

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Jyväsjärvi, Jussi; Aroviita, Jukka; Hämäläinen, Heikki

Title: An extended Benthic Quality Index for assessment of lake profundal macroinvertebrates: addition of indicator taxa by multivariate ordination and weighted averaging

Year: 2014

Version:

Please cite the original version:

Jyväsjärvi, J., Aroviita, J., & Hämäläinen, H. (2014). An extended Benthic Quality Index for assessment of lake profundal macroinvertebrates: addition of indicator taxa by multivariate ordination and weighted averaging. *Freshwater science*, 33(3), 995-1007. <https://doi.org/10.1086/676914>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

An extended Benthic Quality Index for assessment of lake profundal macroinvertebrates: addition of indicator taxa by multivariate ordination and weighted averaging

Jussi Jyväsjärvi^{1,2,4}, Jukka Aroviita^{3,5}, and Heikki Hämäläinen^{1,6}

¹University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FIN-40014 Jyväskylä University, Finland

²University of Oulu, Department of Biology, P.O. Box 3000, FI-90014, Oulu University, Finland

³Finnish Environment Institute, Freshwater Centre, Monitoring and Assessment, P.O. Box 413, FI-90014 Oulu University, Finland

Abstract: The chironomid Benthic Quality Index (BQI) is a widely used metric in assessments of lake status. The BQI is based on 7 indicator taxa, which like most profundal fauna, often occur sporadically in low densities. Hence, a major weakness of the index is that it cannot be calculated when indicator taxa are not captured. Thus, an extension of the BQI that incorporates more macroinvertebrate taxa is desirable. We used 2 statistical approaches (Detrended Correspondence Analysis and Weighted Averaging) to estimate new benthic quality indicator scores for profundal macroinvertebrate taxa and to construct modified BQIs called Profundal Invertebrate Community Metrics (PICMs). We calibrated the PICMs and evaluated their bioassessment performance with macroinvertebrate and environmental data from 735 lake basins in Finland. Both PICMs included 70 taxa and could be calculated for a substantially greater proportion (99.5%) of sites than the original BQI (83.5%). Compared to the BQI, the PICMs were more strongly correlated with whole-community variation and were more predictable from environmental factors independent of human activities in undisturbed reference lakes. PICMs were more specific in identifying undisturbed lakes and more sensitive in discriminating nonreference from reference lakes. The strength of relationships to total P concentration was equal among indices. These results suggest that the extension of BQI to incorporate more taxa will increase generality, accuracy, and representativeness of lake profundal macroinvertebrate assessment.

Key words: benthic invertebrates, bioassessment, O/E ratio, indicator species, lakes, weighted averaging

Profundal macroinvertebrate assemblages are frequently used in biological assessment and monitoring of lakes. Their use originated with Thienemann (1922) and Brundin (1951), who empirically described a strong association between profundal chironomid species composition and lake trophic type. Largely based on Thienemann's and Brundin's lake typology concept, Wiederholm (1980) set quality indicator values for a selection of profundal chironomid and Oligochaeta taxa and developed a separate Benthic Quality Index (BQI) for both groups. These indices are structurally similar to most bioindices (Cairns and Pratt 1993) and are calculated as abundance-weighted averages of indicator scores (integers from 1 to 5) of taxa to indicate nutrient enrichment and associated conditions in lakes.

The chironomid BQI has been widely adopted in Europe as a profundal habitat-monitoring tool (Gerstmeier

1989, Kansanen et al. 1990, Johnson 1998, Hämäläinen et al. 2003, Raunio et al. 2007, Verbruggen et al. 2011). BQI also has been used provisionally as a metric in Swedish (Johnson and Goedkoop 2007) and Finnish (Rask et al. 2011, Jyväsjärvi et al. 2012) bioassessments of lakes mandated by the Water Framework Directive of the European Union (European Commission 2000). The BQI is well suited for such assessment because it can be interpreted as measuring the community characteristics, 'ratio of disturbance-sensitive taxa to insensitive taxa' and 'taxonomic composition,' included in the normative definitions of ecological status by the WFD (Jyväsjärvi et al. 2010). The strong response of BQI to increasing and decreasing human disturbance has been well documented in contemporary (Rask et al. 2011, Jyväsjärvi et al. 2012) and palaeolimnological (Ilyashuk et al. 2003, Meriläinen et al. 2003, Hynynen et al. 2004, Verbruggen et al. 2011) studies. Another advantage

E-mail addresses: ⁴Present address: University of Oulu, Department of Biology, P.O. Box 3000, FI-90014, Oulu University, Finland, jussi.jyvasjarvi@oulu.fi; ⁵jukka.aroviita@ymparisto.fi; ⁶heikki.o.hamalainen@ju.fi

DOI: 10.1086/676914. Received 17 April 2013; Accepted 21 November 2013; Published online 2 June 2014.
Freshwater Science. 2014. 33(3):995–1007. © 2014 by The Society for Freshwater Science.

995

of BQI for modern reference-condition-based assessments (Bailey et al. 2004, Stoddard et al. 2006) is that its site-specific reference value can be predicted reliably from environmental variables (Jyväsjärvi et al. 2010). In contrast, previous studies (Jyväsjärvi et al. 2011) suggest that more conventional assessment metrics, such as taxonomic completeness widely used, e.g., for river invertebrates (Hawkins 2006), cannot be applied to profundal invertebrates of boreal lakes. A range of modelling approaches (including River InVertebrate Prediction and Classification System [RIVPACS]) tested to predict profundal macroinvertebrate fauna in near-pristine conditions were imprecise, apparently because of a low number of species (Jyväsjärvi et al. 2011).

However, the BQI is based on only 7 chironomid indicator taxa, and samples may not contain any of these taxa. Wiederholm (1980) suggested that a lack of these taxa indicates the most deteriorated status, but this interpretation can be incorrect because the few indicator species can be missing for natural reasons or simply be undetected. In fact, these taxa are often missing from samples from nearly pristine lakes (Jyväsjärvi et al. 2010) because of their low abundance. This situation creates great uncertainty in BQI estimates (Veijola et al. 1996). Some investigators have attempted to tackle this issue by extending the BQI indicator species list with additional chironomid taxa (Kansanen et al. 1990, Johnson and Goedkoop 2007), but the additions have been limited, subjective, and not formally validated. Thus, a need exists to improve the BQI to make it a more broadly applicable profundal macroinvertebrate assessment metric that: 1) can be calculated for all sites and 2) whose reference values can be modeled with precision at least equal to that of the original, and 3) is sensitive in indicating anthropogenic community changes.

We used a large data set including invertebrates and environmental variables from 735 lake basins in Finland to evaluate 2 statistical approaches to estimating 'benthic quality' indicator scores from the original BQI gradient for all typical profundal macroinvertebrate taxa in the data set. We compared the bioassessment performance of the resulting new indices, the Profundal Invertebrate Community Metric (PICM), with the original BQI. We specifically examined: 1) accuracy and precision in predicting index values for reference sites from environmental variables insensitive to human actions over time and across the study area, and 2) relationships of the status estimates (observed-to-expected index values) to eutrophication among the sites. We expected that the extension of the indicator taxa list would enable us to calculate the PICMs at more sites than the original BQI could be. We postulated that because they incorporated more taxa and invertebrate groups, the PICMs would show less sampling variation, be more predictable, and exhibit more consistent responses to human disturbance than the BQI.

