

# Continuous variation of macroinvertebrate communities along environmental gradients in northern streams

Jani Heino<sup>1)2)</sup>, Jari Ilmonen<sup>3)</sup> and Lauri Paasivirta<sup>4)</sup>

<sup>1)</sup> Finnish Environment Institute, Natural Environment Centre, Ecosystem Change Unit, P.O. Box 413, FI-90014 University of Oulu, Finland

<sup>2)</sup> Department of Biology, P.O. Box 3000, FI-90014 University of Oulu, Finland

<sup>3)</sup> Metsähallitus, Natural Heritage Services, P.O. Box 94, FI-01301 Vantaa, Finland

<sup>4)</sup> Ruuhikoskenkatu 17, FI-24240 Salo, Finland

Received 22 Nov. 2012, final version received 23 Mar. 2013, accepted 4 Mar. 2013

Heino, J., Ilmonen, J. & Paasivirta, L. 2014: Continuous variation of macroinvertebrate communities along environmental gradients in northern streams. *Boreal Env. Res.* 19: 21–38.

Our aim was to examine the nature of macroinvertebrate community variation across a set of streams in three drainage basins in Finland. We found that there were no clearly discrete community types, but rather macroinvertebrate communities varied continuously along environmental gradients. Local environmental factors and geographical location were strongly collinear and both were important in accounting for variation in macroinvertebrate community structure in the multivariate regression tree analysis and based on a combination of *k*-means clustering and discriminant analysis. We conclude (i) that geographical location and local environmental factors are strongly intertwined and both are associated with variation in macroinvertebrate communities across northern streams at the spatial scale of the three drainage basins studied; and (ii) that environmental assessment and conservation studies should not rely too much on delineating “community types”, but rather acknowledge the continuous variation of stream macroinvertebrate communities.

## Introduction

Community ecologists have long debated about the nature of changes in community structure along ecological gradients. This debate started almost one hundred years ago, when prominent vegetation ecologists had somewhat divergent opinions on whether communities formed discrete community types or showed continuous variation along environmental gradients (Shipley and Keddy 1987, Allen and Hoekstra 1992, McIntosh 1995, Leibold and Mikkelsen 2002). Clements (1916) argued that various species in communities respond similarly to underlying environmental gradients, leading to discrete community types.

By contrast, Gleason (1926) stated that community structure varies continuously along environmental gradients due to the independent responses of species to these gradients. The degree to which communities vary continuously or form discrete types is still debated, and the question of variation in community structure along environmental gradients is far from settled (Allen and Hoekstra 1992, McIntosh 1995). Understanding the degree to which community types can be considered clearly definable entities has important implications for environmental assessment and conservation planning across landscapes and wider regions (Anderson and Clements 2000, Heino *et al.* 2003a, Merovich and Petty 2010).

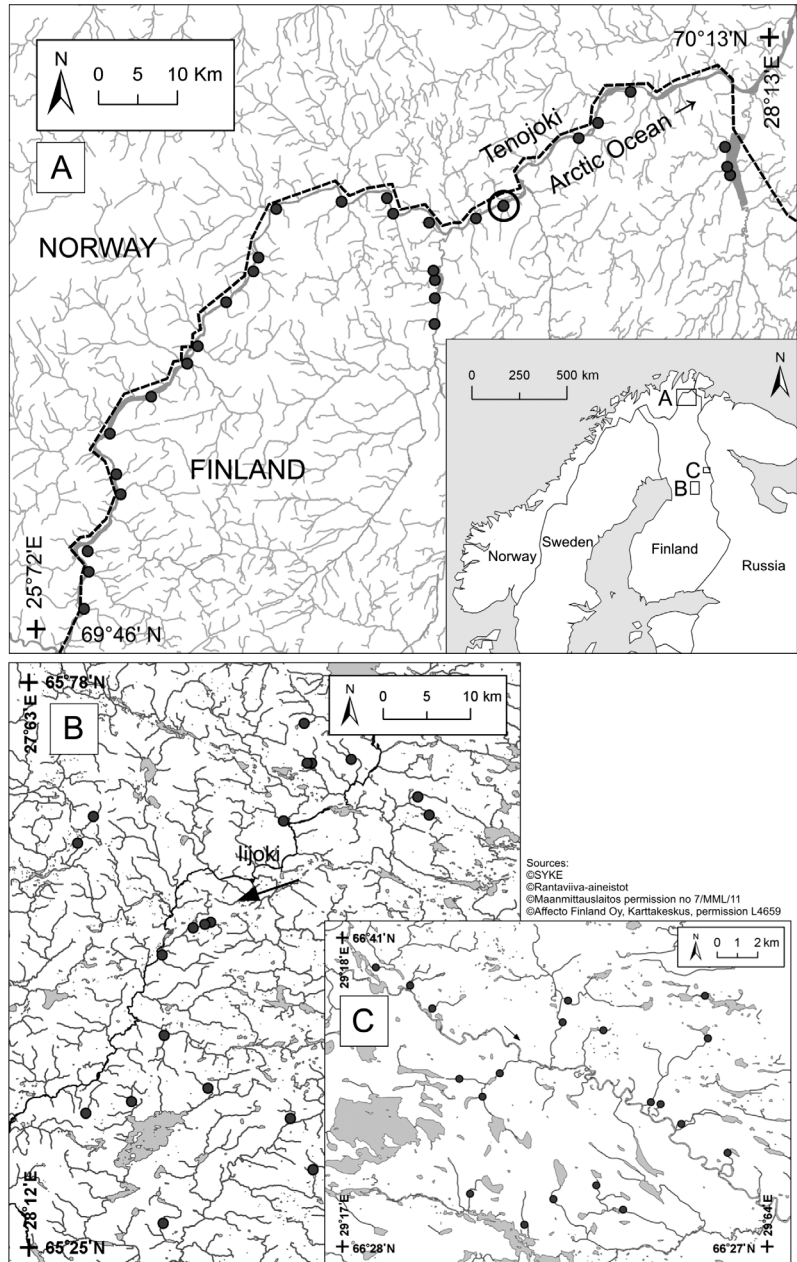
Community types in stream ecosystems have typically been defined using cluster analysis. However, many, if not most, stream studies have not considered the internal homogeneity of resulting community clusters (but *see* Van Sickle and Hughes 2000). This means that researchers are considering “community types” without taking into account how similar community structure among sites is within each cluster. Furthermore, relatively few studies have explicitly considered whether community structure varies continuously or if community types are clearly visible (Heino *et al.* 2003a, Sandin 2003, Merovich and Petty 2010). Understanding both the internal homogeneity of and overlap among community clusters is important for much of environmental assessment in stream ecosystems.

The next step after delineating community clusters is to examine their environmental relationships. These relationships can be assessed by means of discriminant analysis (Legendre and Legendre 1998) or multivariate regression trees (De'ath 2002), whereby the explanatory variables best discriminating among the community clusters can be considered the most important drivers of community variation (Wright *et al.* 1998, Heino *et al.* 2003a, Sandin 2003). In stream ecosystems, typical variables entering discriminant analysis and constrained ordination models include stream size, acidity, nutrients and various physical characteristics of streams (Townsend *et al.* 1983, Malmqvist and Mäki 1994, Malmqvist and Hoffsten 2000, Sandin and Johnson 2004). In addition to these environmental variables, also geographical location could be considered when associating community variation with ecological gradients. It is likely that across a spatial extent broad enough, geographical location supersedes the importance of local environmental variables in accounting for variation in community structure (Heino *et al.* 2003a, Sandin 2003, Mykrä *et al.* 2007). This is because the distributions of species should be restricted to certain regions, and they may thus be absent in some regions irrespective of suitable environmental conditions. Therefore, community variation in streams may be accounted for by either geographical location or local environmental conditions. Although increasing numbers of studies have examined the relative roles of spatial and environmental factors

for variation in community structure (Mykrä *et al.* 2007, Heino *et al.* 2012), fewer studies have explicitly asked if there are discrete “community types” that can be better explained by geographical location or environmental variables (Heino *et al.* 2003a, Sandin 2003).

For conservation studies, the identities of species associated with variation in community structure are also important. The identities of species are nowadays rarely revealed in community ecological studies, where the main focus is on community variation rather than individual species. Although finding generalities in community patterns does not require reporting species names, single species carry rich information about ecological communities and are necessary for planning the conservation of particular species. Thus, we opted for an approach where we reported all indicator species of our community clusters of northern streams.

We examined the nature of macroinvertebrate community variation along ecological gradients of northern streams. First, we attempted to define community types and their indicator species for a set of northern streams spanning a latitudinal gradient of about 500 kilometres in Finland. Second, we associated variation in community structure with environmental variables using multivariate regression trees (De'ath 2002) and a combination of *k*-means clustering and discriminant analyses (Legendre and Legendre 1998). Third, we examined if community structure varied continuously or if there were discrete community types. Finally, we examined the internal homogeneity of community clusters using test of homogeneity of dispersion (Anderson 2006). Based on existing information of northern headwater streams (Heino *et al.* 2003a, Sandin 2003, Heino 2005a), we hypothesised (i) that communities were parts of a continuum; (ii) that there would be much variation within each community cluster and overlap among the community clusters in species composition; and (iii) that the geographical location would be more important than the influences of local environmental factors in accounting for variation in community structure. In contrast to most previous studies, we examined whole stream macroinvertebrate communities, including often-omitted taxonomic groups such as non-biting midges, blackflies and water mites.



