JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE
201

Jukka Aroviita

Predictive Models in Assessment of Macroinvertebrates in Boreal Rivers

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 201

Jukka Aroviita

Predictive Models in Assessment of Macroinvertebrates in Boreal Rivers

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteen salissa YFL230
heinäkuun 3. päivänä 2009 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in Ylistönrinne, Hall YFL230, on July 3, 2009 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2009

Predictive Models in Assessment of Macroinvertebrates in Boreal Rivers

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 201

Jukka Aroviita

Predictive Models in Assessment of Macroinvertebrates in Boreal Rivers

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2009

Editors
Timo Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

Cover photo by Jukka Aroviita

URN:ISBN:9789513936372
ISBN 978-951-39-3637-2 (PDF)

ISBN 978-951-39-3604-4 (nid.)
ISSN 1456-9701

Jyväskylä University Printing House, Jyväskylä 2009

ABSTRACT

Aroviita, Jukka
Predictive models in assessment of macroinvertebrates in boreal rivers
Jyväskylä: University of Jyväskylä, 2009, 45 p.
(Jyväskylä Studies in Biological and Environmental Science,
ISSN 1456-9701; 201)
ISBN 978-951-39-3637-2 (PDF), 978-951-39-3604-4 (nid.)
Yhteenveto: Ennustavat mallit jokien pohjaeläimistön tilan arvioinnissa
Diss.

Biological assessments are widely required to evaluate and improve the condition of freshwater ecosystems that are significantly degraded. The need to assess biological condition across large geographical scales imposes a particular demand for consistent methods. My aims in this thesis were (1) to compare two alternative and widely applied approaches for assessment of river macroinvertebrate communities; simple *a priori* typology and multivariate RIVPACS-type modelling. Specifically, I examined (2) the influence of spatial scale on the performance of the approaches. I also compared (3) concordance of macroinvertebrate and macrophyte assessments, and (4) examined whether conventional site quality classifications based on macroinvertebrates could support preservation of threatened macroinvertebrate species (TS). The compared approaches are both based on the reference condition philosophy and measure biotic assemblage condition by observed-to-expected ratio of taxa (O/E). The RIVPACS-type modelling was generally more precise and more sensitive to anthropogenic impairment than the typology approach. However, the relative superiority of the two approaches was scale-dependent, with the performance of the typology approach increasing with decreasing spatial extent. Assessments of macroinvertebrate and macrophyte communities showed relatively high concordance, although the macrophyte model was strongly inferior to the invertebrate model. Number and abundance of TS showed a significant positive correlation with O/E. The TS were concentrated in sites classified as 'high' biotic quality, with few occurrences in 'good' status. The results suggest that large-scale freshwater bioassessments should use modelling approaches that account for continuous variation of biota and environment along multiple gradients. Simple typologies are likely to be usable only at regional scales and with simple environmental gradients. Biological quality classifications should be established with care to ensure that they agree with the ultimate environmental objectives (e.g. safeguarding TS).

Key words: Macroinvertebrates; spatial scale; streams; taxonomic completeness, typology; RIVPACS; Water Framework Directive.

Jukka Aroviita, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

Author's address Jukka Aroviita
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
jukka.aroviita@jyu.fi

Supervisors Dr. Heikki Hämäläinen
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Professor Roger I. Jones
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Reviewers Dr. Bruce Chessman
New South Wales Department of Environment and Climate Change
Parramatta NSW 2124
Australia

Dr. Jani Heino
Finnish Environment Institute
P.O. Box 413
FI-90014 University of Oulu
Finland

Opponent Professor Richard K. Johnson
Department of Aquatic Sciences and Assessment
Swedish University of Agricultural Sciences
P.O. Box 7050
SE-75007 Uppsala
Sweden

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

1 INTRODUCTION 7
1.1 Natural condition as a ′gold standard′ 8
1.2 Measuring biological condition 8
1.3 Controlling for natural variation 9
1.4 Classifying condition 11
1.5 Study aims 12

2 STUDY AREA AND DATA SOURCES 13
2.1 Study sites 13
2.2 Biological data 16
2.3 Environmental data 17

3 METHODS 18
3.1 Selection of reference sites 18
3.2 Regionalisations 18
3.3 O/E of taxonomic completeness 19
3.4 Estimation of expected taxa 20
3.4.1 Typology 20
3.4.2 RIVPACS 21
3.5 Evaluation of approaches 21
3.6 Performance measures 22
3.7 Classification of condition and threatened species 23

4 RESULTS AND DISCUSSION 24
4.1 Relationships between biota and environment 24
4.2 Performance of typology and RIVPACS at varying scale 25
4.3 Influence of geographical scale on RIVPACS-type models 26
4.4 Concordance of assessments of two organism groups 27
4.5 All or only common taxa 28
4.6 Bioassessment and threatened species 29
4.7 A note on reference condition approach 30

5 CONCLUSIONS 31

Acknowledgements 33

YHTEENVETO (RÉSUMÉ IN FINNISH) 34

REFERENCES 37

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original papers, which are referred to throughout the summary by the Roman numerals I–V. I planned studies I, III and V together with Heikki Hämäläinen, and II and IV together with the co-authors. I contributed to the collection of field data in III–V and to the collation of original datasets used in I. Large part of the original data of all papers were collected by earlier authors. I did data analyses for I–III and V, and those for IV together with Heikki Mykrä. I wrote first manuscript draft of I, III and V. Heikki Mykrä wrote first draft of II and IV. All papers were finished together with the co-authors.

I Aroviita, J., Koskenniemi, E., Kotanen, J. & Hämäläinen, H. 2008. A priori typology-based prediction of benthic macroinvertebrate fauna for ecological classification of rivers. Environmental Management 42: 894–906.

II Mykrä, H., Aroviita, J., Kotanen, J., Hämäläinen, H. & Muotka, T. 2008. Predicting the stream macroinvertebrate fauna across regional scales: influence of geographical extent on model performance. Journal of the North American Benthological Society 27: 705–716.

III Aroviita, J., Mykrä, H., Muotka, T. & Hämäläinen, H. 2009. Influence of geographical extent on typology- and model-based assessments of taxonomic completeness of river macroinvertebrates. Freshwater Biology 54: 1774–1787.

IV Mykrä, H., Aroviita, J., Hämäläinen, H., Kotanen, J., Vuori, K.-M. & Muotka, T. 2008. Assessing stream condition using macroinvertebrates and macrophytes: concordance of community responses to human impact. Fundamental and Applied Limnology 172: 191–203.

V Aroviita, J. Mykrä, H. & Hämäläinen, H. 2009. River bioassessment and the preservation of threatened species: towards acceptable biological quality criteria. Manuscript.

1 INTRODUCTION

'If you can not measure it, you can not improve it.'
Lord Kelvin, 1824–1907

Freshwaters are notorious for human alterations. Pollution, acidification, eutrophication, diversion, regulation, damming and warming of rivers and lakes, among other anthropogenic disturbances, intensified in the 20th century — and continue to the present. Societal concern about these alterations and particularly their deleterious ecosystem effects has led to the establishment of various environmental legislations in the developed world (e.g. Clean Water Act in United States and Water Framework Directive in European Union; Stoddard et al. 2006), aiming to protect and improve the condition of freshwater systems and their biota. These recent mandates differ fundamentally in philosophy from their predecessors in that whereas the old ones mainly required assessment of freshwaters by their chemical properties and from the point of human use, the new legislations are more holistic in nature, assigning value to ecosystems and biota *per se*, 'in their own right'.

The implementation of the new legislations is a two-fold task (e.g. Hawkins 2006a). First, biological assessments are needed to measure biological condition. There is a particular need to assess condition at large geographical scales which emphasises demand for developing comparable methods. Second, to improve any deteriorated condition by restoration, the causes of the deterioration need to be identified. How these goals should be achieved, however, have proven complex with little consensus to date (Hawkins 2006b). In this thesis, my fundamental aim has been to develop and test approaches for the first task; specifically, for measuring condition of benthic invertebrate communities in boreal rivers — in an intuitively meaningful and consistent manner.

1.1 Natural condition as a 'gold standard'

It has become an increasingly common practice to assess freshwater ecosystems by comparing their current condition to their expected natural condition (Hughes et al. 1986, Bailey et al. 2004, Stoddard et al. 2006). In this reference condition approach, 'naturalness' of ecosystems is explicitly valued by using the natural condition as a 'gold standard', both in setting expectations for assessments and objectives for restoration and recovery of impaired sites or regions (Bailey et al. 2004). The naturalness is often further portrayed as 'biological integrity' which Frey (1977) and Karr and Dudley (1981) defined as, in part, a '*community of organisms having a species composition, diversity and functional organization comparable to those of natural habitats within a region*' (see also Stoddard et al. 2006). These components of biotic integrity are now widely regarded as the key ecosystem properties whose preservation is of ultimate concern (e.g. Davies & Jackson 2006). The degree of anthropogenic biotic impairment can typically be indicated by deviation of potentially altered sites' observed values from their expected values (e.g. Moss et al. 1987, Bailey et al. 1998, Pont et al. 2006). These general concepts have gained widespread acceptance, but much less consensus has emerged on 1) how the biological condition should be measured, and 2) how the expected values should be estimated.

1.2 Measuring biological condition

One of the first decisions that needs to be made is what organism group or groups to assess (Resh 2008). Freshwater ecosystems are often assessed using well-known groups such as macroinvertebrates and fishes, with less-known groups like bryophytes receiving less attention (e.g. Diamond et al. 1996). If the condition of the group being assessed functions as a surrogate for the ignored groups there is no cause for concern, and assessing only the surrogate would clearly also be more cost-effective. However, studies on concordance of biotic communities in freshwaters have given mixed results, with some reporting strong concordance (e.g. between benthic invertebrates and fish; Jackson & Harvey 1993, Kilgour & Barton 1999), but others reporting weak concordance (e.g. among benthic invertebrates, fish, and bryophytes; Paavola et al. 2003). Different groups may indeed show differing responses to environmental conditions (Paavola et al. 2006) and thus, also to different kinds of human disturbances (Hering et al. 2006, Feio et al. 2007). Therefore, legislative mandates like the EU Water Framework Directive (WFD; Anon. 2000) have adopted a multi-taxon approach.

Another decision that needs to be made is what properties of the organism groups to assess. Biological assessments in aquatic environments have traditionally rested heavily on the use of indicator species or indicator community metrics (e.g. Rosenberg & Resh 1993, Hewitt et al. 2005). The indicators are typically based on species-specific sensitivities to different environmental conditions so that each indicator is calibrated to detect one type of impact at a time (e.g. nutrients; Kelly & Whitton 1995, acidification; Davy-Bowker et al. 2005, or organic pollution; Armitage et al. 1983). Although possibly being successful in indicating presence and intensity of particular stressors, or potentially valuable tools for seeking causes of impairment (Kilgour et al. 2004, Clews and Ormerod 2009), the stressor-specific indices might not necessarily be relevant for assessing general biological condition. Multiple stressors are typically simultaneously present in the environment, and a stressor-specific approach would require use of separate indicators for each type of stress, which might become unfeasible in practice. Furthermore, all stressors are not always known *a priori*.

Biological communities can also be assessed without any *a priori* assumption of the type of anthropogenic stress. Multivariate ordinations (James & McCulloch 1990, McCune et al. 2002) and community similarity measures (Green 1980, Novak & Bode 1992, Van Sickle 2008) are commonly used for this purpose. An increasingly popular approach is to use multivariate modelling to predict lists of taxa expected to be present in the absence of human-caused stress (i.e. in reference conditions; e.g. Moss et al. 1987, Resh et al. 2000, Oberdorff et al. 2001). The similarity of the observed taxa list to the predicted one can then be compared with an observed-to-expected ratio (O/E) of 'taxonomic completeness' (Moss et al. 1987, Clarke et al. 1996, Hawkins 2006). The models are often used to predict probabilities of taxa capture, like in River InVertebrate Prediction and Classification System (RIVPACS) (Wright et al. 2000), where O is the number of captured taxa that reach a predetermined capture probability threshold (p_t), and E is the sum of taxa probabilities ≥ p_t. Non-detection of the predicted taxa indicates anthropogenic impairment, but there area no other assumptions about the direction of anthropogenic change of communities. O/E of taxonomic completeness is thus stressor-nonspecific and is a universal index that allows direct comparisons of assessments across large geographical scales (e.g. Turak et al. 1999, Wright et al. 2000, Anon. 2006); accordingly, its use as an indicator for the general condition of biological communities can be supported.