METHODS

Data and assignment of sites

We gathered all available lake profundal-macroinvertebrate data and corresponding environmental data collected in Finland during 1986–2011. The data originated from various sources, including the national database (HERTTA) of the Finnish Environmental Institute (Suomen ympäristökeskus [SYKE]), other archives of SYKE, various monitoring reports, and theses. The data set includes 735 discrete lake basins (sites; Fig. 1). Based on the criteria suggested by European Union guidance (European Commission 2003), the study sites were screened for possible anthropogenic pressures by Finnish environmental authorities and experts and classified into 2 groups: reference sites with minimal human influence (REF; $n = 79$) and sites not meeting the reference criteria that are potentially impacted by human activity of varying intensity (IMP; $n = 656$). We subdivided the REF sites further into 2 data sets: 75% of the sites ($n = 57$) were randomly selected for calibration of linear regression models for the prediction of reference values (CAL data; see below) and the remaining 22 sites were assigned to a data set for validation of the regression models (VLD data).

Macroinvertebrate sampling

Macroinvertebrate samples were collected (September–October) from the deepest point of each lake basin using a



Figure 1. Geographic locations of the 735 study lakes in Finland.

standard method (SFS 5076; 5–8 Ekman grab replicates, area = 250–300 cm²/replicate, 400–600- μ m sieve). For sites with multiple sampling years, we used the most recent sample from each site for the analyses. In addition, we used temporal monitoring data from 3 REF sites in validation (see below). All macroinvertebrates were sorted from the ethanol-preserved samples on a white background, identified to the lowest possible taxon, and counted. We harmonized species data to achieve taxonomic consistency (e.g., some species identifications were reduced to genus). We pooled grab replicates to 1 sample for each lake basin and converted counts to densities (individuals/m²). We omitted the rarest taxa (observed in <1% of the 735 samples; $n = 117$) from the analyses, and we removed taxa identified to levels coarser than genus (Nematoda, Ostracoda, Turbellaria, Ceratopogonidae, Hydrachnellae).

Environmental data

The environmental data included geographic (altitude, longitude, latitude), water-quality (epilimnetic [0–2 m depth] total P concentration [TP], total N [TN], chlorophyll a [chl a], pH, water color, hypolimnetic [1 m above bottom] water temperature, dissolved O₂ concentration) and morphometric (lake surface area, lake mean depth, basin maximum depth) information (Table 1). The water-quality data were measured with standard methods (APHA 1998) and were collected from the same site and year of macroinvertebrate sampling. In a few cases when year- or site-specific TP, TN, chl a , or pH data were not available, we used information from the preceding year or the closest sampling site. We calculated averages of all available measurements from the ice-free period (May–September) for

Table 1. Summary of geographic, morphometric, and water-chemistry variables for the 735 lake basins.

Variable	Mean	Minimum	Maximum
Latitude ^a	62.32	59.96	69.74
Longitude ^a	27.03	20.77	31.37
Altitude (m asl) ^a	99	0	473
Lake area (km ²) ^a	179	0.03	1377
Mean depth (m) ^a	6.9	0.6	21.1
Sampling depth (m) ^a	22	5	90
Hypolimnetic temperature (°C) ^a	11.6	4.3	23.4
Water color (mg Pt/L) ^a	61	4	360
Hypolimnetic O ₂ (mg/L)	4.9	0	11.6
Chlorophyll a (μ g/L)	9.5	0.5	73.6
Total P (μ g/L)	19.5	1.9	200.0
Total N (μ g/L)	555	100	3467
pH	7.0	5.1	8.9

^a Variable considered insensitive to human actions

TP, TN, chl a , color, and pH. Measurements of hypolimnetic O₂ concentration and temperature were taken during the late summer stagnation period (mid or late August), representing the respective summer minimum and maximum.

Index calculation

BQI We calculated the original BQI (Wiederholm 1980) for all samples as the abundance-weighted average of indicator taxon scores k_i :

$$\text{BQI} = \frac{\sum_{i=0}^7 n_i k_i}{N} \quad (\text{Eq. 1})$$

in which k_i is an integer from 1 (representing preference for eutrophy) to 5 (oligotrophy) for each indicator taxon i , n_i is the numerical abundance of taxon i and N is the sum of n_i . The 7 indicator taxa with their corresponding scores are: *Chironomus plumosus*-type ($k = 1$), *Chironomus anthracinus*-type (2), *Sergentia coracina* (3), *Stictochironomus rosenscholdi* (3), *Micropsectra* spp. (4), *Paracladopelma nigrutulium* (4), and *Heterotrissocladius subpilosus* (5). According to Wiederholm (1980), absence of indicator taxa indicates the most disturbed conditions and BQI is given a value of 0. However, we omitted these observations from statistical analyses because they can also arise from sampling error (see above), in these cases the index is formally undefined, and inclusion of 0 values would make the BQI scale noncontinuous.

Ordination approach Wiederholm's (1980) BQI is based on 7 key profundal chironomid taxa with distinctive preferences for temperature, O₂, and nutrient status. Thus, the BQI taxa have well defined ecological niches manifested by their systematic positioning along primary community and environmental gradients in multivariate ordination space (Johnson et al. 1990, Kansanen et al. 1990, Mousavi 2002, Jyväsjärvi 2011). The main gradient of profundal invertebrate community turnover among lakes can be expected to coincide with benthic quality, so the indicator scores corresponding to the original BQI might be inter- and extrapolated for all taxa by using their position in the 1st ordination dimension. Therefore, we ran Detrended Correspondence Analysis (DCA; Hill and Gauch 1980) on log₁₀(x)-transformed species abundance data from the 735 lake basins without the option of down-weighting rare taxa in the *decorana* function of the *vegan* package (Oksanen et al. 2008) in R (R Project for Statistical Computing, Vienna, Austria). DCA produces scores for both samples (DCA_S) and taxa (DCA_T) for ordination axes. These scores are the locations of the sites and taxa along the corresponding ordination dimensions. We extracted the DCA_T scores of the ordination axis 1, and re-

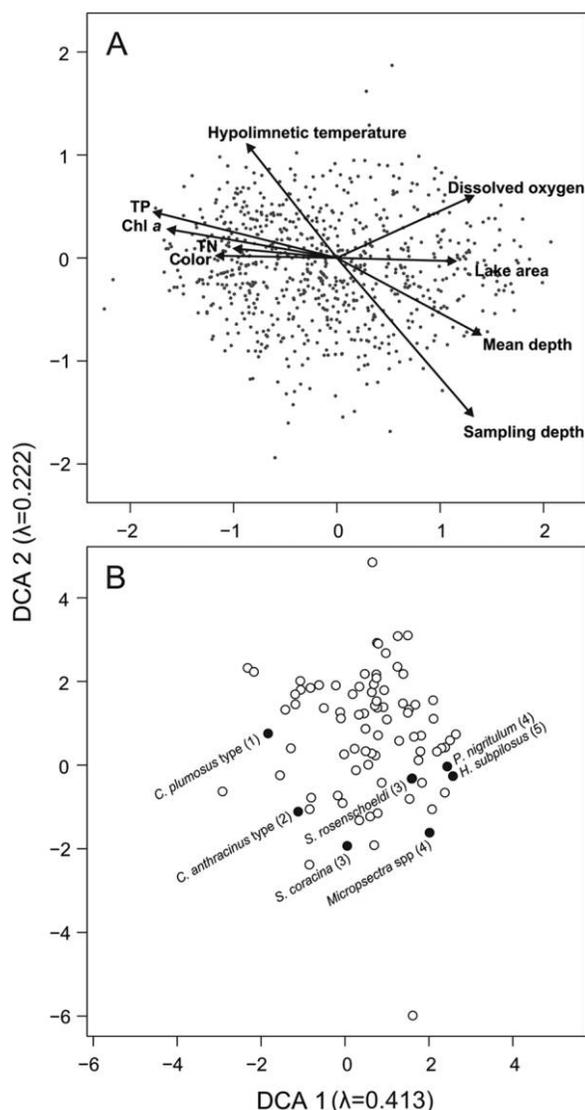


Figure 2. Detrended Correspondence Analysis (DCA) ordination showing distribution of 735 study sites (A) and 77 profundal macroinvertebrate taxa (B). In A, the vector arrows point in the direction of increasing values for the environmental variables. In B, only the chironomid taxa included in the Benthic Quality Index are labeled, and their indicator values are shown in parentheses.