**Fig 1.** Locations of the three study regions in Finland and the study sites in each region. Regions: (A) Tenojoki drainage basin, (B) Iijoki drainage basin, and (C) Koutajoki drainage basin. Note that the sites in the Tenojoki basin are tributary streams close to the main stem river.

## Material and methods

### Study areas

The test data set comprised 70 stream riffle sites sampled in three drainage basins in northern Finland (Fig. 1). The details of the study areas and field methods will be reiterated here to facili-

tate understanding the ecological context of the three drainage basins, although these details have been previously described elsewhere (Heino *et al.* 2012, Heino 2013a). We sampled 20 streams in the Iijoki drainage basin (centred on 65°N, 27°E), 20 streams in the Koutajoki drainage basin (centred on 66°N, 29°E) and 30 streams in the Tenojoki drainage basin (centred on 70°N, 27°E)

in late May 2009, late May 2008 and early June 2010, respectively. As our resources did not allow sampling all the sites within a short period of time in a single year, we had to sample the sites in different regions in different years. We argue that it is more important to sample the sites in the same season (i.e. soon after the snowmelt in the spring) rather than in the same year. If the sites are not sampled within a short period of time in the same season, the results may not portray spatial differences but seasonal differences in stream macroinvertebrate communities due to the timing of aquatic insect life cycles.

All the stream sites sampled were near-pristine with regard to channel morphology and water chemistry, although some forestry practices were visible in the catchments of some streams in the two southerly regions. However, previous findings showed that biotic communities did not differ among draining-protected streams *vs.* managed forest catchments (Heino *et al.* 2009).

### Environmental variables

Several riparian, in-stream habitat and water chemistry variables were measured at each site. Percentage cover of deciduous trees of all trees was assessed in a 50-m section on both banks directly upstream of the sampling site. Shading was estimated visually as percentage canopy cover at 20 locations along transects (the number of which depended on stream width, i.e., there were more transects in smaller than larger streams) at the whole study section. Current velocity (at  $0.6 \times$  depth) and depth were measured at 30 locations along cross-stream transects, the number of which depended on stream width. Stream wetted width was measured at each site based on five cross-stream transects. Moss cover (%) and substratum particle class cover (%) were assessed at ten random randomly spaced  $50 \times 50$  cm plots. Visual estimates of the percentage cover of five particle size classes were made for each plot using a modified Wentworth scale: (i) sand (diameter 0.25–2 mm), (ii) gravel (2–16 mm), (iii) pebble (16–64 mm), (iv) cobble (64–256 mm), (v) boulder (256–1024 mm). Water samples were collected simultaneously with the

field sampling, and they were analysed for pH, conductivity, water colour, and total phosphorus using Finnish national standards (National Board of Waters and the Environment 1981). Water colour and total phosphorus were not measured in the Tenjoki drainage basin, as there is rather little variability in colour, and total phosphorus is typically below easily detectable limits (Heino *et al.* 2003b). The north and east coordinates of each site were recorded in the field using Garmin GPS.

### Macroinvertebrate data

Stream macroinvertebrates were sampled between late May and early June, depending on the latitude of a drainage basin. This is the season when the majority of macroinvertebrates in high latitude streams are still in the larval stage. The timing of sampling also facilitated the identification of aquatic insect larvae, most of which are close to their maximum size at this time of the year. At each site, we took a two-minute kick-net (net mesh size 0.3 mm) sample covering most microhabitats present in a riffle of approximately 100 m<sup>2</sup>. This spatial extent of a riffle site was visually estimated in the field, and the microhabitats present in the area were attempted to be included in the sample. This sampling effort typically yields more than 70% of species occurring at a northern headwater site in a given season, mainly missing sporadic species that occur only rarely in stream riffles (Mykrä *et al.* 2006). All macroinvertebrates sampled were identified to the lowest feasible level, this being species, species group or genus.

### Statistical methods

Our main statistical method was multivariate regression tree analysis (MRT; De'ath 2002). We used MRT based on Euclidean distance of Hellinger-transformed multivariate species abundance data (response variables) and all continuous environmental variables, north and east coordinates, and the categorical variable "region" (independent variables). MRTs are a method of constrained clustering which is highly

suitable for examining community–environment relationships (De’ath 2002). It forms clusters of sites by repeatedly splitting the data, each split is defined by a simple rule based on the values of explanatory variables, and the splits are chosen to minimise the dissimilarity of sites within clusters. MRT results in a tree whose terminal site groups (also called “leaves”) are composed of subsets of sites that are chosen to minimise the within-group sums of squares. Each successive partitioning of data is defined by a threshold value or a state of one of the explanatory variables. This method is particularly useful, as it retains a solution with the greatest predictive power and can handle a wide variety of situations, including a situation where community–environment relationships are non-linear (Borcard *et al.* 2011). The computation of MRTs consists of (1) constrained partitioning of the data and (2) cross-validation of the results; these computations are fully explained elsewhere (De’ath 2002, Borcard *et al.* 2011). In this study, we picked the “best” tree with the minimum cross-validated relative error (CVRE). MRTs were constructed using the R packages “mvp” (Therneau and Atkinson 2012) and “MVPARTwrap” (Ouellette and Legendre 2012).

Second, we used indicator species analysis (IndVal; Dufrene and Legendre 1997) to detect statistically significant ( $p < 0.05$ ; based on 1000 permutations) indicator species for each node and the leaves of the “best” multivariate regression tree. IndVal is a method that combines a species mean abundance and frequency of occurrence in each group. A high indicator value results when a species is both abundant (“specificity”) and occurs in most sites (“fidelity”) belonging to a group. The indicator value ranges from 0 to 1, with 1 referring to a perfect indicator species in terms of both “specificity” and “fidelity” to a group. IndVal was run in association with multivariate regression trees using the R package “MVPARTwrap” (Ouellette and Legendre 2012).

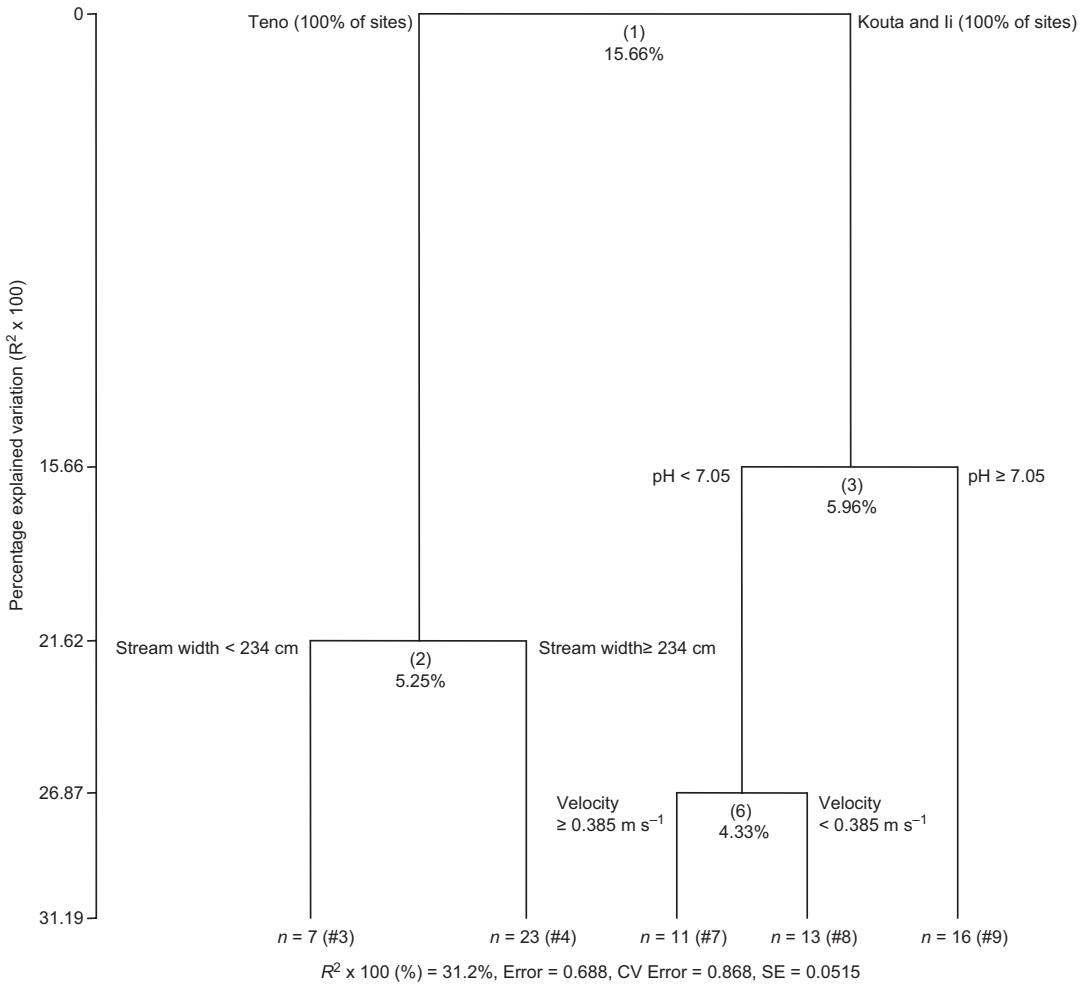
Third, we tested statistical differences in the environmental variables that were important in the MRT among the final leaves of the best tree. This was carried out using a non-parametric Kruskal-Wallis test, as the environmental variables were not always normally distributed. The differences in environmental variables among the final leaves

were visualised using boxplots. A Kruskal-Wallis test was performed and boxplots drawn using the R package “Rcmdr” (Fox 2005).