1.3 Controlling for natural variation

Commonly a sample of near-natural reference sites is used to estimate the expected conditions, and the sites typically cover wide geographical areas and encompass substantial biological variation (e.g. Moss et al. 1987, Pont et al. 2006, Paulsen et al. 2008). A great challenge is to find efficient ways to control for the natural variation to allow detection of human-caused impairment (Johnson 1998, Cao et al. 2007). Commonly either *a priori* typologies (classifications) (e.g. Barbour et al. 1995, Hawkins et al. 2000a), or multivariate predictive modelling (e.g. RIVPACS; Wright et al. 2000) are used to control for natural biotic variation in freshwater bioassessments. Both approaches use groupings of similar sites, but they differ in their assumptions and methodology when selecting and controlling biologically meaningful environmental gradients that are insensitive to human activities (Bowman & Somers 2005). Typologies categorise sampling units into types *a priori* by environmental characteristics that are assumed to be important determinants of natural variation in biological parameters (e.g. Hawkins et al. 2000a). In contrast, the RIVPACS-type models first use clustering to place biologically similar samples into groups, and then linear discriminant function analysis to select *a posteriori* the key environmental determinants (predictors) of the biotic groupings (Wright et al. 1984). Expected type- or site-specific reference values for the biological properties can then be estimated with both approaches using either the *a priori*- or *a posteriori* -selected environmental variables, respectively.

Which of the two approaches should one use? Perhaps surprisingly, outcomes from typology- and RIVPACS-based assessments have rarely been compared directly, probably because the former have traditionally been used in developing stressor-specific multimetric indices (e.g. Barbour et al. 1995), and the latter for O/E based assessments (e.g. Hawkins et al. 2000b). However, the relative performance of the two approaches is important to know, not least because typologies have a pivotal role in the EU WFD and typology-based assessment systems are increasingly developed in Europe (Hering et al. 2004, Sandin & Verdonschot 2006), as well as elsewhere (e.g. Snelder et al. 2004, Turak & Koop 2008). Theoretically, the RIVPACS-type models should perform better than the categorical typologies, because they simultaneously take into account multiple and continuous environmental gradients that control the biological variation (Reynoldson et al. 1997, Joy & Death 2002). Most studies have indeed suggested that single-variable *a priori* typologies are inferior to biotic clustering in controlling for natural assemblage variation (Hawkins et al. 2000a, Heino et al. 2002). However, only direct comparison of bioassessments that measure same biotic properties would show the relative merits of the two approaches. The O/E index can be easily calculated type-specifically (Hämäläinen et al. 2002), and recent comparisons with O/E have also indicated better performance by the RIVPACS-approach (Davy-Bowker et al. 2006, Mazor et al. 2006, Neale & Rippey 2008).

A factor that might fundamentally influence the relative performance of typologies and models is the spatial scale at which the assessments are developed. Over large geographical extents, within-type biologic variation may become large and predictions by typologies imprecise (e.g. Hawkins et al. 2000a, Mazor et al. 2006). In contrast, RIVPACS-type models might perform better at large-scale because 1) location can be used as a continuous predictor variable (e.g. Smith et al. 1999) and 2) the number of predictor variables is not limited. The RIVPACS-type models might be influenced by scale essentially if importance of different predictors varies between different scales (Townsend et al. 2003, Mykrä et al. 2007). For example, large-scale models might primarily reflect large-scale relationships in species distributions and environment (e.g. catchment-scale variables as selected predictor variables), whereas smaller scale models might particularly reflect local-scale relationships (with local habitat variables as selected predictor variables).

Last, the relative performance of typology- and RIVPACS-based assessments could be further influenced by the subset of predicted taxa (all or only common) used in the O/E index calculations (Hawkins et al. 2000b, Clarke & Murphy 2006, Van Sickle et al. 2007). Particularly if geographical patterns in species distributions are strong, geographical extent could influence the pool of expected taxa. For example, if only common taxa were included in O/E calculations, large-scale typologies might exclude taxa with regionally restricted distributions, whereas RIVPACS-type models and spatially more detailed typologies should predict more accurately the occurrence of regionally common taxa.

1.4 Classifying condition

Finally, whatever organisms, indices and methods are used to evaluate condition of freshwater ecosystems, the final output is usually a verbal summary (e.g. ’good’ or ’fair’) for management purposes and for easy public communication. The boundary values for status classes are largely political (Stoddard et al. 2006), but are nevertheless of fundamental importance because they determine the areas considered to be in need of restoration. Impairment is typically suggested to have taken place if an observed value differs ’reasonably’ from the one expected in natural conditions, for example if it falls outside a threshold of the 10th percentile of the distribution of the reference values (e.g. Clarke et al. 1996, Kilgour et al. 1998). Minimal acceptable condition can then be anchored to this boundary value. However, the statistical quality class boundaries are arbitrary (Simpson & Norris 2000, Hawkins 2006), and concern may arise as to whether they are consistent with the ultimate environmental objectives. An alternative approach could be to select parameters indicative of the objectives, and to set boundary values according to the compliance with the actual targets. For example, a central objective of environmental management is to maintain biodiversity, including populations of threatened species. If a level of condition of biota that still supports the occurrence of threatened populations could be found, the critical limit would provide a meaningful boundary value for the assessment metric corresponding to minimum acceptable condition, e.g. the ’good’ status in WFD bioassessments.

1.5 Study aims

My main aim in this thesis was to evaluate alternative approaches for assessing condition of benthic macroinvertebrate communities in boreal rivers. A reference condition approach was applied in all studies and the O/E ratio of 'taxonomic completeness' was used as a measure of the biological condition. Analyses were based on data from 345 Finnish boreal river riffle sites. The study had four main themes. First, I compared bioassessment performance (predictive accuracy, precision, and sensitivity to detect anthropogenic impairment) of two alternative methods, *a priori* typologies and RIVPACS-type models, used to predict the river fauna (I, III). I specifically examined the influence of spatial scale on the relative performance of the two approaches (II, III). I anticipated that with increasing spatial extent the RIVPACS-models would allow for the control of continuous spatial variation and would thus perform better. Second, I examined whether assessments based on all taxa or on a subset of more common taxa (I–V) would perform better. I expected that exclusion of rare taxa would increase precision of O/E, but decrease sensitivity of detecting impairment. I specifically examined whether an optimal balance between taxa inclusion and exclusion could be found. Third, concordance between assessments of macroinvertebrate and macrophyte communities were studied (IV). It was specifically examined, whether these two groups could be used as surrogates for each other in river bioassessment. Fourth, I examined whether an assessment of macroinvertebrate assemblages (III) was associated with occurrence and abundance of threatened macroinvertebrate species (V). I specifically evaluated whether conventional assessments based on deviation from reference condition could support preservation of threatened species.

2 STUDY AREA AND DATA SOURCES

This thesis is based on analyses of biological and environmental data from 345 river sites from the boreal zone in Finland (61°40'–68°15' N, 21°31'–30°27' E; Fig. 1). The data consisted of samples of benthic macroinvertebrates (I–V), field survey of macrophytes (IV), field and GIS (Geographic Information System)-derived information on channel and catchment characteristics (I–V) and water quality measurements (I–IV). A large part of the data was obtained from previous surveys conducted in 1979–2003 (Table 1). New data were collected in 2004–2005.

Using previously published data had its pros and cons. The major benefit was that a relatively large dataset was obtained in short time with reasonable costs. A potential pitfall was that the varying sources and long time span might introduce uncontrolled variation to the analyses. Avoidance of biases was attempted by controlling for sample quality when collating the original data sets (see below). The main results of this study were probably unaffected by any remaining bias, which should have been equal in all the approaches developed and compared. Nevertheless, issues of data comparability certainly deserve particular consideration whenever 'final' bioassessments are development (see e.g. Olsen & Peck 2008).

2.1 Study sites

All study sites are swiftly flowing riffle sections representing probably the most commonly studied habitat in lotic ecology. The sampling sites represent a large size gradient (mean catchment area 579 km^2, range 0.4–9744 km^2) of humic and clearwater (mean colour value 171 mg Pt l^{-1}, range 10–600 mg Pt l^{-1}) streams and rivers draining both organic and mineral catchments (mean peatland cover 30 %, range 0–60 %). Partly different subsets of sites were included in the five substudies (Fig. 1), which encompassed Western Finland (I), Western and Northern Finland (II), Western and Central Finland (III, V), and Central Finland (IV).

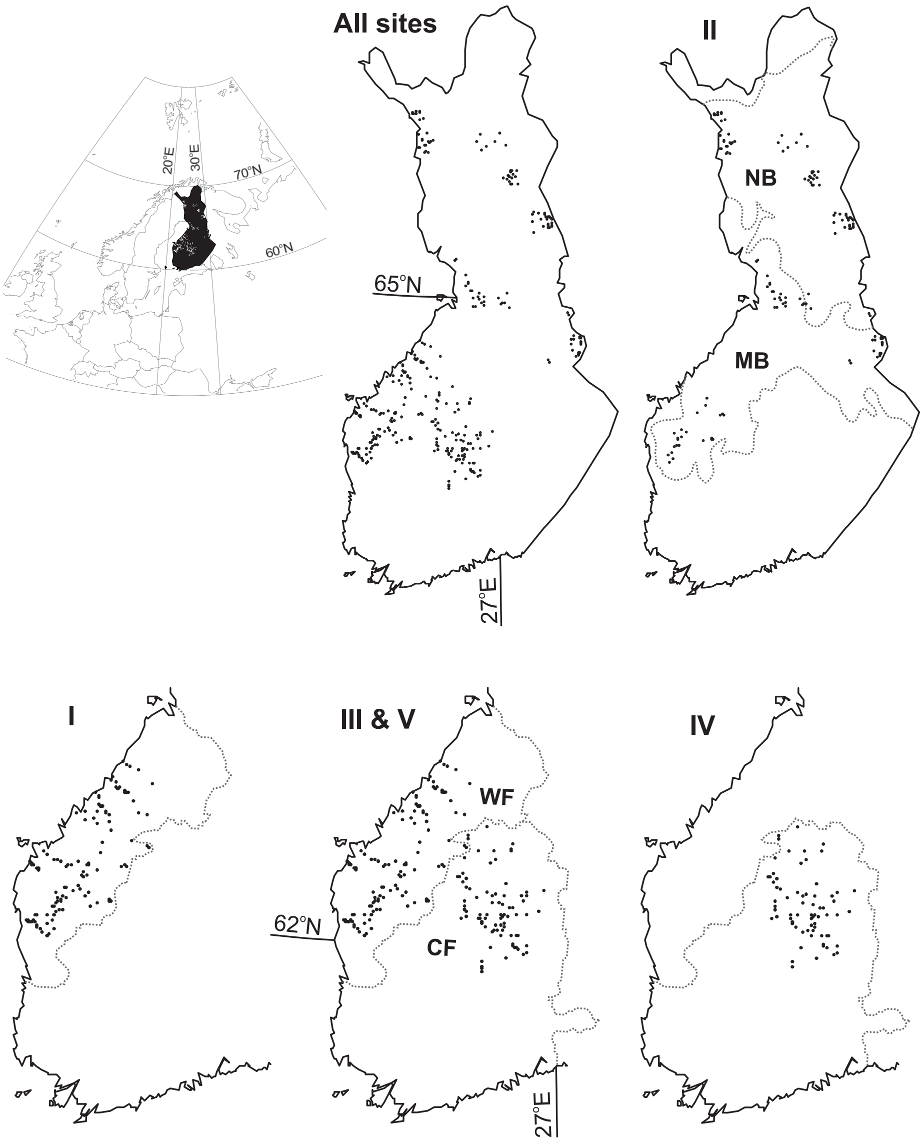

FIGURE 1 Location of the 345 river sites of studies I−V in Finland. The dashed lines delineate regionalisations used in II (ecoregions) and III (combinations of drainage basins; see text for details). NB=North Boreal ecoregion, MB=Middle Boreal ecoregion, WF=Western Finland, CF=Central Finland.