valued them by setting the minimum DCA_T value to 0. We then fitted linear regression equations to predict the original indicator scores of the 7 BQI taxa from their revalued DCA 1 axis scores:

$$BQI\ score = a + b + DCA_T. \quad (Eq. 2)$$

We used this regression model to estimate indicator values on a continuous scale (instead of integers) for all taxa on the basis of their DCA_T axis 1 scores. We linearly rescaled estimates to range from 0 to 5 to allow achieve-

ment of observed/expected (O/E) values (see below) of 0 as required by the WFD. We calculated the new index, Profundal Invertebrate Community Metric ($PICM_{DCA}$), with Eq. 1 but summed across all the taxa with an estimated indicator value. The DCA_T scores were based on logarithmic data, so we used $\log_{10}(x)$ -transformed abundances instead of the original abundance values (n_i) for metric calculation. This approach helped downweight occasionally extremely dominant taxa, such as *Chaoborus flavicans*, and decreased unpredictable metric variation. The species data were strongly dominated by 2 taxa identified only to genus (the chironomid *Procladius* spp. and the oligochaete worms *Tubifex tubifex*/*Potamothrix hammoniensis*). These taxa occurred in most samples (86 and 66%, respectively), were often highly abundant, and each included >1 species and had wide environmental tolerances. According to preliminary analyses, these 2 taxa caused considerable noise in the index values. Therefore, we omitted them from the indicator species list.

WA approach We also explored weighted averaging (WA; ter Braak and Prentice 1988) to estimate optima (indicator scores) directly along the BQI gradient for all invertebrate taxa and subsequently to infer the benthic quality of sites from the species data. WA modelling is an efficient method to predict environmental gradients from species data and is widely used in palaeolimnological (e.g., Hall and Smol 1992) and contemporary (Hämäläinen and Karjalainen 1994, Hämäläinen and Huttunen 1996, Raunio et al. 2010) assessment and monitoring by invertebrates. The BQI itself can be considered a simplified special case

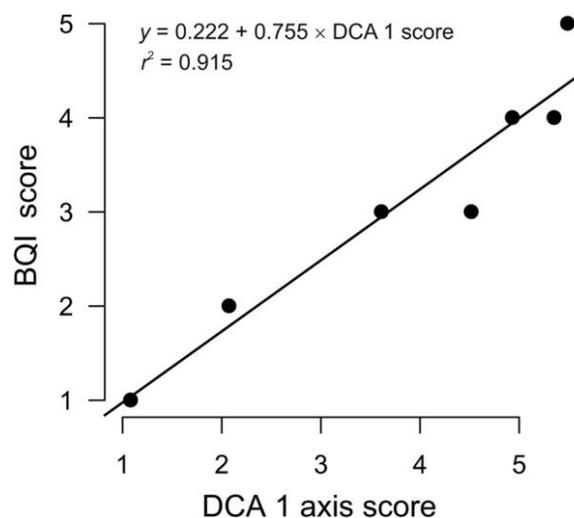


Figure 3. Relationship of the indicator scores of the 7 Benthic Quality Index (BQI) indicator taxa (see Fig. 2B) to revalued Detrended Correspondence Analysis for taxa (DCA_T) axis 1 scores. Fit of a linear regression model is indicated by solid line.

of WA modelling where the environmental gradient to be inferred is not explicit and directly measurable and, therefore, (necessarily) the indicator values are subjective and not formally estimated.

We calculated the original BQI for all possible sites. Using the data from those sites, we estimated the BQI optimum (k), corresponding to the indicator score k , for each taxon by weighted averaging

$$\hat{k}_i = \frac{\sum y_{ij} x_j}{\sum y_{ij}} \quad (\text{Eq. 3})$$

where y is $\log_{10}(x)$ -transformed numerical abundance of the taxon i at site j and x is the environmental variable

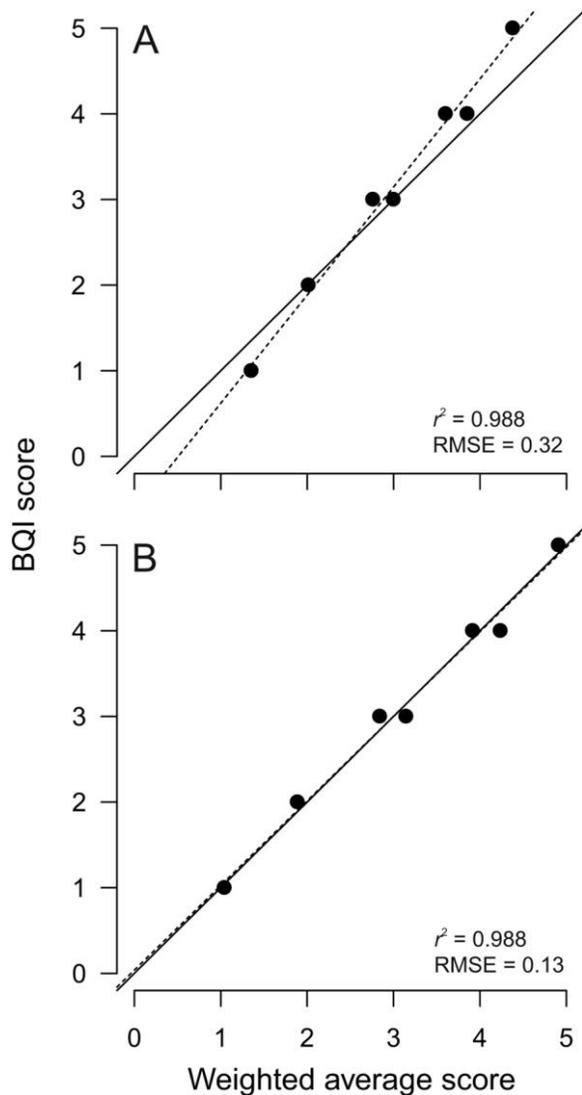


Figure 4. Relationship of the original indicator scores of the 7 Benthic Quality Index (BQI) chironomid taxa to initial weighted average scores (A) and those corrected by inverse deshinking (B). Solid line indicates 1:1 relationships and dashed line denotes fit of a linear regression model.