Fourth, to examine if the groups generated by MRT showed discrete Clementsian community types or continuous Gleasonian community variation, we ordinated the Hellinger-transformed species abundance data (Legendre and Gallagher 2001) using principal components analysis (PCA). We also tested the internal homogeneity of each community cluster and tested among-community cluster differences in internal homogeneity using test of multivariate homogeneity of groups dispersions (PERMDISP; Anderson 2006). This test is a multivariate analogue of Levene’s test for homogeneity of variances, and it can be used with any distance or dissimilarity index to examine variation in community composition. We used Hellinger transformation of species abundances followed by calculation of Euclidean distance to facilitate direct comparisons with the other analyses in this study. We also used multiresponse permutation procedure (MRPP; Mielke and Berry 2001) to test for average differences ( $p < 0.05$ ; based on 1000 permutations) in community structure between the community clusters. Hellinger-transformed species abundance data were again used prior to constructing Euclidean distance matrices for the MRPP analysis. PCA, PERMDISP and MRPP were run using the R package “Vegan” (Oksanen *et al.* 2012).

Finally, we used a combination of an unconstrained clustering method and discriminant analysis to assess the robustness of the MRT results. We thus used  $k$ -means partitioning to cluster the sites into five groups (Legendre and Legendre 1998). The five  $k$ -means groups were used to guarantee that the number of groups was comparable to that of the MRT groups in the “best” predictive model.  $K$ -means partitioning works by forming groups through identifying high density regions in the data and, contrary to other clustering methods, it is based on a pre-determined number of final groups. We used Hellinger-transformed species abundance data as the input matrix. A thorough overview of the method can be found elsewhere (Borcard *et al.* 2011). We compared the similarities between the MRT and  $k$ -means clusters by plotting sites in each grouping in PCA ordination





**Fig. 2.** A summary tree of the MRT analysis. There are four nodes (1, 2, 3 and 6) and five leaves (#3, #4, #7, #8 and #9) in the “best” regression tree. Shown are the four variables discriminating each node and the threshold values for the environmental variables. The first node was divided by the categorical variable region, with Tenojoki and Iijoki–Koutajoki sites being divided between the two clusters in the first node. CV Error = cross-validated mean error; SE = standard error.

space and by testing the relationship between the two groupings using a  $\chi^2$ -test (Borcard *et al.* 2011). *K*-means clustering was done using the R package “Rcmdr” (Fox 2005) and discriminant analysis based on Wilks’  $\lambda$  with the R package “klaR” (Weihs *et al.* 2005). All analyses were done in the R environment for statistical analysis (R Core Team 2012).

## Results

When we used MRT to examine for discrete

organisation in community structure, the best classifying scheme found by MRT was one of five groups (but this classification was extremely weak), and community structure was best described as varying continuously among study sites. The final MRT comprised four nodes (1, 2, 3 and 6) and five final leaves (#3, #4, #7, #8 and #9). This was the best predictive model, accounting for 31.2% of variation in community structure and having a cross-validated error of 0.868 (Fig. 2). The first node was related to the categorical variable “region”, the second node was related to stream width, the third node was

related to pH, and the sixth node was related to current velocity (Fig. 2). Surprisingly, north and east coordinates were not related to community structure prior to the variables mentioned above in the MRT analysis, although the first node largely divided the data into northern streams (with a high proportion of riparian deciduous trees) and more southern streams (with a lower proportion of riparian deciduous trees). Thus, geographically, the streams in leaves #3 and #4 were from the Tenjoki basin, most streams in leaves #7 and #8 were from the Iijoki basin, and the streams in leaf #9 were from the Koutajoki basin. The exceptions to the pure groupings by basins were the following: One stream from the Koutajoki basin was in leaf #7; and three streams from the Koutajoki basin were in leaf #8. Note that the same MRT result was provided by a model where the categorical variable “region” was replaced by the continuous variable deciduous trees. This was because of very high collinearity between the two variables.

A Kruskal-Wallis test showed that there were significant differences in environmental variables among the final leaves of the best multivariate regression tree. This finding was especially true for the four environmental variables important in discriminating the nodes of MRT, with deciduous trees ( $\chi^2 = 56.0465$ ,  $p < 0.001$ ), stream width ( $\chi^2 = 30.2944$ ,  $p < 0.001$ ), pH ( $\chi^2 = 38.0862$ ,  $p < 0.001$ ) and current velocity ( $\chi^2 = 28.7642$ ,  $p < 0.001$ ) all differing statistically among the final leaves. Thus, environmentally, leaf #3 was characterised by a high amount of deciduous trees in the riparian zone and small stream width; leaf #4 was typified similarly by a high amount of deciduous trees but larger stream width than in leaf #3; leaf #7 was characterised by high current velocity; leaf #8 was characterised by low current velocity; and leaf 9 was characterised by high water pH (Fig. 3). There was also some variation in other environmental variables among the final leaves, including moss cover and shading (Table 1).

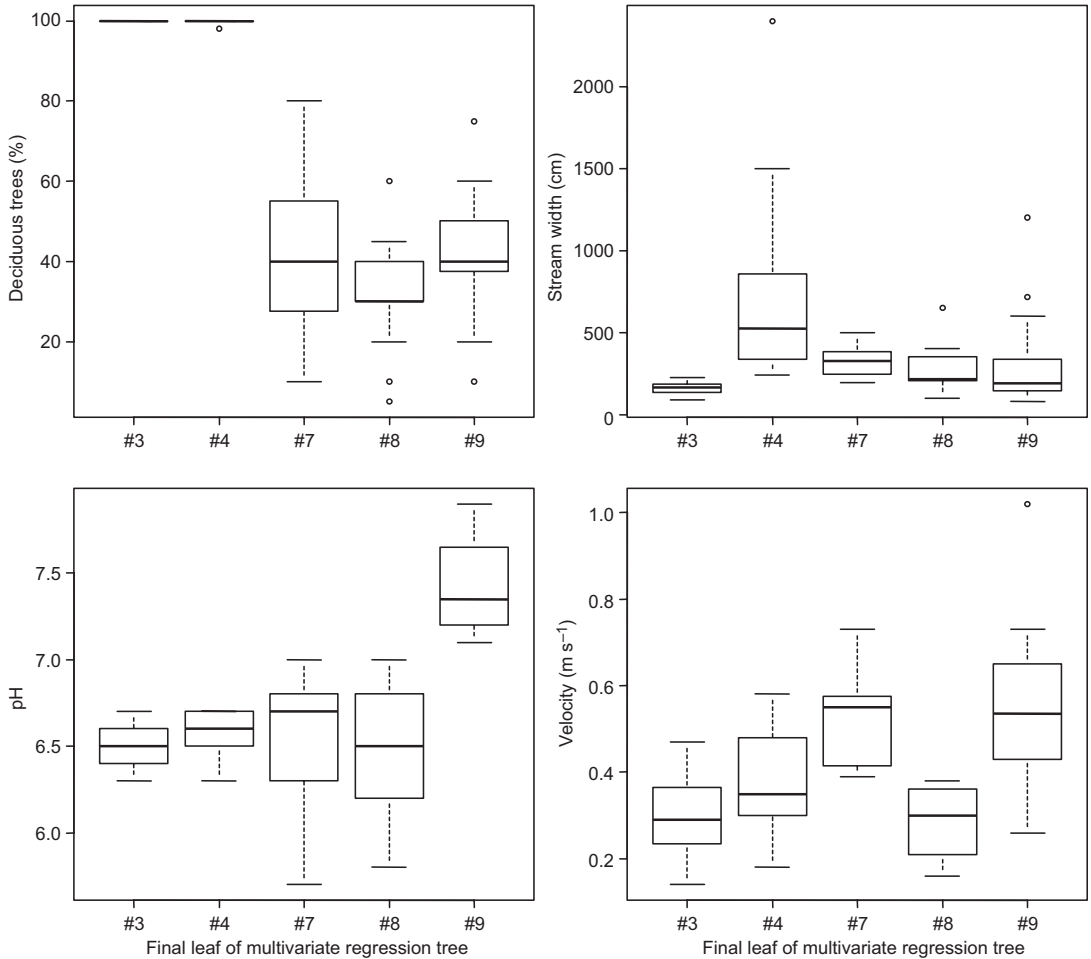
Of the total of 228 taxa detected, the first node of MRT was characterised by 16 and 49 significant indicator species in the left and right directions in IndVal analyses (Appendix 1). The best discriminating species were *Baetis rhodani* and *Orthocladius rivicola* (left) and *Simulium venum* gr. and *Elmis aenea* (right). The second

node was characterised by three (left) and seven (right) significant indicator species. The best discriminating species were *Nemoura* sp. (left) and *Baetis muticus* (right). The third node was characterised by 11 (left) and 15 (right) significant indicator species, with *Protonemura meyeri* (left) and *Isoperla grammatica* (right) being the most important indicators of this division. The sixth node was characterised by eight (left) and seven (right) indicator species, with *Prosimulium hirtipes* and *Baetis rhodani* being the best indicators for the left hand side group and *Thienemanimyia* ag. and *Micrasema gelidum* being the best indicators for the right hand side group.