Western Finland sampling sites (n = 141; mean catchment area 1002 km^2, range 2–4835 km^2; Fig. 1) are in the drainage systems of rivers Lapväärtinjoki, Maalahdenjoki, Kyrönjoki, Lapuanjoki, Kovjoki, Purmonjoki, Ähtävänjoki, Kruunupyynjoki, Perhonjoki, Lestijoki and Kalajoki, which drain into the Gulf of Bothnia and mainly belong to the middle boreal ecoregion (Alalammi & Karlson 1988). Most of the sites were originally sampled for regional biological monitoring (Nyman et al. 1986, I), and study sites were selected among those from which good quality riffle macroinvertebrate data from autumn were available (see below). Catchments are peatland-dominated and are typically heavily used for agricultural purposes, but also for forestry. Large hydromorphological changes are evident in many rivers in the area (e.g. Nyman 1995). The 141 sites were studied in I, III and V (Table 1). Twenty-three of the sites (sampled by Heino et al. 2002) were also studied in II.

TABLE 1 Sources and number of sampling sites of the three datasets for studies I−V. Number of reference (REF) and impacted (IMP) sites among the five substudies is also given.

Dataset	Source	Total	I	II	III	IV	V
Various[1]	See table 1 in I	116	116	-	116	-	116
FIBRE[2]	Heino et al. (2002)	165	25	143	45	20	45
MOSSE[3]	III, IV	64	-	-	64	51	64
Total		345	141	143	225	71	225
REF[4]		216*	51*	143	96*	36	96*
IMP[5]		134	95	-	134	35	134

[1] Sampling by Nyman et al. (1986) and various monitoring projects in Western Finland in 1979−2000. [2] 'Finnish Biodiversity Research Programme' (subproject: 'Biodiversity and its conservation in boreal streams'; Anon. 2003a), sampling from Western, Central and Northern Finland in 1998-2000. [3] 'Monimuotoisuuden Tutkimusohjelma' (subproject: 'Predictive modeling in the classification of freshwater habitats, with implications to impact assessment and management'; Anon. 2008a), sampling from Central Finland in 2003-2005 (see also Heino et al. 2007) and includes data from two theses (Ruokonen 2004, Majuri 2008). [4] Reference sites that are minimally altered by human activities. [5] Impacted sites where biota was considered potentially impaired by human activities. * The numbers include historical reference data from five IMP-sites (see text for details).

Sampling sites in Central Finland (n = 84; mean catchment area 670 km^2, range 4–9744 km^2; Fig. 1) belong to drainage systems Kymijoki and Kokemäenjoki that drain into the Gulf of Finland and the Bothnian Sea, respectively, and belong mainly to south boreal ecoregion (Alalammi & Karlson 1988). Stratification was applied so that both least-disturbed and impacted sites with approximately similar catchment size and proportion of peatland in the catchment were selected. The area is part of the Finnish lake-district, where rivers typically originate from headwater lakes or connect larger lakes. Both humic and clearwater rivers are present. Catchments are used for both agriculture and forestry. Studies III and V shared all 84 sites and 71 of them were included in IV (Table 1).

Sampling sites from Northern Finland (n = 120; mean catchment area 21 km^2, range 0.4–111 km^2; Fig. 1) drain into the Gulf of Bothnia and are either within the catchments of the rivers Kiiminkijoki and Oulujoki, belonging to the middle boreal ecoregion, or else are within the catchments of the rivers Muonionjoki, Kemijoki and Koutajoki, belonging to the north boreal ecoregion. The sites were randomly selected among least-impacted headwater streams (Heino et al. 2002) and were studied in II (Table 1).

2.2 Biological data

The macroinvertebrate community data (I–V) were sourced from Nyman et al. (1986), Heino et al. (2002), monitoring reports, theses and new stream surveys of 47 sites in Central Finland (see Table 1). Larger data sets were initially collated but only samples with consistent sampling procedure and sufficient taxonomic resolution were included in this study. All samples were taken from fast-flowing riffle sections in autumn (late August to October) in 1979–2005. A sample consisted of a 1.5-min (Nyman 1995) or 2-min (Mykrä et al. 2006) composite kick sample taken with a hand-net (mesh size 0.3–0.5 mm) and aimed to cover most micro-habitats within each riffle section. This sampling effort has been shown to collect >70 % of species present at a site in a given season (Mykrä et al. 2006). Samples were preserved in ethanol in the field and sorted in the laboratory, where all animals were identified and counted. Chironomids were not identified beyond family level and they were omitted from all analyses. Taxonomy was harmonised for each substudy to avoid duplicate taxon records. The macroinvertebrate data contained 409375 non-chironomid individuals distributed to 186 identified taxa (111 species, 52 genera, 20 families and three higher groups). The average number of taxa and individuals per sample were 23 (range 5–47) and 1170 (16–18345), respectively.

The macrophyte community data (IV) were collected from 71 of the same riffle sections as macroinvertebrates in Central Finland in autumn 2005 (by Mr. Jukka Salmela, University of Jyväskylä). The macrophytes were studied from ten 50 cm × 50 cm quadrates randomly placed to those areas of the streambed that remain wetted during base flow. This sampling method allows detection of 70–75 % of bryophyte species present in a riffle section (IV). All macrophyte species were identified in the field and their percentage cover recorded (mean 39.6 %, range 0.1–81.8 %). The flora consisted of 30 aquatic bryophyte species and four vascular plant species.

2.3 Environmental data

Several characteristics of the catchment of each macroinvertebrate sampling site were measured from maps with GIS (by Mr. Juho Kotanen, South Savo Regional Environment Centre). These included latitude, longitude, site altitude, channel gradient, catchment area, distance to source and various land use and land cover features like percentage of peatland and cultivated land. Catchment characteristics were used for river typology (I, III) as candidate model predictors (I–V), or as variables indicative of human influence (I–V). Sites subject to regulation for hydropower production or flood protection were also identified from Western Finland (I).

A set of site-specific physical characteristics were measured concurrently with the macroinvertebrate sampling (II and IV, Heino et al. 2002). The variables included riparian (tree species composition, riparian integrity, shading) and in-stream (stream width, depth, current velocity, moss cover and substratum particle size) habitats. Some of these variables were also used as candidate model predictors in II and IV.

Data on water chemistry were obtained for each macroinvertebrate sampling site from the Hertta database of the Finnish Environment Institute (SYKE) or by taking water samples at the time of the biotic sampling (FIBRE and MOSSE data-sets; Table 1). Finnish national standards were used in all analyses. The macroinvertebrate sampling sites were associated with the nearest water chemistry sampling sites found in Hertta (median distance 3.0 km). Acceptably consistent data were available for alkalinity, total phosphorus concentration, total nitrogen concentration, pH, iron concentration, conductivity and water colour. Water chemistry variables were used to indicate human influence on the studied streams (I–V). Alkalinity was used as a candidate model predictor in II and IV.

3 METHODS

3.1 Selection of reference sites

This study was based on the reference condition approach (Bailey et al. 2004, Stoddard et al. 2006) where expectation for the biota is estimated from a sample of reference (REF) sites considered minimally altered by human activities. Catchment characteristics and expert judgement by local environmental authorities were used to select 211 REF-sites (no point-source pollution or marked habitat alteration, <15 % of catchment area cultivated, ≤5 % of catchment area used for peat production). All available sites from large rivers were slightly disturbed so those sites were among the best available ones (*sensu* Stoddard et al. 2006). Historical samples from five sites taken before their loss of reference status were additionally assigned to the REF data set (I, III and V; Table 1), which thus contained observations from 216 sites. No biological data were used in selection of the REF-sites to avoid any circular reasoning.

All remaining 134 sites were subject to considerable organic or nutrient pollution, forestry, peat mining, acidification caused by intensive land use in sulphite-rich, acid soils of the coastal area, or to hydrological or habitat alteration, and they were assigned to a group of impacted sites (IMP, Table 1) where the biota was considered to be potentially impaired by human activities.

3.2 Regionalisations

Influence of regionalisation on bioassessment performance was studied in II and III. Two spatial extents were used in both studies. The larger extent encompassed the entire study area and the smaller extent comprised two distinct regions, delineated either by forest vegetation ecoregions according to Alalammi & Karlson 1988 (II, north and middle boreal ecoregion, Fig. 1) or by combinations of river basins (III, Western and Central Finland, Fig. 1). Both options are commonly used and have appeared biologically meaningful approaches for spatial division of freshwater assemblage data (e.g. Frissell et al. 1986, Heino et al. 2002). A single model (i.e. null model and RIVPACS-type model in II, and also a typology model in III, see below) was developed for the whole study area, and the same models separately for each of the two regions. Results from the regional models were pooled and performance measures (see below) calculated over all sites of the two regions, which allowed a direct comparison of smaller and larger scale assessments using the same set of sites. Regionalisation was expected to increase particularly the performance of the typology approach.

3.3 O/E of taxonomic completeness

Biological condition (macroinvertebrates in I–V and macrophytes in IV) of the sampling sites was assessed with the O/E index of 'taxonomic completeness' (Moss et al. 1987, Hawkins 2006) which compares observed (O) and expected (E) lists of taxa. The expected list was those taxa that were predicted to be observed in the absence of anthropogenic alteration. The O/E ratio was calculated as follows. First, for each site, biological and environmental data from REF-sites were used to estimate probabilities of taxa capture at reference conditions. The probabilities are called capture probabilities, because they are conditional to the sampling procedure used. Typology and RIVPACS-method were used and compared (I, III) as two alternative methods to estimate the probabilities (see below for details). The observed value (O) was then calculated as the number of captured taxa that reached a predetermined taxa probability threshold, p_t; and the expected number of taxa (E) was the sum of the estimated probabilities $\geq p_t$ (Moss et al. 1987). The O/E index should theoretically be close to one in reference conditions when expected numbers of predicted taxa are observed. In contrast, when anthropogenic impairment leads to partial disappearance of the predicted taxa, O decreases and O/E ratios fall below one.

An important decision that needs to be made with the O/E index is whether to include all taxa, including also rare taxa with any probability of capture > 0 (as specified by p_t = '0+'), or only a subset of taxa predicted to be locally more common (e.g. $p_t = 0.5$) in the calculation and consequently in assessment of the biological condition (Moss et al. 1987, Hawkins et al. 2000b). Inclusion of rare taxa with small p may cause undesirable 'noise' and hinder detection of impact, whereas more common 'core' taxa with large p might have wide environmental tolerances and therefore also be relatively inert to human-caused disturbances. Therefore, discrimination between IMP and REF-sites was expected to decrease with large p_t. The rate of these changes with p_t, however, could not be anticipated and attempts were made to estimate an optimal p_t. To fully examine the behaviour of the O/E index in the different situations (biotic groups, methods to estimate E, spatial scales), varying p_t-levels were examined to evaluate bioassessment performance of the O/E index (I–V). These were either the whole p_t-range (from '0+' to 0.9 at 0.1 intervals) or all *vs* common taxa (II, IV, V).

3.4 Estimation of expected taxa

Two methods were used to estimate the taxa capture probabilities for the O/E calculations. These were a simple *a priori* typology (I, III) and RIVPACS-type modelling (I–V). The approaches were directly comparable and their relative merits were assessed in I and III. Both approaches are essentially predictive models, but whereas the typology-approach estimates type-specific occurrence probabilities for taxa, the RIVPACS-approach estimates site-specific occurrence probabilities.

