the data

Note that these secondary objectives are largely investigated as they relate to the primary objectives. Detailed investigations of the impact of liming (e.g. a species- or stream-level assessment, or the extent of deleterious effects) or longer term trends (e.g. community persistence and longer term responses to liming) were beyond the scope of the current assessment, but are worthy goals for future analyses, given the wealth of data available.

Methods

Data compilation

Species data were initially delivered in several separate files, as compiled by the different agencies responsible . The first set comprised data from annual autumn sampling conducted over the period 1994-1999. The second set included both autumn data and some spring data for a subset of streams, covering the period 2000-2002. The third set comprised extensive data collected in both autumn 2004 and spring 2005. In the first two data sets (collectively covering 1994-2002), benthic macroinvertebrates were sampled using two methods (M42 and Surber), with data for these methods generally stored in separate worksheets or files. Macroinvertebrate sampling over 2004-05 utilised the M42 method only.

Data was compiled for 23 streams, for which data was generally available from 1998 on (the first year a full set of limed and reference streams were sampled). Appendix one lists the chosen

streams, with a summary of some important physico-chemical features. Additionally, for streams with longer records (most limed streams and the acid reference Laxbäcken), data were compiled back to and including 1994.

Because many different workers performed species identification over several years, it was necessary to screen the data during the compilation process, and either remove or harmonise obvious inconsistencies among the various files. Accordingly, the following modifications were made during compilation:

1) Meiofauna and fish: representatives of these groups sometimes occurred in the identified samples, but the extent to which they were enumerated and identified varied. Since the samples were collected for macroinvertebrate biomonitoring, and given that neither the M42 nor Surber method is designed for sampling these elements, all meiofauna (mostly microcrustaceans) and fish were excluded from the data file.
2) Mutually exclusive identification: in several cases, identical taxa were identified differently in different files. For example, before 2000, all *Pisidium* sphaeriids were identified as *Pisidium*, whereas after this date none were identified to this level, but rather to Sphaeriidae, a category that does not exist in the earlier file. As these categories never overlapped and clearly correspond to the same taxon, they were pooled in the compilation file.
3) Uncertain identifications: from year to year there were differences in the level of certainty with which some species were identified. For example, the following categories were used for uncertain *Leuctra* in different years:
    a. Leuctra fusca-digitata-other
    b. Leuctra fusca-digitata-hippopus
    c. Leuctra other

    As none of these categories were used in every year, and almost never co-occurred within a single year (which would indicate consistent separation of distinctive taxa), and as *fusca* and *digitata* are notoriously difficult to distinguish when smaller, all these categories were pooled (as "*Leuctra* other").
4) Difficult groups: some groups are difficult to identify without specialist specimen preparation and identification. An example is the Chironomidae, which can be recognized to subfamily using a light microscope, but require slide preparation for identification to species. In all samples, most chironomids were identified to subfamily, but some individuals in some years were identified to species, indicating processing by a chironomid specialist. However, as this expertise was clearly not available in all years, and because the set of species distinguished in a given year was small, most chironomid species were pooled at the subfamily level. The exception is *Stenochironomus*, which is distinctive

morphologically, allowing easy identification without slide preparation, and which appeared to be distinguished in all years.

Despite these problems, the number of changes necessitated was actually very small, and the vast majority of original identifications were retained. In the resultant file "IKEU compiled species data", it was necessary to store the Surber and M42 data on separate spreadsheets, as the number of species distinguished exceeded the capacity of a single excel sheet.

Physico-chemical data were also delivered in several separate files, but these did not have the same problems of inconsistency as the species data. All available variables (e.g. discharge, water chemistry, depth, temperature, cations, forest type and land use) were incorporated, with missing values left blank. In all cases, mean data for the 12 months prior to the relevant macroinvertebrate sampling date were compiled. Thus the compiled data file is to be used in conjunction with the macroinvertebrate data. In many cases, maxima and minima for relevant variables were compiled also.

Biotic indices and community metrics were delivered already calculated for all years except 2004 and 2005, which were consequently calculated from the compiled data file. Indices can be time consuming to calculate, and so only the following were calculated for the 2004/05 data:

<u>Acidification indices</u>

a. Medin's acidification index
b. The ratio of total Baetis abundance (a) to total Plecoptera abundance (B:Pa)
c. The ratio of total Ephemeroptera richness (r) and abundance to total Plecoptera richness and abundance (E:Pr and E:Pa)

<u>Other communitiy metrics:</u>

a. The sumof Ephemeroptera, Plecoptera and Trichoptera richness (EPTr) and abundance (EPTa)
b. Shannon,s diversity index (H')
c. Total abundance
d. Total species richness

These indices and metrics were focused on in subsequent analyses. Because different workers calculated indices from different data sets, there may be some variation in how the indices were calculated from year to year. Unfortunately, wholesale recalculation of earlier indices was beyond the scope of this work. For those indices relying on the number of individuals and taxa in a defined subset of the assemblage (e.g. Medins) this should not be overly problematical, as the relevant taxa were generally identified consistently from year to year. Nevertheless, this is a shortcoming, and it is advisable that at least those metrics calculated from the entire assemblage (Shannon's diversity, total species richness and abundance), and preferably all indices, be recalculated for all years prior to any serious analysis of long-term trends in the data, to ensure consistency.

Analysis

I) Data preparation

The compiled species data was little modified prior to detailed multivariate analyses. Uncertain species identifications that were unlikely to represent novel taxa were removed (e.g. *Nemoura* sp. was removed, since multiple species from this genus were identified from each stream, and it is unlikely the unidentified individuals represented distinct new taxa). Otherwise all clearly distinguished taxa were retained, because in distinguishing between two sample methods, especially in relation to acidity, the collection of rare (potentially acid sensitive) taxa is just as important as collection of common taxa.

II) Exploratory Data Analysis

Exploratory data analysis was undertaken to describe major patterns in the data, prior to more detailed investigation of the main objectives outlined earlier. Species abundances were ordinated separately for each year using **non-metric multidimensional scaling (nMDS)** with Bray-Curtis dissimilarities (Clarke 1993). Plots from these ordinations are presented with both sample method (M42 or Surber) and lime treatment (lime or reference stream) categories overlaid. Similarities in faunal composition according to sample method and lime treatments were assessed additionally using **unweighted pair group method with arithmetic mean (UPGMA) cluster analysis** (Sneath & Sokal 1973). These analyses were conducted using *PC-ORD for Windows* (Version 4.0, © 1999 MjM software, Oregon USA).

Divergences in faunal composition according to sampling technique were tested statistically using **Analysis of Similarities (ANOSIM)**, available in *Primer for Windows* (Version 5.2.9, © 2002 Primer-E Ltd.). ANOSIM tests for differences between similarity matrices (generated using Bray-Curtis similarities), through calculation of the global test-statistic "Rho". Rho varies between 0 and 1, with a Rho value closer to 1 indicating that similarities are greater within than between sets of replicates, while a value of zero indicates uniform similarities between and within sets (Clarke & Gorley 2001). Sampling method and stream were both fitted within one ANOSIM, with sampling method nested within stream. Accordingly, the test for the sampling method effect averages across pairwise tests within stream groups, with different years comprising the replicates within each stream. Whilst the effect of method on assemblage composition could be tested separately for each year, the approach used here, utilizing data combined across years, is the only way to properly account for the presence of a stream blocking factor using ANOSIM. Following ANOSIM, the **SIMPER (similarity percentages – species composition)** procedure, also available in *Primer,* was used to investigate the contribution of each species to mean Bray-Curtis dissimilarity between sites grouped by method treatment (Clarke & Gorley 2001).

Annual mean data for the major acidity related chemistry variables (pH, minimum pH, alkalinity, Calcium, TOC and inorganic Aluminium) and some biotic metrics (taxa richness,

abundance, Shannon diversity and Medins index) were also plotted, in order to identify aberrant streams and to gain a first impression of interannual variation.

III) Objective two: assessment of the M42 and Surber sampling methods

In order to distinguish between the quality of the Surber and M42 sampling methods, two questions were addressed (note that the word "quality" in this context relates to the ability of the sampling methods to distinguish between streams of differing acid status, and not to the rigour with which the data was collected.):

1) Which method is best for collecting an assemblage of invertebrate taxa that reflects the acid status of the environment?
2) Which method better samples a range of acid sensitive taxa?

A third question addressed the ability of the two methods to detect a treatment effect:

3) Which method is better able to distinguish between limed and reference sites?

Because no single analysis is adequate for assessing all aspects of data quality, several different approaches were employed: Canonical Correspondence Analysis, Mantel's test, Weighted Averaging, non-parametric correlation, and the occurrence of acid sensitive taxa. A brief summary of each method follows. In all cases, separate analyses were carried out for each sampling method (Surber and M42) within each year. Note that although jointly collected M42 and Surber sample data were available from 1994-2003, the time-consuming nature of file preparation for most of these analyses necessitated a focus on the years 1998-2003, when a full set of reference and limed streams were available.

1) **Canonical Correspondence Analysis (CCA)**, using *PC-ORD,* was employed to assess the fit of the species data to four acidity related variables in multivariate space. CCA ordination forces ordination axes (eigenvalues) constructed from abundance data to be expressed in terms of a set of measured environmental variables (ter Braak 1986). This allows assessment of how the measured variables have influenced the distribution of stream sites within the ordination space. Four acidity-related variables were chosen for use in CCA analyses: pH, Calcium (Ca), total organic Carbon (TOC) and inorganic Aluminium (inorg. Al) concentration. These variables were generally not strongly autocorrelated (though inorg. Al and pH were correlated in some years), and reflect different aspects of the acid status of streams: low pH and high inorganic Aluminium are stressful for acid sensitive taxa, whilst TOC (strongly correlated with dissolved organic carbon) and Ca reflect the capacity of the water to buffer such stressful effects. Ca was chosen rather than alkalinity as this is the variable directly manipulated by the liming program, and Ca was less correlated with pH than alkalinity. These chemistry variables were range-standardised prior to analysis so that

all varied on the same scale, whilst species abundance data were natural log transformed. The following parameters were recorded from each analysis:

a. Number of significant axes extracted: a Monte-Carlo randomization test was used to evaluate the null hypothesis of no relationship between a given ordination axis extracted during the CCA and the acidity data.
b. Percentage of variance in the species data explained by the significant axes: the percentage explained only by those axes selected as significant by the Monte-Carlo test.

A CCA ordination of species data that reflects well the acidity data should extract significant axes, and explain a relatively high percentage of variance.

2) A **Mantel Test**, also available in PC-Ord was used to assess the null hypothesis of no relationship between the species data for each sampling method and the same four acidity-related environmental variables assessed in the CCA (pH, Ca, TOC and inorg. Al). From these tests, the standardized Mantel statistic (r, analogous to a Pearson correlation coefficient) and the associated significance level were recorded.

3) **Weighted averaging (WA)** was used to model the relationship between species abundances and each of the acidity variables pH, Ca, TOC and inorg. Al. WA modeling is based on the idea of the ecological niche (ter Braak & Looman 1987) in that it assumes (i) that the response of a given species to a given environmental variable is unimodal, with an optimal point at which abundance is maximal, and (ii) that species are segregated according to the environmental variable (i.e. each species will have a different optimum). Such modeling can be used to describe a community's responses to a set of environmental variables, and consequently to predict environmental variables from species composition data, a widespread application of WA in paleaoecology (Birks 1998). Here, WA was used to predict pH, Ca, TOC and inorganic Al for each stream from the multivariate species data. Following this, the correlation (r^2) between the observed and predicted values for each acidity variable was calculated separately for both methods from each year. This correlation coefficient is higher when a data set is better able to predict the given environmental variable. Both WA models and correlations were generated using C2 (Version 1.4.2, © 2003-05, Steve Juggins and the University of Newcastle, UK), with application of tolerance down-weighting to control for the likelihood that niche widths are not equal for all species.

4) Separate Non-parametric **Kendall's tau-b correlations** between four acidity indices (Medins index, B:Pa, E:Pr, E:Pa) and selected acidity-related variables (pH, Ca, TOC and inorg. Al) were calculated for each sampling method. Non-parametric methods were used because of variability in the form of the relationships among the indices and variables.

5) The **acid sensitivity** of each taxa was scored according to the scheme used in the formulation of Medins index (current specifications: www.naturvardsverket.se). This scheme ranks taxa from 0-3, with 0 being acid insensitive and 3 highly sensitive. Taxa ranked 1-3 are hereafter termed "acid-sensitive taxa", with those ranked 2 or above (and thus having a strong influence on the value of Medins index) are additionally termed "highly sensitive taxa". For each year it was also noted whether each acid sensitive taxon was more common in M42 or Surber samples.

The ability of the two sampling methods to distinguish between limed and reference streams was investigated using univariate ANOVA and multivariate ANOSIM and SIMPER techniques. Indices calculated from the two sampling methods for each year were subjected to one-way ANOVA, with liming treatment fitted as the between subjects factor, and streams as replicates. Indices were transformed where necessary to satisfy parametric assumptions. Differences in species assemblages attributable to liming were assessed using ANOSIM, with streams treated as replicates, while SIMPER was used to quantify percent dissimilarity between limed and control stream assemblages, and identify those species explaining most of the dissimilarity.

IV) Objective 3: assessment of seasonal effects on sampling

The same analyses used to distinguish the quality and performance of data collected using Surber and M42 sampling were also used to distinguish data collected in autumn and spring. Data were available for a variable subset of streams from Spring 2000-2002, but this comprised 4-7 streams with only 1-2 reference streams, and is thus not comparable with data collected during autumn over those years. Accordingly, seasonal analyses focused only on the autumn 2004 and spring 2005 data, when a full set of limed and reference streams were sampled. In ANOVA models, "season" was fitted as a repeated measures factor within streams (subjects).

V) Significance levels

The emphasis of analyses presented here is on distinguishing the performance of different sampling methods, rather than strictly testing hypotheses about differences between groups. Many hypotheses tests conducted as part of these analyses (e.g. the CCA Monte-Carlo test, Mantel and Kendall's tau correlation tests, ANOSIM and ANOVA) lacked power because (i) most of the data were relatively noisy, (ii) the number of replicates was often low, and (iii) many tests involved relatively insensitive non-parametric statistics. Accordingly, for all analyses involving hypothesis testing, tests with significance levels ranging up to the 10% level ($p = 0.1$) are emphasised. This is not to suggest that significance levels greater than 0.5 should be used to reject null hypotheses, but is done merely so that cases in which strong but marginally significant (>0.05-0.1) trends exist are not overlooked in assessing the success of the sampling methods.

Results

Water chemistry: interannual variation

A detailed analysis of water chemistry data is beyond the scope of this report. Nevertheless, some aspects of the water chemistry data are highlighted, in order to set the context for later analyses:

1) Two reference streams, Laxbäcken (sampled 1994-2005) and Lillån Bosgård (sampled 2000-2005) were chronically acid (low pH and alkalinity, high inorganic Al), in contrast with the remaining reference streams, which were circumneutral (see Appendix 1). Such extremely low pH streams could be of value in testing the performance of the two sampling methods over a broader range of environmental variation, but the effect of having only two streams of this type (only one prior to 2000) was to create strong outliers, which distorted correlations and suppressed patterns elsewhere in the data. Accordingly, these two streams were excluded from all analyses, in order that the reference streams be as uniform as possible in important acidity related characteristics.
2) In most years, mean pH in limed streams was comparable to that in reference streams, and was never below 6 (Fig. 1a), indicating the success of liming in maintaining non-acid conditions. However, minimum pH values were sometimes below 6, both in some reference and limed streams (see Appendix 1). Interannual variation in pH was not marked. Similar generalizations can be made about other pH-related variables (alkalinity, TOC). However, inorganic Al concentrations, though differing little between limed and reference streams, did vary more from year to year (Fig 1b). Note liming appeared to reduce variation in most acid-related variables (e.g. Fig. 1).

Interannual variation: community metrics and indices

There was little interannual variation in the number of individuals (Fig. 2a) or number of species (Fig. 2b) collected from Surber samples, but there was substantial variation for M42 samples. From 1994-99 and in 2004, more individuals and species were collected from M42 than Surber samples. The reverse was true from 2000-2001, with richness and abundance collected from M42 samples much lower than observed in previous years (Fig. 2). Consequently, abundances and richness collected in M42 samples dropped below the numbers collected in Surber samples during 2000-02, which remained constant through these years. More taxa and individuals were collected from M42 samples in 2002, leading to greater equality between the two sample methods (Fig 2). The extremely high mean for abundance in 1994 (Fig. 2a) is attributable to two unusually high readings for chironomid abundance in two streams.

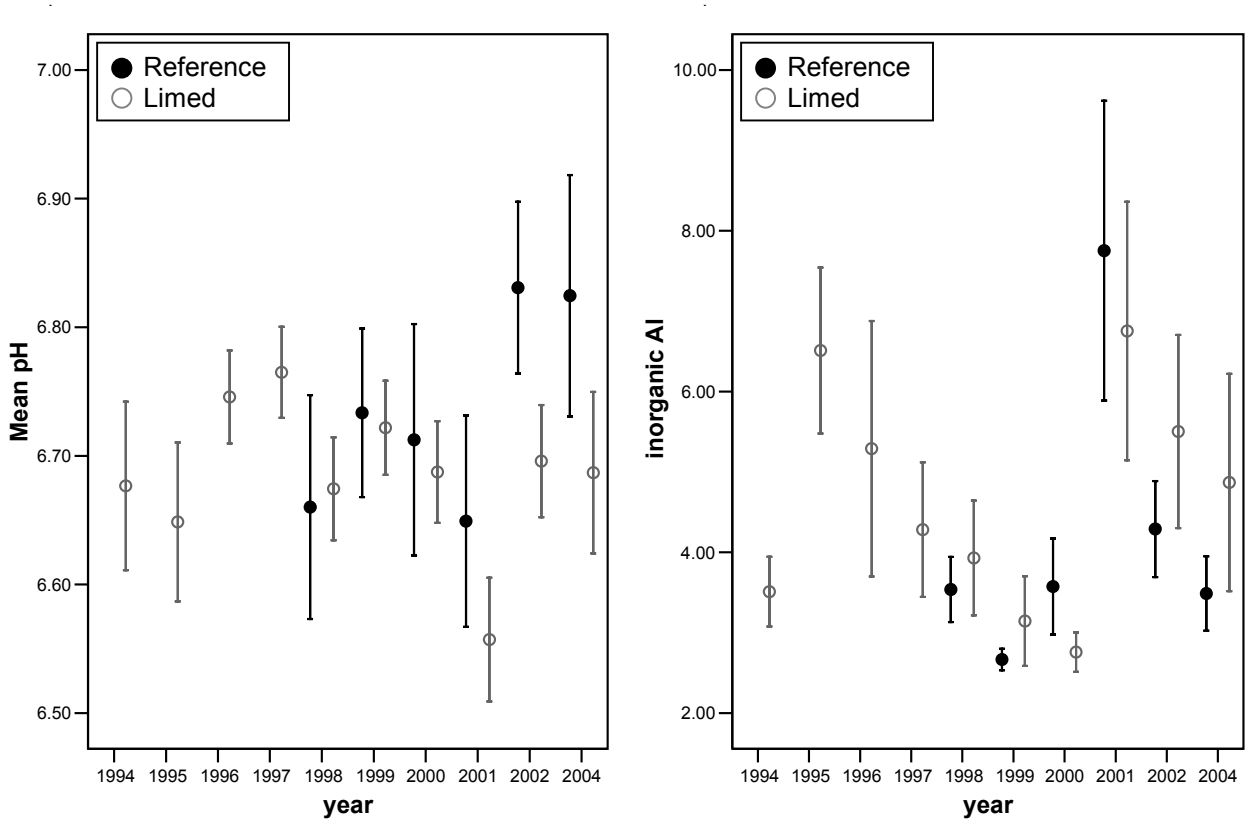

Figure 1. interannual variation in water chemistry from limed and reference streams: (a) mean pH; (b) mean inorganic aluminium concentrations (mean ± SE plotted).

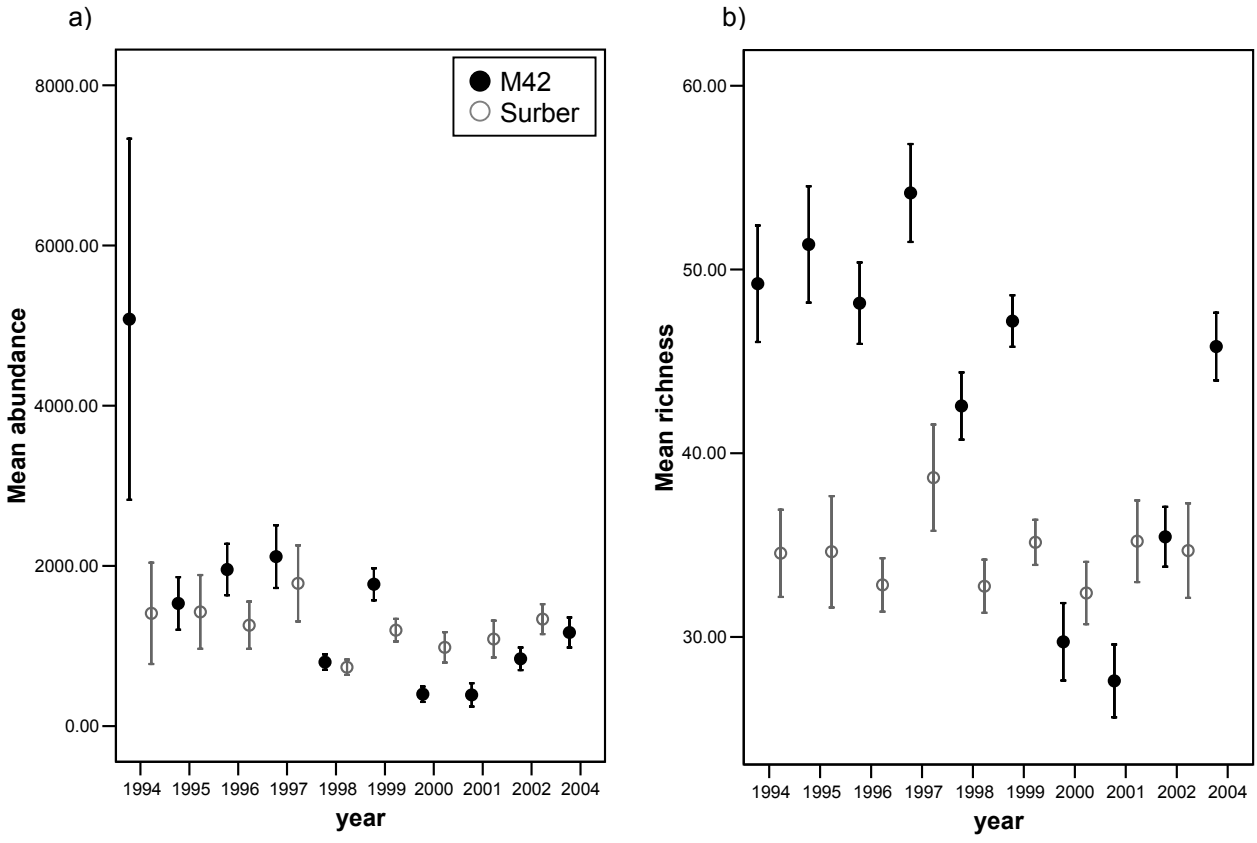

Figure 2. interannual variation in (a) mean animal abundance and (b) mean species richness per stream from the M42 (closed circles) and Surber (open circles) samples (mean ± SE plotted).