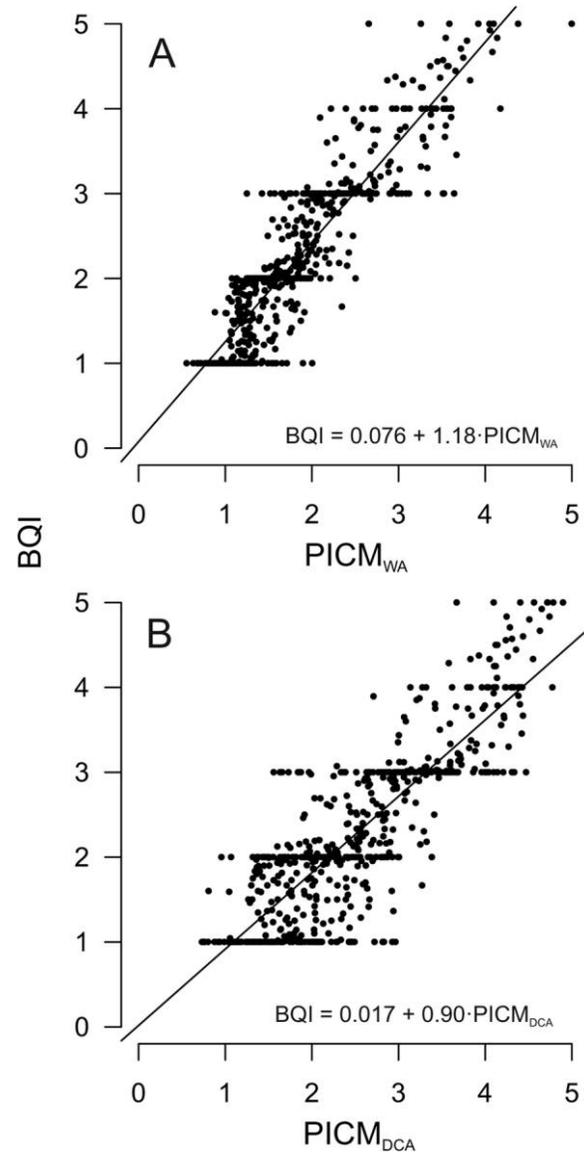


Figure 5. Relationships between Benthic Quality Index (BQI) scores and Profundal Invertebrate Community Metrics (PICM) scores based on weighted averaging (WA) (PICM_{WA}) (A) and detrended correspondence analysis (DCA) (PICM_{DCA}) (B) among all sites with calculated BQI values. Solid line denotes fit of a linear regression.

(here BQI) at that site. We calculated an estimate of ‘BQI’ for all sites as the abundance (y_{ij})-weighted average of species scores and used an inverse deshinking method (Birks et al. 1990) to correct for the inherent prediction bias of WA. We adjusted the optimum estimates according to the deshinking parameters (Marchetto 1994). We rescaled the indicator values to the original BQI range (see above). We calculated site-specific PICM_{WA} index values with the BQI formula on $\log_{10}(x)$ -transformed abundance data. We omitted the 2 problematic taxa (*Procladius* spp. and *Tubifex/Potamothrix*) from the indicator species list.

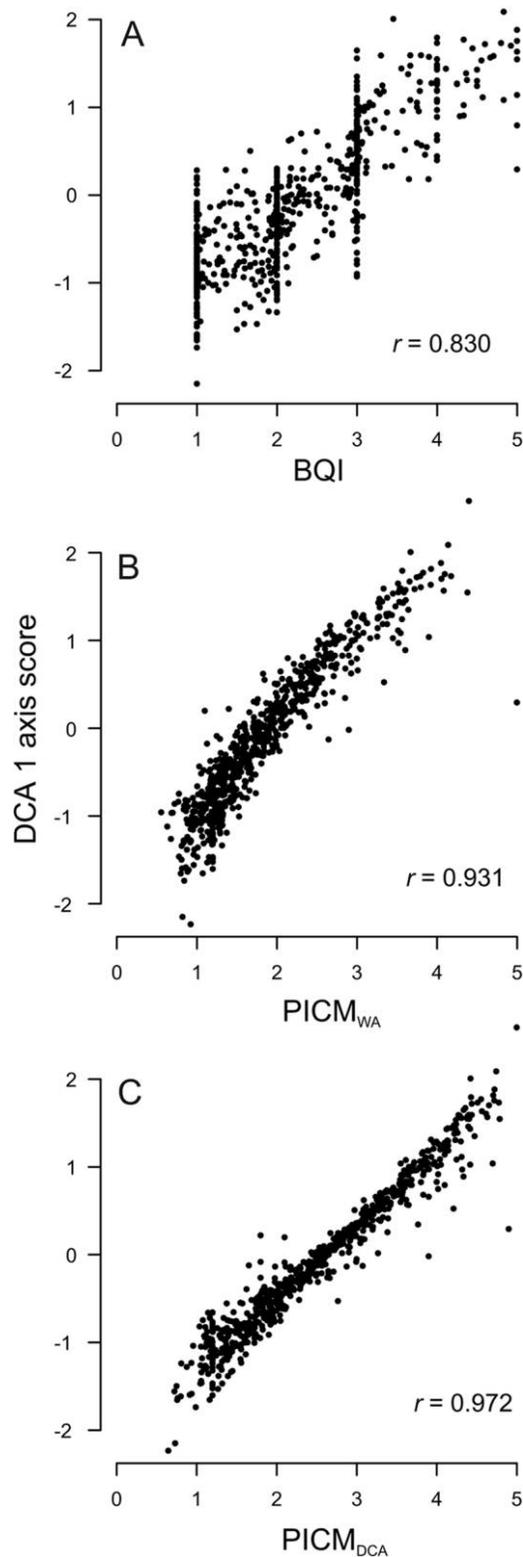


Figure 6. Relationships of DCA 1 axis scores and the BQI (A), $PICM_{WA}$ (B), and $PICM_{DCA}$ (C) among all sites with calculated BQI values. Pearson correlation coefficients are shown in each panel. See Fig. 5 for index definitions.

Evaluation of index performance We determined the total number and proportion of sites for which each index could be calculated. After this, we reduced the data by omitting sites without BQI indicator taxa. We used linear regression analysis to evaluate the relationship between the BQI and the PICM indices. We also correlated (Pearson product–moment correlation) the index values with site-specific DCA_S axis 1 scores to assess the relationship of indices with general community turnover of the profundal macroinvertebrate assemblages.

We developed multiple linear regression (MLR) models (see Jyväsjärvi et al. 2010) based on CAL data to estimate site-specific reference values for each index on the basis of environmental variables insensitive to human activity (Table 1). We also included water color among the candidate predictors because it is used as a proxy for geology in Finnish lake typology, but we acknowledge that the natural color value expected in the absence of human disturbance should be used. The candidate predictor variables (Table 1) with skewed distributions (all but longitude) were $\log_{10}(x)$ -transformed to obtain normality. We used the *leaps* package (Lumley 2004) in R to evaluate all possible candidate MLR models. The *leaps* package uses branch-bound algorithms (Gatu and Kontoghiorghes 2006) to select a specified number of optimal models. We used this routine to select the single best (based on the lowest Bayesian Information Criterion) regression model for each index. We did not model interactions among the predictors. We assessed the performance of the MLR models by predicting index values for VLD sites from their environmental characteristics. We plotted observed index values against the predictions and assessed goodness of fit with root mean squared error (RMSE; Wallach and Goffinet 1989).

For each index and site, we determined the deviation of the observed (O) index from the modeled site-specific expected (E) reference value as the ratio O/E, equivalent to the Ecological Quality Ratio demanded by the WFD (European Commission 2000). We examined accuracy and precision of the indices by calculating averages and standard deviations (SD) of the resulting O/E ratios for CAL and VLD sites. We estimated specificity (F) of the indices as the proportion of VLD sites with an O/E ratio higher than the lower quartile of the reference values (CAL) distribution, a frequently used lower boundary for ‘high status’ (European Commission 2003, Ostermiller and Hawkins 2004, Aroviita et al. 2010, Jyväsjärvi et al. 2011). F estimates the proportion of correct classifications and $1 - F$ the rate of false alarms for sites in reference status (with no human impact). Sensitivity in detecting anthropogenic influence was estimated as the proportion of IMP sites with an O/E ratio lower than high status. An assessment metric that performs well should produce O/E ratios that are precise (low SD) and accurate (close to 1) at reference sites and, hence, be able to identify biologically minimally

Table 2. Results of multiple linear regression analyses for the 3 indices. See Fig. 5 for index definitions.