3.4.1 Typology

The *a priori* typology-approach in bioassessment is based on the assumption that environmental characteristics that are selected *a priori* to define the types are important determinants of natural variation of biological parameters (Hawkins et al. 2000a). The main guideline for present bioassessment systems in Europe (WFD; Anon. 2000) strongly emphasises the typology-approach. Simple catchment size categories according to WFD 'System A' typology were used here to group study sites into river types (I and III). The represented types were 'very small' (catchment area <10 km^2), 'small' (10–100 km^2), 'medium sized' (100–1000 km^2) and 'large' (1000–10000 km^2) rivers. The smallest rivers are not included in System A, but they were included here to incorporate a wider size spectrum. Other factors of System A were either not relevant (all sites of I and III were 'lowland rivers' <200 m above sea level from 'Fenno-Scandian shield' ecoregion), or the distribution and low number of sites among geology types in Western (organic) and Central Finland (organic and siliceous) did not allow further stratification. Numerous studies have shown that river size is a key factor explaining variation of lotic macroinvertebrate assemblages (e.g. Vannote et al. 1980, Malmqvist & Mäki 1994), and the size typology was thus expected to explain at least part of the variation. As a first step (I), the performance of the typology in accounting for the natural variation in macroinvertebrate assemblages was evaluated using non-metric multidimensional scaling ordination (NMS; Kruskal 1964, McCune et al. 2002).

The type-specific probability of capture for each taxon in the absence of human-caused stress was estimated as the ratio of the number of REF-sites at which the taxon was recorded to the total number of studied REF-sites in that type (I, III). This calculation is analogous to that applied in RIVPACS-type predictive models (Moss et al. 1987, see below) when for a given site the probability of a type (cluster) membership is either 1 (for one type) or 0 (for all other types). In total 12 type-models were developed; four types in each of

Western Finland (I, III), Central Finland (III), and combined Western and Central Finland (III).

3.4.2 RIVPACS

Environmental characteristics that are important determinants of natural variation of the biota are derived *a posteriori* in the River InVertebrate Prediction and Classification System (RIVPACS; Wright et al. 2000) type models. The environmental characteristics are thus specific to the biota-environment -relationship present in each models' calibration data. The methodology of RIVPACS-type models that combine clustering and discriminant function (DF) analysis is well described elsewhere (e.g. Moss et al. 1987, Wright et al. 2000, Ostermiller & Hawkins 2004) and only essential details of the model construction are given here.

First, the REF-sites were grouped with the Flexible-β clustering algorithm (Agglomerative Nesting; Kaufman & Rousseeuw 1990) using the Bray-Curtis dissimilarity measure (Legendre & Legendre 1998) on the biological datasets. The cluster dendrograms were used to differentiate biologically similar site groups. Second, the DF analyses were used to identify the best predictors of biological groupings from a set of candidate environmental variables insensitive to human influence (Wright et al. 1984). The predictors were identified with 'all-possible-subsets' (Van Sickle et al. 2006, I–III, V) or stepwise selection procedure (IV), avoiding overfitted models. Third, the resulting DF models were used to predict an *a posteriori* probability of cluster membership for all sites. The membership probabilities were then multiplied by frequencies of reference taxa occurrences within each cluster. Finally, the products were summed to obtain the site-specific probability of occurrence for each taxon in the absence of human-caused stress (Moss et al. 1987).

In total eight different RIVPACS-type models were developed: one for each of Western Finland (I, III), Central Finland (III), and combined Western and Central Finland (III); one for each of the middle boreal ecoregion (II), the north boreal ecoregion (II) and the combined middle and north boreal ecoregions (II); and one for Central Finland for both macroinvertebrates and macrophytes (IV). Clustering and DF analysis were performed in R-program (I–III and V; Anon. 2008b), or (IV) in PC-Ord (McCune & Mefford 1999) and SPSS (Anon. 1999), respectively.

3.5 Evaluation of approaches

To provide a baseline for comparisons, the performance (see below) of all approaches was evaluated by comparing them with corresponding null models (Van Sickle et al. 2005). In a null model, all sites belong to one group and a single probability of occurrence is predicted for each taxon. A null model thus does not account for any variability among REF-sites and provides a simple way to evaluate the benefit of more complex approaches. Separate null models were constructed in each case (the same eight in total as in RIVPACS).

The performance of any statistical model should ideally be evaluated with independent validation data that have not been used in model calibration (e.g. Hastie et al. 2001). In practice, reference sites are often too scarce to set aside for completely independent validation set and this study was no exception in this respect. The number of REF-sites allowed independent validation sites only in II, where predictive accuracy and precision (see below) of the models was estimated by fitting the models to 30 REF-sites (across-ecoregions, 15 in each ecoregion) set aside from model calibration. In III and V, internal leave-one-out cross-validation was used to evaluate performance. The cross-validation procedure was applied by first re-estimating E for each REF-site using a set of sites from which that particular site was excluded. The respective O-value was then recalculated to obtain a cross-validated O/E for that site. A stable model structure was maintained in the RIVPACS-approach so that both clustering and DF predictor variables were kept unchanged. The validation and cross-validations were run separately for all null models, typology models, and RIVPACS-models at each spatial extent and p_t. In I and IV, model quality was assessed by calculating O/E back to the calibration data. The cross-validation (III) and particularly the back-fitting (I, IV) might result in overoptimistic results of model quality. However, since the main goal of these studies was to compare relative merits of the different approaches, this was not regarded a major problem.

3.6 Performance measures

The bioassessment performances of the different approaches were evaluated by their precision (I–V) and sensitivity (I, II–V) of the O/E. A measure of precision essentially indicates how well a given approach controls for the natural reference variation of biota.

Root mean squared error (RMSE) of reference site O/E (II) was used to combine accuracy and precision in a single measure (Wallach & Groffinet 1989). Precision was estimated by standard deviation of the cross-validated O/E (III and V) or standard deviation (SD) of the reference site O/E (I, IV). Smaller RMSE and SD indicate greater precision (Ostermiller & Hawkins 2004, Van Sickle et al. 2006).

Sensitivity of the assessments to detect human impairment was evaluated using O/E ratios from IMP-sites (I and III–V; Table 1). Sensitivity was measured by the percentage of IMP-sites that had O/E < the 10th (I, IV) or 25th (III, V) percentile of the REF-O/E -distribution (cross-validated in III and V). The percentages indicate departure of the IMP-sites from the reference variation. The percentiles are arbitrary and do not necessarily reflect any particular threshold of community response, but they provide an objective means for method comparisons (e.g. Hawkins 2006b). Typology-based O/E ratios were evaluated by percentiles of the corresponding type-specific distributions (I, III, Anon. 2003b).

To evaluate whether the estimated condition of macroinvertebrate assemblages was attributable to the observed intensity of human disturbance, the relationship of O/E with environmental variables indicative of human influence was examined either with regression analysis (I) or with Pearson correlation (III) by examining the strength of the relationships between O/E and Principal Components of summarised impact gradients.

3.7 Classification of condition and threatened species

Finally, the percentile cut-points were further used to develop a narrative classification of biological condition (V). Following a common practice (e.g. Anon. 2003b, Paulsen et al. 2008, Feio et al. 2009), the percentiles were used as alternative boundary values for the 'high' (equal to undisturbed reference conditions) and 'good' (equal to acceptable condition) status classes that the WFD demands. The good status, and three lower condition classes ('moderate', 'poor' and 'bad') in WFD, represent unacceptable condition that require management actions. These classes were derived by dividing the O/E range between the high/good boundary (HG, 10th or 25th percentile) and 0 (= lowest possible value of O/E) to four even parts. Accordingly, sites with O/E values >HG were considered to be in high, the remaining sites with O/E >¾HG in good, with O/E >½HG in moderate, with O/E >¼HG in poor, and the remaining sites with O/E <¼HG in bad biological condition.

Threatened macroinvertebrate species (hereafter TS) were selected from the Red Lists of Finnish species (Rassi et al. 2001) by considering all threat categories for those groups (Ephemeroptera, Odonata, Plecoptera, Coleoptera and Trichoptera) with sufficient taxonomic resolution in the data set (V). Additionally, also species classified as threatened by the EU Habitats directive (Anon. 1992) were considered. Spearman rank order correlation was then used to assess if the number of TS and number of TS individuals were associated with the O/E (V). All analyses were conducted separately with the two percentile thresholds to evaluate whether the acceptable 'good' status is sufficient to protect the threatened species or, whether a more acceptable boundary could be found.

4 RESULTS AND DISCUSSION

4.1 Relationships between biota and environment

The natural variation of macroinvertebrate communities was most strongly associated with river size. This pattern was evident in all analyses throughout the thesis: the reference sites grouped in accordance with their catchment size in NMS-ordinations (I), the catchment size types and the biotic clusters showed strong concordance (III), and the predictor variables (catchment area, stream slope, current velocity) selected to the RIVPACS-type models are strongly associated with river size (I–V). The difference between faunas of small and large rivers is not a novel finding (Vannote et al. 1980, Malqvist & Mäki 1994), but, however, supports the use of the *a priori* catchment area categories in the typology (I and III). Other variables that accounted for a considerable amount of variation in macroinverterbrate species composition and richness were latitude, altitude, peatland cover and % of lakes in the catchment (as identified by DFA in the RIVPACS-type models; I, III, IV), agreeing with earlier studies reporting the important drivers of macroinvertebrate community structure in running waters (Malmqvist & Mäki 1994, Heino et al. 2002). Variables insensitive to human activity explained a reasonable amount of natural variation in macroinvertebrate community composition (cross-validation success of predicting the correct biological groupings in RIVPACS-type models ranged from 46 % to 89 %), but the models are far from perfect. Much of the unexplained variation was likely associated with factors that were not measured. In general, the estimated human impairment of the biota (measured by O/E) was related to the measured variables indicative of human disturbance (I, III).

4.2 Performance of typology and RIVPACS at varying scale

The relative performance of the typology- and the RIVPACS-approaches was scale-dependent (III). At the larger scale (across Western and Central Finland), the RIVPACS performed in all aspects better than the typology, but at the smaller, regional scale the two approaches showed strikingly similar predictive accuracy and precision, as well as sensitivity to detect anthropogenic impairment of macroinvertebrate fauna (I, III). For example, at the larger scale the RIVPACS-approach categorised many more (at best 75 % of the total 134 IMP sites, $O/E_{0.4}$; III) impacted sites as biologically impaired than the typology-approach (at best 60 %, $O/E_{0.3}$), whereas regional assessments revealed similar sensitivities (74 % and 72 % sites categorised as impaired with RIVPACS [$O/E_{0.3}$] and typology [$O/E_{0.6}$], respectively).

The good performance of the typology-approach was quite surprising, particularly in the light of recent direct comparisons which have indicated a distinctly better performance of RIVPACS-type models over the typology-approach (e.g. Davy-Bowker et al. 2006, Mazor et al. 2006). An easy explanation for the contrasting results could be poor performance of the present models, but actually that was not the case: the predictive accuracy and precision of the models (I, III) were similar to those reported for similar models elsewhere (e.g. Moss et al. 1987, Hawkins et al. 2000b, Davy-Bowker et al. 2006). Most likely, the following three factors explain the differences. First, the geographical extent in III was smaller than in the earlier comparisons of Davy-Bowker et al. (2006) and Mazor et al. (2006). Western and Central Finland probably incorporated less environmental and biological variation, allowing the typology a better control of them. This conjecture is supported by the results which showed increasing performance of typology with decreasing spatial extent, a result probably at least partly associated with geological differences between the regions (III). Second, the study sites spanned a large size gradient from headwater streams to lowland rivers, an important gradient shaping running water communities (e.g. Vannote et al. 1980, Malmqvist & Mäki 1994), whose incorporation was crucial for the good typology performance. The catchment size categories of WFD System A were effective in this (I, III, Lorenz et al. 2004). Third, other large-scale factors important in governing the composition of river fauna, like altitude (Rundle et al. 1993), varied little in the study area.