The extent to which these contrasts are reflected in data for other indices and metrics varies. For example, the B:Pa ratio calculated from M42 data was little affected by the lower numbers of individuals and species collected from 2000-2002 (Fig. 3a), but Medin's index was reduced for M42 samples from these years (Fig 3b). Mean values for the E:Pr index were also reduced from 2000-2002, but E:Pa was unaffected, whilst Shannon diversity was increased (data not plotted).

NMDS Ordinations and cluster analyses

Ordinations and cluster analyses gave broadly similar results from year to year, though there were differences in details. The analyses presented in Figs 4-5 from 2000 are typical. Whilst there is no distinct separation of limed and unlimed streams in the ordination (Fig. 4), there is a tendency for the reference streams to occur towards the top left hand corner of the chart (See Fig. 25 for further examples of this pattern). In both the ordination and cluster analyses, there is little consistency in the similarity of M42 and Surber samples from single streams (Figs. 4-5). Thus whilst the two samples from Stråfulån (Stråf in Fig. 5) are indistinguishable on the same terminal branch in the cluster analysis, samples from other streams (e.g. Gnyltån, "Gnylt" in Fig. 5) occur on widely separated terminal branches. However, in the ordination, the direction of offset between paired M42 and Surber samples follows the bottom right-top left diagonal in most cases (Fig. 4). ANOSIM results confirm a general difference in the composition of M42 and Surber samples (Rho = 0.195, $p < 0.001$).

Sampling method assessment – SIMPER analysis

Output from a preliminary SIMPER analysis of the effects of sampling method on assemblage composition are presented in Table 1. Notable are the greater abundances of several acid sensitive mayflies in the Surber samples (Table 1).

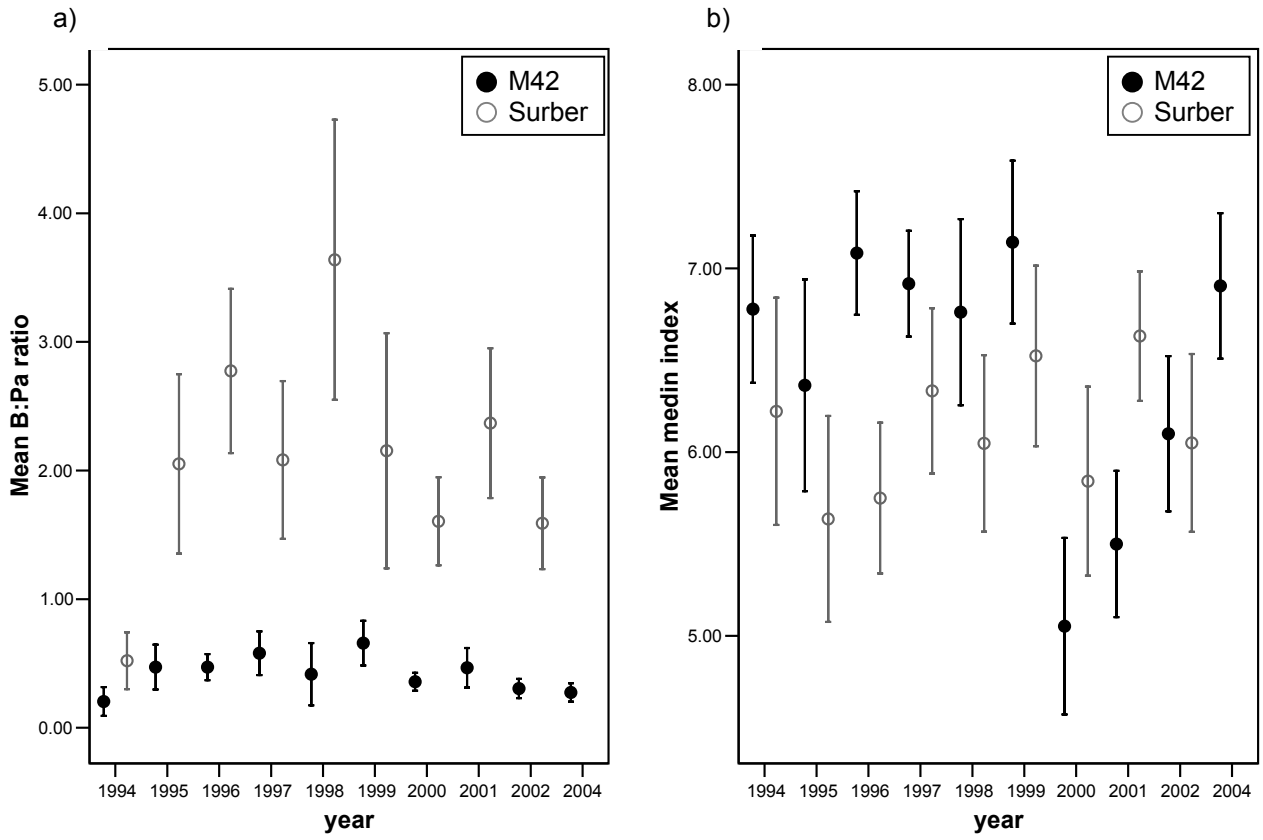

Figure 3. Interannual variation in the mean value of (a) the B:Pa ratio and (b) Medin's index per stream from the M42 and Surber samples (mean ± SE plotted).

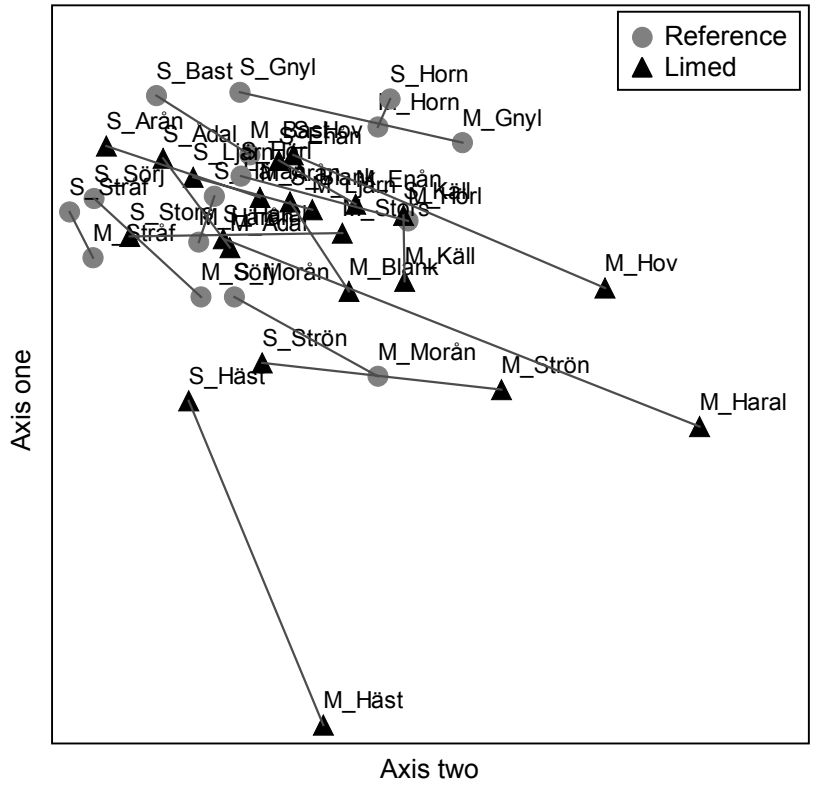

Figure 4. nMDS ordination of benthic invertebrate data from autumn 2000, with liming categories superimposed. The letters "M" and "S" preceding the stream names refer to whether the data was collected using M42 or Surber samples respectively. The lines join M42 and and Surber samples from a single stream. Ordination in 3 dimension, axes 1 and 2 plotted. Stress = 11.89.

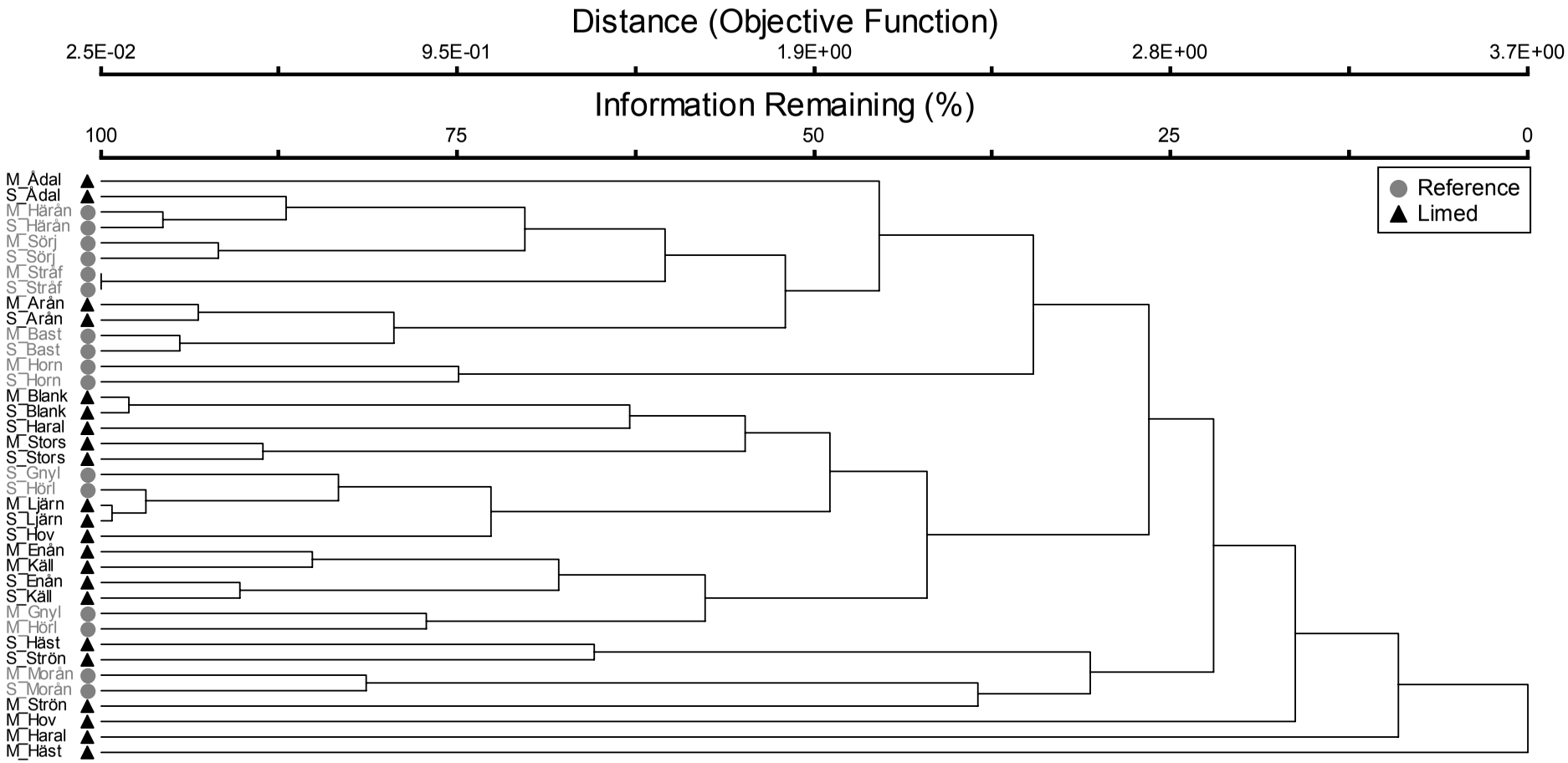

Figure 5. UPGMA cluster analysis of benthic invertebrate data from autumn 2000, with liming categories superimposed. The letters “M” and “S” preceding the stream names refer to whether the data was collected using M42 or Surber samples respectively.

Table 1. Output from SIMPER analysis of the difference in species composition between M42 and Surber samples, with data pooled across years. Listed are taxa that collectively explain 75% of the dissimilarity between sample groups, together with their acid sensitivity rank (as scored for Medins index), their mean abundance from the two sample types, and their contribution to the dissimilarity. Mean dissimilarity 57.44.

Taxon	Medins rank	M42 mean abundance	Surber mean abundance	% contribution to dissimalarity	Cumulative %
Elmis aenea	1	43.92	78.19	2.2	2.2
Nigrobaetis niger	1	13.35	51.13	2.18	4.38
Limnius volckmari	1	29.41	48.99	2.16	6.54
Tanytarsini		121.52	168.62	2.12	8.65
Baetis rhodani	1	41.13	130.4	2.12	10.77
Chironomini		288.17	71.79	2.09	12.86
Leuctra hippopus		25.8	35.35	2.02	14.88
Amphinemura borealis	1	31.74	57.83	1.98	16.86
Nemoura avicularis	1	31.66	3.99	1.87	18.74
Heptagenia sulphurea	1	18.19	17.67	1.85	20.59
Leptophlebia marginata		37.57	4.52	1.83	22.42
Protonemura meyeri		42.08	31.34	1.83	24.25
Limnephilidae		31.98	5.69	1.82	26.07
Simuliidae		23.9	55.97	1.81	27.88
Orthocladiinae		129.01	107.31	1.81	29.7
Tanypodinae		50.73	33.53	1.75	31.45
Polycentropus flavomaculatus		24.51	15.4	1.73	33.18
Hydracarina		9.08	17.24	1.71	34.89
Hydropsyche siltalai		17.65	12.99	1.71	36.6
Agapetus ochripes	2	16.03	20.44	1.7	38.3
Sericostoma personatum	1	13.43	15.9	1.68	39.98
Sphaeriidae	1	28.34	9.93	1.64	41.63
Oulimnius tuberculatus	1	8.75	13.99	1.54	43.16
Taeniopteryx nebulosa	1	15.69	6.01	1.53	44.69
Empididae		7.69	10.26	1.42	46.12
Naididae		3.47	13.06	1.42	47.53
Oxyethira sp.	1	18.7	5.28	1.39	48.93
Isoperla sp.		6.9	9.84	1.36	50.28
Ephemerella aurivillii		19.2	5.2	1.35	51.63
Rhyacophila nubila		10.5	6.3	1.35	52.98
Lepidostoma hirtum	1	7.29	10.56	1.33	54.32
Ceratopogonidae		10.64	9.22	1.3	55.61
Diura nanseni	1	7.55	2.97	1.29	56.9
Enchytraeidae		4.31	6.2	1.27	58.17
Lumbriculidae		4.63	6.59	1.25	59.43
Hydraena sp.		0.57	6.64	1.21	60.64
Hexatominae		5.35	0.7	1.16	61.8
Pediciinae		4.57	3.99	1.14	62.95
Psychodidae		3.24	5.13	1.14	64.08
Hydropsyche pellucidula	1	4.54	2.91	1.11	65.19
Leuctra fusca/digitata		6.8	8.88	1.1	66.29
Asellus aquaticus	1	12.91	2.15	1.09	67.38
Hydroptila sp.	2	21.55	4.5	1.05	68.43
Brachyptera risi		2.46	7.92	1.02	69.45
Nematoda		4.01	1.85	0.99	70.43
Centroptilum luteolum	1	11.63	3.48	0.95	71.38
Capnopsis schilleri		2.93	4.32	0.94	72.32
Heptagenia fuscogrisea		11.13	1.47	0.92	73.24
Eiseniella tetraedra		2.39	1.82	0.87	74.11

Sampling method assessment – data quality

Because of the variation in M42 data apparent between years (see Fig. 2), graphical data relating to method assessment is here presented in two ways. First, a line chart plots differences between the M42 and Surber samples from year to year, and second, a mean chart averages data for the two methods across all years.

I) Canonical correspondance analysis. Detailed output from CCAs is given in Table 2. CCA of M42 data extracted more significant components for 1998 and 1999; thereafter the same number was extracted from both Surber and M42 data (Fig. 6a). Averaged across years, more significant components were extracted from M42 data (fig. 6b). Similar patterns were apparent for the percentage of significant variance explained. A higher percentage of variance was explained by significant components extracted from M42 data for 1998 and 1999, thereafter the differences between the methods was smaller – Surber data performed slightly better from 2000-01, whereas M42 data performed slightly better in 2002 (Fig. 7a). Averaged across all years, mean percent variance explained by the significant axes was higher for M42 data (Fig. 7b).

In most CCA ordinations for both M42 and Surber samples, there was a general gradient apparent along opposing biplot axes for pH and TOC (e.g. Fig 8, for another example see Fig. 27a). For all years except 2001, there was little general difference between M42 (Fig. 8a) and Surber (Fig. 8b) sampling, based on CCA ordinations.

II) Mantel correlation. Signed correlations and significance levels for Mantel's test are given in Table 2. In most years, the Mantel correlation was greater for M42 than Surber data (Fig. 9a). Averaged across all years, the correlation was greater for M42 (Fig. 9b). However, all coefficients were low, and none were statistically significant (Table 2).

III) Weighted averaging. Correlation coefficients between observed values for pH, Ca, TOC and inorganic aluminium and values modelled from macroinvertebrate data are given in Table 2. The capacity of M42 and Surber sample data to model pH data appeared similar (Fig. 10a-b), but for the remaining three variables M42 generally performed better – M42 correlations were higher for Ca after 1999 (Fig. 11), for TOC after 1998 (Fig. 12), and for inorganic Al in all years (Fig. 13)

Table 2. Method Assessment: output from Canonical Correspondance Analyses (the number of significant components, the associated significance values, and proprtion of variance in the species data explained by the significant components), Mantel Correlations (correlation coefficient and significance level) and Weighted Averaging correlations (correlations between observed values of the variable and values modeled from the species data).

		Canonical Correspondance Analysis			Mantel Correlation		Weighted Averaging Correlations			
Year	Method	# Significant components	Significance levels	Significant variance explained	R	Sig	pH	Ca	TOC	Inorganic Al
1998	M42	2	0.013, 0.0933	17.1	-0.046	0.395	0.50	0.78	0.56	0.48
1998	Surber	1	0.0867	11.3	-0.01	0.493	0.49	0.77	0.70	0.45
1999	M42	2	0.033,0.033	17.3	0.1021	0.255	0.55	0.72	0.66	0.86
1999	Surber	0		0	-0.041	0.463	0.45	0.72	0.52	0.66
2000	M42	1	0.013	6.5	0.015	0.379	0.56	0.80	0.85	0.84
2000	Surber	1	0.04	11.2	-0.009	0.543	0.63	0.71	0.63	0.75
2001	M42	2	0.026, 0.0067	14.4	0.161	0.159	0.87	0.76	0.83	0.90
2001	Surber	2	0.0067. 0.004	18.4	0.027	0.38	0.84	0.72	0.55	0.83
2002	M42	1	0.067	8.5	0.084	0.316	0.49	0.80	0.70	0.75
2002	Surber	1	0.08	7	-0.092	0.367	0.72	0.65	0.69	0.72

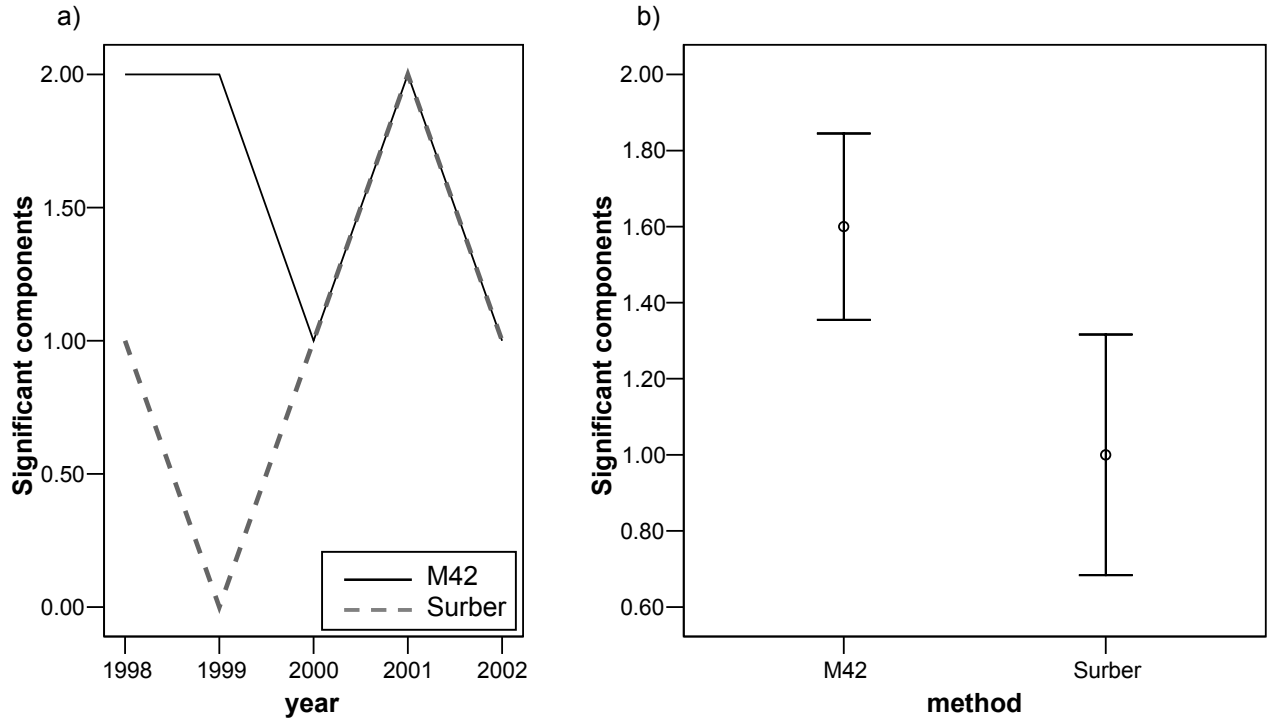

Figure 6. Method assessment CCA analysis: the number of significant components extracted from M42 and Surber sample data (a) per year and (b) averaged across all years (mean ± SE plotted).