Index	Variable	Unstandardized coefficient	Standard error	<i>p</i>	<i>r</i> ²
BQI	Constant	-1.056	0.435	0.019	0.585
	Sampling depth (log ₁₀)	2.767	0.314	<0.001	
	Constant	-1.277	0.323	<0.001	
PICM _{WA}	Sampling depth (log ₁₀)	2.593	0.233	<0.001	0.692
	Constant	-0.904	0.395	0.026	
	Mean depth (log ₁₀)	1.759	0.846	0.042	
PICM _{DCA}	Sampling depth (log ₁₀)	1.764	0.589	0.004	0.657

disturbed and impaired sites effectively and correctly (high specificity and sensitivity, respectively). We compared responsiveness of the indices to anthropogenic stress by correlating (Pearson product-moment correlation) the O/E ratios with TP concentration as an indicator of eutrophication, the main stressor among the study lakes.

Last, we examined temporal variation of the indices using monitoring data from 3 reference lakes sampled by standard protocols (see above). The lakes were Lake Pääjärvi (lat 61°05'N, long 25°08'E), a deep (mean depth = 14.8 m), moderately large oligomesotrophic (mean TP = 13 µg/L), mesohumic (mean water color = 66 mg Pt/L) lake; Lake Vuohijärvi (lat 61°18'N, long 26°71'E), a moderately large, oligotrophic (mean TP = 5 µg/L), oligohumic (mean color = 22 mg Pt/L) lake; and Lake Pyhäjärvi (lat 63°58'N, long 25°91'E), a very large, relatively shallow (mean depth 6.3 m), oligotrophic (mean TP = 10 µg/L) lake. The time-series for these lakes span ~20 y and include macroinvertebrate data from 17, 10, and 12 y, respectively. We used the MLR models to predict expected index values for the 3 lakes, and we calculated O/E index ratios for all sampling occasions. We plotted the O/E ratios along time (y) and calculated averages and SDs of the O/E ratios within each lake to evaluate the precision and accuracy of the indices in temporal scale. We also tested for correlations of O/E ratios with TP concentration to explore whether the indices covaried with water quality in each time series.

RESULTS

After the removal of rare taxa (occurring in <1% of sites) the data set contained 77 taxa. The original BQI could be calculated for 83.5% of the samples (*n* = 604). The 131 sites at which BQI taxa were not captured represented a wide range of conditions, from ultraoligotrophic (TP = 2 µg/L) to hypereutrophic (TP = 129 µg/L) lakes.

In the detrended correspondence analysis (DCA), the total inertia of the species data was 7.20. The length of the DCA 1 axis (4.83 SD units, eigenvalue 0.413) indicated a

major community overturn along the 1st axis. DCA 1 represented the trophic gradient. Productivity-related variables, such as nutrients and chlorophyll *a*, were strongly correlated with DCA_S 1 scores (Fig. 2A). The 7 chironomid BQI indicator taxa were fairly consistently positioned along DCA 1 (Fig. 2B), resulting in strong correlation between the axis and BQI scores (Fig. 3). The minimum DCA_T 1 score (-2.97) (Fig. 2B), set to 0, belonged to chironomid *Prosilocerus jacuticus*, which seemed to be the species most tolerant of eutrophy. As an alternative to DCA, we used WA to estimate indicator scores for new taxa. The estimated initial WA optimum values for the 7 BQI chironomid taxa corresponded well with the original BQI scores (Fig. 4A) particularly when corrected by deshrinking (Fig. 4B).

We removed *Procladius* spp., *Potamothenix/Tubifex*, and taxa higher than genus (*n* = 5) and estimated the DCA- and WA-based indicator values for 70 taxa (Appendix S1). Therefore, we were able to calculate the expanded PICM indices for nearly all sites (*n* = 731; 99.5%). The 4 sites with no PICM indicator taxa present included 3 meso- to eutrophic lake basins (TP = 14–57 µg/L) and 1 oligotrophic lake basin (TP = 6 µg/L). Both PICM_{WA} and PICM_{DCA} had strong relationships with the original BQI (Fig. 5A, B). The samples that deviated the most had a low number of specimens (≤50) indicating that the discrepancy between BQI and PICMs was mainly related to sampling variability. The PICMs were more strongly correlated than the BQI with the 1st DCA_S axis scores (Fig. 6A–C) and, thus, were more congruent with the composition of the entire macroinvertebrate community.

We examined all possible subsets of the environmental predictor variables. The best multiple regression model to predict reference values of the indices included only log₁₀ (*x*)-transformed sampling depth (= maximum depth) to account for the variation of BQI (*r*² = 0.585; Table 2) and PICM_{WA} (*r*² = 0.692; Table 2). Mean depth also was included in the best model for PICM_{DCA} (*r*² = 0.657; Table 2). All 3 models predicted index values for the 22 VLD sites reasonably accurately, but the coefficient of determi-

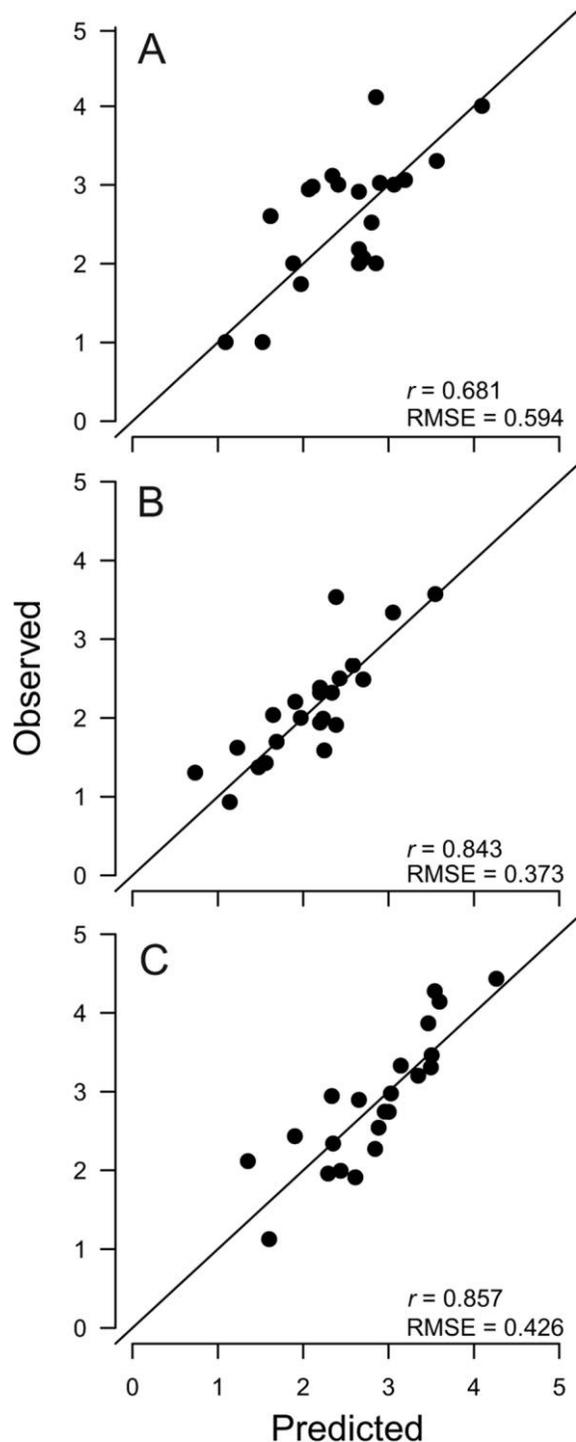


Figure 7. Relationship between observed and predicted values of Benthic Quality Index (A), Profundal Invertebrate Community Metrics scores based on weighted averaging (B), and detrended correspondence analysis (C) among the 22 validation sites. Solid line denotes 1:1 relationship. Pearson correlation coefficients and root mean squared errors (RMSE) are shown in each panel.

nation was lowest and RSME highest for the original BQI (Fig. 7A–C).