Furthermore, based on IndVal, the final leaves of the multivariate regression tree were characterised by 49 significant indicator species (Table 2). For leaf #3, the best two indicator taxa were *Nemoura* sp. and *Corynoneura celtica*; for leaf #4, the best two indicator species were *Leuctra hippopus* and *Baetis rhodani*; for leaf #7, the best indicators were *Protonemura meyeri* and *Hydraena gracilis*; for leaf #8, the best indicators were *Limnephilus* sp. and *Asellus aquaticus*; and for leaf #9, the best indicators were *Isoperla grammatica* and *Habrophlebia lauta*.

The eigenvalues of the first two PCA axes were 0.142 and 0.077, accounting for 22.67 and 12.30 % of variation, respectively, in the species data. PCA showed that the groups generated by MRT were not discrete, but rather showed continuous variation and some overlap in the ordination space (Fig. 4A). The MRT groups also showed statistical differences in internal homogeneity (PERMDISP:  $F = 8.993$ ,  $p = 0.001$ ), with group #9 being the least homogeneous and group #4 the most homogeneous with regard to average distance to centroid (Fig. 5). Despite this internal heterogeneity, the groups were statistically different with regard to average differences in community structure (MRPP:  $A = 0.154$ ,  $p = 0.001$ ).

A comparison of the MRT and *k*-means analyses showed largely similar main patterns. Although there were some differences in the group memberships between MRT and *k*-means analyses, the similarity of the main patterns was clearly visible when comparing the two PCA plots (Fig. 4A and B) and supported by a significant outcome of the  $\chi^2$ -test between the



**Fig. 3.** Boxplots of the most important environmental variables for the nodes and final leaves of the MRT analysis. Horizontal line = median, box = 25th and 75th quartiles, whiskers = the highest and lowest values, excluding outliers, circles = outliers.

two groupings ( $\chi^2 = 153.821$ ,  $p < 0.001$ , based on 2000 permutations). Both plots showed the general pattern of continuous variation of community structure, although the  $k$ -means clusters showed more discrete variation in community structure than the MRT groups (Fig. 4A and B). PCA based on the environmental variables also showed that the environmental template was continuous (Fig. 4C and D). Furthermore, there were no differences in environmental conditions among the MRT or  $k$ -means clusters, but the clusters overlapped considerably in the ordination space (Fig. 4C and D).

Discriminant analysis showed that east and north coordinates entered first and second, respectively, in the model accounting for differ-

ences between the  $k$ -means groups, and were followed by a set of local environmental variables (Appendix 2). Similarly, north and east coordinates entered the Wilks'  $\lambda$  based model prior to local environmental factors when the MRT's final leaves were considered (Appendix 2).

## Discussion

### Main patterns of community variation

We detected five main community clusters using the constrained clustering of MRT analysis. These MRT community clusters were to some degree different from those shown by  $k$ -means



clustering, and the *k*-means groups showed slightly more structured “community types” than the MRT leaves when the sites of different groups were delimited by hulls in the ordination diagrams (Fig. 4A and B). However, the main patterns were similar between the two clustering methods. Thus, community variation followed predominantly a Gleasonian continuum, and there was an extensive overlap among the community types in ordination space. The present and previous findings from northern streams (Heino *et al.* 2003a, Sandin 2003, Heino 2005a) thus clearly show that there are no discrete Clementsian community types in these systems. The continuous nature of community variation across

northern streams in Finland most likely results from the independent responses of macroinvertebrate species to environmental gradients, as different species have highly different environmental niches (Heino 2005b). Similar was found for near-pristine streams from Sweden, where macroinvertebrate community types varied continuously along broad geographical and wide environmental gradients (Sandin 2003). The absence of discrete community types across our near-pristine sites was also not surprising in light of findings from strongly anthropogenically altered streams, where community variation was continuous despite discrete changes in environmental conditions (Merovich and Petty 2010).

**Table 1.** Mean, standard deviation (SD) and coefficient of variation (CV) for environmental variables in each final leaf of the MRT analysis.

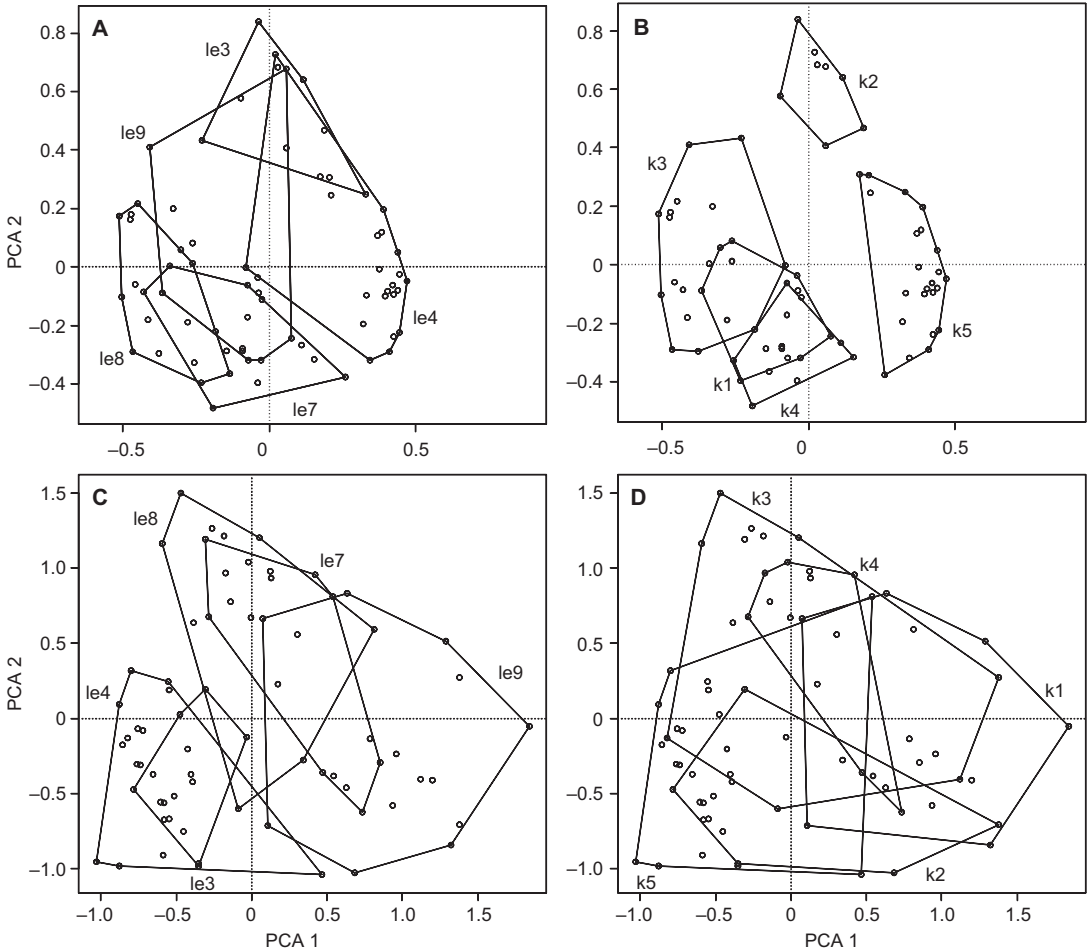
Variable	Leaf	Mean	SD	CV	Variable	Leaf	Mean	SD	CV
<b>Conductivity</b> (mS m <sup>-1</sup> )	#3	1.7	0.28	0.16	<b>Moss cover</b> (%)	#3	6.8	5.27	0.76
	#4	1.8	0.33	0.17		#4	2.9	3.95	1.33
	#7	2.3	1.04	0.44		#7	28.9	18.43	0.63
	#8	2.5	1.16	0.45		#8	46.4	28.07	0.60
	#9	7.6	3.82	0.49		#9	12.1	15.48	1.27
<b>pH</b>	#3	6.5	0.14	0.02	<b>Sand</b> (%)	#3	5.1	9.08	1.76
	#4	6.5	0.13	0.02		#4	0	0.20	4.79
	#7	6.5	0.38	0.05		#7	11	14.37	1.30
	#8	6.4	0.36	0.05		#8	13.8	19.74	1.42
	#9	7.4	0.27	0.03		#9	8.3	11.78	1.40
<b>Shading</b> (%)	#3	21.4	13.86	0.64	<b>Gravel</b> (%)	#3	2.0	3.14	1.57
	#4	13.8	14.21	0.02		#4	1.5	5.23	3.29
	#7	44.0	22.45	0.5		#7	7.5	11.60	1.53
	#8	27.6	15.08	0.54		#8	4.2	4.04	0.95
	#9	45.3	26.23	0.57		#9	10.3	8.27	0.79
<b>Deciduous trees</b> (%)	#3	100	0	0	<b>Pebble</b> (%)	#3	18.6	7.74	0.41
	#4	99.9	0.41	0		#4	13.3	12.53	0.93
	#7	42.6	21.71	0.50		#7	16.9	20.53	1.21
	#8	32.3	14.80	0.45		#8	10.0	10.41	1.03
	#9	42.5	15.16	0.35		#9	34.0	19.02	0.55
<b>Stream width</b> (cm)	#3	160.5	46.54	0.28	<b>Cobble</b> (%)	#3	49.2	21.43	0.43
	#4	700.8	515.44	0.73		#4	44.2	19.26	0.43
	#7	319.0	92.75	0.29		#7	29.2	15.40	0.52
	#8	272.0	145.85	0.53		#8	29.7	13.94	0.46
	#9	312.4	295.51	0.94		#9	25.0	15.62	0.62
<b>Depth</b> (cm)	#3	16.0	4.24	0.26	<b>Boulder</b> (%)	#3	25.0	19.92	0.79
	#4	19.6	5.39	0.27		#4	40.9	21.49	0.52
	#7	24.8	7.09	0.28		#7	35.2	27.87	0.79
	#8	23.9	9.36	0.39		#8	42.0	25.55	0.60
	#9	24.0	8.29	0.34		#9	21.9	25.87	0.17
<b>Velocity</b> (m s <sup>-1</sup> )	#3	0.3	0.10	0.36					
	#4	0.3	0.10	0.28					
	#7	0.5	0.11	0.21					
	#8	0.2	0.08	0.28					
	#9	0.5	0.19	0.35					