Geographical extent influenced the pool of expected taxa in the typology-approach, provided a large p_t was used (III). With $p_t = 0.5$, for example, an average of 2.3 more taxa were expected in the smaller than in larger scale. Two taxa might not seem a large difference, but the results indicate that the accuracy of typologies might decrease with increasing geographical extent, when fewer regional taxa reach the selected p_t in each river type. This might have substantial effects on impact detection, as suggested by the poorer sensitivity of the larger-scale typology (III). In contrast, the expected pool of taxa in the

RIVPACS-approach was not susceptible to changes in extent, suggesting that these models are more accurate over large spatial scales.

The outcome that the RIVPACS-approach performed well irrespective of the regionalisation decision (III) clearly implies a wider applicability of RIVPACS -type predictive models for freshwater bioassessment. These models use multiple predictor variables to model the continuous natural assemblage variation (Reynoldson et al. 1997, Joy & Death 2002), forming a theoretically sound basis to favour them over categorical typologies. Recent advances in predictive modelling that do not require any form of site classification appear particularly promising (Linke et al. 2005, Chessman et al. 2008). In contrast, precision and sensitivity of typologies depends directly on how well the reference variation can be controlled by the *a priori* defined types (e.g. Hawkins et al. 2000a, Dodkins et al. 2005). The great challenge of typologies is to divide the environment into small enough 'parts' in biologically relevant dimensions, which is likely to succeed only at small scales and with relatively simple environmental matrix (e.g. in Western Finland, I). Then, however, well-established typologies can perform approximately as well as the RIVPACS-type models, as shown in I and III. In environmentally complex settings, however, well-performing typologies are likely to become an unfeasible endeavour as a sufficient number of reference sites needs to be found for each type. Then, typologies could be useful as a simple means of evaluating the gain of the more complex predictive modelling approaches that are likely to prove indispensable for assessment of freshwater biota.

4.3 Influence of geographical scale on RIVPACS-type models

Unlike typology, the performance of RIVPACS-models was only moderately influenced by the spatial scale (II, III). Generally (a few inconsistencies notwithstanding), the regionalisation by ecoregions (II) or river basins (III) slightly improved precision (II, III) and sensitivity (III). For example, RMSE of O/E_{0+} for validation sites indicated greater accuracy and precision with regional (north+middle boreal ecoregion models; RMSE = 0.27) than with across-ecoregions models (0.32, II), and in most cases regional models distinguished more sites impaired than the larger scale model (III).

These findings agree with other recent studies studying RIVPACS at varying scales (Ode et al. 2008, Yuan et al. 2008, Feio et al. 2009). Subdivisions of data sets influence the bioassessment performance of RIVPACS likely through varying importance and distribution of predictors at different scales (II, Ode et al. 2008, Yuan et al. 2008). For example, geographical location was less important in the NB model than in the MB model, probably because the total area of NB was smaller (60 %) than that of MB, reducing the likelihood of importance of biogeographical factors for macroinvertebrate assemblage variability within NB (II). Ode et al. (2008) found in California that regional models accounted for two local natural gradients (% slope and % fast-water habitat) that more spatially extensive models did not adjust for. They further supposed that at the large scale map-derived variables (catchment area) were not consistently associated with the local condition. Inability to reflect gradients could also result if the variables do not fulfil the assumption of linear relationship of the discriminant function models (Ode et al. 2008, Yuan et al. 2008). According to general ecological theory, species are distributed unimodally along environmental gradients, and thus linearly only at either end of the species range along any gradient. Nonlinearity should thus increase with increasing scale, and it might be that models that do not assume linearity (see Ode et al. 2008) might outperform RIVPACS-type models at large scales.

Last, we expected that local variables would be more important than catchment-scale variables at the regional scale and *vice versa*. However, we found no evidence for this hypothesis. In contrast, catchment-scale variables were important predictors generally at both scales (II). The results indicate that equal precision of RIVPACS-type models could be obtained irrespective of scale by catchment-scale variables. The use of catchment variables is also supported by their temporal invariability and cost-efficient measurement.

4.4 Concordance of assessments of two organism groups

Assessments of macroinvertebrate and macrophyte communities were compared in Central Finland (IV). The RIVPACS-type model based on macroinvertebrates was more precise and also more sensitive to anthropogenic impairment than the corresponding macrophyte-model. The macrophyte-model performed poorly, showing virtually no superiority to the corresponding null model (IV).

Why did the macrophyte-model fail? One likely explanation is that the measured environmental variables were mainly those most important for structuring macroinvertebrate communities. The macroinvertebrate clusters were best discriminated by larger scale variables (catchment area, longitude and peatland cover), whereas only alkalinity discriminated the macrophyte clusters, and with lower success (IV). Current velocity and % of lakes were variables that were closest to being included in the macrophyte model, but were rejected as their inclusion did not increase model power (not shown). It is thus likely that the poor precision of the macrophyte model indicates that the factors most influential to natural variation of macrophyte communities were not among the candidate set of predictors (see also Mazor et al. 2006). Macrophyte communities in streams have been typically found to be controlled by habitat stability and associated in-stream variables (Muotka & Virtanen 1995, Suren & Ormerod 1998, Paavola et al. 2003, Heino et al. 2005), which were not well represented in the present study. Poor performance of the macrophyte model might have been partly also due to high distinctiveness of the site clusters, which may have enhanced impact of site misclassifications in DFA (IV). Last, the small number of expected and observed macrophyte taxa (e.g. with $p_t = 0.4$

only one taxon observed and 2.2 expected for some reference sites) probably further weakened the impact detection of the group.

Use of multiple organism groups in bioassessment is particularly appropriate if different groups respond differently to different disturbance types (Barbour et al. 1999). Recent evidence has supported this, showing non-concordant response of multiple lotic groups to natural or stressor gradients (Paavola et al. 2003, Johnson et al. 2006, Feio et al. 2007, Carlisle et al. 2008). Also, macroinvertebrates and macrophytes in Central Finland responded somewhat differently to different stressors (IV), in agreement with the above studies. On the other hand, the actual assessments of the two assemblages showed relatively strong concordance. Most sites categorised as impaired by macrophytes were also categorised impaired by macroinvertebrates, indicating that the former provided little information independent from the latter, and being in line with the documented correlation between species richness of boreal mosses and macroinvertebrates (Heino et al. 2003, Paavola et al. 2003). However, the present evidence is inconclusive, as the impact detection of macrophytes was hindered by the imprecise model. Only improvement of the predictive power for macrophyte species can provide a more conclusive view of the potential redundancy of macroinvertebrate and macrophyte assessments.

4.5 All or only common taxa

Exclusion of taxa with the lowest estimated capture probabilities generally improved the precision and sensitivity of the O/E index (I–V). Performance was optimal at intermediate p_t values (c. 0.2–0.6), when the rarest taxa were excluded (I, III). At the highest p_t values, when only the most common taxa were included, predictive accuracy and precision of O/E were highest, but impact detection was lower than at intermediate levels of pt.

The observed patterns agree with previous similar studies (e.g. Johnson & Sandin 2001, Ostermiller & Hawkins 2004, Clarke & Murphy 2006, Van Sickle et al. 2007). In general, it seems that there is a trade-off between the gain in precision and loss in sensitivity when p_t is increased and fewer and only higher-probability taxa are included in the expected taxa pool (Cao et al. 2001, Clarke & Murphy 2006, Van Sickle et al. 2007). However, any generalisations of optimal p_t may be unwarranted, as the outcomes seem to be case-specific, depending on, for example, sampling methodology (Ostermiller & Hawkins 2004), species distributions, study extent (II, III), or the prediction approach (I, III).

The low precision of O/E at small p_t values is likely caused by the inability of the models to predict the locally or temporally rarer taxa that are often less abundant (Hämäläinen et al. 2003, Resh et al. 2005), and thus occur sporadically in samples. Nevertheless, O/E can perform well with low p_t, as shown for original RIVPACS (Clarke & Murphy 2006), where intensive sampling and low taxonomic resolution (family level) probably weights the accurate and precise prediction of the low-p taxa. At higher p_t values the included naturally common and abundant taxa, that are likely to be ubiquitous and intolerant to environmental variation, could also be expected to be insensitive to anthropogenic disturbances. Contrasting with this conjecture, it was rather surprising to find that many of the common taxa, and those with intermediate probabilities in particular, were consistently absent from IMP-sites (I). These results are in line with those of Van Sickle et al. (2007), who noticed a reduced frequency of occurrence of common taxa among test sites for at least half of the 10 separate predictive models that they explored in the USA. It thus seems that, probably particularly with high taxonomic resolution, exclusion of the rare taxa seems to be a convenient option to increase predictive accuracy of the O/E index, without losing sensitivity to detect impact. Recently proposed measures of compositional dissimilarities of assemblages seem to circumvent the problem (Van Sickle 2008). Last, the fact that sensitivity and precision were not fully concordant strongly suggests when alternative approaches are evaluated for bioassessment; one should always pay attention also to sensitivity, not only to the accuracy and precision of the given models.

4.6 Bioassessment and threatened species

Eleven insect species categorised as threatened (TS) were found from Western and Central Finland (V). Both the number and abundance of TS showed a significant positive correlation with the O/E ratios. TS were generally concentrated to those sites that were classified to high condition by the O/E. Generally, the results imply that O/E based assessments could be useful for identifying sites which are in good enough condition to have the potential to harbour populations of threatened species and thus sites of special conservation value (Wright et al. 1993, Linke & Norris 2003). However, regarding the protection of TS, the results from the biological quality classification exercises were perhaps somewhat alarming.

If the 10th percentile of the reference O/E distribution was used as the upper boundary for the good condition class, the TS were captured only sporadically from sites representing the good class. The outcome of the classification exercise thus indicates that application of similar classification schemes for management might have poor ability to safeguard threatened species, and stricter classification criteria might be needed. The results also suggest that for safeguarding threatened species the maintenance of the suggested ‘high’ condition sites is the most critical. In the exercise the 25th percentile of the reference distribution seemed a more acceptable threshold value and using it alleviated the identified problems.

The main implication of the classification exercise is that conventional class boundaries may not be efficient enough for preservation of threatened river macroinvertebrates. However, the approach could bring an objective, meaningful and societally acceptable means for setting site quality criteria in freshwater bioassessment. The exercise demonstrates an approach to set for biological metrics objective and meaningful critical boundary values, which are relevant to the actual targets of freshwater protection and management. In addition to the occurrence of TS, in a similar manner other ecosystem properties of ecological and/or societal values might be used as a decisive factor in setting target values for the actual bioassessment metrics and for biological quality in general.

4.7 A note on reference condition approach

This work was based on the reference condition approach, whereby expected biotic assemblages are modelled with reference data considered minimally disturbed by human activity. The concept has gained widespread recognition in freshwater bioassessment, mainly because often little or no biological information from a given site prior to human alterations is available. Then the 'space-for-time approach' (i.e. the use of a sample of similar, but non-altered sites for reference) circumvents the lack of historical reference data. The quality of these reference sites is thus fundamental, and affects the ability to detect impact. In today's world it is virtually impossible to find sites or areas unaffected by human activity, and therefore most works, including this one, have used near-natural, or best-available sites for reference. The degree of human activity that is allowed to be included at reference sites sets a baseline for the degree of human activity whose potential effect on biota can be detected. For example, only sites with catchments having <15 % of cultivated area were accepted as reference sites in this study. The present models thus cannot distinguish any potential effects of agricultural practices smaller than that extent. The complex issues of definition of reference condition are currently being considered in many parts of the world. In lotic environments, a particularly challenging issue will be assessing biological condition in large lowland rivers, for which reference conditions are lacking in most parts of the world.

5 CONCLUSIONS

Community-level attributes have become standard measures in environmental assessment mainly because multiple species are more likely to exhibit responses to multiple environmental stresses than are single indicator species. However, whole communities also vary more than single species along environmental gradients, which challenges efforts to distinguish human-caused impacts from the natural variation. Despite considerable methodological development in measuring community-level impacts in aquatic ecosystems in recent decades, comparability of assessments over large scales have often been hampered by the great diversity of indicators that have been developed. This situation may often have left managers unsure of the relative merits of the various alternatives. Therefore in this study, for the sake of simplicity, only one community measure (observed-to-expected ratio of taxa, O/E) was applied to compare the alternative approaches.