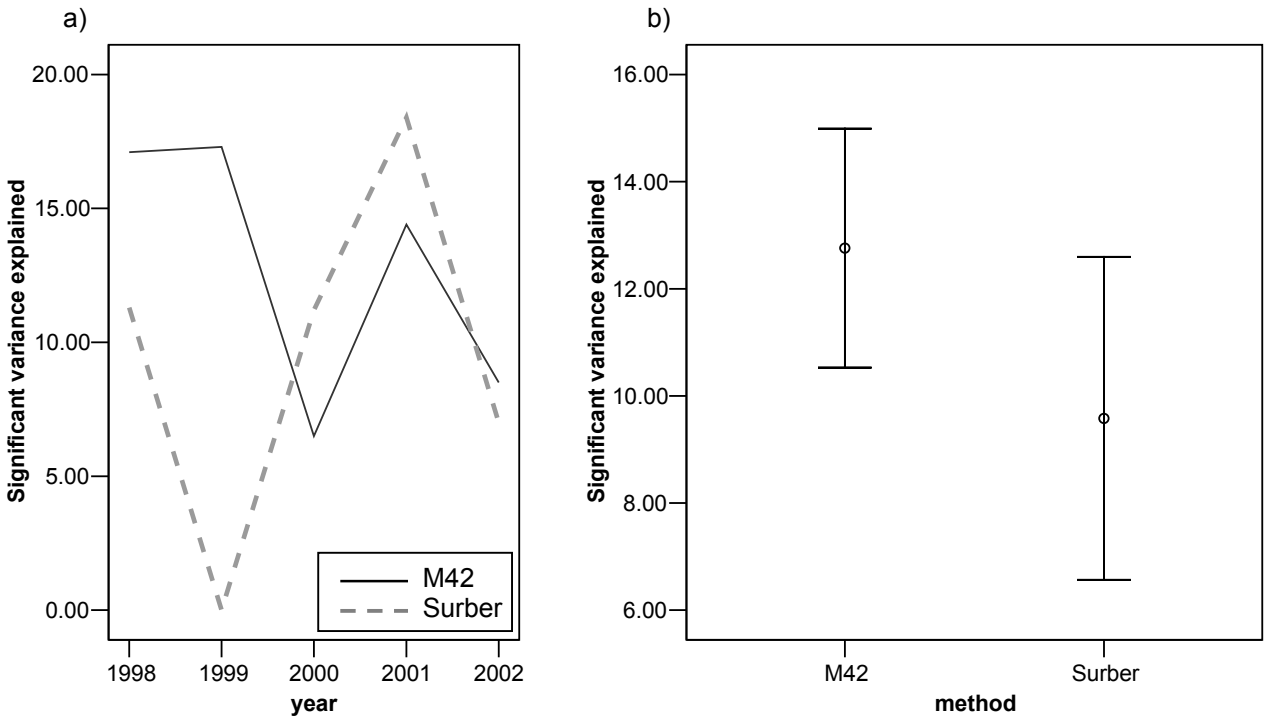

Figure 7. Method assessment CCA analysis: percentage variance explained by significant components extracted from M42 and Surber sample data (a) per year and (b) averaged across all years (mean ± SE plotted).

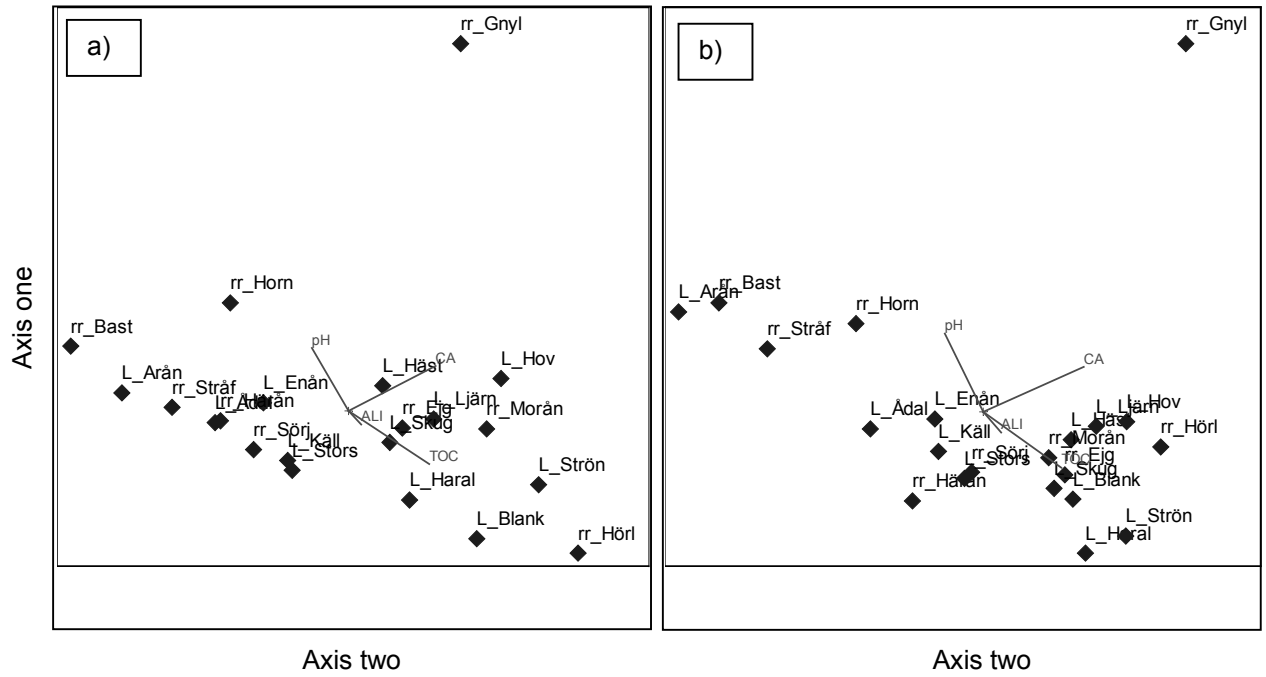

Figure 8. CCA ordination of benthic invertebrate data from autumn 1999, from the (a) M42 and (b) Surber samples. The letters “L” and “rr” preceding the stream names refer to whether the stream was limed or unlimed (reference) respectively. Percent variance explained: (a) 17.3 (both axes significant), (b) 17.5 (neither axis significant)

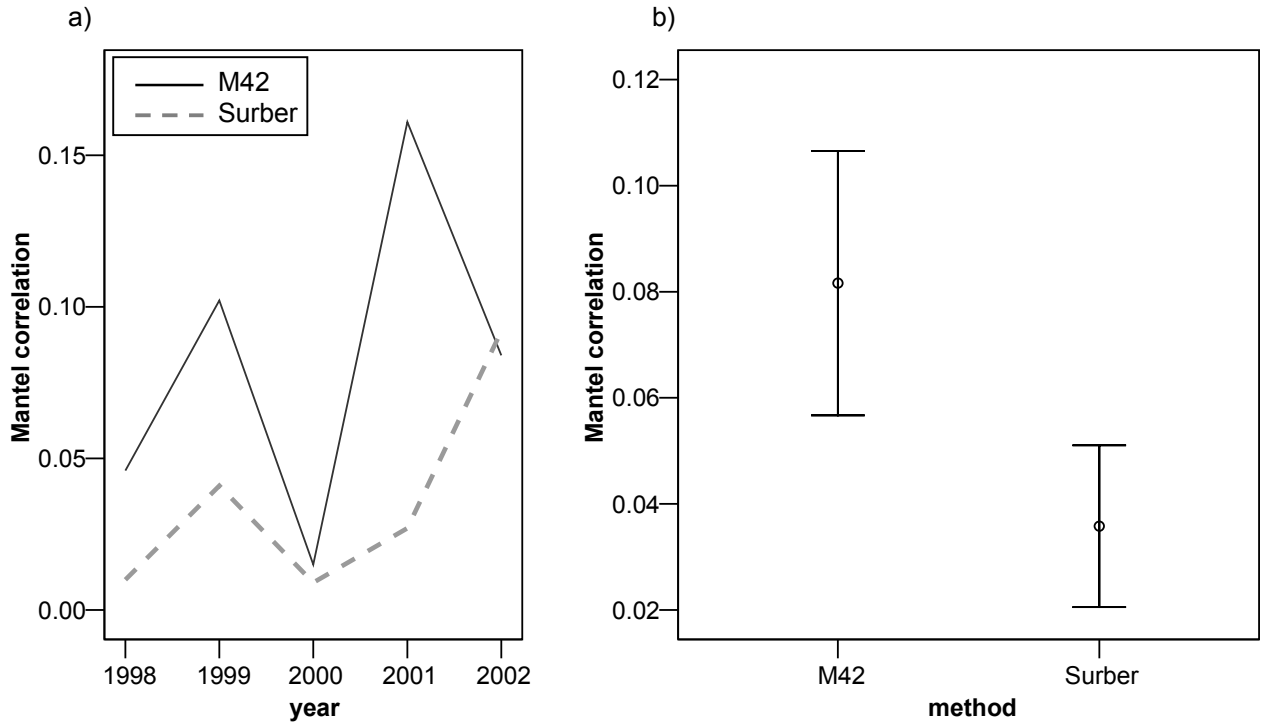

Figure 9. Method assessment Mantel test: correlation between acidity characteristics and multivariate species data from M42 and Surber samples (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 1.

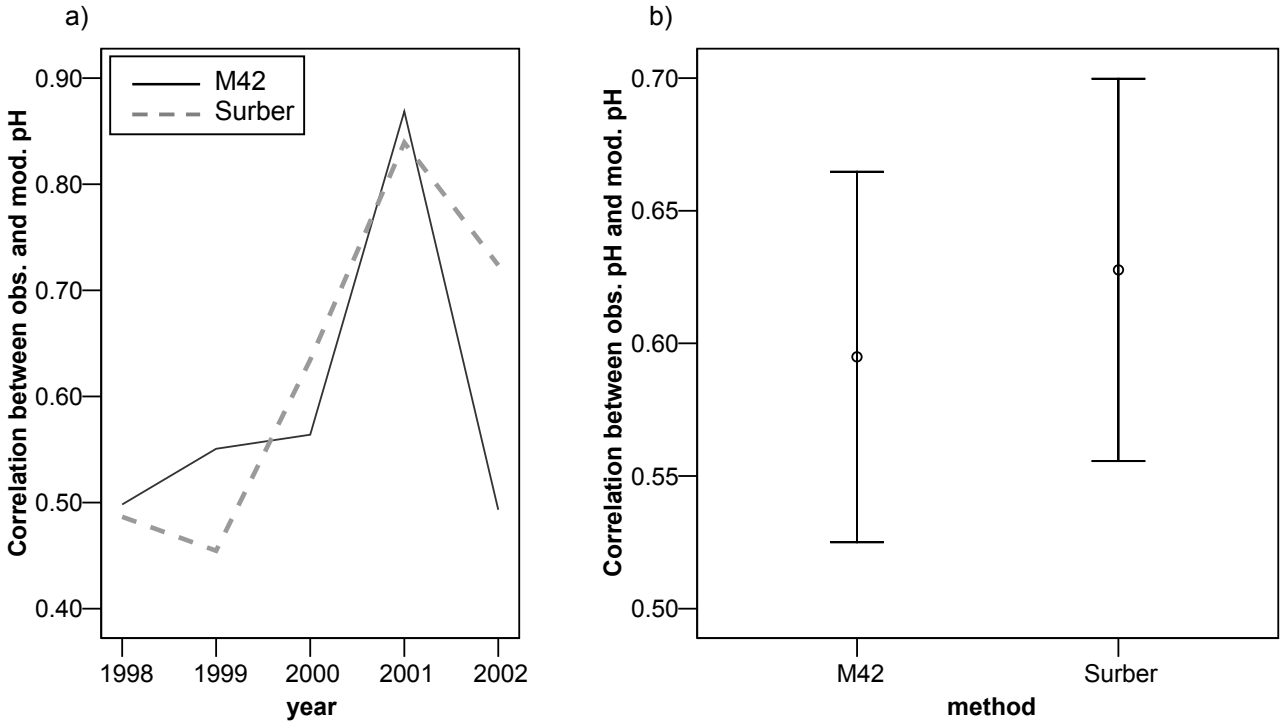

Figure 10. Method assessment Weighted Averaging: correlation between observed (obs.) pH and pH modeled (mod.) from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted).

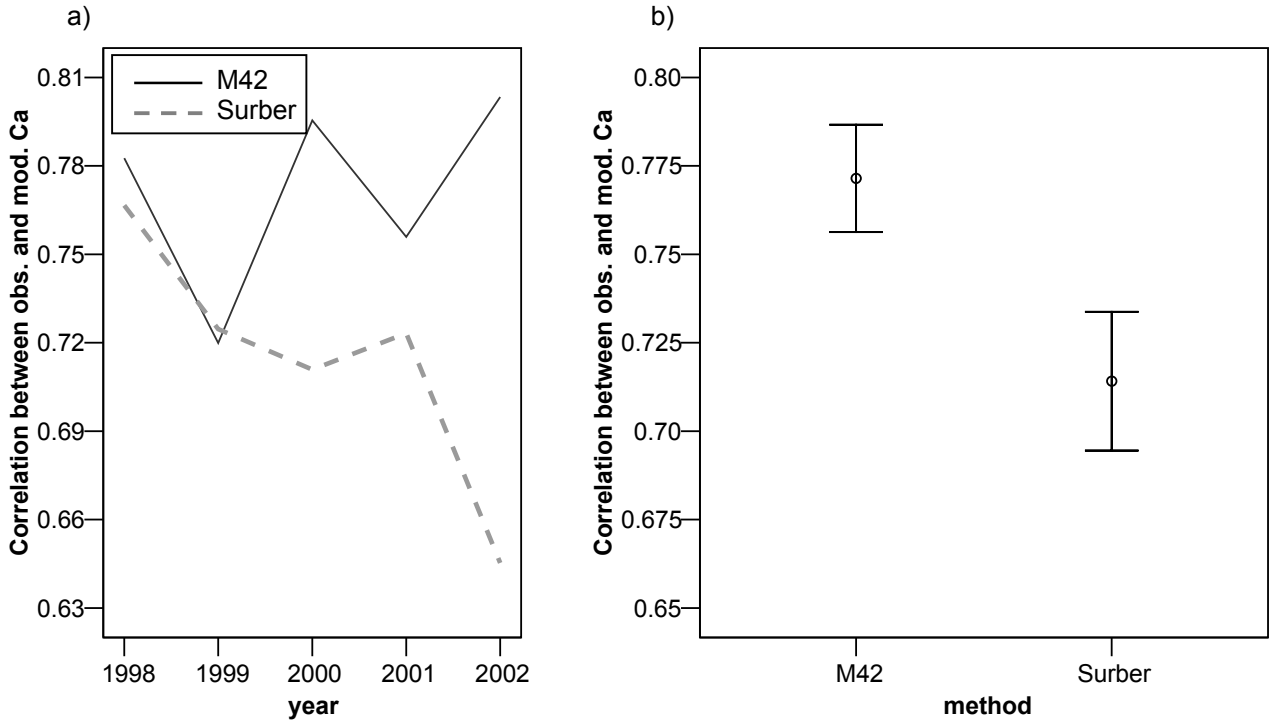

Figure 11. Method assessment Weighted Averaging: correlation between observed (obs.) Ca and Ca modeled (mod.) from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted).

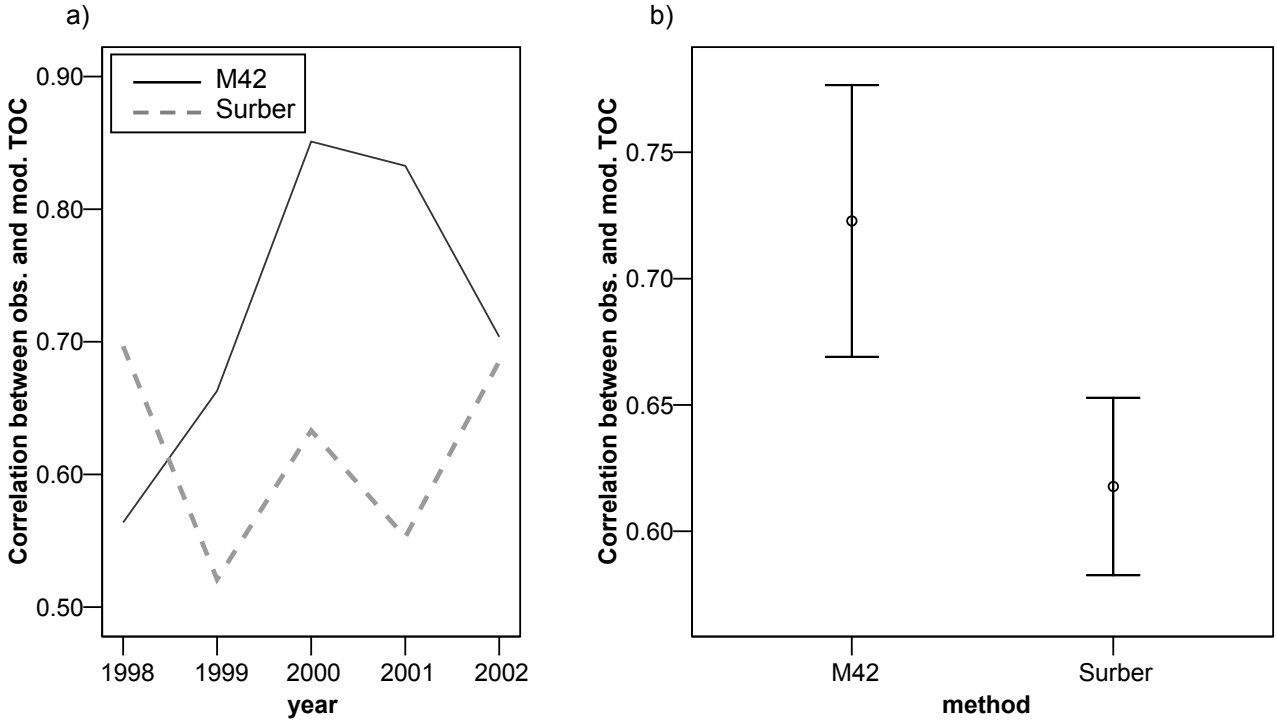

Figure 12. Method assessment Weighted Averaging: correlation between observed (obs.) TOC and TOC modeled (mod.) from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted).

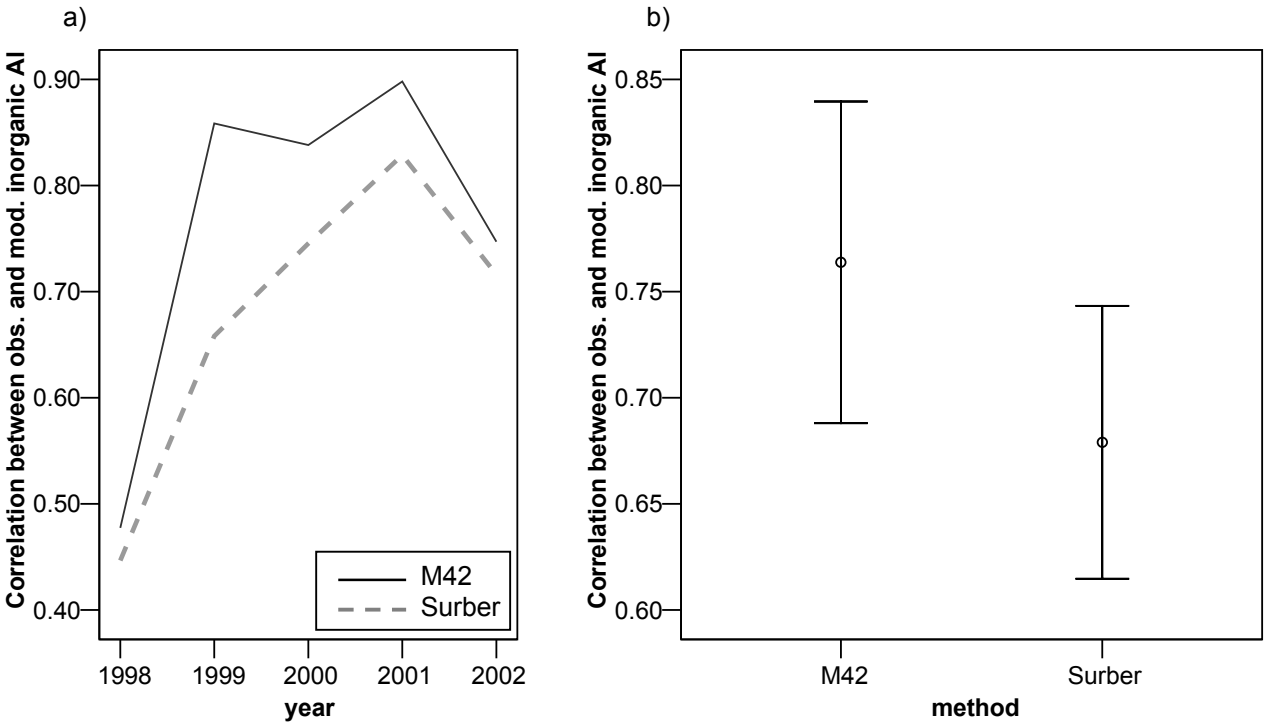

Figure 13. Method assessment Weighted Averaging: correlation between observed (obs.) inorganic Al and inorganic Al modeled (mod.) from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted).