The different environmental niches of species, continuous variation and overlap among the community clusters were also suggested by considerable variation in community structure

across the sites within each community cluster. Although the multi-response permutation procedure showed statistically significant average differences among the community types, these

**Table 2.** Final IndVal results for leaves #3 to #9 of the MRT analysis. Nomenclature follows Fauna Europaea (<http://www.faunaeur.org/>). IV = Indicator value.

Taxon	Order	Leaf	IV	<i>p</i>
<i>Nemoura</i> sp.	Plecoptera	#3	0.4624	0.001
<i>Corynoneura celtica</i>	Diptera	#3	0.2703	0.010
<i>Leuctra hippopus</i>	Plecoptera	#4	0.5591	0.001
<i>Baetis rhodani</i>	Ephemeroptera	#4	0.4186	0.001
<i>Heptagenia dalecarlica</i>	Ephemeroptera	#4	0.4095	0.001
<i>Protonemura intricata</i>	Plecoptera	#4	0.3949	0.002
<i>Baetis muticus</i>	Ephemeroptera	#4	0.3763	0.008
<i>Simulium monticola</i>	Diptera	#4	0.3662	0.010
<i>Orthocladius rivicola</i>	Diptera	#4	0.3126	0.022
<i>Rhyacophila nubila</i>	Trichoptera	#4	0.3105	0.022
<i>Tokunagaia</i> sp.	Diptera	#4	0.2526	0.032
<i>Protonemura meyeri</i>	Plecoptera	#7	0.6559	0.001
<i>Hydraena gracilis</i>	Coleoptera	#7	0.4596	0.001
<i>Prosimulium hirtipes</i>	Diptera	#7	0.4558	0.001
<i>Isoperla difformis</i>	Plecoptera	#7	0.4415	0.001
<i>Silo pallipes</i>	Trichoptera	#7	0.3996	0.002
<i>Elmis aenea</i>	Coleoptera	#7	0.3638	0.005
<i>Rhyacophila obliterated</i>	Trichoptera	#7	0.3456	0.003
<i>Baetis niger</i>	Ephemeroptera	#7	0.3277	0.021
<i>Eukiefferiella devonica</i>	Diptera	#7	0.3156	0.028
<i>Micropsectra junci</i>	Diptera	#7	0.2875	0.038
<i>Elodes</i> sp.	Coleoptera	#7	0.2606	0.018
<i>Simulium tuberosum</i>	Diptera	#7	0.1818	0.033
<i>Limnephilus</i> sp.	Trichoptera	#8	0.5669	0.001
<i>Asellus aquaticus</i>	Isopoda	#8	0.4956	0.002
<i>Leptophlebia marginata</i>	Ephemeroptera	#8	0.4831	0.001
<i>Simulium venum</i>	Diptera	#8	0.4566	0.001
<i>Micrasema gelidum</i>	Trichoptera	#8	0.4335	0.002
<i>Thienemannimyia</i> ag.	Diptera	#8	0.4130	0.003
<i>Stempellinella brevis</i>	Diptera	#8	0.3826	0.002
<i>Corynoneura</i> sp.	Diptera	#8	0.3411	0.003
<i>Rheocricotopus fuscipes</i>	Diptera	#8	0.3383	0.005
<i>Micropsectra pallidula</i>	Diptera	#8	0.2660	0.027
<i>Hygrobates longipalpis</i>	Hydracarina	#8	0.2381	0.031
<i>Polypedilum convictum</i>	Diptera	#8	0.2308	0.023
<i>Synorthocladius semivirens</i>	Diptera	#8	0.2129	0.021
<i>Cricotopus festivellus</i>	Diptera	#8	0.2121	0.030
<i>Leptophlebia vespertina</i>	Ephemeroptera	#8	0.2036	0.049
<i>Isoperla grammatica</i>	Plecoptera	#9	0.7848	0.001
<i>Habrophlebia lauta</i>	Ephemeroptera	#9	0.5196	0.001
<i>Eiseniella tetraedra</i>	Oligochaeta	#9	0.4585	0.001
<i>Berdeniella</i> sp.	Diptera	#9	0.4391	0.004
<i>Hydropsyche angustipennis</i>	Trichoptera	#9	0.4375	0.001
<i>Leuctra digitata</i>	Plecoptera	#9	0.3922	0.005
<i>Oligochaeta</i> sp.	Oligochata	#9	0.3202	0.012
<i>Isoperla obscura</i>	Plecoptera	#9	0.3125	0.009
<i>Halesus</i> sp.	Trichoptera	#9	0.3115	0.008
<i>Bezzia</i> sp.	Diptera	#9	0.2773	0.036
<i>Parametriocnemus stylatus</i>	Diptera	#9	0.2709	0.022
<i>Hydropsyche saxonica</i>	Trichoptera	#9	0.2282	0.024

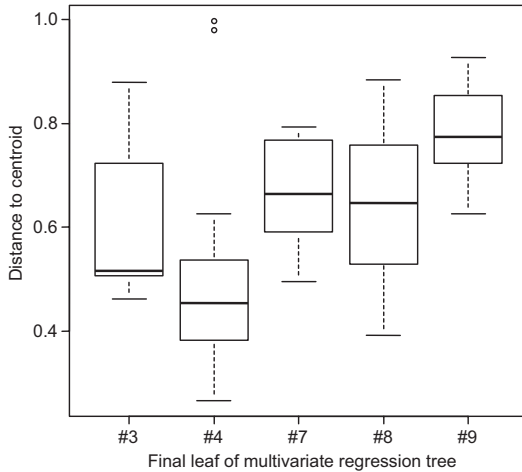


**Fig. 4.** PCA ordination plots of community data, showing the study sites delimited (A) based on the final leaves of the multivariate regression tree, and (B) based on  $k$ -means clusters. Also, shown are PCA plots of environmental data, with sites delineated (C) by the final leaves of the multivariate regression, and (D)  $k$ -means clusters.

differences were rather weak due to within-community type variation as shown by the test of homogeneity of dispersion. The most homogeneous community cluster comprised the sites in MRT leaf #4, although even in this community cluster two sites had large distances to centroid. This means that, even in the most homogeneous community cluster, there was much variation in community structure among sites. This was likely due to variation in stream size among the sites comprising this community cluster. The most heterogeneous community cluster comprised the sites in MRT leaf #9, sites of which were from the environmentally and biologically heterogeneous Koutajoki drainage basin (Heino 2013a; M. Grönroos unpubl. data). Thus, mac-

roinvertebrate communities are (i) highly variable across northern headwater streams and (ii) that true community types are not easily discernible.

The most important variables accounting for differences among the community clusters (at the nodes of the MRT analysis) were the categorical variable region (or riparian deciduous trees), stream width, pH and current velocity. All these variables have been formerly identified as important correlates of variation in macroinvertebrate communities in northern regions (Malmqvist and Hoffsten 2000, Heino *et al.* 2003a, Sandin 2003, Sandin and Johnson 2004, Heino and Mykrä 2008). Thus, it was not surprising that they were also important in the



**Fig. 5.** Boxplots of average internal homogeneity of community clusters, measured as mean Hellinger distance to centroid. Horizontal line = median, box = 25th and 75th quartiles, whiskers = the highest and lowest values, excluding outliers, circles = outliers. Group #4 is the most homogeneous = least variation in community composition among sites, and group #9 is the least homogeneous = most variation in community structure among sites.

nodes of the MRT analysis. Percentage of riparian deciduous trees may be related to stream type in general, as leaves #3 and #4, which comprised the sites with very high proportion of deciduous riparian trees, were northernmost streams from the Tenojoki basin. In addition to a high proportion of deciduous trees in the riparian zone, these northernmost streams also had low conductivity and low moss cover (Table 1). The second node was typified by stream width, with the northernmost streams being divided between small and large sites. Stream width has been shown to be one of the “master” variables, explaining variation in stream macroinvertebrate communities (Malmqvist and Mäki 1994, Vinson and Hawkins 1998, Heino 2009), and thus its importance in the present context was expected. The third node was characterised by water pH, with mostly slightly acidic Iijoki sites being located in the left hand cluster and alkaline Koutajoki sites being located in the right hand cluster. Although water pH did not vary very much across the present study sites, it nevertheless accounted for variation in community clusters. This finding suggests that macroinvertebrates respond to

relatively small changes in acidity and, had there been larger gradients in pH, water acidity might have even resulted in more discrete variation of community structure than detected in the present study. This suggestion is controversial, however, as Merovich and Petty (2010) did not find discrete community types across a large acidity gradient. The sixth node was related to current velocity, which divided slowly-flowing and more fast-flowing sites into two clusters. Current velocity is also one of the key variables accounting for variation in stream macroinvertebrate communities due to the fact that different species may prefer highly contrasting flow conditions (Allan and Castillo 2007). Although the mentioned environmental variables were important in discriminating among the nodes of the MRT analysis, it has to be emphasised that there was no discrete variation in environmental factors other than riparian deciduous trees among the final clusters. Thus, given the continuous nature of environmental variation, similarly continuous variation in community structure was the expected result.