O/E of the indices averaged exactly or close to 1 for CAL and VLD sites (Table 3). O/E of BQI varied most among the VLD sites ($SD = 0.268$), whereas O/E of $PICM_{DCA}$ was the most precise ($SD = 0.199$), followed by O/E of $PICM_{WA}$ ($SD = 0.234$) (Table 3). $PICM_{DCA}$ outperformed the BQI and $PICM_{WA}$ in specificity by classifying 13.7% of the VLD sites in lower than high status (27.3% for BQI and 18.2% for $PICM_{WA}$). O/Es among the nonreference IMP sites averaged between 0.876–0.940 and were lowest for $PICM_{DCA}$ and highest for $PICM_{WA}$ (Table 3). $PICM_{DCA}$ was the most able to separate IMP sites from the reference sites, classifying ~57% of the IMP sites in nonreference status. The BQI was the least sensitive, assigning only ~45% of the IMP sites into the nonreference category (Table 3). All indices showed a moderate negative correlation with TP concentration with small differences between BQI ($r = -0.323$), $PICM_{WA}$ ($r = -0.253$), and $PICM_{DCA}$ ($r = -0.297$) (Fig. 8A–C).

Examination of temporal variation of the indices among the 3 reference lakes demonstrated the main defect of the original BQI. PICMs were obtained for all observations, but the BQI could be calculated only for some of the years, regardless of the supposed high status of the study lakes (Fig. 9A–C). BQI was available for 15 of 17 y for Lake Pääjärvi. Lake Vuohijärvi and Lake Pyhäjärvi were more problematic, with only 4 of 10 y, and 5 of 8 y, respectively, with ≥ 1 captured BQI indicator species. However, when calculable, the original BQI produced rather accurate (mean O/E = 0.971–1.169) and precise (SD O/E = 0.093–0.236) estimates of status of macroinvertebrate assemblages in these 3 reference lakes (Table 4). Similar accuracy was reached with the both PICMs, except for Lake Pääjärvi, where the mean O/E deviated markedly from 1 (~0.86–0.87). Both PICMs were more precise than the BQI, and $PICM_{DCA}$ had the smallest random error (Table 4). O/E ratios and TP were not correlated in Lake Vuohijärvi and Lake Pyhäjärvi. In Lake Pääjärvi, $PICM_{WA}$ and $PICM_{DCA}$ were negatively correlated with TP, whereas the BQI was not (Table 4).

DISCUSSION

We demonstrated that the lake profundal Benthic Quality Index (Wiederholm 1980) comprising only 7 chironomid indicator taxa can be objectively extended by including all common profundal macroinvertebrate taxa with 2 simple statistical methods. Substantial extension of the indicator species list solved the problem of missing values for the original BQI and enabled index calculations for almost all sites. The new PICM indices also seemed to outperform the original BQI in that the reference conditions could be more accurately and precisely predicted from environmental variables. Consequently, the new indices were

Table 3. Summary statistics of the observed/expected ratios of the 3 indices among calibration (CAL), validation (VLD), and impacted (IMP) sites. SD = standard deviation, REF = reference condition. See Fig. 5 for index definitions.

Index	CAL sites			VLD sites			IMP sites	
	Mean	SD	25 th quantile	Mean	SD	25 th quantile	Mean	% nonREF
BQI	1.003	0.271	0.859	1.027	0.268	0.834	0.907	45.2
PICM _{WA}	1.011	0.231	0.843	1.046	0.234	0.911	0.940	54.0
PICM _{DCA}	1.006	0.239	0.844	1.008	0.199	0.886	0.876	56.8

more specific in assigning fewer reference sites as being impacted and were apparently more sensitive in detecting more nonreference sites as correctly impacted. Moreover, the response to anthropogenic disturbance remained equal in strength. The PICMs were strongly correlated with species ordination scores (DCA 1 axis) and, hence, effectively captured the main gradient of community turnover in a single number and matched up to 'taxonomic composition', which is a key feature in WFD-based bioassessment of European lakes.

The use of ordination analysis or WA to derive indicator scores for macroinvertebrates is not particularly novel. Many investigators have applied multivariate ordination methods to generate a synthetic pollution gradient and then used species scores along that ordination dimension to indicate species tolerance to the measured stressor or, alternatively, used constrained ordination (CCA) to estimate species stressor optima. Such approaches have been developed for marine (Smith et al. 2001) and stream benthic invertebrates (Davy-Bowker et al. 2005), freshwater phytoplankton (Phillips et al. 2013), and macrophytes (Dodkins et al. 2005). An approach somewhat analogous to our WA method was developed by Rossaro et al. (2007), who modified Wiederholm's (1980) BQI to include multiple lentic macroinvertebrate taxa for Italian lakes and more recently for lakes in the central European Alpine region (Rossaro et al. 2012). All of these investigators reported strong correlations between the index values and supposed pollution status (an obvious outcome considering that the indices were calibrated to do so). In contrast, our PICMs were calibrated against a general 'benthic quality' gradient which equaled the main overturn gradient in the benthic invertebrate assemblages, rather than against any specific single environmental stressor directly.

The O/Es of both PICMs and the original BQI were negatively correlated ($r = -0.25$ to -0.32) with TP concentration of the ice-free period, although the correlations were weaker than those reported for recently developed assessment metrics of other lentic communities, such as fish (Argillier et al. 2013), phytoplankton (Phillips et al. 2013), and macrophytes (Lyche-Solheim et al. 2013). Many of these metrics or their components were intentionally selected to be responsive to eutrophication, and strong cor-

relation between metric variation and TP was expected. Moreover, compared to primary producers, profundal communities are not directly dependent on nutrient conditions and depend instead on eutrophication-driven alteration of food conditions and particularly O₂ depletion (Jyväsjärvi et al. 2013). In the present data, after accounting for index variation by TP content in the linear regression model, hypolimnetic O₂ concentration slightly but significantly explained the residual O/E variation of all indices (addition in $r^2 = 0.01$ – 0.06). A further explanation for weak correlation between TP and the O/E ratios may arise from morphometry-related lake productivity. Lake morphometry, lake depth in particular, and nutrient status often naturally strongly covary in dimictic lakes so that shallow lakes are typically more eutrophic than deeper ones (Moss 1980, Cardoso et al. 2007). Accounting for morphometry-related natural-lake productivity by using depth or other morphological parameters as predictors, as we have done here for macroinvertebrate metrics, should produce more accurate estimates of the reference communities and, thus, the whole status assessment in general (see Jyväsjärvi et al. 2009). A similar depth-related approach might be worth applying to other biological components because the literature indicates strong relationships between lake morphometry and assemblage/metric variation of fish (Olden 2003, Mehner et al. 2007), macrophytes (Alahuhta et al. 2012), and phytoplankton (Maileht et al. 2013). A final explanation for the weak observed relationship between TP and O/E may be that our data encompassed a wide range of environmental gradients and included very shallow basins (Table 1), which appear to be problematic for profundal macroinvertebrate assessment (Jyväsjärvi et al. 2012). After removal of the 40 shallowest (maximum depth < 7 m) sites, the relationship between indices and TP was substantially stronger ($r = -0.40$ to -0.42).