IV) Biotic indices: correlations with acidity-related variables. Full output from correlation analyses are presented in Table 3. The best performing index (most significant correlations using both M42 and Surber data) was B:Pa (Table 3). Correlations between pH and B:Pa were similar for both M42 and Surber data from year to year (Fig 14a) and averaged across years (Fig. 14b). M42 performed better in 4 of 5 years for Ca (Fig 15a), in 2 of 5 years for TOC (Fig. 16a), and 3 of 5 years for inorganic Aluminium (Fig. 17a), resulting in higher means for M42 average across all years for all three variables (Figs 15b-17b). Medins index was well correlated with Ca in 5 of 10 cases (Table 3). In 3 of 5 years, correlations between Ca and Medin's index were greater when calculated from M42 data (Fig. 18a), though when averaged across all years, there was little to differentiate the two methods (Fig 18b). E:Pr was well correlated with TOC and inorganic Al (significant in 6-7 of 10 cases). The correlation between E:Pr and TOC was not well differentiated according to method (Fig. 19). For inorganic Al, the correlation was greater for Surber sample data in 3 of 5 years (Fig. 20a), with a slightly greater mean correlation when averaged over all years (Fig. 20b). Correlations between inorganic Al and the E:Pa index were significant in 5 of 10 cases, but there was no clear differentiation between the M42 and Surber methods (Table 3). The correlation between E:Pa and pH was significant in four of 10 cases, with Surber data tending to be better correlated (Table 3, means similar to Fig. 20b).

Across all four acidity-related response variables, there were 18 cases where correlations for both M42 and Surber data were significant (Table 3). In 12 of these 18 cases, the correlation coefficient was greater for M42 data (Table 3).

Table 3. Method assessment: non parametric Kendall's Tau-b correlations between four biotic indices and four acidity related variables. For each variable, both the coefficient and significance level are indicated, with p-values < 0.01 highlighted in bold.

			n	pH		Ca		TOC		Al inorg	
Index	Year	Method		tau	sig.	tau	sig.	tau	sig.	tau	sig.
Medins	1998	M42	21	0.167	0.311	0.57	**0.001**	0.198	0.231	0.068	0.688
	1998	Surber	21	0.093	0.578	0.479	**0.004**	0.195	0.24	0.221	0.191
	1999	M42	21	-0.01	0.951	0.638	**<0.001**	0.391	**0.019**	-0.215	0.21
	1999	Surber	21	0.137	0.407	0.421	**0.011**	0.056	0.735	-0.074	0.663
	2000	M42	19	0.068	0.695	0.205	0.239	0.081	0.643	-0.179	0.323
	2000	Surber	19	0.118	0.498	0.28	0.109	0.019	0.915	-0.186	0.305
	2001	M42	18	-0.03	0.873	0.12	0.522	0.045	0.81	0.015	0.936
	2001	Surber	19	0.205	0.248	0.27	0.129	-0.116	0.516	-0.303	**0.09**
	2002	M42	20	0.158	0.354	0.327	**0.055**	0.079	0.643	0.056	0.74
	2002	Surber	20	-0.011	0.947	0.257	0.129	0.112	0.509	0.129	0.448
E:Pr	1998	M42	21	0.332	**0.071**	0.406	**0.028**	0.407	**0.027**	0.388	**0.035**
	1998	Surber	21	0.145	0.364	0.126	0.431	0.179	0.262	0.218	0.173
	1999	M42	21	0.26	0.155	0.26	0.155	0.439	**0.016**	0.439	**0.016**
	1999	Surber	21	0.191	0.236	0.259	0.107	0.347	**0.031**	0.259	0.107
	2000	M42	19	0.259	0.181	0.418	**0.031**	0.337	**0.081**	0.322	**0.096**
	2000	Surber	19	0.373	**0.027**	0.399	**0.019**	0.55	**0.001**	0.456	**0.007**
	2001	M42	18	0.033	0.866	0.025	0.899	-0.017	0.933	-0.099	0.612
	2001	Surber	19	0.334	**0.049**	0.282	**0.098**	0.299	**0.079**	0.143	0.399
	2002	M42	20	0.278	0.133	0.293	0.114	0.414	**0.026**	0.444	**0.017**
	2002	Surber	20	0.07	0.671	0.2	0.227	0.113	0.493	0.297	**0.072**
E:Pa	1998	M42	21	0.349	0.042	0.159	0.356	-0.109	0.524	-0.129	0.461
	1998	Surber	21	0.41	**0.009**	0.081	0.608	-0.076	0.629	-0.332	**0.039**
	1999	M42	21	0.086	0.621	-0.109	0.531	0.017	0.921	-0.167	0.349
	1999	Surber	21	0.248	0.116	-0.181	0.251	-0.276	**0.08**	0.015	0.927
	2000	M42	19	0.142	0.457	-0.158	0.409	-0.237	0.215	-0.051	0.799
	2000	Surber	19	0.164	0.327	-0.282	**0.093**	-0.469	**0.005**	-0.031	0.857
	2001	M42	18	0.301	0.104	0.287	0.122	-0.007	0.968	-0.457	**0.014**
	2001	Surber	19	0.532	**0.001**	0.135	0.421	-0.111	0.506	-0.34	**0.042**
	2002	M42	20	0.436	**0.018**	0.248	0.178	0.127	0.489	-0.417	**0.024**
	2002	Surber	20	0.326	**0.044**	0.105	0.516	0.032	0.846	-0.47	**0.004**
B:Pa	1998	M42	21	0.435	**0.006**	-0.228	0.154	-0.134	0.41	-0.444	**0.005**
	1998	Surber	21	0.453	**0.004**	0.019	0.904	-0.289	**0.072**	-0.158	0.319
	1999	M42	21	0.301	**0.057**	-0.243	0.123	-0.279	**0.086**	-0.224	0.156
	1999	Surber	21	0.272	**0.085**	-0.31	**0.05**	-0.05	0.759	-0.291	**0.065**
	2000	M42	19	-0.043	0.805	-0.47	**0.006**	0.241	0.179	-0.494	**0.004**
	2000	Surber	19	0.131	0.44	-0.321	**0.058**	0.025	0.885	-0.487	**0.004**
	2001	M42	18	0.578	**0.001**	0.234	0.191	-0.338	**0.059**	-0.055	0.758
	2001	Surber	19	0.571	**0.001**	0.159	0.344	-0.342	**0.042**	-0.147	0.381
	2002	M42	20	0.414	**0.015**	-0.235	0.166	-0.544	**0.001**	-0.269	0.113
	2002	Surber	20	0.34	**0.037**	0.096	0.558	-0.421	**0.01**	-0.085	0.603

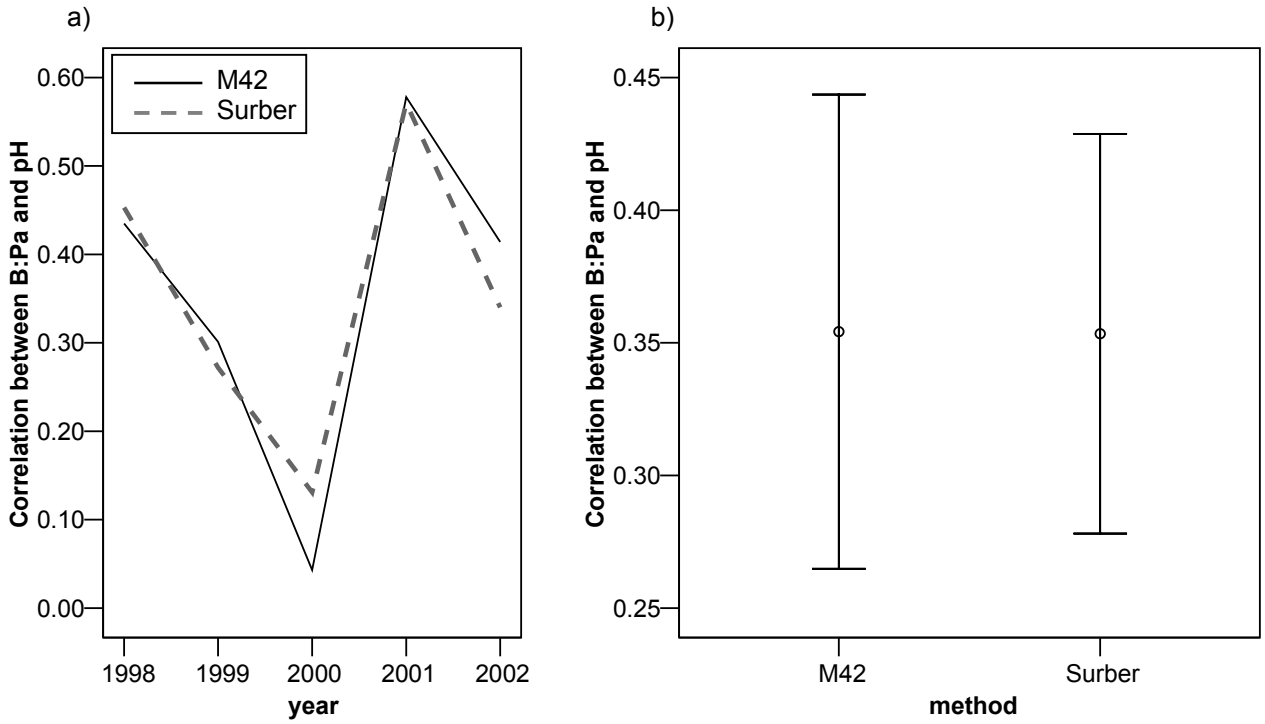

Figure 14. Method assessment Tau correlation: Tau correlation between pH and B:Pa calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted).). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

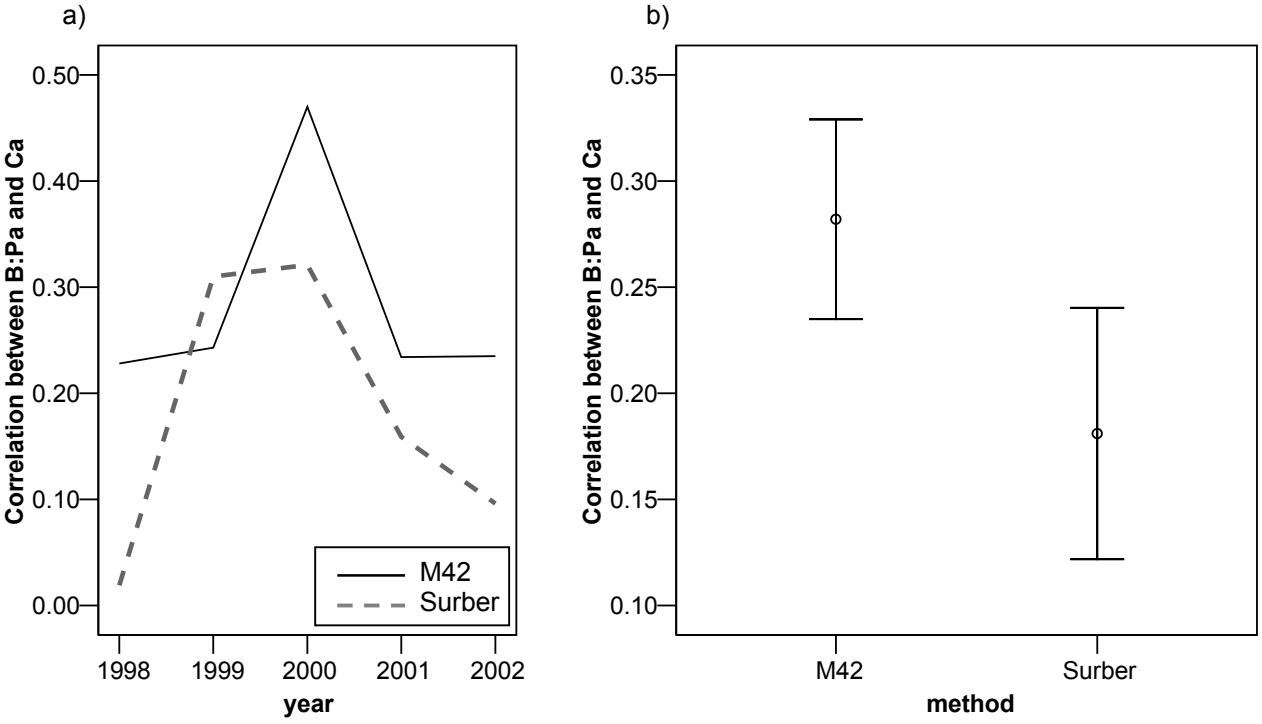

Figure 15. Method assessment Tau correlation: Tau correlation between Ca and B:Pa calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

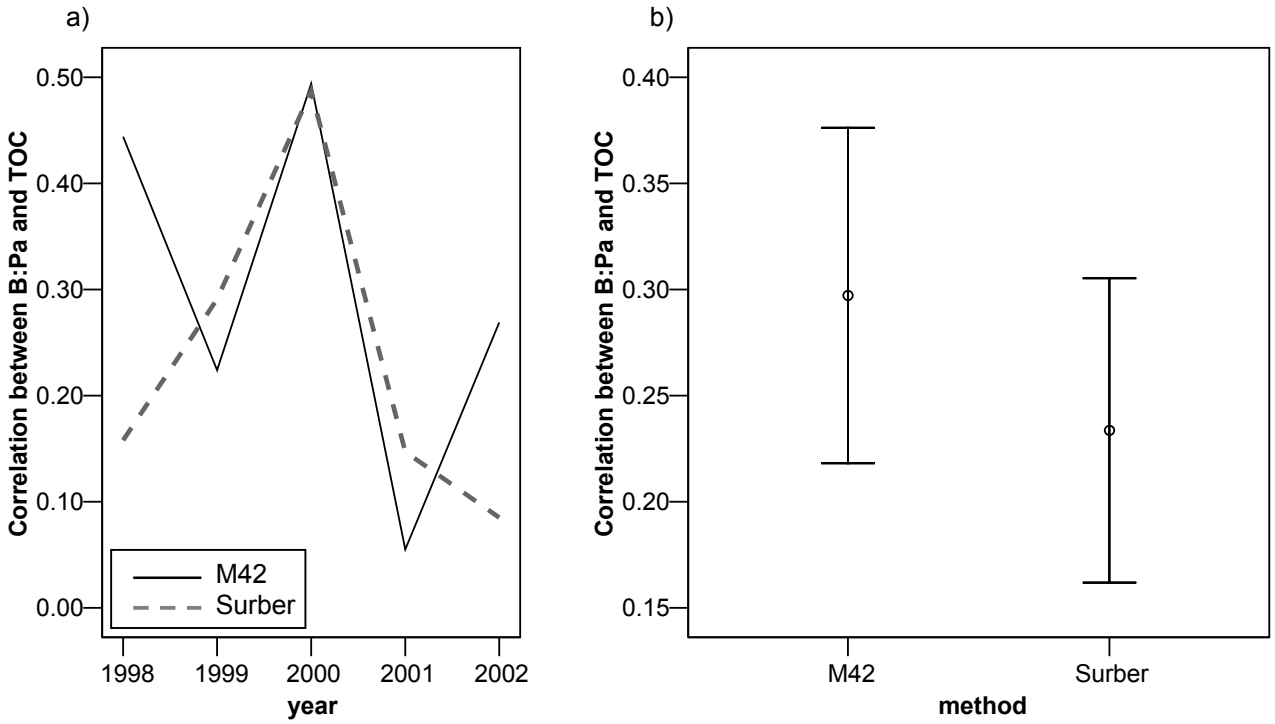

Figure 16. Method assessment Tau correlation: Tau correlation between TOC and B:Pa calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

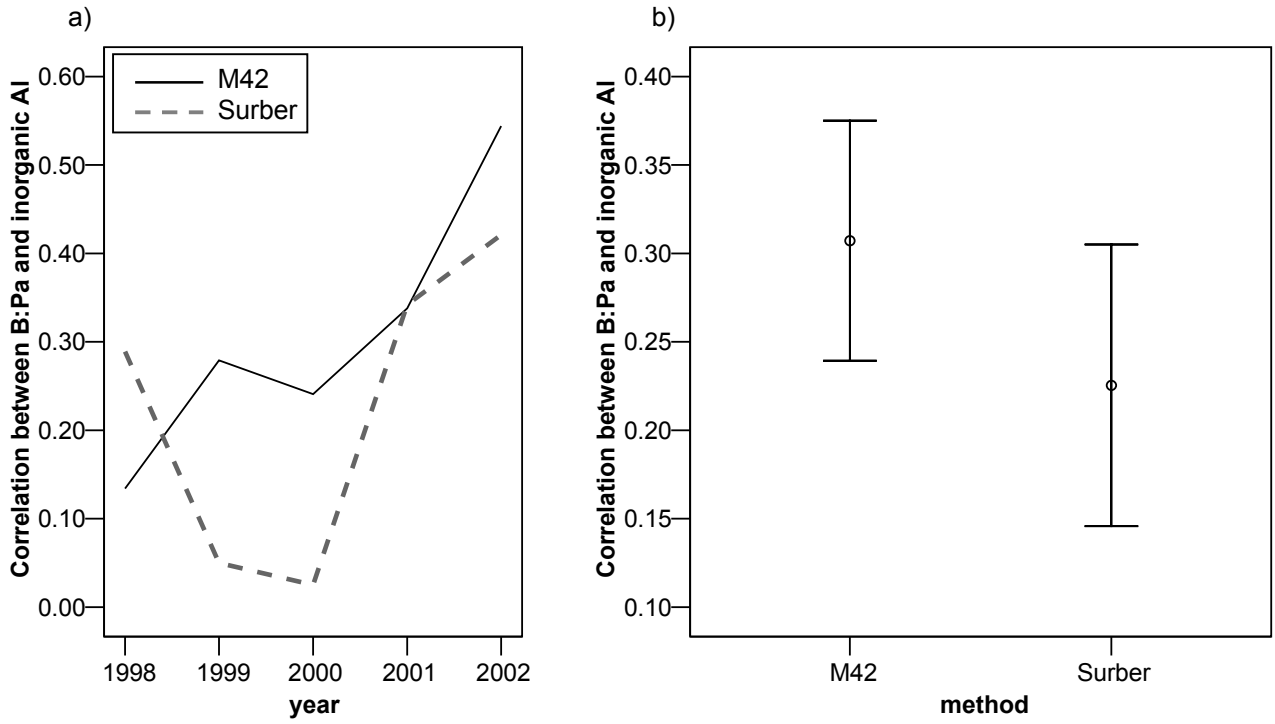

Figure 17. Method assessment Tau correlation: Tau correlation between inorganic Al and B:Pa calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

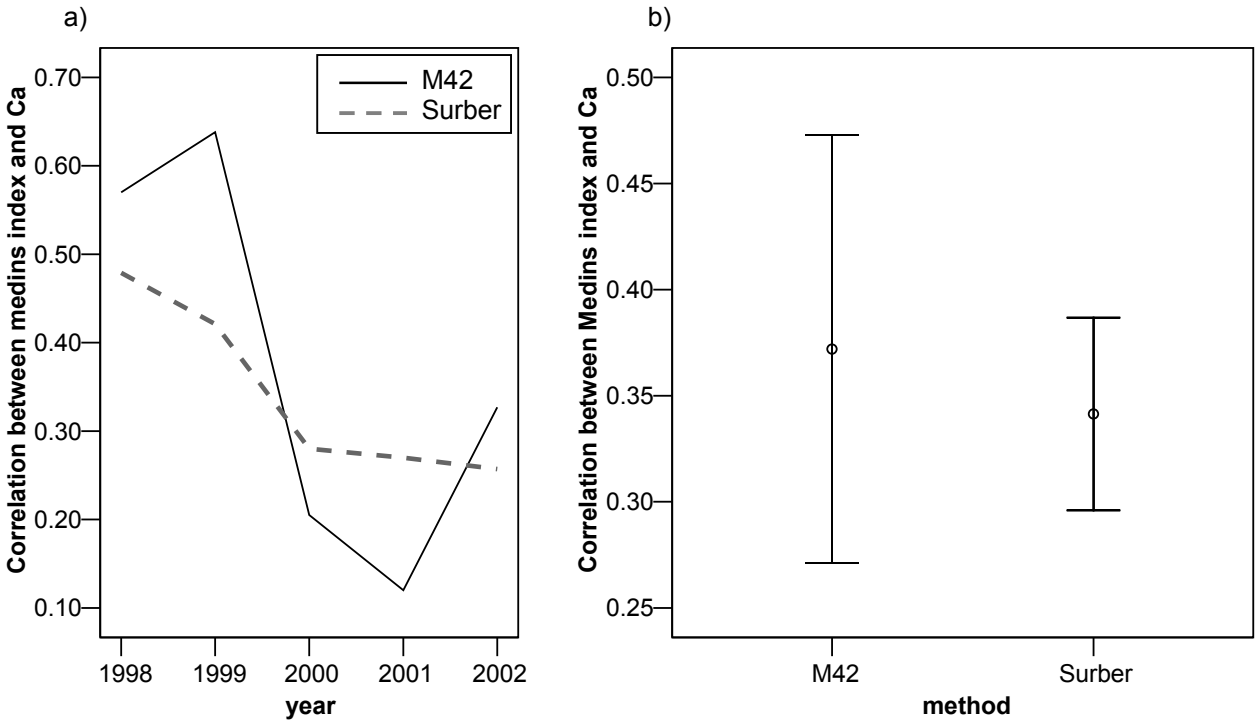

Figure 18. Method assessment Tau correlation: Tau correlation between Ca and Medins index calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 4.