The three drainage basins were separated by relatively large geographical distances. Thus, one might expect that geographical location supersedes the importance of local environmental conditions in determining the community types. This scenario was supported by MRT, as the categorical variable region entered first the best model. This finding was not surprising given the relatively large latitudinal gradient of the study that spanned three ecoregions (Heino *et al.* 2002), suggesting that geographical location is more important than local environmental factors in explaining variation in community structure across large geographical extents (ca. 500 km). This reasoning is also suggested by a previous study on northern headwater streams at a larger spatial extent in Finland (Heino *et al.* 2003), a study combining clustering and ordination methods across a large spatial extent in Sweden (Sandin 2003), and a study on the distance decay of macroinvertebrate communities in Finland (Astorga *et al.* 2012).

There are, however, statistically significant differences in environmental conditions among the three drainage basins studied here (D. Schmera unpubl. data). Thus, we cannot show

it decisively whether the identity of a drainage basin or local environmental conditions are the most important determinants of community variation. Furthermore, partial constrained ordination has shown that a large proportion of variation in community structure is shared between geographical location and local environmental variables across the three drainage basins (M. Grönroos unpubl. data). The high shared fraction in community variation also shows that environmental variables vary among the regions, and thus the categorical variable region, geographical coordinates or environmental variables may, by chance, enter first the MRT model. This suggests that careful consideration and simultaneous use of both constrained and unconstrained clustering methods is preferable in studies on community variation.

### Indicator species and their ecological characteristics

Many, if not most, statistically significant indicator taxa were either weak indicators of the MRT community clusters ( $IV < 0.5$ ) or, based on our previous knowledge of northern streams, are common species that occur across extensive environmental gradients (e.g. *Nemoura* sp., *Baetis rhodani*, *Rhyacophila nubila*, *Elmis aenea*, *Prosimulium hirtipes*, *Simulium vernalis*). These species appeared as significant indicators of the community clusters because their abundance varied among these clusters (i.e. high “specificity”), although they occur in most northern streams (i.e. low “fidelity”). The weak indicator status of most other species also suggests that species are distributed individually along the environmental gradients, and their distributions may be sporadic due to frequent disturbances typical in streams and due to subsequent extinction-colonisation dynamics (Heino and Mykrä 2008, Merovich and Petty 2010, Brown *et al.* 2011, Swan and Brown 2011).

In the following, we will address the ecological characteristics of the most important indicator species for the final leaves of the MRT analysis. For leaf #3 (i.e. smaller streams in the Tenojoki drainage basin), IndVal analysis detected two significant indicator taxa: *Nemoura* sp.

and *Corynoneura celtica*. The first taxon is a nearly ubiquitous shredding stonefly in northern streams, and its high indicator value for leaf #3 was due to its high abundance in these streams. It is likely that it attained high abundances in these small streams draining deciduous riparian zones, as coarse leaf material is an important food resource in this type of streams. *Corynoneura celtica* is a species that is increasingly more common in the northern parts of the study region (L. Paasivirta pers. obs.), and thus its affinity to the northern stream group was not surprising.

The most important indicator species for leaf #4 (i.e. larger streams in the Tenojoki drainage basin) were the mayflies *Baetis rhodani* and *Heptagenia dalecarlica*. Both these species are mostly scrapers, and thus association with larger, less-shaded streams is not surprising, as algae should be important resources there (Allan and Castillo 2007). However, *Baetis rhodani* is also present at most sites in northern streams (Heino 2005b), although its abundance may be higher in larger than smaller streams. *Heptagenia dalecarlica* is less of an environmental generalist than the above species, being typically found in the riffle sites of circumneutral-alkaline streams (Heino 2005b).

The shredding stonefly *Protonemura meyeri* and the scraping beetle *Hydraena gracilis* were the most important indicator species for leaf #7 (i.e. high-velocity streams in the Iijoki drainage basin). These species were associated with relatively high-velocity stream sites in our study, suggesting their preference for well-oxygenated and fast-flowing environmental conditions.

The shredding caddisfly *Limnephilus* sp. and the shredding water louse *Asellus aquaticus* were the best indicator species for leaf #8 (i.e. low-velocity streams in the Iijoki drainage basin). The sites comprising this community cluster were slowly-flowing, and thus it was not surprising that these two taxa with affinities to standing waters were rather common there. Although these taxa are almost ubiquitous in standing waters, they seem to be more restricted in distribution across the riffle sites of northern streams.

Finally, leaf #9 (i.e. alkaline streams in the Koutajoki drainage basin) was typified by the predatory stonefly *Isoperla grammatica* and

the gathering mayfly *Habrophlebia lauta*. These two species are not very common in northern streams, and showed affinity to high pH streams in this study. These species may thus be good early warning indicators of acidification in northern stream systems.

### Implications for environmental assessment and conservation planning

Defining community types is an important approach in community classification for applied purposes (Anderson *et al.* 2000, Heino *et al.* 2003a). This is because discrete community types of near-pristine streams could be regarded as potential benchmarks for environmental assessment or conservation planning. In this regard, our present findings were unconvincing, as we found mostly continuous variation in community structure. We thus suggest that applied studies should not take it for granted that communities can be divided into clear entities that are recurring in space and time. Rather, applied researchers should acknowledge that community variation is continuous and that modelling such variation requires using analytical methods that do not divide the data in separate clusters. Furthermore, as both geographical position and local environmental variables are typically important in affecting variation of community structure (Sandin 2003, Mykrä *et al.* 2007, Brown *et al.* 2011, Swan and Brown 2011), applied researchers cannot rely on the assumption that only niche-based processes affect natural variation and anthropogenic changes in community structure but that spatial processes may also be important (Siqueira *et al.* 2012, Heino 2013b). Applied researchers would thus benefit from applying ideas from metacommunity research to environmental assessment and conservation planning.

### Conclusions

We found that macroinvertebrate community variation in northern streams was continuous, showed much within-community cluster variation and exhibited clear among-community

cluster overlap. The continuous nature of macroinvertebrate communities was not surprising given the continuous nature of the underlying environmental template. This finding is strengthened by knowledge of the independent responses of species to environmental gradients, as was also suggested by the low indicator values for the species detected in our study. Our results thus strongly suggest that environmental assessment and conservation planning should not rely exclusively on “community types” as study entities, but rather acknowledge the continuous variation and high biological heterogeneity of stream ecosystems.

*Acknowledgements:* We thank Mira Grönroos, Tommi Karhu and Heikki Mykrä for help with the field work, and two anonymous reviewers for comments on this paper. Financial support was provided by the Academy of Finland, Kone Foundation and Maj and Tor Nessling Foundation.

### References

- Allan J.D. & Castillo M.M. 2007. *Stream ecology: structure and function of running waters*. Springer, New York.
- Allen T.F.H. & Hoekstra T.W. 1992. *Toward a unified ecology*. Columbia University Press, New York.
- Anderson M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62: 245–253.
- Anderson M.J. & Clements A. 2000. Resolving environmental disputes: a statistical method for choosing among competing cluster models. *Ecol. Appl.* 10: 1341–1355.
- Astorga A., Oksanen J., Luoto M., Soininen J., Virtanen R. & Muotka T. 2012. Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecol. Biogeogr.* 21: 365–375.
- Borcard D., Gillet F. & Legendre P. 2011. *Numerical ecology with R*. Springer, New York.
- Brown B.L., Swan C.M., Auerbach D.A., Campbell Grant E.H., Hitt N.P., Maloney K.O. & Patrick C. 2011. Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems. *J. N. Amer. Benthol. Soc.* 30: 310–327.
- Clements F.E. 1916. *Plant succession. An analysis of the development of vegetation*. Carnegie Institution, Washington, DC.
- De'ath G. 2002. Multivariate regression trees: a new technique for modelling species–environment relationships. *Ecology* 83: 1105–1111.
- Dufrene M. & Legendre P. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* 67: 345–366.
- Fox J. 2005. *The R Commander: a basic statistics graphical*