This study adds to the growing evidence that O/E of taxonomic completeness is a useful proxy for measuring biological integrity of freshwaters. The index is appealing because it is transparent and simple, and it seems to give meaningful results. For example, the index showed a positive correlation with the occurrence and abundance of threatened species (V), opening an interesting avenue to link important conservational values to bioassessment. Even though the O/E index is a unifying measure of the condition of a community that is not calibrated to respond to any particular stressor, the index gave meaningful results: O/E showed a consistent decrease with increasing intensity of human activity (I, III). The index might thus be particularly useful to assess the quality of freshwater ecosystems in situations where multiple and diffuse anthropogenic stresses persist, currently a common situation in the developed world.

The results showed that regionalisation improves performance of assessments of river macroinvertebrates, particularly when *a priori* typologies instead of RIVPACS-type models are used to predict expected biota (I–III). This documentation of the phenomenon by an actual bioassessment exercise supports previous more theoretical studies (e.g. Hawkins et al. 2000a, Heino et al. 2003), and the results imply that regionalisation should be used in large-scale typology-based assessments of river macroinvertebrate communities; otherwise the bioassessments are likely to suffer from imprecise predictions and poor sensitivity to detect human-caused biotic impairment. Such unreliable bioassessment is expected ultimately to increase the likelihood of management actions that are either unnecessary or insufficient to improve biological integrity of freshwater ecosystems. The RIVPACS-approach was much more invariant of regionalisation, probably because these models take into account the continuous biotic variation along unlimited number of environmental gradients. RIVPACS or similar modelling approaches are thus likely to prove superior to, and have wider applicability than, categorical typologies for bioassessment of freshwaters. This general inference is supported by the fact that in the typology approach a considerable number of reference sites is needed for each type — a requirement that is likely to be unrealistic in environmentally complex regions. The continuous models are not constrained by such limitations.

The results indicated that a simple typology-approach can perform approximately as well as a more complex RIVPACS-type modelling to predict expected biota, given the typology is well-established. This might be good news for many European countries that are currently developing typology-based bioassessment systems. Typology-based assessment might prove a robust alternative in areas with relatively simple gradients, like in this study, or maybe in boreal rivers in general. Typologies could be particularly appealing for environmental agencies conducting regional assessments, because typology-based O/E ratios or other metrics are easy to calculate in a user-friendly spreadsheet form. In contrast, multivariate RIVPACS-type models require more expertise to be established, and an interface to be developed for end-users. Also other practical reasons, like lack of predictor data or ease of communication may appeal the choice of typologies to predict expected biota.

Finally, effective environmental management requires accurate diagnosis of the anthropogenic stressors that have caused degradation of biological integrity. The observed correlative evidence between O/E and human activity does not yet show causality and direct measurements of the stressors and stressor-specific diagnostic tools are evidently needed. It might be, however, that measures of assemblage condition are particularly difficult to use for accurate diagnosis (Pollard & Yuan 2006). Analysis and prediction of characteristics of biota that are stressor-specific (Gayraud et al. 2003) might be especially useful in this, as well as combination of ecological classification with ecotoxicological modelling (De Zwart et al. 2006). Such approaches could help to disentangle the mechanistic basis of interactions between specific human activities, stressors and biotic responses, and enable ecologically sound management of freshwater ecosystems.

Acknowledgements

I am the most grateful to my supervisor Heikki Hämäläinen for his full support, many visions shared and all the opportunities given. It has been a privilege to follow Heikki's way to science. He always spurred me to think for 'ultimate purpose' of our doings. Often in practice, that also meant dismissal of an alluring short-cut and, reanalysing data, once again. However, Heikki has provided inspiring years filled with uncompromising training that only a highly professional supervisor can provide. I thank my second supervisor Roger Jones for all his support, many general advices and help with English. Roger also freely let me bioassess without any isotopes involved! When I arrived to Jyväskylä, Heikki Mykrä introduced me to many ways of doing stream research that I had much been unaware of. Hessu's peculiar definitions (e.g. '*seula-spede*'; a lake benthos guy) made work and other fishing trips great fun. I acknowledge his expertise and slick teamworking; particularly in II & IV, which much assured my defense for this year. I thank co-authors Esa Koskenniemi, Timo Muotka and Kari-Matti Vuori for discussions that have widened my perspective. Special thanks to all three; to Esa for important support in the early days, to Timo for several vigorous science-wise advices, and to Kari-Matti especially for understanding my occasional absent-mindedness in the final months. I thank John Van Sickle for improving 2 papers by precise refereeing and by kindly sharing his R-scripts for RIVPACS. This study would not have been possible without the invaluable effort of the researchers that collected the original data (A. Bonde, J. Heino, I. Kananen, J. Kotanen, H.-G. Lax, P. Majuri, C. Nyman, J. Latvala, R. Paavola, T. Ruokonen, J. Salmela, A. Teppo, among others). I have been privileged to use these data. I thank many colleagues (J. Heino, S. Hellsten, A. Keto, M. Marttunen, J. Muotka, K. Meissner, J. Miettinen, T. Sutela, S. Rekolainen, A. Tarvainen, T. Vehanen, among others) and people at deparment (Heather, Jari, Jukka, Kimmo, Mikko, Sami, Saija, Timo, Uche, among others) for inspiring discussions. Jussi provided good lunch company and important peer support for several years. Last, my friends Atso, Katja, Marianne, Mikko, Mirko, Riikka, Riku, Saija, Sari, Sime and two Villes, among others, have always provided excellent leisure content and true disinterest of invertebrates — best cure for any stress caused by this book.

My sister and brother, Laura and Pekka, and their families, supported me by their dear presence. At the early steps, Pekka showed the way and gave valuable brotherly advice. Finally, I am forever grateful to my parents for their presence and love. Importantly, they provided support with discipline at times when I had not the intellect yet. E.g., making it clear to a 9-year-old kid that 4 was an unacceptable score from an English exam was likely among the most important prerequisites to see this book coming. The very last, I thank Henna for widening my understanding of the societal processes and structures. However, it is her endless support and love, for which I owe her the very most.

This thesis was funded by Ministry of Agriculture and Forestry, Maa- ja vesitekniikan tuki ry and Finland's environmental administration.

YHTEENVETO (RÉSUMÉ IN FINNISH)

Ennustavat mallit jokien pohjaeläimistön tilan arvioinnissa

Ihmisen toiminta on aiheuttanut mittavia muutoksia makean veden ekosysteemeille. Yhteiskunnallinen huoli näistä muutoksista on sisällytetty teollisuusmaissa ympäristösäädöksiin (esim. Clean Water Act Yhdysvalloissa ja Vesipuitedirektiivi Euroopan Unionissa), joiden keskeinen tavoite on suojella ja parantaa järvien ja jokien tilaa kokonaisvaltaisesti ja eliöstöä korostaen. Tässä kunnianhimoisessa tavoitteessaan uudet säädökset eroavat merkittävästi edeltäjistään, jotka painottivat vesien tilan arviointia pääosin ihmisen näkökulmasta.

Tavoitteiden uudelleenmäärittely ja niiden toteuttaminen edellyttävät uusia menetelmiä, etenkin vesistöjen ekologisen tilan mittaamiseen eliöstön perusteella. Tila-arvioiden tulisi olla mahdollisimman tarkkoja, sillä niillä on juridisesti määräytyviä yhteiskunnallisia ja taloudellisia seuraamuksia. Suuri tarve on etenkin laajojen maantieteellisten alueiden vesistöjen yhdenmukaisille arvioinneille, mikä korostaa vertailukelpoisten menetelmien tärkeyttä. Valtaosa käytössä olevista menetelmistä soveltuu vain tietyn ihmistoiminnan vaikutusten tunnistamiseen, jolloin muuntyyppisen ihmistoiminnan vaikutukset voivat jäädä havaitsematta etenkin, kun vesistöjen tilaa huonontavat usein useat eri tekijät samanaikaisesti. Vesien tilaa tulisikin arvioida yhtäläisin perustein vaikuttavan ihmistoiminnan laadusta riippumatta.

Nykyään tila-arvioinneissa käytetään yleisesti ns. vertailuololähestymistapaa, jossa eliöstön tilaa mitataan biologisten muuttuja-arvojen poikkeamana luonnontilaisista tai lähes luonnontilaisista arvoista, eli vertailuarvoista, jotka määrittelevät vertailuolot. Luotettavan tila-arvion saamiseksi on oleellista pystyä erottamaan ihmistoiminnan aiheuttamat biologiset muutokset eliöstön luonnollisesta taustavaihtelusta. Yksinkertaisimmillaan vertailuololähestymistapaa sovelletaan EU:n vesipuitedirektiivissä, jossa keinona taustavaihtelun hallitsemiseksi on vesistöjen ryhmittely luonnollisilta piirteiltään mahdollisimman samankaltaisiin vesimuodostumatyyppeihin. Kullekin tyypille arvioidaan omat vertailuolonsa. Biologinen vaihtelu on kuitenkin luonteeltaan jatkuvaa, ja kategorista tyypittelyä luotettavampi lähestymistapa saattaisi olla luonnon jatkuvuuden paremmin huomioiva ennustava mallinnus, jossa vertailuolot tuotetaan kullekin vesimuodostumalle erikseen.

Tässä väitöskirjatyössä kehitin ja testasin menetelmiä erityisesti pohjoisten virtavesien pohjaeläinyhteisöjen tilan mittaamiseen. Yhtenä päätavoitteena oli vertailla yksinkertaisen jokityypittelyn ja monimuuttujaisen ns. RIVPACS-tyyppisen mallinnuksen toimivuutta pohjaeläimistön tilan arvioinnissa (I, III). Lähestymistapojen toimivuutta arvioin lajiston ennustamistarkkuuden (eli vertailuolojen määrittelytarkkuuden) sekä ihmistoiminnan aiheuttamien yhteisömuutosten tunnistamisherkkyyden perusteella. Selvitin erityisesti maantieteellisen alueen laajuuden vaikutusta lähestymistapojen toimivuuteen (II, III), ja tarkastelin, tulisiko arviointien perustua kaikkiin ennustettuihin lajeihin vai vain yleisempien lajien osajoukkoon (I–V). Tarkastelin myös mallinnukseen perustuvien pohjaeläin- ja vesikasviyhteisöjen tila-arvioiden yhteneväisyyttä (IV). Jos eri eliöryhmien tila-arviot vastaisivat toisiaan, jokien biologisen tilan arviot voitaisiin perustaa vain yksittäisiin ryhmiin.

Vesienhoidon tarpeita varten tila-arviot esitetään yleensä luokituksina, joissa tietyn luokan, esim. vesipuitedirektiivin mukaisen "hyvän ekologisen tilan", katsotaan vastaavan minimitavoitetta eli heikointa hyväksyttävissä olevaa tilaa, joka ei edellytä kunnostustoimia. Luokitukset voivat kuitenkin olla mielivaltaisia, eivätkä ne välttämättä vastaa ympäristölle asetettuja konkreettisia laatutavoitteita, kuten monimuotoisuuden ja uhanalaisten lajien säilymistä. Osatyössä V tutkin pohjaeläimistön tila-arvioiden suhdetta uhanalaisiksi luokiteltujen pohjaeläinlajien esiintymiseen ja runsauteen. Tarkastelin erityisesti, kuinka hyvin tavanomainen pohjaeläimistöön perustuva paikkojen laatuluokitus ja sen "hyvä tila" turvaisi uhanalaislajien esiintymisen.

Työ perustuu Länsi-, Keski- ja Pohjois-Suomesta kerättyyn aineistoon 345 koskipaikan pohjaeläinnäytteistä, valuma-aluetiedoista, vedenlaatutiedoista ja arvioista ihmistoiminnan aiheuttamista muutoksista. Havaintopaikat jaettiin lähinnä luonnontilaa oleviin vertailupaikkoihin — joiden perusteella muodostettiin vertailuolot — ja ihmistoiminnan eriasteisesti muuttamiin paikkoihin. Pohjaeläimistön tilaa mitattiin "taksonomisena eheytenä" eli havaittuna osuutena niistä taksoneista (lajeista tai suvuista), joiden kullakin paikalla ennustettiin tyypin tai mallin perusteella esiintyvän ihmistoiminnan aiheuttaman häiriön puuttuessa (ns. O/E-indeksi).