Evaluation of temporal variation of the metrics suggested that PICMs perform better than the original BQI. The number of samples with indicator species present was considerably higher, and O/Es were, in most cases, more accurate and precise. Lake Pääjärvi was an exception to this pattern. The BQI indicated high status (mean O/E = 1.07), but O/E of the 2 PICMs deviated appreciably from 1 (O/E = 0.86–0.87). We relate this divergence to high

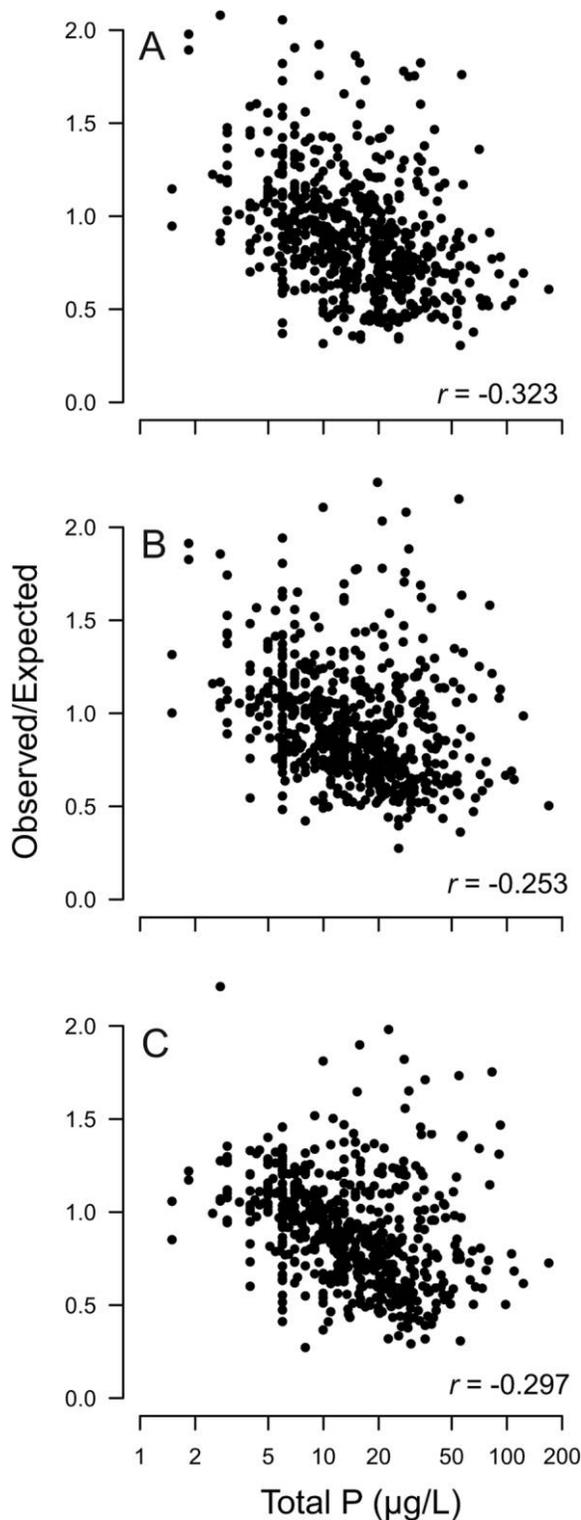


Figure 8. Relationship between observed/expected ratios of Benthic Quality Index (A), Profundal Invertebrate Community Metrics scores based on weighted averaging (B), and detrended correspondence analysis (C) and epilimnetic total P (TP) concentration among all sites with calculated BQI values. Pearson correlation coefficients are shown in each panel.

abundance of the oligochaete *Limnodrilus hoffmeisteri* in Lake Pääjärvi. This eutrophy-tolerant species (Aston 1973, Wiederholm 1980) is typical of shallow boreal lakes (Jyväsjärvi et al. 2012), and therefore, its occurrence in deep oligomesotrophic Lake Pääjärvi was rather unexpected. Compared to other medium-sized deep reference lakes in our data set, the slightly raised TP content (average = 13.3 µg/L) and, in particular, the elevated total N concentration (1292 µg/L), suggest agricultural impacts from the catchment (17% of the catchment area is cultivated). These observations suggest that the signal given by the PICM may not be erroneous, and that Lake Pääjärvi might

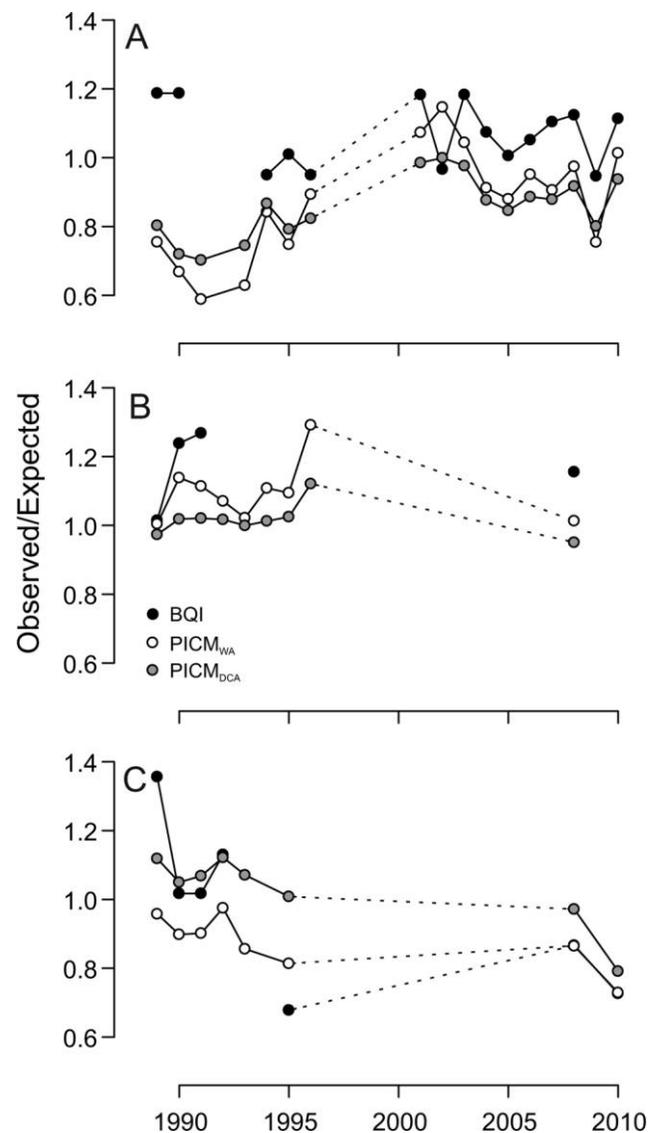


Figure 9. Temporal variation of the observed/expected ratios of the indices in Lakes Pääjärvi (A), Vuohijärvi (B), and Pyhäjärvi (C). Gap in the BQI line indicates that no indicator species were found from the samples. Dashed line represents a period with missing data. See Fig. 5 for index definitions.

Table 4. Means and standard deviations (SD) of the observed/expected (O/E) ratios and correlations of O/E and total P concentration among the 3 reference lakes. Statistically significant ($p < 0.05$) correlations are denoted in bold. See Fig. 5 for index definitions.