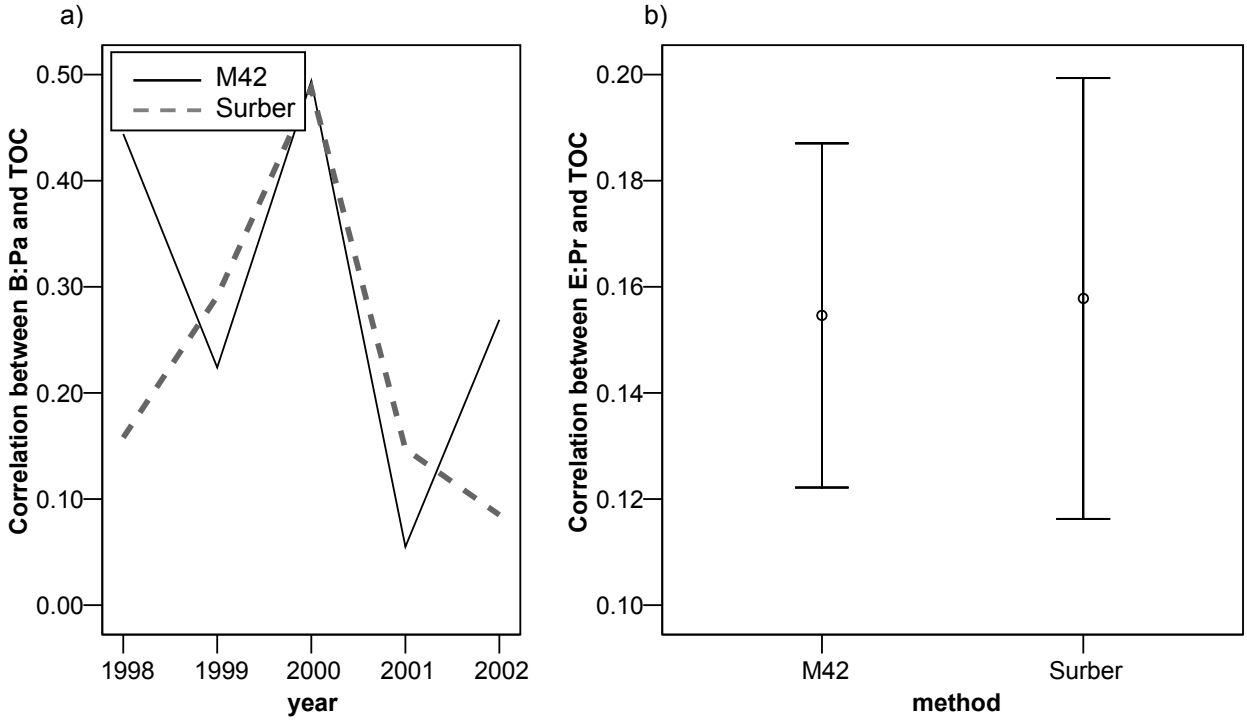

Figure 19. Method assessment Tau correlation: Tau correlation between TOC and the E:Pr calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

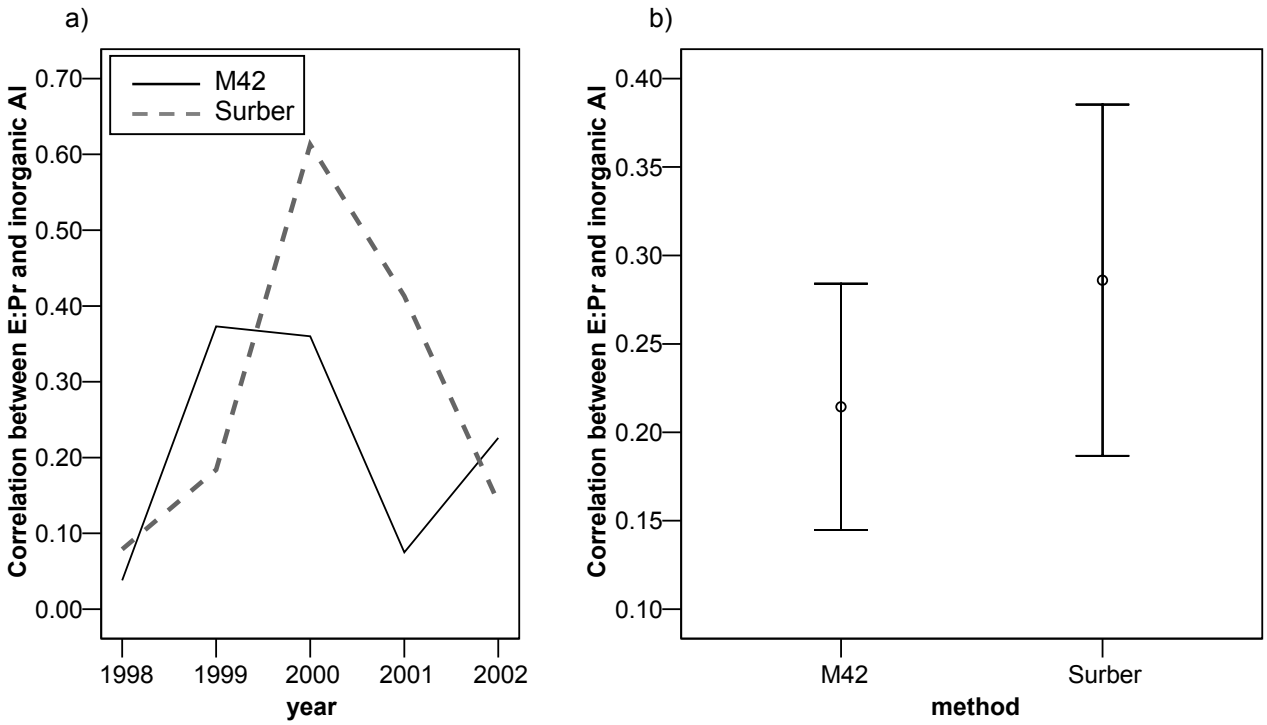

Figure 20. Method assessment Tau correlation: Tau correlation between inorganic Al and the E:Pr calculated from M42 and Surber species data (a) per year and (b) averaged across all years (mean ± SE plotted). Note that the correlation coefficients are plotted as absolute values. For signed values, see Table 3.

V) Acid sensitive taxa. In four out of five years, the number of acid sensitive taxa was greater in M42 samples (Fig. 21a), with the mean number of acid sensitive taxa averaged over all years also greater from M42 samples (Fig 21b). For 3 of 5 years, acid sensitive taxa were more likely to be most abundant in Surber samples (Fig. 22). However, highly sensitive taxa (Medin's rank 2 or 3) were more likely to be most abundant in M42 samples in most years, and when averaged across all years (Fig. 23)

Method Assessment – detection of difference

I) ANOSIM and SIMPER. Output from ANOSIM and SIMPER analyses of the differences between limed and reference streams are given in Table 4. ANOSIM detected differences in assemblage structure at the 5% level only once – for M42 streams in 2002 (Table 4). Borderline significant tests occurred in two further cases: M42 in 1999 and Surber in 2002 (Table 4). Detailed species-level output from the SIMPER analysis of 2002 M42 data is presented in Table 5. Of 20 taxa ranked as acid sensitive in the calculation of Medin's index, 11 have higher abundances in limed streams, and 9 have higher abundances in reference streams (Table 5). Mean dissimilarity between limed and reference streams from SIMPER analyses was always greater for M42 data (Fig 24).

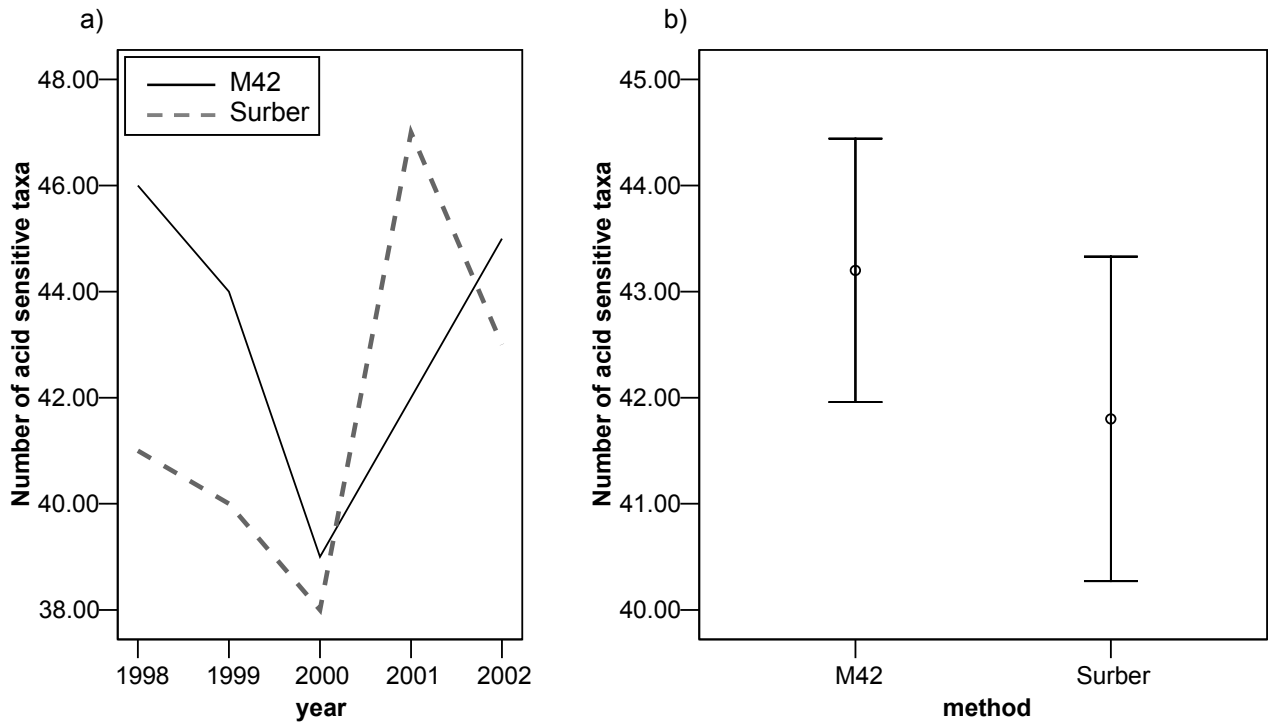

Figure 21. Method assessment acid sensitive taxa: number of acid sensitive taxa (medins rank 1-3) was more abundant for each sampling method (a) per year and (b) averaged across all years (mean ± SE plotted).

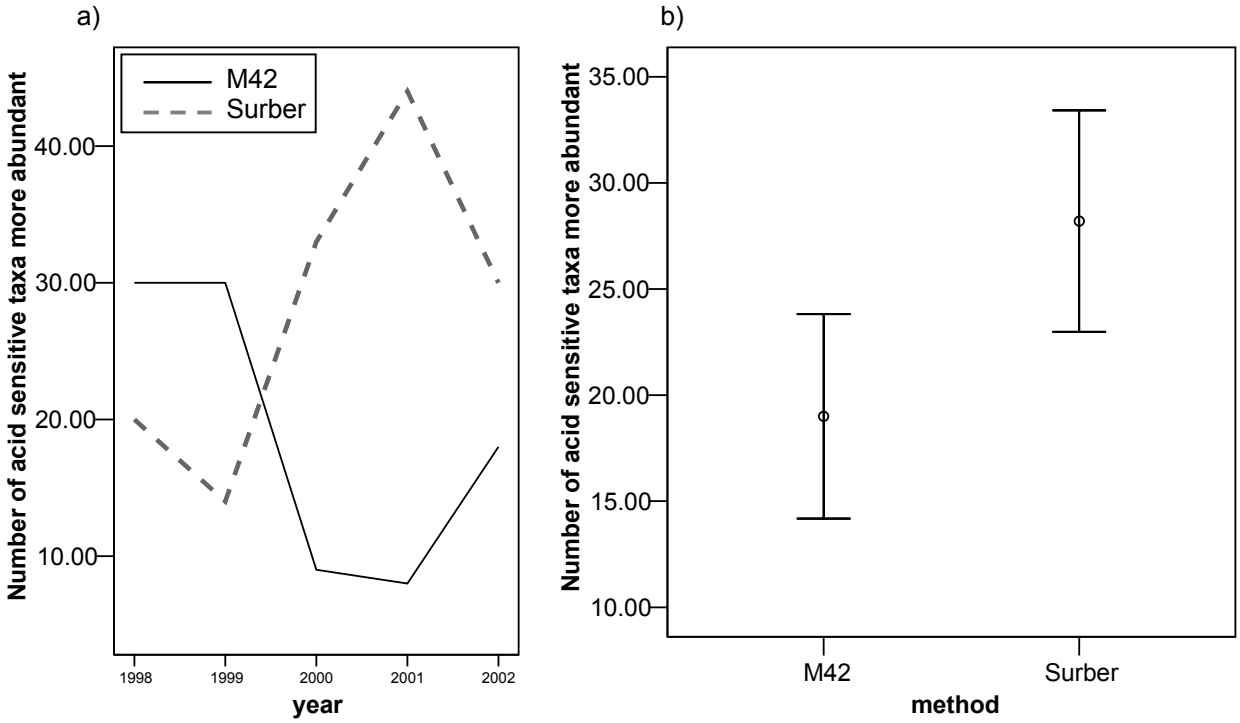

Figure 22: Method assessment acid sensitive taxa: number of times an acid sensitive taxon (medins rank 1-3) was more abundant for each sampling method (a) per year and (b) averaged across all years (mean ± SE plotted).

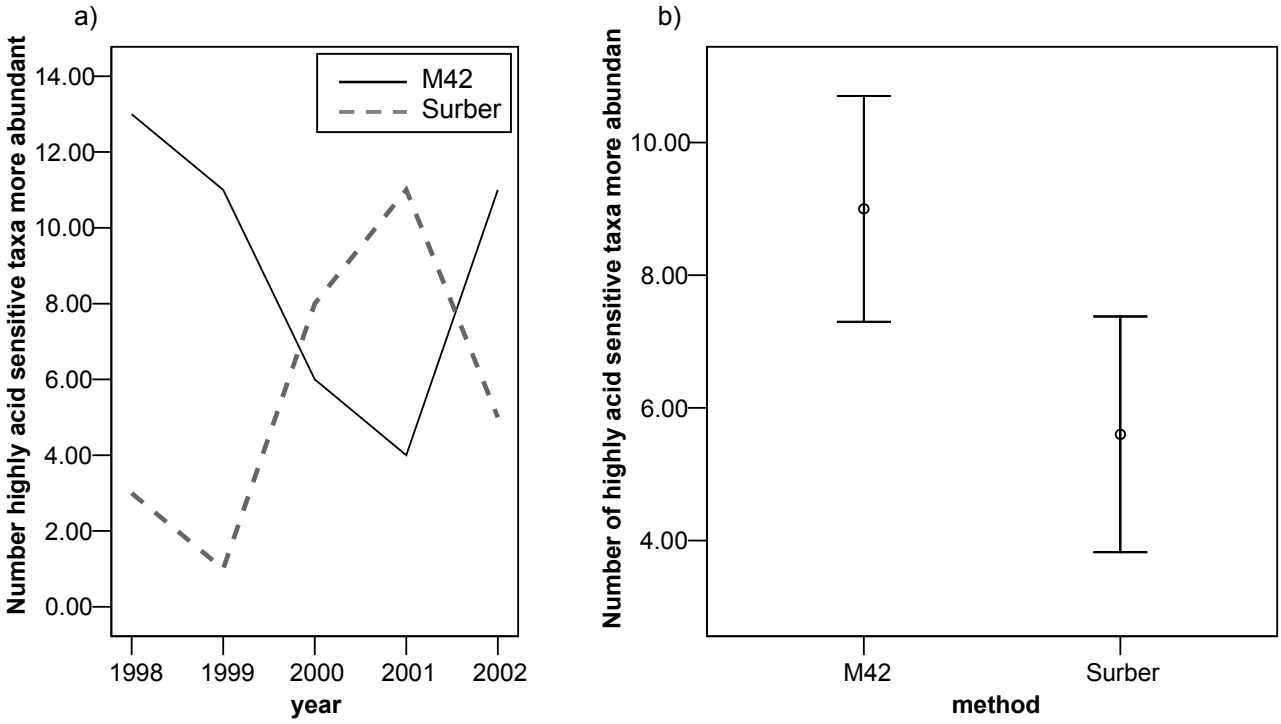

Figure 23: Method assessment acid sensitive taxa: number of times a highly acid sensitive taxon (medins rank 2-3) was more abundant for each sampling method (a) per year and (b) averaged across all years (mean ± SE plotted).

Table 4. Method assessment: output from ANOSIM and SIMPER analyses of assemblage differences between limed and reference streams. Rho statistics and significance values (Sig.) from ANOSIM analyses are tabulated (Sig. values < 0.1 are highlighted in bold), along with the mean dissimilarity between limed and reference streams from SIMPER analyses.

		ANOSIM		SIMPER
Year	Method	Rho	Sig.	Mean Dissimilarity
1998	M42	0.06	0.171	56.96
1998	Surber	0.077	0.123	51.17
1999	M42	0.103	**0.076**	51.4
1999	Surber	0.022	0.34	48.1
2000	M42	-0.046	0.692	56.11
2000	Surber	-0.005	0.488	51.27
2001	M42	-0.064	0.833	59.82
2001	Surber	0.068	0.154	54.94
2002	M42	0.288	**0.005**	52.76
2002	Surber	0.081	**0.086**	52.38

Table 5. Output from SIMPER analysis of the difference in species composition between limed and reference streams in Autumn 2002. Listed are taxa that collectively explain 75% of the dissimilarity between sample groups, together with their acid sensitivity rank (as scored for Medins index), their mean abundance from the two seasons, and their contribution to the dissimilarity. Mean dissimilarity 52.76.

Taxon	Medins rank	Limed mean abundance	Reference mean abundance	% contribution to dissimalarity	Cumulative %
Protonemura meyeri		49	27.78	2.39	2.39
Baetis rhodani	1	21.09	92.56	2.23	4.62
Limnius volckmari	1	23.45	7.67	2.14	6.76
Elmis aenea	1	21.36	31	2.13	8.89
Agapetus ochripes	2	5.18	49.89	2.05	10.94
Amphinemura borealis	1	45.27	93.11	2.04	12.98
Hydropsyche siltalai		14.82	4.11	2.01	14.99
Heptagenia sulphurea	1	10.55	12.44	2	16.99
Polycentropus flavomaculatus		21.64	9	1.94	18.93
Chironomini		12.91	9.67	1.94	20.87
Lepidostoma hirtum	1	16.91	3.33	1.93	22.8
Sericostoma personatum	1	14.36	14.56	1.88	24.68
Limnephilidae		13.73	32	1.84	26.52
Hexatominae		3.09	12.89	1.72	28.25
Orthocladiinae		49.82	97.33	1.71	29.96
Tanypodinae		57.36	51.78	1.71	31.66
Pediciinae		3.18	12.89	1.7	33.37
Simuliidae		7.82	20.33	1.69	35.06
Sphaeriidae	1	9.09	4.67	1.68	36.74
Leuctra hippopus		42.91	33.67	1.6	38.34
Taeniopteryx nebulosa	1	3.73	5.11	1.6	39.95
Nigrobaetis niger	1	13.91	7.67	1.6	41.54
Isoperla sp.		5	12.89	1.59	43.14
Asellus aquaticus		4.45	13.33	1.58	44.72
Oulimnius spp.	1	4.91	1.22	1.58	46.3
Leptophlebia marginata		13.18	2.78	1.56	47.86
Oxyethira sp.	1	10.09	1.67	1.55	49.41
Diura nanseni	1	1.82	12	1.53	50.94
Rhyacophila nubila		2.09	9.11	1.47	52.41
Ephemerella aurivillii		3.27	5.78	1.46	53.87
Nemoura avicularis	1	7	6.33	1.46	55.33
Psychodidae		2.36	5.56	1.44	56.77
Hydracarina		4.64	6.22	1.42	58.19
Ceratopogonidae		7	15.33	1.4	59.6
Leptophlebia vespertina		4.55	0.11	1.38	60.98
Tanytarsini		93.18	88.78	1.36	62.34
Capnopsis schilleri		4	4.56	1.36	63.7
Hydropsyche pellucidula	1	4.45	0.67	1.34	65.03
Empididae		4.09	3.89	1.25	66.28
Brachyptera risi		1	7.44	1.21	67.49
Oecetis testacea	2	2.73	0	1.16	68.65
Hydraena sp.		2.73	2.78	1.13	69.77
Ceratopsyche silfvenii	2	0.73	3.44	1.11	70.88
Rhyacophila sp.		0.91	2.78	1.06	71.94
Caenis rivulorum	3	3.45	3	1.06	73
Ithytrichia sp.	2	8.55	0.22	1.02	74.01
Lumbriculidae		1.64	1.89	0.98	75
Hydroptila sp.		5	0.56	0.96	75.96

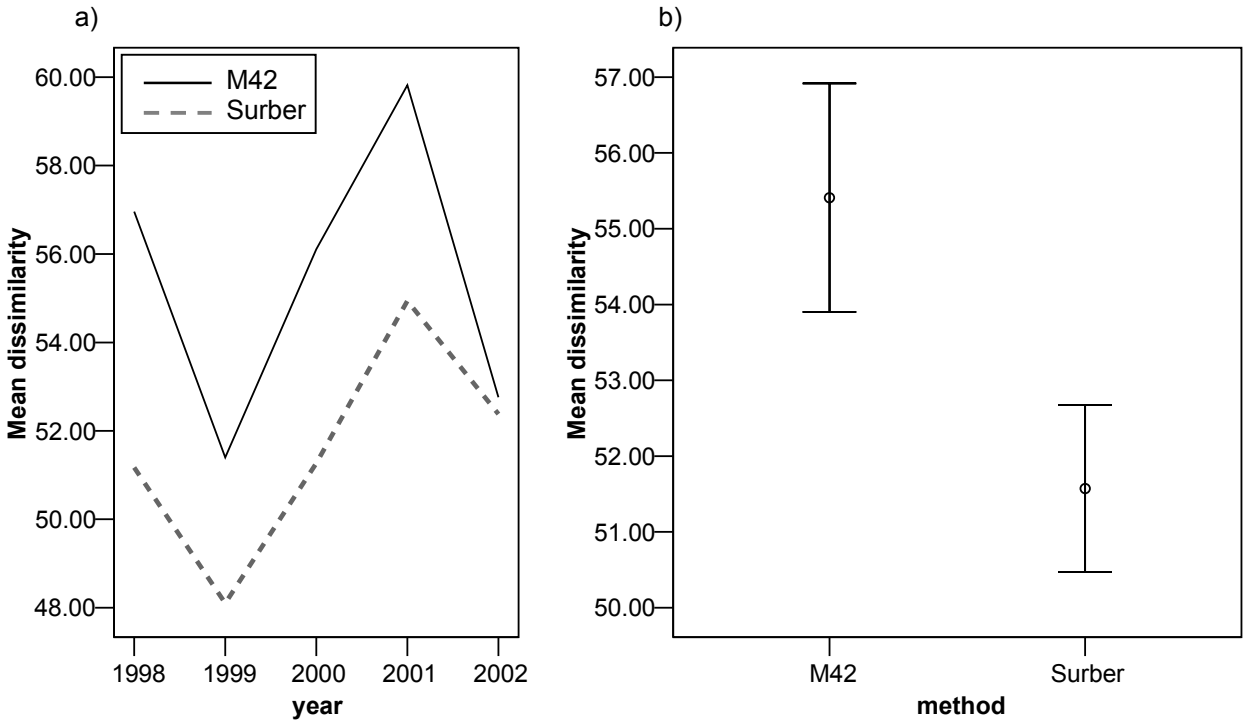

Figure 24. Method assessment simper analysis: mean dissimilarity between limed and reference streams (a) per year and (b) averaged across all years (mean ± SE plotted).