Alueellisella tasolla (Länsi- ja Keski-Suomi) tyyppi- ja malliperusteiset tila-arviot eivät eronneet merkittävästi toisistaan. Laajemmalla maantieteellisellä alueella (Länsi- ja Keski-Suomen yhdistetty aineisto) mallinnukseen perustuvan O/E-indeksin vertailuolovaihtelu oli kuitenkin pienempi ja myös ihmistoiminnan heikentämät virtavedet voitiin tunnistaa paremmin. Paikallisesti harvinaiseksi ennustettujen taksonien poisto tarkastelusta paransi molempien lähestymistapojen toimivuutta. Erityisesti useat yleisyydeltään keskimääräiset lajit olivat herkkiä ympäristöpaineille ja siten merkityksellisiä vaikutusten tunnistamisessa. Pohjaeläin- ja vesikasviyhteisöjen tila-arviot olivat suhteellisen voimakkaasti yhteydessä toisiinsa. Pohjaeläimiin perustuva malli oli kuitenkin selkeästi tarkempi, ja se tunnisti muutokset herkemmin kuin vesikasvimalli, jonka ennustuskyky oli heikko. Havaintoaineistosta mahdollisesti puuttui vesikasveille tärkeitä ympäristömuuttujia, eikä tulosta voida siten välttämättä yleistää. Länsi- ja Keski-Suomen aineistossa esiintyi 11 vaarantuneeksi tai silmälläpidettäväksi luokiteltua hyönteislajia. Näiden uhanalaisiksi luokiteltujen lajien esiintyminen ja runsaus korreloivat positiivisesti O/E-indeksin kanssa. Uhanalaislajit keskittyivät koskiin, jotka pohjaeläinyhteisöjen perusteella luokittuivat tilaluokkaan "erinomainen", kun taas vain muutamia esiintymisiä oli luokkaan "hyvä" luokittuneilla paikoilla.

Tulosten perusteella luonnon jatkuvan vaihtelun ja moniulotteisuuden huomioivat mallit ovat kategorista tyypittelyä soveliaampia vertailuarvojen ennustamiseen. Yksinkertainen tyypittely näyttäisi kuitenkin tarjoavan käyttäjäystävällisen vaihtoehdon, mikäli sitä sovelletaan alueellisella mittakaavalla ja ympäristögradientit ovat suhteellisen yksinkertaisia. Alueellisten tyyppikohtaisten vertailuolojen määrittelyä voivat tosin haitata käytännön seikat, kuten riittämätön vertailupaikkojen määrä. Monimuuttujaisen mallinnuksen mahdollisuuksia koko Suomen kattavien vertailuolojen muodostamisessa tulisikin selvittää kattavammin ja myös muilla eliöryhmillä. O/E-indeksi osoittautui hyvin käyttökelpoiseksi tilamuuttujaksi. Vaikka indeksiä ei ole kalibroitu ilmentämään mitään tiettyä ympäristöstressiä, muutokset pystyttiin havaitsemaan hyvin ja vaikuttavan ihmistoiminnan laadusta riippumatta. Pohjaeläinyhteisöihin perustuvat koskien laatuluokituskokeilut osoittivat, etteivät tavanomaiset, mielivaltaiset luokitustavat välttämättä tue uhanalaislajien suojelua. Tavoitetilaa ("hyvä" ekologinen tila) vastaavat mielekkäät luokittelumuuttujien arvot olisi kuitenkin mahdollista asettaa ainakin osin uhanalaislajien esiintymisen perusteella. Myös muita, yksilöityjä ja konkreettisia ympäristön laatutavoitteita voitaisiin käyttää samalla tavalla luokkarajojen objektiiviseen määrittelyyn.

REFERENCES

Alalammi, P. & Karlson, K. P. 1988. Atlas of Finland, Folio 141-143. Biogeography and nature conservation. 32 p., National Board of Survey and Geographical Society of Finland, Helsinki.

Anon. 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities L 206: 7-50.

Anon. 1999. SPSS base 10.0 applications guide. 426 p., SPPS Inc., Chicago.

Anon. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities L 327: 1-73.

Anon. 2003a. Finnish Biodiversity Research Programme 1997-2002. Evaluation report. Publications of the Academy of Finland 3/03: 1-34.

Anon. 2003b. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) Guidance document No. 10 River and lakes - Typology, reference conditions and classification systems. 87 p., Office for Official Publications of the European Communities, Luxembourg.

Anon. 2006. Wadeable Streams Assessment: a collaborative survey of the nation's streams. EPA 841-B-06-002. 113 p., Office of Research and Development and Office of Water, US Environmental Protection Agency, Washington, DC.

Anon. 2008a. MOSSE Monimuotoisuuden tutkimusohjelma (2003-2006) Loppuraportti ja itsearviointi. 57 p., Maa- ja metsätalousministeriö ja ympäristöministeriö, Vammalan kirjapaino. (In Finnish).

Anon. 2008b. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org. 25.5.2009.

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

Bailey, R. C., Kennedy, M. G., Dervish, M. Z. & Taylor, R. M. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwat. Biol. 39: 765-774.

Bailey, R. C., Norris, R. & Reynoldson, T. B. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. 184 p., Kluwer Academic Publishers, Boston, MA.

Barbour, M. T., Gerritsen, J., Snyder, B. D. & Stribling, J. B. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, Second Edition. 339 p., United States Environmental Protection Agency, Washington, D.C.

Barbour, M. T., Stribling, J. B. & Karr, J. R. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. In: Davis, W. S. & Simon, T. P. (eds), Biological assessment and criteria: Tools for water resource planning and decision making: 63-77. Lewis Publishers, Boca Raton, Florida.

Bowman, M. F. & Somers, K. M. 2005. Considerations when using the reference condition approach for bioassessment of freshwater ecosystems. Water Qual. Res. J. Canada 40: 347-360.

Cao, Y., Larsen, D. P. & Thorne, R. S. 2001. Rare species in multivariate analysis for bioassessment: some considerations. J. N. Am. Benthol. Soc. 20: 144-153.

Cao, Y., Hawkins, C. P., Olson, J. & Kosterman, M. A. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. J. N. Am. Benthol. Soc. 26: 566-585.

Carlisle, D. M., Hawkins, C. P., Meador, M. R., Potapova, M. & Falcone, J. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. J. N. Am. Benthol. Soc. 27: 16-37.

Chessman, B., Muschal, M. & Royal, M. J. 2008. Comparing apples with apples: use of limiting environmental differences to match reference and stressor-exposure sites for bioassessment of streams. River. Res. Applic. 24: 103-117.

Clarke, R. T., Furse, M. T., Wright, J. F. & Moss, D. 1996. Derivation of a biological quality index for river sites: Comparison of the observed with the expected fauna. J. Appl. Stat. 23: 311-332.

Clarke, R. T. & Murphy, J. F. 2006. Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. Freshwat. Biol. 51: 1924-1940.

Clews, E. & Ormerod, S. J. 2009. Improving bio-diagnostic monitoring using simple combinations of standard biotic indices. River. Res. Applic. 25: 348-361.

Davies, S. P. & Jackson, S. K. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16: 1251-1266.

Davy-Bowker, J., Murphy, J. F., Rutt, G. R., Steel, J. E. C. & Furse, M. T. 2005. The development and testing of a macroinvertebrate biotic index for detecting the impact of acidity on streams. Arch. Hydrobiol. 163: 383-403.

Davy-Bowker, J., Clarke, R. T., Johnson, R. K., Kokes, J., Murphy, J. F. & Zahradkova, S. 2006. A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates. Hydrobiologia 566: 91-105.

De Zwart, D., Dyer, S. D., Posthuma, L. & Hawkins, C. P. 2006. Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. Ecol. Appl. 16: 1295-1310.

Diamond, J. M., Barbour, M. T. & Stribling, J. B. 1996. Characterizing and comparing bioassessment methods and their results: A perspective. J. N. Am. Benthol. Soc. 15: 713-727.

Dodkins, I., Rippey, B., Harrington, T. J., Bradley, C., Ni Chathain, B., Kelly-Quinn, M., McGarrigle, M., Hodge, S. & Trigg, D. 2005. Developing an optimal river typology for biological elements within the Water Framework Directive. Water Res. 39: 3479-3486.

Feio, M. J., Almeida, S. F. P., Craveiro, S. C. & Calado, A. J. 2007. Diatoms and macroinvertebrates provide consistent and complementary information on environmental quality. Fundam. Appl. Limnol. 169: 247-258.

Feio, M. J., Norris, R. H., Graçam, M. A. S. & Nichols, S. 2009. Water quality assessment of Portuguese streams: Regional or national predictive models? Ecol. Indicat. 9: 791-806.

Frey, D. G. 1977. Biological integrity of water - a historical approach. In: Ballentine, R. K. & Guarria, L. J. (eds), The integrity of water. Proceedings of a symposium: 127-140. U.S. Environmental Protection Agency, Washington, D.C.

Frissell, C. A., Liss, W. J., Warren, C. E. & Hurley, M. D. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. Environ. Manage. 10: 199-214.

Gayraud, S., Statzner, B., Bady, P., Haybachp, A., Scholl, F., Usseglio-Polatera, P. & Bacchi, M. 2003. Invertebrate traits for the biomonitoring of large European rivers: an initial assessment of alternative metrics. Freshwat. Biol. 48: 2045-2064.

Green, R. H. 1980. Multivariate approaches in ecology: the assessment of ecologic similarity. Ann. Rev. Ecol. Syst. 11: 1-14.

Hastie, T., Tibshirani, R. & Friedman, J. 2001. The elements of statistical learning: data mining, inference, and prediction. 533 p., Springer, New York.

Hawkins, C. P. 2006a. Maintaining and restoring the ecological integrity of freshwater ecosystems: Refining biological assessments. Ecol. Appl. 16: 1249-1250.

Hawkins, C. P. 2006b. Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecol. Appl. 16: 1277-1294.

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K. & Stevenson, R. J. 2000a. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. J. N. Am. Benthol. Soc. 19: 541-556.

Hawkins, C. P., Norris, R. H., Hogue, J. N. & Feminella, J. W. 2000b. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol. Appl. 10: 1456-1477.

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. Am. Benthol. Soc. 21: 397-413.

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. & Paasivirta, L. 2003b. Among-taxon congruence in biodiversity patterns: can stream insect diversity be predicted using single taxonomic groups? Can. J. Fish. Aquat. Sci. 60: 1039-1049.

Heino, J., Paavola, R., Virtanen, R. & Muotka, T. 2005. Searching for biodiversity indicators in running waters: do bryophytes, macroinvertebrates, and fish show congruent diversity patterns? Biodivers. Conserv. 14: 415-428.

Heino, J., Mykrä, H., Hämäläinen, H., Aroviita, J. & Muotka, T. 2007. Responses of taxonomic distinctness and species diversity indices to anthropogenic impacts and natural environmental gradients in stream macroinvertebrates. Freshwat. Biol. 52: 1846-1861.

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

Hering, D., Johnson, R. K., Kramm, S., Schmutz, S., Szoszkiewicz, K. & Verdonschot, P. F. M. 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. Freshwat. Biol. 51: 1757-1785.

Hewitt, J. E., Anderson, M. J. & Thrush, S. F. 2005. Assessing and monitoring ecological community health in marine systems. Ecol. Appl. 15: 942-953.

Hughes, R. M., Omernik, J. M. & Larsen, D. P. 1986. Regional reference sites: a method for assessing stream potentials. Environ. Manage. 10: 629-635.