- user interface to R. *J. Stat. Soft.* 14: 1–42.
- Gleason H.A. 1926. The individualistic concept of the plant association. *Bull. Tor. Bot. Club* 53: 7–26.
- Heino J. 2005a. Metacommunity patterns of highly diverse stream midges: gradients, checkerboards and nestedness, or is there only randomness? *Ecol. Entomol.* 30: 590–599.
- Heino J. 2005b. Positive relationship between regional distribution and local abundance in stream insects: a consequence of niche breadth or niche position? *Ecography* 28: 345–354.
- Heino J. 2009. Biodiversity of aquatic insects: spatial gradients and environmental correlates of assemblage-level measures at large scales. *Freshwat. Rev.* 2: 1–29.
- Heino J. 2013a. Environmental heterogeneity, dispersal mode and co-occurrence in stream macroinvertebrates. *Ecol. Evol.* 3: 344–355.
- Heino J. 2013b. The importance of metacommunity ecology for environmental assessment research in the freshwater realm. *Biol. Rev.* 88: 166–178.
- Heino J., Grönroos M., Soininen J., Virtanen R. & Muotka T. 2012. Context dependency and metacommunity structuring in boreal headwater streams. *Oikos* 121: 537–544.
- Heino J., Ilmonen J., Kotanen J., Mykrä H., Paasivirta L., Soininen J. & Virtanen R. 2009. Surveying biodiversity in protected and managed areas: algae, macrophytes and macroinvertebrates in boreal forest streams. *Ecol. Ind.* 9: 1179–1187.
- Heino J., Muotka T., Mykrä H., Paavola R., Hämäläinen H. & Koskenniemi E. 2003a. Defining macroinvertebrate assemblage types of headwater streams: implications for bioassessment and conservation. *Ecol. Appl.* 13: 842–852.
- Heino J., Muotka T. & Paavola R. 2003b. Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. *J. Anim. Ecol.* 72: 425–434.
- Heino J., Muotka T., Paavola R., Hämäläinen H. & Koskenniemi E. 2002. Correspondence between regional delineations and spatial patterns in macroinvertebrate assemblages of boreal headwater streams. *J. N. Amer. Benthol. Soc.* 21: 397–413.
- Heino J. & Mykrä H. 2008. Control of stream insects assemblages: roles of spatial configuration and local environmental variables. *Ecol. Entomol.* 33: 614–622.
- Legendre P. & Gallagher E.D. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271–280.
- Legendre P. & Legendre L. 1998. *Numerical ecology*. Elsevier, Amsterdam.
- Leibold M.A. & Mikkelsen G.M. 2002. Coherence, species turnover, and boundary clumping: elements of metacommunity structure. *Oikos* 97: 237–250.
- Malmqvist B. & Hoffsten P.-O. 2000. Macroinvertebrate taxonomic richness, community structure and nestedness in Swedish streams. *Arch. Hydrobiol.* 150: 29–54.
- Malmqvist B. & Mäki M. 1994. Benthic macroinvertebrate assemblages in north Swedish streams: environmental relationships. *Ecography* 17: 9–16.
- McIntosh R.P. 1995. H. A. Gleason's "individualistic community concept" and theory of animal communities: a continuing controversy. *Biol. Rev.* 70: 317–357.
- Merovich G.T.Jr. & Petty J.T. 2010. Continuous response of benthic macroinvertebrate assemblages to a discrete disturbance gradient: consequences for diagnosing stressors. *J. N. Amer. Benthol. Soc.* 29: 1241–1257.
- Mielke P.W. & Berry K.J. 2001. *Permutation methods: a distance function approach*. Springer, New York.
- Mykrä H., Heino J. & Muotka T. 2007. Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. *Global Ecol. Biogeogr.* 16: 149–159.
- Mykrä H., Ruokonen T. & Muotka T. 2006. The effect of sample duration on the efficiency of kick-sampling in two streams with contrasting substratum heterogeneity. *Ver. Int. Ver. Theoret. Angew. Limnol.* 29: 1351–1355.
- National Board of Waters and the Environment 1981. Vesi-hallinnon analyysimenetelmät. *Publications of the National Board of Waters Finland* 213: 1–136.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H. & Wagner H. 2012. *Vegan: community ecology package*. R package version 2.0-4. Available at <http://CRAN.R-project.org/package=vegan>.
- Ouellette M.-H. & Legendre P. 2012. *MVPARTwrap: Additional functionalities for package mvpert*. R package version 0.1-9. Available at <http://CRAN.Rproject.org/package=MVPARTwrap>.
- R Core Team 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. [Available at <http://www.R-project.org/>].
- Sandin L. 2003. Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. *Ecography* 26: 269–282.
- Sandin L. & Johnson R.K. 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. *J. N. Amer. Benthol. Soc.* 19: 462–474.
- Sandin L. & Johnson R.K. 2004. Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. *Lands. Ecol.* 19: 501–511.
- Shiple B. & Keddy P.A. 1987. The individualistic and community unit concepts as falsifiable hypotheses. *Vegetatio* 69: 47–55.
- Siqueira T., Bini L.M., Roque F.O. & Cottenie K. 2012. A metacommunity framework for enhancing the effectiveness of biological monitoring strategies. *PLoS ONE* 7: e43626, doi:10.1371/journal.pone.0043626.
- Swan C.M. & Brown B.L. 2011. Advancing theory of community assembly in spatially structured environments: local vs. regional processes in river networks. *J. N. Amer. Benthol. Soc.* 30: 232–234.
- Therneau T.M. & Atkinson B. 2012. *Mypart: Multivariate partitioning*. R package version 1.6-0. Available at <http://CRAN.R-project.org/package=mypart/>.
- Townsend C.R., Hildrew A.G. & Francis J. 1983. Community structure in some southern English streams: the influence of physicochemical factors. *Freshwat. Biol.* 13: 521–544.
- Van Sickle J. & Hughes R.M. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for

- aquatic vertebrates in Oregon. *J. N. Amer. Benthol. Soc.* 19: 370–384.
- Vinson M.R. & Hawkins C.P. 1998. Biodiversity of stream insects: variation at local, basin, and regional scales. *Annu. Rev. Entomol.* 43: 271–293.
- Weihs C., Ligges U., Luebke K. & Raabe N. 2005. Analyzing German business cycles. In: Baier D., Decker R. & Schmidt-Thieme L. (eds.). *Data analysis and decision support*, Springer-Verlag, Berlin, pp. 335–343.
- Wright J.F., Furse M.T. & Moss D. 1998. River classification using invertebrates: RIVPACS applications. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 8: 617–631.

**Appendix 1.** Significant ( $p < 0.05$ ) indicator taxa at each node of MRT analysis. The indicator value of each taxon is in the direction of either “left” or “right” cluster. Nomenclature follows Fauna Europaea (<http://www.faunaeur.org/>).

	Taxon	Order	Cluster	IV	$p$
<b>Node 1</b>	<i>Baetis rhodani</i>	Ephemeroptera	left	0.7234	0.001
	<i>Orthocladius rivicola</i>	Diptera	left	0.7000	0.001
	<i>Simulium monticola</i>	Diptera	left	0.6806	0.001
	<i>Leuctra hippopus</i>	Plecoptera	left	0.6180	0.001
	<i>Nemoura</i> sp.	Plecoptera	left	0.6089	0.010
	<i>Micropsectra atrofasciata</i>	Diptera	left	0.5508	0.001
	<i>Rhyacophila nubila</i>	Trichoptera	left	0.5457	0.001
	<i>Protonemura intricata</i>	Plecoptera	left	0.5105	0.001
	<i>Baetis muticus</i>	Ephemeroptera	left	0.4096	0.032
	<i>Heptagenia dalecarlica</i>	Ephemeroptera	left	0.3842	0.004
	<i>Tvetenia discoloripes</i>	Diptera	left	0.3812	0.005
	<i>Paratrichocladius skirwithensis</i>	Diptera	left	0.2683	0.001
	<i>Simulium murmanum</i>	Diptera	left	0.2228	0.035
	<i>Tokunagaia</i> sp.	Diptera	left	0.2128	0.011
	<i>Thienemanniella majuscula</i>	Diptera	left	0.2000	0.006
	<i>Corynoneura lobata</i>	Diptera	left	0.1667	0.014
	<i>Simulium vernalis</i>	Diptera	right	0.7160	0.001
	<i>Elmis aenea</i>	Coleoptera	right	0.6301	0.001
	<i>Rhyacophila oblitterata</i>	Trichoptera	right	0.6250	0.001
	<i>Oligochaeta</i> sp.	Oligochaeta	right	0.5027	0.001
	<i>Hydraena gracilis</i>	Coleoptera	right	0.5005	0.001
	<i>Thienemannimyia</i> ag.	Diptera	right	0.4887	0.001
	<i>Limnephilus</i> sp.	Trichoptera	right	0.4750	0.001
	<i>Baetis niger</i>	Ephemeroptera	right	0.4738	0.004
	<i>Micrasema gelidum</i>	Trichoptera	right	0.4573	0.001
	<i>Protonemura meyeri</i>	Plecoptera	right	0.4500	0.001
	<i>Simulium ornatum</i> gr.	Diptera	right	0.4500	0.014
	<i>Potamophylax cingulatus</i>	Trichoptera	right	0.4411	0.001
	<i>Micropsectra pallidula</i>	Diptera	right	0.4250	0.001
	<i>Leuctra digitata</i>	Plecoptera	right	0.4185	0.001
	<i>Plectrocnemia conspersa</i>	Trichoptera	right	0.4016	0.003
	<i>Bezzia</i> sp.	Diptera	right	0.3897	0.001
	<i>Eukiefferiella brevicar</i>	Diptera	right	0.3820	0.002
	<i>Leptophlebia marginata</i>	Ephemeroptera	right	0.3750	0.002
	<i>Micropsectra junci</i>	Diptera	right	0.3750	0.001
	<i>Eiseniella tetraedra</i>	Oligochaeta	right	0.3700	0.003
	<i>Berdeniella</i> sp.	Diptera	right	0.3663	0.011
	<i>Isoperla grammatica</i>	Plecoptera	right	0.3500	0.001
	<i>Habrophlebia lauta</i>	Ephemeroptera	right	0.3000	0.003
	<i>Corynoneura</i> sp.	Diptera	right	0.3000	0.002
	<i>Rheocricotopus fuscipes</i>	Diptera	right	0.3000	0.002
	<i>Rheotanytarsus</i> sp.	Diptera	right	0.2969	0.005
<i>Eloeophila</i> sp.	Diptera	right	0.2889	0.004	
<i>Oulimnius tuberculatus</i>	Coleoptera	right	0.2884	0.005	
<i>Wiedemannia</i> sp.	Diptera	right	0.2762	0.013	