II) ANOVA of community metrics (total abundance, richness, Shannon diversity, EPTr and EPTa). Output from ANOVA of community metrics, with tests of the effect of sampling method and liming, are given in Table 6. Total abundance differed between method groups in 1999, when M42 samples collected more individuals, and from 2000-2001, with significantly more individuals occurring in Surber samples (Table 6). Species richness differed between sample methods from 1998-99 only, with M42 collecting more species. Shannon diversity (H') differed from 1998-99, when H' was greater in Surber samples, and in 2002, when the opposite was true (Table 6). EPTr differed significantly between methods in 1998 only, when richness of EPT taxa was greater from M42 samples, though tests were borderline significant for 1999 (M42 greater) and 2001 (Surber greater) also (Table 6). EPTa differed significantly in most years, with abundance of EPT taxa greater from M42 samples from 1998, but greater from Surber samples during 2000-02 (Table 6).

Liming did not affect any community metric at the 5% level of significance, though borderline significant cases were observed in 3 cases, all based on M42 data (Table 6).

Table 6. Method assessment: output from ANOVA tests of differences between method and lime groups for five community metrics. The mean ± SE for each category are tabulated, along with significance values (Sig.) for each test (Sig. values < 0.1 are highlighted in bold).

Metric			Method test			Lime test				
	Year	Method	Mean	SE	Sig	Mean lime	SE	Mean ref.	SE	Sig
Total abundance	1998	M42	800.14	95.41		857.83	144.73	723.22	115.80	0.652
	1998	Surber	736.67	95.49	0.355	751.67	131.75	716.67	145.89	0.793
	1999	M42	1771.86	200.17		1771.50	197.80	1772.33	402.50	0.685
	1999	Surber	1198.65	141.23	**0.004**	1251.91	222.44	1133.56	169.43	0.893
	2000	M42	399.74	95.64		424.91	160.30	365.13	71.35	0.528
	2000	Surber	983.89	189.61	**0.002**	965.70	274.60	1006.63	274.11	0.967
	2001	M42	391.00	144.64		416.60	247.44	359.00	125.11	0.515
	2001	Surber	1087.68	230.30	**<0.001**	1034.82	272.18	1160.38	420.83	0.917
	2002	M42	841.05	141.02		790.00	170.84	903.44	243.88	0.843
	2002	Surber	1336.90	187.23	0.257	1256.64	243.04	1435.00	304.29	0.606
Species richness	1998	M42	42.57	1.83		43.50	2.22	41.33	3.18	0.523
	1998	Surber	32.76	1.45	**<0.001**	32.33	1.79	33.33	2.51	0.799
	1999	M42	47.19	1.40		49.50	1.53	44.11	2.26	**0.064**
	1999	Surber	35.15	1.23	**<0.001**	34.91	1.92	35.44	1.51	0.760
	2000	M42	29.74	2.11		29.64	3.36	29.88	2.23	0.711
	2000	Surber	32.39	1.69	0.207	33.00	1.71	31.63	3.29	0.535
	2001	M42	27.61	1.98		27.00	2.63	28.38	3.19	0.782
	2001	Surber	35.21	2.22	0.024	35.73	2.44	34.50	4.27	0.583
	2002	M42	35.45	1.63		37.00	2.07	33.56	2.58	0.296
	2002	Surber	34.70	2.57	0.516	35.82	3.76	33.33	3.59	0.816
Shannon diversity	1998	M42	2.85	0.07		2.81	0.10	2.90	0.07	0.536
	1998	Surber	3.49	0.13	**<0.001**	3.46	0.16	3.52	0.24	0.845
	1999	M42	2.61	0.11		2.66	0.13	2.54	0.19	0.603
	1999	Surber	3.27	0.11	**<0.001**	3.28	0.14	3.26	0.20	0.916
	2000	M42	3.47	0.15		3.37	0.22	3.60	0.21	0.481
	2000	Surber	3.38	0.13	0.537	3.41	0.19	3.33	0.20	0.754
	2001	M42	3.66	0.09		3.61	0.11	3.72	0.14	0.546
	2001	Surber	3.58	0.12	0.580	3.61	0.14	3.55	0.21	0.817
	2002	M42	3.62	0.13		3.75	0.12	3.46	0.26	0.295
	2002	Surber	3.33	0.16	**0.040**	3.47	0.09	3.15	0.33	0.335
EPTr	1998	M42	21.62	0.90		22.50	1.10	20.44	1.49	0.239
	1998	Surber	19.29	1.07	**0.020**	19.50	1.48	19.00	1.61	0.815
	1999	M42	23.48	0.81		24.83	1.12	21.67	0.90	**0.059**
	1999	Surber	21.76	0.76	**0.052**	22.33	1.21	21.00	0.75	0.500
	2000	M42	19.21	1.42		18.82	2.24	19.75	1.58	0.538
	2000	Surber	20.42	1.02	0.274	20.64	1.18	20.13	1.88	0.706
	2001	M42	17.50	1.19		17.50	1.76	17.50	1.68	0.954
	2001	Surber	22.32	1.48	**0.070**	22.82	1.83	21.63	2.58	0.562
	2002	M42	22.00	1.04		23.27	1.47	20.44	1.38	0.199
	2002	Surber	21.80	1.26	0.689	23.27	1.62	20.00	1.92	0.209
EPTa	1998	M42	485.52	57.94		524.67	92.45	433.33	57.59	0.703
	1998	Surber	333.86	44.23	**0.002**	335.92	50.19	331.11	82.42	0.552
	1999	M42	609.95	57.07		682.42	68.55	513.33	91.34	**0.088**
	1999	Surber	511.05	60.07	0.119	511.25	87.63	510.78	83.50	0.763
	2000	M42	178.84	32.98		181.27	55.51	175.50	24.26	0.346
	2000	Surber	790.68	139.81	**<0.001**	748.00	187.42	849.38	222.56	0.834
	2001	M42	201.11	64.67		210.50	113.62	189.38	45.49	0.471
	2001	Surber	1119.21	288.58	**<0.001**	1136.55	360.91	1095.38	502.58	0.758
	2002	M42	465.60	99.28		435.64	102.03	502.22	189.23	0.789
	2002	Surber	1129.35	229.71	**<0.001**	1017.36	165.67	1266.22	481.98	0.697

III) ANOVA of acidity indices (medins, B:Pa, E:Pr, E:Pa). Output from ANOVA of community metrics, with tests of the effect of sampling method and liming, are given in Table 6. Medins index differed according to sampling method in 3 of 5 years, with values higher from M42 data in 1998-99, and higher from Surber data in 2001 (Table 7). The B:Pa ratio differed in all years, with a higher ratio from Surber samples (Table 7). Similar results were observed for E:Pa (Table 7). However, E:Pr differed according to sampling method only in 1998 (M42 value higher), though a marginally significant difference was found for 2000 also (Surber higher).

In only one case did an acidity index differ according to liming (the B:Pa ratio was higher in Surber samples from reference streams in 1999), with borderline significant results were observed in two other cases (both Surber). However, for the 1999 E:Pa result, data transformation failed to normalise the residuals, and so this result should be treated with caution (Table 7).

Season assessment: ordination, cluster, and simper analyses

In both nMDS ordination (Fig. 25) and UPGMA cluster analysis (Fig. 26) of benthic macroinvertebrate data (collected using the M42 method only), there is a clear separation of sites according to season. There is no clear separation of sites according to liming treatment, although there is a tendency for limed sites to occur towards the top left hand corner of the ordination space. Output from a SIMPER analysis of dissimilarity between autumn and spring assemblages is given in Table 8. Several acid sensitive beetles and Baetid mayflies were more abundant in the spring, whilst acid sensitive caddis, bivalves and stoneflies tended to be more abundant in the autumn (Table 8).

Season assessment – data quality

I) Canonical correspondance analysis. Detailed output from CCAs is given in Table 9. CCA extracted one significant component from the autumn data, explaining 9.9% of the variance, and none from the spring data. Axes one and two from the CCA ordinations are plotted in Fig. 27. The placement of sites with respect to the environmental variables differs substantially in several cases.

II) Mantel correlation. Signed correlations and significance levels for Mantel's test are given in Table 9. Mantel's correlation was stronger and slightly negative in the spring, whereas in the autumn it was almost zero. In neither case was it significant.

III) Weighted averaging. Correlation coefficients between observed values for pH, Ca, TOC and inorganic aluminium and values modelled from macroinvertebrate data are given in Table 9. Modelling of pH, Ca and inorganic Al appeared to be better in the spring, whilst modelling of TOC appeared better in the autumn.

Table 7. Method assessment: output from ANOVA tests of differences between method and lime groups for four acidity indices. The mean ± SE for each category are tabulated, along with significance values (Sig.) for each test (Sig. values < 0.1 are highlighted in bold).

			Method test			Lime test				
	Year	Method	Mean	SE	Sig	Mean lime	SE	Mean ref.	SE	Sig
Medins	1998	M42	6.76	0.51		6.42	0.61	7.22	0.88	0.783
	1998	Surber	6.05	0.48	**0.062**	5.67	0.45	6.56	0.96	0.745
	1999	M42	7.14	0.44		7.33	0.40	6.89	0.92	0.276
	1999	Surber	6.52	0.49	**0.049**	6.17	0.32	7.00	1.08	0.834
	2000	M42	5.05	0.48		4.91	0.61	5.25	0.82	0.626
	2000	Surber	5.84	0.51	0.135	5.82	0.55	5.88	1.01	0.299
	2001	M42	5.50	0.40		5.40	0.34	5.63	0.82	0.405
	2001	Surber	6.63	0.35	**0.002**	6.45	0.31	6.88	0.74	0.508
	2002	M42	6.10	0.42		6.27	0.52	5.89	0.72	0.210
	2002	Surber	6.05	0.48	0.824	6.36	0.54	5.67	0.87	0.152
B:Pa	1998	M42	0.42	0.24		0.16	0.04	0.76	0.56	0.201
	1998	Surber	3.64	1.09	**<0.001**	2.25	0.64	5.49	2.33	0.355
	1999	M42	0.66	0.17		0.58	0.20	0.76	0.31	0.715
	1999	Surber	2.15	0.91	**0.001**	0.65	0.10	4.15	1.99	**0.029**
	2000	M42	0.36	0.07		0.25	0.06	0.50	0.13	0.108
	2000	Surber	1.61	0.34	**<0.001**	1.12	0.31	2.28	0.65	**0.085**
	2001	M42	0.47	0.15		0.27	0.07	0.71	0.32	0.215
	2001	Surber	2.37	0.58	**<0.001**	1.90	0.49	3.01	1.23	0.517
	2002	M42	0.31	0.07		0.21	0.06	0.42	0.15	0.166
	2002	Surber	1.59	0.36	**<0.001**	1.22	0.27	2.04	0.71	0.356
E:Pr	1998	M42	1.05	0.08		1.08	0.08	1.00	0.17	0.397
	1998	Surber	0.77	0.08	**0.041**	0.73	0.10	0.82	0.13	0.597
	1999	M42	1.00	0.10		1.08	0.08	0.89	0.20	0.162
	1999	Surber	0.83	0.09	0.538	0.83	0.06	0.84	0.20	0.734
	2000	M42	0.68	0.13		0.82	0.18	0.50	0.19	0.286
	2000	Surber	0.75	0.12	**0.081**	0.81	0.17	0.66	0.14	0.534
	2001	M42	0.89	0.25		0.90	0.38	0.88	0.35	0.959
	2001	Surber	0.79	0.11	0.346	0.77	0.14	0.83	0.19	0.824
	2002	M42	0.95	0.15		1.09	0.21	0.78	0.22	0.228
	2002	Surber	0.70	0.08	0.691	0.76	0.09	0.61	0.15	0.225
E:Pa	1998	M42	1.67	0.55		1.42	0.50	2.00	1.14	0.704
	1998	Surber	4.16	1.12	**0.003**	2.84	0.66	5.92	2.42	0.441
	1999	M42	1.29	0.23		1.17	0.24	1.44	0.44	0.775
	1999	Surber	2.55	0.96	**0.024**	1.02	0.13	4.59	2.12	**0.052***
	2000	M42	0.95	0.19		1.18	0.30	0.63	0.18	0.188
	2000	Surber	1.89	0.38	**<0.001**	1.42	0.35	2.55	0.73	0.155
	2001	M42	1.67	0.54		2.00	0.79	1.25	0.73	0.424
	2001	Surber	2.86	0.59	**0.023**	2.58	0.57	3.24	1.21	0.819
	2002	M42	0.85	0.15		0.82	0.18	0.89	0.26	0.951
	2002	Surber	1.87	0.37	**0.001**	1.50	0.29	2.32	0.75	0.420

*unusual residuals distribution

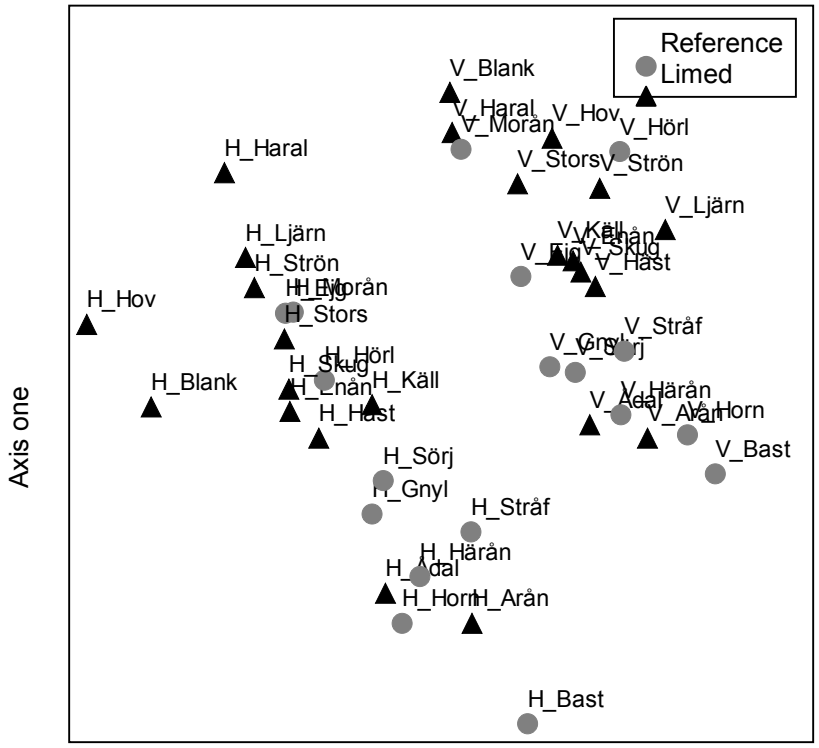

Figure 25. nMDS ordination of benthic invertebrate data from autumn 2004 and Spring 2005, with liming categories superimposed. The letters "H" and "V" preceding the stream names refer to whether the data was collected in Autumn (Swedish "Höst") or spring (Swedish "Vår") respectively. Ordination in 3 dimensions, axes 2 and 3 plotted. Stress = 14.38.

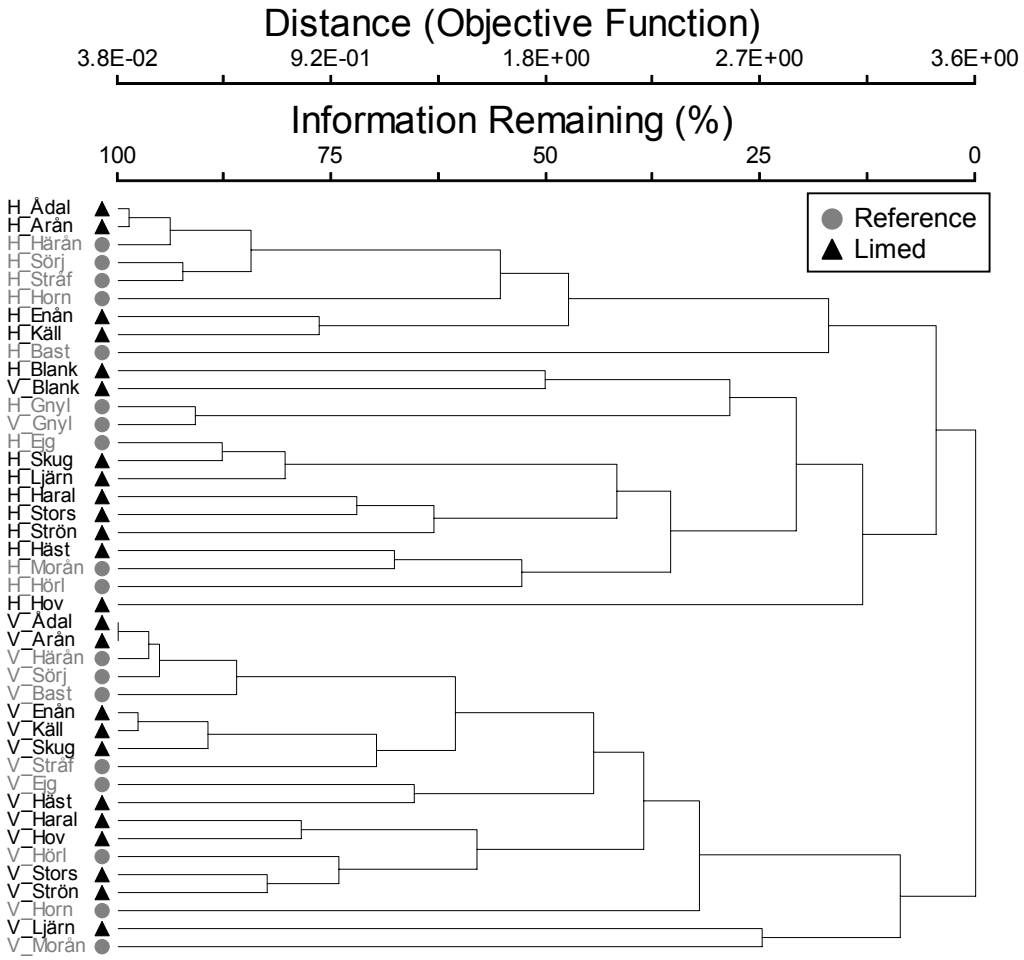

Figure 26. UPGMA cluster analysis of benthic invertebrate data autumn 2004 and Spring 2005, with liming categories superimposed. The letters "H" and "V" preceding the stream names refer to whether the data was collected in Autumn (Swedish "Höst") or spring (Swedish "Vår") respectively.

Table 8. Output from SIMPER analysis of the difference in species composition between Autumn 2004 and Spring 2005. Listed are taxa collectively explaining 68% of dissimilarity between sample groups, together with their acid sensitivity rank (as scored for Medins index), their mean abundance from the two seasons, and their contribution to the dissimilarity. Mean dissimilarity 56.95.