Hämäläinen, H., Koskenniemi, E., Kotanen, J., Heino, J., Paavola, R. & Muotka, T. 2002. Benthic invertebrates and the implementation of WFD: sketches from Finnish rivers. Tema Nord 566: 55-58.

Hämäläinen, H., Luotonen, H., Koskenniemi, E. & Liljaniemi, P. 2003. Inter-annual variation in macroinvertebrate communities in a shallow forest lake in eastern Finland during 1990-2001. Hydrobiologia 506: 389-397.

Jackson, D. A. & Harvey, H. H. 1993. Fish and benthic invertebrates: community concordance and community-environment relationships. Can. J. Fish. Aquat. Sci. 50: 2641-2651.

James, F. C. & McCulloch, C. E. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box. Ann. Rev. Ecol. Syst. 21: 129-166.

Johnson, R. K. 1998. Spatiotemporal variability of temperate lake macroinvertebrate communities: Detection of impact. Ecol. Appl. 8: 61-70.

Johnson, R. K. & Sandin, L. 2001. Development of a prediction and classification system for lake (littoral, $SWEPAC_{LLI}$) and stream (riffle $SWEPAC_{SRI}$) macroinvertebrate communities. Department of Environmental

Assessment, Swedish University of Agricultural Sciences, Rapport 2001 (23): 1-66.
Johnson, R. K., Hering, D., Furse, M. T. & Verdonschot, P. F. M. 2006. Indicators of ecological change: comparison of the early response of four organism groups to stress gradients. Hydrobiologia 566: 139-152.
Joy, M. K. & Death, R. G. 2002. Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. Freshwat. Biol. 47: 2261-2275.
Kaufman, L. & Rousseeuw, P. J. 1990. Finding groups in data: an introduction to cluster analysis. 368 p., Wiley-Interscience, New York.
Kelly, M. G. & Whitton, B. A. 1995. Trophic Diatom Index: a new index for monitoring eutrophication in rivers. J. Appl. Phycol. 7: 433-444.
Kilgour, B. W., Somers, K. M. & Matthews, D. E. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience 5: 542-550.
Kilgour, B. W. & Barton, D. R. 1999. Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. Freshwat. Biol. 41: 553-566.
Kilgour, B. W., Somers, K. M. & Barton, D. R. 2004. A comparison of the sensitivity of stream benthic community indices to effects associated with mines, pulp and paper mills, and urbanization. Environ. Toxicol. Chem. 23: 212-221.
Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. Psychometrika 29: 115-129.
Legendre, P. & Legendre, L. 1998. Numerical ecology. 853 p., Elsevier Science BV, Amsterdam.
Linke, S. & Norris, R. 2003. Biodiversity: bridging the gap between condition and conservation. Hydrobiologia 500: 203-211.
Linke, S., Norris, R. H., Faith, D. P. & Stockwell, D. 2005. ANNA: A new prediction method for bioassessment programs. Freshwat. Biol. 50: 147-158.
Lorenz, A., Feld, C. K. & Hering, D. 2004. Typology of streams in Germany based on benthic invertebrates: Ecoregions, zonation, geology and substrate. Limnologica 34: 379-389.
Majuri, P. 2008. Habitat measurements and benthic invertebrates on evaluating ecological status of running waters. MSc Thesis. 23 p., University of Jyväskylä. (In Finnish, English summary).
Malmqvist, B. & Mäki, M. 1994. Benthic macroinvertebrate assemblages in north Swedish streams: environmental relationships. Ecography 17: 9-16.
Mazor, R. D., Reynoldson, T. B., Rosenberg, D. M. & Resh, V. H. 2006. Effects of biotic assemblage, classification, and assessment method on bioassessment performance. Can. J. Fish. Aquat. Sci. 63: 394-411.
McCune, B. & Mefford, M. J. 1999. PC-Ord. Multivariate analysis of ecological data. MjM Software Design, Gleneden Beach, Oregon, USA.

McCune, B., Grace, J. B. & Urban, D. L. 2002. Analysis of ecological communities. 304 p., MJM Software Design, Gleneden Beach, Oregon, USA.

Moss, D., Furse, M. T., Wright, J. F. & Armitage, P. D. 1987. The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwat. Biol. 17: 41-52.

Muotka, T. & Virtanen, R. 1995. The stream as a habitat templet for bryophytes - species distributions along gradients in disturbance and substratum heterogeneity. Freshwat. Biol. 33: 141-160.

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. Verh. Internat. Ver. Limnol. 29: 1351-1355.

Mykrä, H., Heino, J. & Muotka, T. 2007. Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. Glob. Ecol. Biogeogr. 16: 149-159.

Neale, M. W. & Rippey, B. 2008. A comparison of environmental and biological site classifications for the prediction of macroinvertebrate communities of lakes in Northern Ireland. Aquatic Conserv: Mar. Freshw. Ecosyst. 18: 729-741.

Novak, M. A. & Bode, R. W. 1992. Percent model affinity - a new measure of macroinvertebrate community composition. J. N. Am. Benthol. Soc. 11: 80-85.

Nyman, C. 1995. Macrozoobenthos in some rapids in a lowland river in Finland before and after the construction of a hydroelectric power-plant. Regul. Rivers: Res. Manage. 10: 199-205.

Nyman, C., Anttila, M.-E., Lax, H.-G. & Sarvala, J. 1986. The bottom fauna of rapids as a measure of the quality of running waters. Vesi- ja ympäristöhallinnon julkaisuja 3: 1-76. (In Finnish, English summary).

Oberdorff, T., Pont, D., Hugueny, B. & Chessel, D. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. Freshwat. Biol. 46: 399-415.

Ode, P. R., Hawkins, C. P. & Mazor, R. D. 2008. Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. J. N. Am. Benthol. Soc. 27: 967-985.

Olsen, A. R. & Peck, D. V. 2008. Survey design and extent estimates for the Wadeable Streams Assessment. J. N. Am. Benthol. Soc. 27: 822-836.

Ostermiller, J. D. & Hawkins, C. P. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. J. N. Am. Benthol. Soc. 23: 363-382.

Paavola, R., Muotka, T., Virtanen, R., Heino, J. & Kreivi, P. 2003. Are biological classifications of headwater streams concordant across multiple taxonomic groups? Freshwat. Biol. 48: 1912-1923.

Paavola, R., Muotka, T., Virtanen, R., Heino, J., Jackson, D. & Mäki-Petäys, A. 2006. Spatial scale affects community concordance among fishes, benthic macroinvertebrates, and bryophytes in streams. Ecol. Appl. 16: 368-379.

Paulsen, S. G., Mayio, A., Peck, D. V., Stoddard, J. L., Tarquinio, E., Holdsworth, S. M., Van Sickle, J., Yuan, L. L., Hawkins, C. P., Herlihy, A. T., Kaufmann, P. R., Barbour, M. T., Larsen, D. P. & Olsen, A. R. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. J. N. Am. Benthol. Soc. 27: 812-821.

Pollard, A. I. & Yuan, L. 2006. Community response patterns: Evaluating benthic invertebrate composition in metal-polluted streams. Ecol. Appl. 16: 645-655.

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N. & Schmutz, S. 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. J. Appl. Ecol. 43: 70-80.

Rassi, P., Alanen, A., Kanerva, T. & Mannerkoski, I. 2001. The 2000 red list of Finnish species. The II committee for the monitoring of threatened species in Finland (In Finnish, English summary). 432 p., Ministry of the Environment and Finnish Environment Institute, Helsinki, Finland.

Resh, V. H., Rosenberg, D. M. & Reynoldson, T. B. 2000. Establishing the reference conditions in the Fraser River catchment, British Columbia, Canada, using the BEAST (BEnthic Assessment of SedimenT) predictive model. In: Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (eds), Assessing the biological quality of fresh water: RIVPACS and other techniques: 181-194. Freshwater Biological Association, Cumbria, UK.

Resh, V. H., Beche, L. A. & McElravy, E. P. 2005. How common are rare taxa in long-term benthic macroinvertebrate surveys? J. N. Am. Benthol. Soc. 24: 976-989.

Resh, V. H. 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environ. Monit. Assess. 138: 131-138.

Reynoldson, T. B., Norris, R. H., Resh, V. H., Day, K. E. & Rosenberg, D. M. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. N. Am. Benthol. Soc. 16: 833-852.

Rosenberg, D. M. & Resh, V. H. 1993. Freshwater biomonitoring and benthic macroinvertebrates. 488 p., Chapman and Hall, New York.

Rundle, S. D., Jenkins, A. & Ormerod, S. J. 1993. Macroinvertebrate communities in streams in the Himalaya, Nepal. Freshwat. Biol. 30: 169-180.

Ruokonen, T. 2004. Brown trout (Salmo trutta L.) as a biodiversity indicator - Can stream macroinvertebrate species richness be predicted by using single species? MSc Thesis. 23 p., University of Jyväskylä. (In Finnish, English summary).

Sandin, L. & Verdonschot, P. F. M. 2006. Stream and river typologies - major results and conclusions from the STAR project. Hydrobiologia 566: 33-37.

Simpson, J. C. & Norris, R. H. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. In: Wright, J. F., Sutcliffe,

D. W., & Furse, M. T. (eds), Assessing the biological quality of fresh waters: RIVPACS and other techniques: 125-142. Freshwater Biological Association, Ambleside, UK.

Smith, M. J., Kay, W. R., Edward, D. H. D., Papas, P. J., Richardson, K. S., Simpson, J. C., Pinder, A. M., Cale, D. J., Horwitz, P. H. J., Davis, J. A., Yung, F. H., Norris, R. H. & Halse, S. A. 1999. AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia. Freshwat. Biol. 41: 269-282.

Snelder, T. H., Cattaneo, F., Suren, A. M. & Biggs, B. J. E. 2004. Is the River Environment Classification an improved landscape-scale classification of rivers? J. N. Am. Benthol. Soc. 23: 580-598.

Stoddard, J. L., Larsen, D. P., Hawkins, C. P., Johnson, R. K. & Norris, R. H. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecol. Appl. 16: 1267-1276.

Suren, A. M. & Ormerod, S. J. 1998. Aquatic bryophytes in Himalayan streams: testing a distribution model in a highly heterogeneous environment. Freshwat. Biol. 40: 697-716.

Townsend, C. R., Doledec, S., Norris, R., Peacock, K. & Arbuckle, C. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. Freshwat. Biol. 48: 768-785.

Turak, E., Flack, L. K., Norris, R. H., Simpson, J. & Waddell, N. 1999. Assessment of river condition at a large spatial scale using predictive models. Freshwat. Biol. 41: 283-298.

Van Sickle, J., Hawkins, C. P., Larsen, D. P. & Herlihy, A. T. 2005. A null model for the expected macroinvertebrate assemblage in streams. J. N. Am. Benthol. Soc. 24: 178-191.

Van Sickle, J., Huff, D. D. & Hawkins, C. P. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwat. Biol. 51: 359-372.

Van Sickle, J., Larsen, D. P. & Hawkins, C. P. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. J. N. Am. Benthol. Soc. 26: 319-331.

Van Sickle, J. 2008. An index of compositional dissimilarity between observed and expected assemblages. J. N. Am. Benthol. Soc. 27: 227-235.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E. 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37: 130-137.

Wallach, D. & Goffinet, B. 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecol. Model. 44: 299-306.

Wright, J. F., Moss, B., Armitage, P. D. & Furse, M. T. 1984. A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data. Freshwat. Biol. 14: 221-256.

Wright, J. F., Furse, M. T., Armitage, P. D. & Moss, D. 1993. New procedures for identifying running-water sites subject to environmental stress and for evaluating sites for conservation, based on the macroinvertebrate fauna. Arch. Hydrobiol. 127: 319-326.

Wright, J. F., Sutcliffe, D. W. & Furse, M. T. 2000. Assessing the biological quality of fresh waters. RIVPACS and other techniques. 400 p., The Freshwater Biological Association, Ambleside.

Yuan, L. L., Hawkins, C. P. & Van Sickle, J. 2008. Effects of regionalization decisions on an O/E index for the US national assessment. J. N. Am. Benthol. Soc. 27: 892-905.