continued

## Appendix 1. Continued.

	Taxon	Order	Cluster	IV	p
	<i>Asellus aquaticus</i>	Isopoda	right	0.2750	0.003
	<i>Diura nanseni</i>	Plecoptera	right	0.2750	0.004
	<i>Sericostoma personatum</i>	Trichoptera	right	0.2750	0.001
	<i>Thienemanniella vittata</i>	Diptera	right	0.2750	0.005
	<i>Eukiefferiella claripennis</i>	Diptera	right	0.2578	0.018
	<i>Parametriocnemus stylatus</i>	Diptera	right	0.2500	0.005
	<i>Stempellinella brevis</i>	Diptera	right	0.2500	0.006
	<i>Hygrobates longipalpis</i>	Hydracarina	right	0.2250	0.003
	<i>Pisidium</i> sp.	Bivalvia	right	0.2250	0.006
	<i>Silo pallipes</i>	Trichoptera	right	0.2250	0.009
	<i>Nanocladius rectinervis</i>	Diptera	right	0.2055	0.038
	<i>Limnius volckmari</i>	Coleoptera	right	0.2000	0.025
	<i>Sialis fuliginosa</i>	Megaloptera	right	0.2000	0.014
	<i>Hydropsyche saxonica</i>	Trichoptera	right	0.2000	0.024
	<i>Halesus</i> sp.	Trichoptera	right	0.2000	0.019
	<i>Elodes</i> sp.	Coleoptera	right	0.1750	0.030
	<i>Polycentropus flavomaculatus</i>	Trichoptera	right	0.1750	0.038
	<i>Hydropsyche angustipennis</i>	Trichoptera	right	0.1750	0.042
	<i>Cricotopus</i> sp.	Diptera	right	0.1750	0.030
	<i>Capnopsis schilleri</i>	Plecoptera	right	0.1500	0.044
<b>Node 2</b>	<i>Nemoura</i> sp.	Plecoptera	left	0.7283	0.006
	<i>Oligochaeta</i> sp.	Oligochaeta	left	0.4121	0.011
	<i>Corynoneura celtica</i>	Diptera	left	0.2703	0.044
	<i>Baetis muticus</i>	Ephemeroptera	right	0.8696	0.001
	<i>Baetis rhodani</i>	Ephemeroptera	right	0.6871	0.001
	<i>Heptagenia dalecarlica</i>	Ephemeroptera	right	0.6522	0.004
	<i>Ameletus inopinatus</i>	Ephemeroptera	right	0.6309	0.018
	<i>Protonemura intricata</i>	Plecoptera	right	0.6226	0.037
	<i>Leuctra hippopus</i>	Plecoptera	right	0.5898	0.027
	<i>Amphinemura sulcicollis</i>	Plecoptera	right	0.4986	0.032
<b>Node 3</b>	<i>Protonemura meyeri</i>	Plecoptera	left	0.7500	0.001
	<i>Simulium venum</i> gr.	Diptera	left	0.7048	0.002
	<i>Limnephilus</i> sp.	Trichoptera	left	0.6646	0.002
	<i>Isoperla difformis</i>	Plecoptera	left	0.6380	0.002
	<i>Amphinemura sulcicollis</i>	Plecoptera	left	0.5833	0.001
	<i>Sperchon</i> sp.	Hydracarina	left	0.5722	0.003
	<i>Baetis niger</i>	Ephemeroptera	left	0.5461	0.033
	<i>Hygrobates longipalpis</i>	Hydracarina	left	0.3750	0.011
	<i>Asellus aquaticus</i>	Isopoda	left	0.3701	0.020
	<i>Lebertia</i> sp.	Hydracarina	left	0.3333	0.012
	<i>Leuctra nigra</i>	Plecoptera	left	0.2500	0.049
	<i>Isoperla grammatica</i>	Plecoptera	right	0.7996	0.001
	<i>Berdeniella</i> sp.	Diptera	right	0.6444	0.001
	<i>Eiseniella tetraedra</i>	Oligochaeta	right	0.5783	0.001
	<i>Habrophlebia lauta</i>	Ephemeroptera	right	0.5651	0.001
	<i>Oligochaeta</i> sp.	Oligochaeta	right	0.5472	0.022
	<i>Micropsectra atrofasciata</i>	Diptera	right	0.5162	0.001
	<i>Baetis muticus</i>	Ephemeroptera	right	0.5096	0.003
	<i>Leuctra digitata</i>	Plecoptera	right	0.4957	0.016
	<i>Hydropsyche angustipennis</i>	Trichoptera	right	0.4375	0.002
	<i>Rhyacophila nubila</i>	Trichoptera	right	0.4091	0.045
	<i>Heptagenia dalecarlica</i>	Ephemeroptera	right	0.3601	0.004
	<i>Halesus</i> sp.	Trichoptera	right	0.3407	0.002

continued

## Appendix 1. Continued.

	Taxon	Order	Cluster	IV	<i>p</i>
	<i>Isoperla obscura</i>	Plecoptera	right	0.3125	0.007
	<i>Tvetenia bavarica</i>	Diptera	right	0.3124	0.020
	<i>Amphinemura borealis</i>	Plecoptera	right	0.2545	0.030
<b>Node 6</b>	<i>Prosimulium hirtipes</i> gr.	Diptera	left	0.8463	0.001
	<i>Baetis rhodani</i>	Ephemeroptera	left	0.7053	0.009
	<i>Protonemura meyeri</i>	Plecoptera	left	0.6559	0.019
	<i>Hydraena gracilis</i>	Coleoptera	left	0.6447	0.011
	<i>Rheopelopia</i> sp.	Diptera	left	0.5249	0.010
	<i>Ephemerella aurivillii</i>	Ephemeroptera	left	0.5124	0.020
	<i>Silo pallipes</i>	Trichoptera	left	0.4461	0.032
	<i>Protonemura intricata</i>	Plecoptera	left	0.4288	0.031
	<i>Elodes</i> sp.	Coleoptera	left	0.3636	0.032
	<i>Thienemannimyia</i> ag.	Diptera	right	0.7036	0.001
	<i>Micrasema gelidum</i>	Trichoptera	right	0.7028	0.003
	<i>Simulium venum</i> gr.	Diptera	right	0.6631	0.006
	<i>Limnephilus</i> sp.	Trichoptera	right	0.5863	0.048
	<i>Leptophlebia marginata</i>	Ephemeroptera	right	0.5514	0.048
	<i>Asellus aquaticus</i>	Isopoda	right	0.5287	0.018
	<i>Stempellinella brevis</i>	Diptera	right	0.4400	0.039

**Appendix 2.** Results of linear discriminant function analysis with Wilks'  $\lambda$  of *k*-means clusters and the final multivariate regression tree (MRT) leaves (Note that this test is not completely valid, as the MRT leaves are affected by environmental variables). The variable minimizing Wilks'  $\lambda$  is shown at each step. SD = standard deviation.

Order	Variables	Wilks' $\lambda$	overall <i>F</i>	overall <i>p</i>
Results based on <i>k</i> -means clusters				
1	East	0.245	49.812	< 0.001
2	North	0.117	30.610	< 0.001
3	pH	0.075	22.965	< 0.001
4	Velocity	0.056	18.520	< 0.001
5	Velocity SD	0.041	16.353	< 0.001
6	Macrophytes	0.034	14.298	< 0.001
7	Boulder	0.029	12.703	< 0.001
8	Gravel	0.024	11.618	< 0.001
Results based on the final MRT leaves				
1	North	0.0120	1340.160	< 0.001
2	East	0.0017	368.806	< 0.001
3	pH	0.0007	197.769	< 0.001
4	Velocity	0.0005	127.034	< 0.001
5	Sand	0.0004	94.295	< 0.001
6	Macrophytes	0.0004	76.153	< 0.001
7	Shading	0.0003	64.595	< 0.001
8	Stream width	0.0003	56.452	< 0.001