Taxon	Medins rank	Autumn mean abundance	Spring mean abundance	% contribution to dissimalarity	Cumulative %
Leuctra hippopus		49.38	1.19	2.75	2.75
Protonemura meyeri		39.95	0.38	2.55	5.29
Sphaeriidae	1	56.81	43.05	1.95	7.24
Leuctra fusca/digitata		0	16.29	1.88	9.12
Agapetus ochripes	2	64.19	7.81	1.86	10.99
Naididae		0.86	17.24	1.83	12.81
Nemoura avicularis	1	11.19	0	1.81	14.62
Amphinemura borealis	1	33.67	34.33	1.7	16.32
Halesus sp.		0	7.67	1.68	18.01
Limnius volckmari	1	22.52	38.57	1.67	19.68
Baetis rhodani	1	38.05	88.71	1.65	21.33
Asellus aquaticus		32.1	12.33	1.6	22.93
Empididae		7.67	15.9	1.56	24.49
Leptophlebia marginata		14.9	1.24	1.55	26.04
Simuliidae		61.29	98.24	1.53	27.57
Elmis aenea	1	22	30.33	1.5	29.07
Heptagenia sulphurea	1	16.48	6.71	1.47	30.53
Nigrobaetis niger	1	5.14	15.81	1.45	31.99
Hydropsyche siltalai		36.67	6.86	1.4	33.39
Lepidostoma hirtum	1	10.43	10.29	1.35	34.73
Diamesinae		17.19	0	1.34	36.08
Tanypodinae		20.86	39.24	1.32	37.4
Chironomini		14.29	5.43	1.31	38.71
Enchytraeidae		0.43	6.76	1.31	40.02
Isoperla sp.		11.52	5.81	1.31	41.33
Sericostoma personatum	1	14.19	10.95	1.31	42.64
Amphinemura sulcicollis		0.43	7.81	1.26	43.89
Limnephilidae		73.71	28.67	1.21	45.11
Baetis fuscatus gr.		0	25.52	1.21	46.32
Ephemerella aurivillii		18.57	5.57	1.2	47.51
Polycentropus flavomaculatus		13.14	9.29	1.18	48.69
Oligochaeta other		6.86	0	1.15	49.84
Hydracarina		3.43	6.9	1.13	50.97
Lumbriculidae		1.71	7.24	1.13	52.1
Hydroptila sp.	2	19.29	3.24	1.13	53.23
Ceratopogonidae		13.9	13.71	1.11	54.34
Psychodidae		9.48	2.57	1.08	55.42
Oulimnius tuberculatus	1	4.52	5.43	1.07	56.49
Orthocladiinae		150.9	140.1	1.06	57.55
Centroptilum luteolum	1	6.05	4.9	1.05	58.59
Oxyethira sp.	1	7.29	1.9	1.04	59.64
Alainites muticus		0.19	14.95	1.03	60.67
Hydraena sp.		3.05	5.76	1.01	61.68
Radix peregra/ovata	1	22.52	9.76	0.97	62.66
Tanytarsini		82.05	136.24	0.94	63.6
Diura nanseni	1	4.9	0.67	0.92	64.52
Limoniidae		0.1	2.71	0.9	65.42
Leptophlebia vespertina		1.76	6	0.9	66.32
Hexatominae		4.05	4.76	0.89	67.21
Pediciinae		4.05	4.76	0.89	68.1

Table 9. Season Assessment: output from Canonical Correspondance Analyses (the number of significant components, the associated significance values, and proprtion of variance in the species data explained by the significant components), Mantel Correlations (correlation coefficient and significance level) and Weighted Averaging correlations (correlations between observed values of the variable and values modeled from the species data).

		Canonical Correspondance Analysis			Mantel Correlation		Weighted Averaging Correlations			
Season	Year	# Significant components	Significance levels	Significant variance explained	R	Sig	pH	Ca	TOC	Inorganic Al
Autumn	2004	1	0.03	9.9	0.002	0.488	0.54	0.83	0.73	0.53
Spring	2005	0		0	-0.016	0.123	0.60	0.88	0.42	0.95

IV) Biotic indices: correlations with acidity-related variables. Full output from correlation analyses are presented in Table 10. The best performing index (most significant correlations in both spring and autumn data) was Medins index (Table 10). Medins was not significantly correlated with pH in either season, but the coefficient was slightly greater, and negative, in the autumn. The correlation between Medins index and Ca was stronger in the autumn than the spring, with the same true for the correlation with inorganic Al (Table 10). However, the reverse was true for the correlation with TOC. Results for E:Pr were similar to those for Medins index (Table 10). For E:Pa, correlations with all four response variables appeared stronger in the spring. The B:Pa ratio was not well correlated with any variable in the 2004-05 data set.

V) Acid sensitive taxa. Greater numbers of acid sensitive taxa (Medin's index rank 1-3) were found in the spring (43) than the autumn (40). However, a greater number of acid sensitive taxa were more abundant in the autumn (25) than the spring (22). The same was true for highly sensitive (Medin's rank 2 or 3) taxa (6 were more abundant in the autumn, 4 were more abundant in the spring).

Table 10. Season assessment: non parametric Kendall's Tau-b correlations between four biotic indices and four acidity related variables. For each variable, both the coefficient and significance level are indicated, with p-values < 0.01 highlighted in bold.

			pH		Ca		TOC		Al inorg	
Index	Season	Year	tau	sig.	tau	sig.	tau	sig.	tau	sig.
Medin	Autumn	2004	-0.052	0.757	0.486	**0.004**	0.29	**0.083**	-0.306	**0.067**
	Spring	2005	0.021	0.901	0.331	**0.047**	0.321	**0.055**	-0.01	0.951
E:Pr	Autumn	2006	0.157	0.402	0.246	0.188	0.112	0.549	-0.337	**0.072**
	Spring	2007	0.164	0.356	0.273	0.124	0.334	**0.06**	-0.249	0.161
E:Pa	Autumn	2008	0.006	0.974	-0.099	0.576	-0.052	0.767	-0.117	0.51
	Spring	2009	0.286	**0.083**	0.348	**0.035**	0.348	**0.035**	-0.235	0.155
B:Pa	Autumn	2010	0.034	0.832	-0.097	0.545	-0.216	0.174	0.25	0.116
	Spring	2011	-0.019	0.904	0.106	0.506	-0.211	0.184	0.172	0.277

Season assessment – detection of difference

I) ANOSIM and SIMPER. Assemblage structure did not differ according to liming in either the autumn (ANOSIM rho = 0.042, p = 0.234) or spring (rho = 0.023, p = 0.305). Mean SIMPER dissimilarity between limed and reference streams was greater in the autumn (50.13) than in the spring (46.12).

II) ANOVA of community metrics (total abundance, richness, Shannon diversity, EPTr and EPTa). Output from ANOVA of community metrics, with tests of the effect of sampling season and liming, are given in Table 11. All five metrics differed between seasons. Total abundance and EPT abundance were greater in the autumn, whilst total richness, Shannon diversity and EPT richness were greater in the spring (Table 11). Liming did not affect any metric in either season (Table 11)

III) ANOVA of acidity indices (Medin's, B:Pa, E:Pr, E:Pa). Output from ANOVA of community metrics, with tests of the effect of sampling season and liming, are given in Table 12. All four metrics differed between seasons, with values of all four greater in the spring (Table 12). Liming significantly affected the B:Pa ratio in the autumn, with smaller ratios observed from limed streams (Table 12). Medins ratio was affected at a borderline level of significance in the autumn, with larger ratios observed in limed streams (Table 12).

Table 11. Method assessment: output from ANOVA tests of differences between season (repeated measures) and lime groups for five community metrics. The mean ± SE for each category are tabulated, along with significance values (Sig.) for each test (Sig. values < 0.1 are highlighted in bold).

			Season test			Lime test				
	Season	Year	Mean	SE	Sig	Mean lime	SE	Mean ref.	SE	Sig
Total abundance	Autumn	2004	1169.33	187.30		1249.67	307.87	1062.22	168.39	0.863
	Spring	2005	1065.19	80.75	**<0.001**	1086.33	104.47	1037.00	133.62	0.271
Species richness	Autumn	2004	45.81	1.84		46.33	2.66	45.11	2.60	0.775
	Spring	2005	48.28	1.75	**0.001**	49.50	2.68	46.66	2.04	0.520
Shannon diversity	Autumn	2004	2.89	0.04		2.91	0.05	2.86	0.08	0.632
	Spring	2005	3.83	0.12	**0.001**	3.91	0.17	3.72	0.18	0.470
EPTa	Autumn	2004	580.00	76.00		571.42	101.03	591.44	122.27	0.935
	Spring	2005	398.48	30.89	**0.001**	374.25	32.45	430.78	58.36	0.648
EPTr	Autumn	2004	23.90	1.11		24.58	1.65	23.00	1.42	0.526
	Spring	2005	24.62	1.19	**0.001**	25.17	1.95	23.89	1.12	0.830

Table 12. Season assessment: output from ANOVA tests of differences between season (repeated measures) and lime groups for four acidity indices. The mean ± SE for each category are tabulated, along with significance values (Sig.) for each test (Sig. values < 0.1 are highlighted in bold).

Index			Season test			Lime test				
	Season	Year	Mean	SE	Sig	Mean lime	SE	Mean ref.	SE	Sig
Medins	Autumn	2004	6.90	0.40		7.17	0.34	6.56	0.82	**0.098**
	Spring	2005	8.14	0.43	0.004	8.25	0.58	8.00	0.67	0.249
B:Pa	Autumn	2004	0.27	0.07		0.15	0.04	0.45	0.14	**0.031**
	Spring	2005	0.46	0.18	0.020	0.25	0.10	0.75	0.38	0.105
E:Pa	Autumn	2004	0.76	0.19		0.75	0.22	0.78	0.36	0.849
	Spring	2005	6.00	1.74	0.005	6.25	2.44	5.67	2.57	0.835
E:Pr	Autumn	2004	0.90	0.07		0.92	0.08	0.89	0.11	0.840
	Spring	2005	1.76	0.18	0.001	1.75	0.22	1.78	0.32	0.817

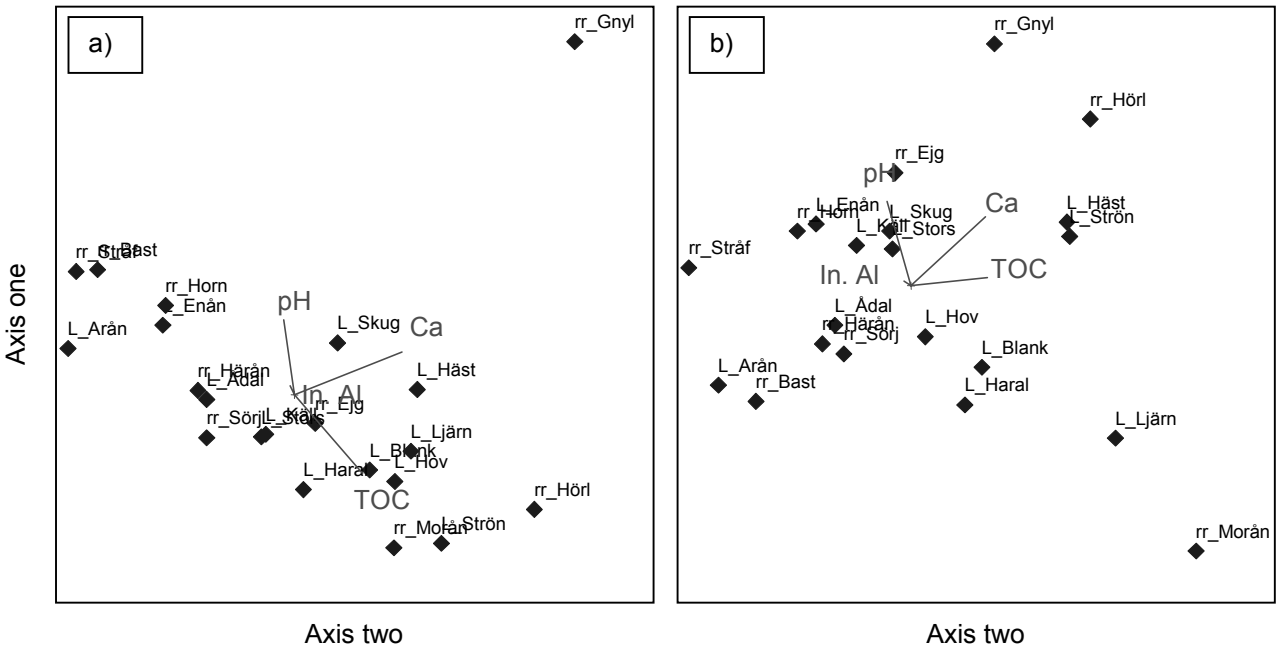

Figure 27. CCA ordination of benthic invertebrate data from (a) autumn 2004 and (b) spring 2005. The letters "L" and "rr" preceding the stream names refer to whether the stream was limed or an unlimed reference respectively. Percent variance explained: (a) 15.9, (b) 14.8 (neither axis significant)

Discussion

Sampling methods

The M42 and Surber methodologies do not sample identical assemblages, as indicated by (i) ANOSIM, (ii) the lack of similarity between samples within some streams apparent in ordination and cluster analyses, and (iii) differences in the values of community metrics (e.g. species richness) and acidity indices between the two methods (especially the B:Pa ratio, attributable to the greater abundances of *Baetis* species sampled by the Surber method). Despite this, the two methods do not appear to yield markedly different information about the stream macroinvertebrate faunas in relation to liming, since:

1) the general gradients uncovered in most NMS and CCA ordinations were similar for both methods
2) Correlations between acid indices and environmental variables were usually of similar magnitude, and almost always of identical direction, for both methods.
3) In most cases, statistical tests of the effect of liming on community structure and acid indices yielded the same outcome regardless of sampling method.

Accordingly, conclusions about the *general* effects of liming drawn from the two methodologies are likely to be similar. However, biomonitoring programmes are often more concerned with *specific* than general effects. A biomonitoring programme for liming needs to be able to identify streams responding poorly to liming in any given year, and distinguish between species that are unaffected or affected positively or negatively by liming. It seems pertinent to focus on only one sampling method, given that the use of two methods does not appear to yield significant extra information. The chosen method should be able to assess not only the general impact of liming, but also accurately highlight details in the responses of particular streams and species.

Markedly low numbers of species and individuals were collected using M42 from 2000-2002 compared with other years. This could reflect either (i) natural interannual variation in field macroinvertebrate populations, or (ii) variation in the intensity of M42 sampling (e.g. samples may have been sorted for different time periods). It seems likely that the second alternative applies, given that species richness and abundance data from the Surber samples remained relatively constant over the period 1998-2002, and that there were no dramatic changes to stream physico-chemistry over this period that could explain a loss of richness and abundance from one sampling method. Nevertheless, the possibility that at least some of the variation in sampled abundance and richness for the M42 method reflects natural variability cannot be completely excluded. Accordingly, this discussion of sampling methodology will consider three time periods: (i) 1998-99, when abundances sampled by M42 were relatively high, in line with data from 1994-97 and 2004;

(ii) 2000-02, when numbers sampled by M42 were low and (iii) the entire five year period (1998-2002). Table 13 summarises the outcome of analyses assessing the performance of the M42 and Surber methods. For the first two time periods (1998-99 and 2000-2002), a method was regarded as performing better for a given parameter if it was superior (yielded a higher value or greater number of significant results) for the bulk of the time period. The method name is written in brackets if it was only clearly superior for one year of the time period, whilst an em dash (—) is used if the methods are indistinguishable, either because they closely tracked one another, or because a high value for one method in one year was counterbalanced by a high value for the alternative method in another year. For the combined years period (1998-2002), a method was entered against a parameter if its mean value over the period was larger (as seen in the "b" panels for Figs. 6-24), and if it performed better for at least 3 of the 5 years. Thus although the mean correlation between B:Pa and TOC was greater for M42 (Fig. 16b), M42 was only superior to Surber sampling in 1998 and 2002 (Fig. 16a), and so in Table 13, an em dash is placed against this parameter.

Table 13. Summary of analyses assessing the performance of the M42 and Surber methodologies. Assessments made over three time periods: 1988-99, 2000-2002 and 1998-2002 ("All", highlighted in bold). See text for further explanation.
Legend: The methods name placed against a parameter indicates better performance for that parameter. A bracketed name indicates either that the method clearly performed better in one year of the time period (Section A) or that significance levels for that method were borderline (Section B).
An em dash (—) indicates the two methods are indistinguishable over that time period.

			Better method			
	Analysis	Parameter assessed	1998-99	2000-02	All	Reference
A: Data quality	CCA and Mantel	More significant axes	M42	—	**—**	Fig. 6
		More variance explained by significant axes	M42	Surber	**M42**	Fig. 7
		Higher Mantel correlation	M42	M42	**M42**	Fig. 9
	WA	Higher correlation: observed & modelled pH	(M42)	(Surb)	**—**	Fig. 10
		Higher correlation: observed & modelled Ca	—	M42	**M42**	Fig. 11
		Higher correlation: observed & modelled TOC	—	M42	**M42**	Fig. 12
		Higher correlation: observed & modelled inorg. Al	M42	M42	**M42**	Fig. 13
	Kendall's Tau	Higher correlation: B:Pa & pH	—	—	**—**	Fig. 14
		Higher correlation: B:Pa & Ca	(M42)	M42	**M42**	Fig. 15
		Higher correlation: B:Pa & TOC	(M42)	—	**—**	Fig. 16
		Higher correlation: B:Pa & inorganic Al	—	M42	**M42**	Fig. 17
		Higher correlation: Medins & Ca	M42	Surber	**—**	Fig. 18
		Higher correlation: E:Pr & TOC	Surber	(M42)	**—**	Fig. 19
		Higher correlation: E:Pr & inorganic Al	(M42)	Surber	**—**	Fig. 20
		Higher correlation: E:Pa & pH	Surber	(Surber)	**Surber**	Table 3
		Higher correlation: E:Pa & inorganic Al	—	(M42)	**—**	Table 3
		Higher correlation: both M42 and Surber p<0.05*	M42	M42	**M42**	Table 3
	Acid-sensitive taxa	Number sampled	M42	(M42)	**M42**	Fig. 21
		More taxa more abundant	M42	Surber	**Surber**	Fig. 22
		More highly sensitive taxa more abundant	M42	—	**M42**	Fig. 23
B Lime test	Separating limed & reference streams	ANOSIM: Detecting difference	(M42)	(M42)	**M42**	Table 4
		SIMPER: Greater dissimilarity	M42	M42	**M42**	Fig. 24
		Community metrics: Detecting difference	(M42)	—	**M42?**	Table 6
		Acidity indices: Detecting difference	Surber	(Surber)	**Surber?**	Table 7

*In cases where correlations for both methods were significant, which correlation coefficient was higher?

Over the full time period, M42 data performed better than Surber data for 10 of 20 data quality parameters (section A in Table 13). In many cases, M42 data performed strongly over the 1998-99 time period, when numbers collected by the M42 method were high, and maintained a superior performance through the 2000-02 period, when numbers sampled were low (e.g. Figs. 9, 13, 15). In other cases, M42 performance dropped towards or below the Surber performance in 2000-01, but recovered in 2002, in concert with a general rise in M42 abundances (see Figs 2, 7, 21, 23 and even Figs. 18 and 22). For a few parameters, a strong performance by the M42 data in 1998-99 was cancelled out by weaker performances over 2000-01 (Figs. 18, 22). Across all years, there was only one parameter for which Surber data consistently performed better: the E:Pa-TOC correlation (Table 3). In other cases, the improvement in performance of Surber sampling over 2000-02 can only be considered relative to the reduced performance of M42 over this period (see Figs 2, 7, 20, 22-23)

The assessment of the capacity of the two methods to detect differences between limed and reference streams is less clear, in large part because the acid indices and community metrics themselves seem to differ little according to liming. ANOVA appeared more likely to detect differences in community metrics with M42 data, but significance levels were borderline, whereas Surber sampling distinguished lime groups for one acidity index in one year, with two further borderline cases. Evidence from the SIMPER analyses is more compelling – in all years, mean dissimilarity between limed and reference stream assemblages was greater for M42 data (Fig. 24), though both methods tracked the same general interannual trends. If it is accepted that M42 data both correlates more strongly with the acidity-related variables (e.g. Figs 7, 9, 15, 16 and Table 3) and also better models those variables (Figs 11-13), then the capacity of M42 to detect greater dissimilarity between lime groups indicates that it may be better suited to isolating streams and taxa when they do not respond as expected to liming. Interestingly, even though the capacity of the M42 method to sample high abundances of acid sensitive taxa closely tracked changes in the general abundance of invertebrates sampled over 2000-02 (compare Figs. 2 and 22), there was only a slight drop in the number of acid sensitive taxa sampled over this period (Fig. 21), and in only one year (2001) were Surber samples better at sampling higher abundances of highly acid sensitive taxa (Fig. 23a). Thus even when a lower number of total individuals were sampled by the M42 method, it was still able to sample good numbers of important acid sensitive taxa. When larger total abundances were sampled, M42 was clearly the superior method for sampling such taxa. Other attributes of M42 are also desirable – it is applicable to a wider range of stream types than the Surber method, which is inefficient in deeper, slow flowing waters or those with coarser, rockier substrates, and produces relatively "clean" samples that are easier to process. Taken together, all these characteristics make the M42 method appropriate for a liming biomonitoring programme.

One drawback of M42 compared with Surber sampling is that it is arguably more difficult to implement in a standardised manner (though the uniformity of Surber sampling can also vary greatly, especially in the degree to which the substrate is stirred up during sampling). Accordingly, if M42 is to be adopted as the main sampling method for the IKEU programme, it is necessary that staff be well trained in the method, to ensure consistency of sampling. In particular, parameters regulating the intensity of sampling (time of sampling, area of coverage, habitats included) need to be clearly established and adhered to. The performance of M42 was generally better when the number of individuals sampled was relatively high, in 1998-99 and 2002. Accordingly, higher intensity sampling is preferable, in order that the number of individuals and species collected approximate levels from 1995-99 (Fig. 2: 1500-2000 individuals, and around 50 species).

Season assessment

Not surprisingly, assemblages sampled in Autumn 2004 and Spring 2005 clearly differed. In both nMDS ordination and UPGMA cluster analyses, samples were markedly differentiated by season, and all community metrics and acidity indices differed between the spring and the autumn. Liming was not found to significantly affect macroinvertebrate assemblage structure or any index in either season, but samples from the two seasons appeared to give differing information about the stream biotas in relation to liming. Groupings in the CCA plots and their relationships with the biplot vectors differing markedly between the seasons (Fig. 27). Correlations between the acid-related variables and acidity indices were often of similar magnitude and in a similar direction in both seasons, but in some cases (eg the Medin-inorg. Al, and all E:Pa correlations) correlation coefficients differed markedly. Weighted averaging analysis indicated that the spring data was slightly better at modelling pH, Ca and inorganic Al, but that the autumn data modelled TOC substantially better. Based on a summary of comparisons between autumn and spring (Table 14), it is not clear that one season's data performs consistently better. Rather, information gained from the two seasons may be complementary, in that different aspects of the responses of the streams to liming are emphasised. This is not surprising, since autumn samples are generally taken prior to the acid episodes associated with winter or spring rains, while the spring samples are generally taken after. Thus it might be expected that spring data would be more likely to reflect the effects of an immediately preceding acid episode, whilst autumn data may be more likely to reflect the chronic status of the stream, particularly as it affects insect oviposition and hatching success early in the season.

However, more years worth of comparisons are required before definitive statements can be made as to what information can be gained from sampling in the spring in addition to, or instead of, the autumn. Furthermore, the advantages and pitfalls of sampling in the two seasons need to be considered. The autumn is generally a more stable period than the spring, both in terms of the abiotic environment (severe autumn storms notwithstanding), and especially in terms of the biota.

Over the course of the spring, different macroinvertebrate species complete larval development and emerge into the adult stage at different times. Consequently, assemblages are very dynamic, complicating comparison between streams when sampling is strongly staggered (e.g. ongoing animal emergence could undermine comparison of streams sampled at the beginning of May with streams sampled late in June), though a well organised paired stream sampling design could minimise this problem. In contrast, assemblages are much more persistent in the autumn months. The main difficulty in the autumn is identification, with many taxa too immature to identify to species at that time.

Table 14. Summary of analyses assessing the performance of M42 samples collected in Autumn 2004 and Spring 2005. Placement of a seasons name against a parameter indicates better performance for that paramater in that season. An em dash (—) indicates that the two seasons are indistinguishable.

	Analysis	Parameter	Better season	Reference
A: Data quality	CCA and Mantel	More significant axes	Autumn	Table 9
		More variance explained by significant axes	Autumn	Table 9
		Higher Mantel correlation	Spring	Table 9
	WA	Higher correlation: observed & modelled pH	Spring	Table 9
		Higher correlation: observed & modelled Ca	Spring	Table 9
		Higher correlation: observed & modelled TOC	Autumn	Table 9
		Higher correlation: observed & modelled inorg. Al	Spring	Table 9
	Kendall's Tau	Higher correlation: Medin & pH	Autumn	Table 10
		Higher correlation: Medin & Ca	Autumn	Table 10
		Higher correlation: Medin & TOC	Spring	Table 10
		Higher correlation: Medin & inorganic Al	Autumn	Table 10
		Higher correlation: E:Pr & TOC	Spring	Table 10
		Higher correlation: E:Pr & inorganic Al	Autumn	Table 10
		Higher correlation: E:Pa & pH	Spring	Table 10
		Higher correlation: E:Pa & Ca	Spring	Table 10
		Higher correlation: E:Pa & TOC	Spring	Table 10
	Acid-sensitive taxa	Number sampled	Spring	Text
		More taxa more abundant	Autumn	Text
		More highly sensitive taxa more abundant	Autumn	Text
B Lime test	Separating limed & reference streams	ANOSIM: Detecting difference	—	Text
		SIMPER: Greater dissimilarity	Autumn	Text
		Community metrics: Detecting difference	—	Table 11
		Acidity indices: Detecting difference	Autumn?	Table 12

Secondary aims: the impact of liming

As emphasised earlier, detailed assessments of the impact of liming and interannual variation in the data were beyond the scope of this report, but some comments may be made, based on the analyses conducted.

In general, it was difficult to distinguish limed stream assemblages from those of reference streams, and this, coupled with the general lack of differentiation in the main acid indices, indicates that liming is minimising any ongoing effects of acidification. However, ordination analyses do indicate that the faunas of limed and reference streams continue to differ, in a similar, albeit subtle, direction (e.g. the left-right gradients apparent in Figs. 4 and 25). This may reflect differences in relative abundance – as highlighted in the SIMPER analysis of the 2002 M42 data (Table 5), several acid sensitive taxa (e.g. *Baetis rhodani, Elmis aenea, Agapetus ochripes*) were less abundant in limed than reference streams, though others were more abundant in limed streams. Further analyses should focus on whether these differences relate primarily to

1) continuing differences in the acid status of the streams
2) consistent differences in other physico-chemical characteristics, or
3) effects on assemblage structure attributable to effects of liming other than the direct amelioration of water acidity (e.g. deleterious effects or trophic effects)

In the longer term, the ability of the IKEU programme to answer these and other questions could be improved by expanding the set of routinely monitored streams. In doing so, three problems with the current set of limed and reference streams need attention:

1) The reference streams are mostly circumneutral, with only two being chronically acid. Tests carried out in this analysis were done with reference only to the circumneutral streams. However, the success of liming might also be assessed relative to reference sites of lower acidity. If, for example, the abundances of acid sensitive taxa are more similar to those of acid rather than circumneutral references, then a different assessment of the success of liming might be warranted.

   Note that may not be necessary to choose new reference streams to facilitate comparison with limed streams – in some cases, suitable reference conditions may exist upstream of liming in the currently monitored streams.

2) More powerful direct tests for the effects of liming could be obtained if background variation (associated with latitude, land use, geology) were better controlled. An effective means to achieve this would be to pair limed and reference sites within regions. Paired sites would be explicitly chosen to be as similar in important physico-chemical characteristics, and differ predominantly only in acid status and liming. Suitable pairs for some IKEU streams may

already exist in the current reference stream set. Otherwise additional reference sites could be selected to pair with a selected subset of the limed sites.

3) More generally, the limed and reference sites are distributed erratically throughout Sweden, with some regions well represented, and others poorly represented. Without a more even distribution of sites, it could be difficult to generalise the results of the IKEU monitoring program.

Note that it might be possible to expand the IKEU data set not only by monitoring more streams as part of the IKEU program, but also by incorporating data collected by other agencies (e.g. regional and kommun government authorities). However, the potential pitfalls of doing so need investigation; not least problems of harmonising kick sample data, favoured by most agencies, with the Surber and M42 methods favoured by the IKEU program.

Secondary aims: inter-annual variation

Over the years for which both reference and limed stream data were available (1998-2005), there appear to be no consistent long term trends relating to liming (e.g. a cumulative improvement from year to year in the acid status of limed streams, as reflected in increasing numbers of acid sensitive taxa, or increasing values for acidity indices). Rather, there is substantial inter year variation (for example, in scores for the acidity indices and in the dissimilarity between limed and reference sites). Identifying the factors driving this variability would have substantial benefits for the future management of liming (e.g. prediction of which streams are likely to require closer attention given particular environmental conditions). Note that Surber sampling may be better suited to assessing long term changes in species abundance, as it is more quantitative than M42.

Recommendations

Following on from this Discussion and the preceding data analyses, several recommendations are made.

Primary recommendations (action is urged on these points):

1) Sample using one method only

Sampling using both the M42 and Surber methods does not appear to substantially increase the information gained on macroinvertebrate assemblage structure in relation to liming and stream acidity.

2) Sample using the M42 method

Overall, M42 data was better correlated with acidity-related environmental variables and also better modelled those variables. M42 data collected more and higher abundances of acid sensitive taxa (especially the most sensitive species) and better discriminated limed from reference sites. This was generally true for most parameters, even during 2000-01, when relatively low invertebrate

abundances were sampled using M42, indicating a certain robustness to altered sampling intensity. The only caveat to this recommendation is that Surber sampling, as a more quantitative method, may arguably be better for detecting longer-term, interannual, shifts in the abundance of specific taxa or functional groups. However, this is only a problem if biomonitoring is focussed more on such abundance changes, rather than on changes in assemblage structure and competition.

3) *Clearly define specifications for M42 sampling and ensure that personnel are well-trained in the method, and that specifications are closely followed.*

M42 performed best when higher abundances (1500-2000) of invertebrates were collected. Parameters regulating the intensity of sampling (time of sampling, area of coverage, habitats included) need to be clearly established and adhered to, and staff should be well-trained in the method to ensure consistency both within and between years.

4) *Use saved resources to expand the breadth of biomonitoring: (I) Expand geographic coverage, and the set of reference streams*

Surber samples are time consuming to collect and process. If this method is dispensed with, the resources saved can be allocated to improving the coverage of the IKEU program. Three areas are worthy of attention:

a) Consider expanding the set of acid reference sites, to allow assessment of the extent to which limed streams have been "restored" from the reference condition. Sites upstream of currently limed sites could be considered.
b) Consider increasing the number of sites in poorly represented regions of Sweden. If the IKEU program is to be truly representative at a national scale, greater regional coverage is required.
c) Consider choosing paired reference sites for at least a subset of the limed sites, to reduce background noise and improve the rigour of statistical tests. Suitable sites for some IKEU streams may exist in the current set of reference streams, otherwise consider selecting new sites

Ideally, every limed stream would have one paired acid reference site and one paired reference site. This may in practice be impossible, but at a minimum, most limed streams could have a paired upstream reference site, provided it is not too divergent in physico-chemical conditions (width, degree of shading, substrate etc.)

5) *Analyse the current data file more deeply*

Species and water chemistry data from over 10 years of IKEU biomonitoring have now been combined into one file for the first time. The current analysis has been largely concerned with questions of methodology, but the resultant file provides a great opportunity for deeper analyses,

focussing on the impact of liming (e.g. a species- or stream-level assessment, or assessment of deleterious effects) and on interannual variation and long-term trends.

Additional recommendations (action is suggested on these points):

6) *Use saved resources to expand the breadth of biomonitoring: (II) Consider both autumn and spring sampling*

If monitoring is to be carried out in only one season, autumn is preferred, as macroinvertebrate assemblages are more stable at that time. However, analyses of the 2004-05 autumn and spring samples indicated that data collected at these different times of year may emphasise different aspects of the effects of liming and acidity on macroinvertebrate faunas. There are also strong *a priori* reasons for expecting this to be true, relating to the typical timing of acid episodes in Sweden. As a minimum, an initial commitment to monitor in both the autumn and spring for a further 2 years would allow a more complete assessment of the benefits of sampling in both seasons.

7) *Consider expanding the IKEU data set by incorporating data from other sources*

The IKEU data set could be further expanded by incorporating data collected by regional and kommun authorities. Data could be incorporated for both limed and reference streams. At a minimum, benefits and difficulties associated with this approach should be investigated.

Statistical references

Birks, H. J. B. (1998). "Numerical tools in paleolimnology - Progress, potentialities, and problems." *Journal of Paleolimnology* 20: 307-332.

Clarke, K. R. (1993). "Non-parametric multivariate analyses of changes in community structure." *Australian Journal of Ecology* 18: 117-143.

Clarke, K. R. and R. N. Gorley (2001). *Primer v5: User manual/tutorial*. Plymouth, UK, PRIMER-E.

Sneath, P. H. A. and R. R. Sokal (1973). *Numerical Taxonomy*. San Francisco, W. H. Freeman.

ter Braak, C. J. F. (1986). "Canonical correspondance analysis: a new eigenvector technique for multivariate direct gradient analysis." *Ecology* 67: 1167-1179.

ter Braak, C. J. F. and C. W. N. Looman (1986). "Weighted averaging, logistic regression and the Gaussian response model." *Vegetatio* 65: 3-11.

Appendix 1: Stream water chemistry

Table A1: Mean data for selected water chemistry variables over the period 1994-2005. See Table A2 for full stream names. Ordering from left-to-right reflects a geographic gradient from north-to-south (stream 1 is northernmost). Shading indicates liming. Other abbreviations: Av. = Average, Max = Maximum, Min = Minimum, Alk = Alkalinity, Tot. = Total, Inorg. = Inorganic, SBC = Sum base cations, TOC = Total organic Carbon, Cond. = Conductivity. Standard abbreviations are used for chemical elements.

			Stream numbers and names																					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20	21	22	23	25
Variable	**Av. L**	**Av. R**	STO	BAS	ARÅ	HSB	ÅDL	STF	KLS	HÄR	ENG	SÖR	HLD	LAX	SKG	EJG	HST	GNY	MOR	LJV	LBG	BLK	HRL	HOV	STR
LIMING	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**Yes**
pH	**6.69**	**6.42**	6.84	6.66	6.81	6.79	6.74	7.00	6.78	6.43	6.90	6.31	6.44	5.18	6.44	6.61	6.90	7.24	6.42	6.46	5.37	6.45	6.56	6.68	6.80
Max pH	**7.06**	**6.96**	7.15	7.07	7.16	7.13	7.05	7.39	7.00	7.05	7.25	6.96	6.87	6.00	6.92	7.05	7.32	7.55	7.14	6.89	6.08	6.95	7.08	6.99	7.15
Min pH	**6.23**	**5.80**	6.41	6.05	6.35	6.44	6.39	6.50	6.50	5.47	6.54	5.44	5.90	4.57	5.60	6.14	6.49	6.81	5.95	6.00	4.59	5.87	5.85	6.29	6.44
Alk (mekv/l)	**0.16**	**0.16**	0.16	0.07	0.11	0.16	0.16	0.20	0.17	0.11	0.18	0.08	0.14	-0.02	0.08	0.25	0.29	0.57	0.17	0.15	0.02	0.08	0.18	0.16	0.23
Max Alk	**0.26**	**0.31**	0.20	0.13	0.16	0.21	0.25	0.32	0.25	0.27	0.33	0.20	0.26	0.04	0.16	0.50	0.52	0.87	0.39	0.29	0.08	0.16	0.39	0.23	0.37
Min Alk	**0.08**	**0.05**	0.10	0.02	0.06	0.09	0.08	0.08	0.10	-0.01	0.09	-0.01	0.03	-0.08	0.01	0.14	0.14	0.28	0.05	0.06	-0.06	0.02	0.03	0.09	0.12
Tot. Al (µg/l)	**170**	**194**	96.0	60.4	35.9	115.3	127.9	43.6	247.3	153.5	170.3	170.7	189.6	323.0	224.7	487.2	147.5	73.7	195.5	168.3	237.5	166.6	278.1	133.0	336.9
Max Tot. Al	**277**	**383**	131.0	114.3	120.3	469.2	200.0	110.8	324.0	297.0	315.6	265.8	255.6	667.0	413.1	1006.0	227.0	203.7	309.5	286.9	363.0	387.6	411.3	214.8	453.6
Min Tot. Al	**94.2**	**101**	63.2	27.6	16.7	37.2	80.1	18.9	164.7	77.2	85.9	97.4	113.1	198.1	108.7	218.2	56.2	20.1	69.7	101.6	163.3	89.2	183.5	63.8	187.8
Inorg. Al (µg/l)	**4.76**	**14.8**	3.42	3.45	2.92	2.52	2.76	2.45	3.70	6.19	3.83	9.49	8.67	71.1	16.1	3.53	2.75	2.70	3.84	3.40	53.6	4.20	4.18	3.17	2.18
Max inorg. Al	**17.6**	**34.5**	6.6	9.9	5.5	6.0	5.8	6.0	16.2	32.3	13.8	36.4	37.3	143.5	79.8	14.4	5.1	6.7	12.5	11.4	90.7	17.7	21.0	8.6	3.7
Min inorg. Al	**2.12**	**5.17**	2.42	2.00	2.48	2.00	2.19	2.00	2.00	2.00	2.00	2.00	2.00	19.34	2.00	2.00	2.37	2.00	2.00	2.00	19.6	2.00	2.00	2.00	2.00
FE (µg/l)	**809**	**941**	265	71	146	217	612	238	532	495	303	1133	450	885	216	612	825	1042	750	729	1828	2184	3075	472	2973
SBC (mekv/l)	**0.53**	**0.53**	0.32	0.14	0.20	0.33	0.36	0.30	0.41	0.32	0.43	0.33	0.40	0.20	0.49	1.00	0.85	1.15	0.65	0.73	0.46	0.53	0.92	0.76	0.92
Si (mg/l)	**2.83**	**3.10**	1.98	1.68	1.95	2.97	3.71	2.60	3.37	4.16	3.07	3.78	2.75	3.19	1.26	2.71	3.56	3.66	3.47	3.19	2.01	2.39	3.83	2.52	4.19
TOC mg/l	**11.2**	**11.7**	9.66	3.9	3.61	7.0	15.24	6.2	14.07	10.7	9.45	14.6	9.34	15.4	8.29	12.1	12.84	7.3	14.4	9.57	14.3	15.33	22.4	6.92	20.00
Q (discharge)	**0.64**	**0.58**	1.00	1.52	1.63	0.64	0.74	0.81	0.24	0.36	0.48	0.31	0.55	0.14	0.35	0.88	0.50	0.36	0.13	0.51	0.56	0.70	0.66	0.63	0.29
Water level	**0.26**	**0.32**	0.24	0.30	0.10	0.20	0.27	0.42	0.16	0.17	0.18	0.46	0.47	0.45	0.26	0.21	0.24	0.23	0.19	0.37	0.49	0.23	0.38	0.33	0.24
Temp. (°C)	**6.30**	**6.52**	5.70	5.11	3.25	6.60	4.56	5.13	5.86	4.32	5.87	6.44	6.98	5.84	6.15	8.06	6.16	6.39	7.49	6.34	8.35	7.55	8.00	8.39	8.78
Max Temp	**15.1**	**15.4**	16.0	12.7	10.7	17.0	12.0	15.1	16.7	13.5	14.6	15.5	18.5	13.6	14.0	16.6	13.9	13.5	17.2	12.9	17.7	14.7	16.7	18.6	19.1
Cond (mS/m2)	**5.65**	**5.58**	3.21	1.55	2.19	3.36	3.31	2.87	3.89	3.28	4.47	3.33	4.32	2.83	5.62	10.6	8.88	11.8	6.79	8.15	5.54	5.69	9.39	8.58	9.51
NH4-N (µg/l)	**20.2**	**20.2**	10.7	5.3	5.5	9.8	8.2	7.4	11.8	11.6	16.2	15.4	17.6	11.7	27.9	31.2	25.1	26.4	18.1	18.7	47.7	21.9	37.4	32.0	46.8
Inorg. N (µg/l)	**176**	**185.6**	34.8	28.4	27.0	52.4	30.2	28.5	82.2	91.8	89.0	83.9	89.1	42.8	104.6	381.5	304.5	516.8	126.5	392.3	191.5	145.9	496.9	402.3	405.0
Tot. P (µg/l)	**11.9**	**14.3**	7.6	4.8	5.9	7.8	10.8	4.3	11.4	7.7	12.4	9.6	8.2	8.4	7.3	37.1	14.3	18.0	9.0	12.3	13.0	13.9	38.0	11.7	26.5
Mn (µg/l)	**118**	**221**	47	22	44	43	94	93	86	108	44	278	79	155	41	151	182	258	213	104	559	186	557	82	423
Cu (µg/l)	**0.84**	**0.55**	0.31	0.14	0.16	0.69	0.40	0.28	0.34	0.20	0.47	0.27	0.64	0.52	0.35	1.09	0.64	0.38	0.54	4.65	1.07	0.68	0.86	0.41	1.06
Zn (µg/l)	**4.35**	**4.30**	2.00	1.26	1.05	1.99	2.38	1.45	2.99	1.84	2.47	3.60	6.36	10.7	5.09	5.21	3.01	3.77	2.23	6.91	8.31	8.26	6.91	5.27	6.46
Cd (µg/l)	**0.03**	**0.03**	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.03	0.05	0.03	0.03	0.02	0.02	0.01	0.04	0.06	0.04	0.05	0.03	0.04
Pb (µg/l)	**0.33**	**0.42**	0.14	0.07	0.05	0.11	0.18	0.09	0.23	0.17	0.18	0.29	0.43	1.04	0.37	0.52	0.21	0.16	0.19	0.37	1.33	0.89	0.64	0.30	0.61

Table A2: Stream abbreviations and sampling period

Number	Abbreviation	Name	Limed?	Years sampled*
1	STO	Storselsån Storsele	Yes	1994-2005
2	BAS	Bastuån	No	1998-2005
3	ARÅ	Arån Arålund	Yes	1994-2005
4	HSB	Hornsjöbäcken	No	1998-2005
5	ÅDL	Ådalsån Lyckemyran (D)	Yes	1994-2005
6	STF	Stråfulan	No	1998-2005
7	KLS	Källsjöån Källsjöklack	Yes	1995-2005
8	HÄR	Härån (Storån)	No	1998-2005
9	ENG	Enångersån V. Lövås	Yes	1994-2005
10	SÖR	Sörjabäcken (Lillån)	No	1998-2005
11	HLD	Haraldssjöån Sandån Övre	Yes	1995-2005
12	LAX	Laxbäcken	No	1994-2005
13	SKG	Skuggälven Ängarna	Yes	1994-2005
14	EJG	Ejgstån	No	1998-2005
15	HST	Hästgångsån Hästgången	Yes	1994-2005
16	GNY	Gnyltån	No	1998-2005
18	MOR	Morån	No	1998-2005
19	LJV	Lillån G:a Järnvägsbron	Yes	1994-2005
20	LBG	Lillån-Bosgårdsån	No	2000-2005
21	BLK	Blankan Ryerna	Yes	1994-2005
22	HRL	Hörlingeån-Rökeå	No	1998-2005
23	HOV	Hovgårdsån Munkhättan	Yes	1994-2005
25	STR	Strönhultsån G. Kvarnen	Yes	1996-2005

*No stream was sampled in 2003