Index	Lake Pääjärvi			Lake Vuohijärvi			Lake Pyhäjärvi		
	Mean	SD	<i>r</i>	Mean	SD	<i>r</i>	Mean	SD	<i>r</i>
BQI	1.069	0.093	-0.204	1.169	0.113	0.254	0.971	0.236	0.012
PICM _{WA}	0.869	0.160	-0.520	1.095	0.088	0.196	0.875	0.079	-0.164
PICM _{DCA}	0.856	0.090	-0.491	1.015	0.046	-0.022	1.025	0.107	-0.233

not have been in reference status, at least during the early 1990s.

To summarize, both of the PICM indices had good assessment performance with minor differences between the indices, and both indices generally outperformed the original BQI. The prediction error was smallest for PICM_{WA}, whereas PICM_{DCA} provided most precise estimation of reference values and recognized most effectively the sites with known reference status (VLD sites) and sites that were potentially degraded by human actions. If modified for regional faunas and conditions, PICMs should also be applicable to other regions and the approach generalizable to other groups of organisms.

ACKNOWLEDGEMENTS

The profundal macroinvertebrate data set sourced from various research projects, lake biomonitoring programs, limnological studies and all parties contributing to the data are gratefully acknowledged. Critical comments of 2 anonymous referees and Associate Editor Lester Yuan helped to improve the manuscript. Roger Jones kindly checked the language of the manuscript.

LITERATURE CITED

Alahuhta, J., A. Kanninen, and K.-M. Vuori. 2012. Response of macrophyte communities and status metrics to natural gradients and land use in boreal lakes. *Aquatic Botany* 103:106–114.

APHA (American Public Health Association). 1998. Standard methods for the examination of water and wastewater. 20th edition. American Public Health Association, American Waterworks Association, and Water Environment Federation, Washington, DC.

Argillier, C., S. Caussé, M. Gevrey, S. Pédrón, J. De Bortoli, S. Brucet, M. Emmrich, E. Jeppesen, T. Lauridsen, T. Mehner, M. Olin, M. Rask, P. Volta, I. J. Winfield, F. Kelly, T. Krause, A. Palm, and K. Holmgren. 2013. Development of a fish-based index to assess the eutrophication status of European lakes. *Hydrobiologia* 704:193–211.

Aroviita, J., H. Mykrä, and H. Hämäläinen. 2010. River bioassessment and the preservation of threatened species: towards acceptable biological quality criteria. *Ecological Indicators* 10: 789–795.

Aston, R. J. 1973. Tubificids and water quality: a review. *Environmental Pollution* 5:1–10.

Bailey, R., R. H. Norris, and T. B. Reynoldson. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. Springer, New York.

Birks, H. J. B., J. M. Line, S. Juggins, A. C. Stevenson, and C. J. F. ter Braak. 1990. Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 327:263–278.

Brundin, L. 1951. The relation of O₂-microstratification at the mud surface to the ecology of the profundal bottom fauna. *Report of the Institute for Freshwater Research* 32:32–43.

Cairns, J., and J. R. Pratt. 1993. A history of biological monitoring using benthic macroinvertebrates. Pages 10–27 *in* D. M. Rosenberg and V. H. Resh (editors). *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, London, UK.

Cardoso, A. C., A. Solimini, G. Premazzi, L. Carvalho, A. Lyche, and S. Rekolainen. 2007. Phosphorus reference concentrations in European lakes. *Hydrobiologia* 584:3–12.

Davy-Bowker, J., J. F. Murphy, G. P. Rutt, J. E. C. Steel, and M. T. Furse. 2005. The development and testing of a macroinvertebrate biotic index for detecting the impact of acidity on streams. *Archiv für Hydrobiologie* 163:383–403.

Dodkins, I., B. Rippey, and P. Hale. 2005. An application of canonical correspondence analysis for developing ecological quality assessment metrics for river macrophytes. *Freshwater Biology* 50:891–904.

European Commission. 2000. Directive 2000/60/EC 2000. Establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities* L327:1–72.

European Commission. 2003. Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 10. River and lakes—typology, reference conditions and classification systems. Office for Official Publications of the European Communities, Luxembourg.

Gatu, C., and E. J. Kontoghiorghes. 2006. Branch-and-bound algorithms for computing the best-subset regression models. *Journal of Computational and Graphical Statistics* 15:139–156.

Gerstmeier, R. 1989. Lake typology and indicator organisms in application to the profundal chironomid fauna of Starnberger See (Diptera, Chironomidae). *Archiv für Hydrobiologie* 116:227–234.

Hall, R. I., and J. P. Smol. 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concen-

- tration from diatoms in British Columbia (Canada) lakes. *Freshwater Biology* 27:417–434.
- Hämäläinen, H., and P. Huttunen. 1996. Inferring the minimum pH of streams from macroinvertebrates using weighted averaging regression and calibration. *Freshwater Biology* 36:697–709.
- Hämäläinen, H., and J. Karjalainen. 1994. An assessment of Harpacticoida and resting stages of Cyclopoida as trophic indicators in Finnish lakes. *Annales Zoologici Fennici* 31:377–388.
- Hämäläinen, H., H. Luotonen, E. Koskenniemi, and P. Liljaniemi. 2003. Inter-annual variation in macroinvertebrate communities in a shallow forest lake in Eastern Finland during 1990–2001. *Hydrobiologia* 506–509:389–397.
- Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. *Ecological Applications* 16:1277–1294.
- Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42:47–58.
- Hynynen, J., A. Palomäki, J. J. Meriläinen, A. Witick, and K. Mäntykoski. 2004. Pollution history and recovery of a boreal lake exposed to a heavy bleached pulping effluent load. *Journal of Paleolimnology* 32:351–374.
- Ilyashuk, B., E. Ilyashuk, and V. Dauvalter. 2003. Chironomid responses to long-term metal contamination: a palaeolimnological study in two bays of Lake Imandra, Kola Peninsula, Northern Russia. *Journal of Paleolimnology* 30:217–230.
- Johnson, R. K. 1998. Spatiotemporal variability of temperate lake macroinvertebrate communities: detection of impact. *Ecological Applications* 8:61–70.
- Johnson, R. K., and W. Goedkoop. 2007. *Bedömningsgrunder för bottenfauna i sjöar och vattendrag? Användarmanual och bakgrundsdokument, Report 4.* Swedish University of Agricultural Sciences, Umeå, Sweden (in Swedish). (Available from: <http://info1.ma.slu.se/IMA/Publikationer/internserie/2007-04.pdf>)
- Johnson, R. K., T. Wiederholm, and L. Eriksson. 1990. The influence of season on the classification and ordination of profundal communities of nutrient poor, oligo-mesohumic Swedish lakes using environmental data. *Verhandlungen der Internationalen Vereinigung für theoretische und angewandte Limnologie* 24:646–652.
- Jyväsjärvi, J. 2011. Environmental drivers of lake profundal macroinvertebrate community variation: implications for bioassessment. PhD Thesis, University of Jyväskylä, Jyväskylä, Finland.
- Jyväsjärvi, J., J. Aroviita, and H. Hämäläinen. 2011. Evaluation of approaches for measuring taxonomic completeness of lake profundal macroinvertebrate assemblages. *Freshwater Biology* 56:1876–1892.
- Jyväsjärvi, J., J. Aroviita, and H. Hämäläinen. 2012. Performance of profundal macroinvertebrate assessment in boreal lakes depends on lake depth. *Fundamental and Applied Limnology* 180:91–100.
- Jyväsjärvi, J., G. Boros, R. I. Jones, and H. Hämäläinen. 2013. The importance of food relative to oxygen and temperature in structuring lake profundal macroinvertebrate assemblages. *Hydrobiologia* 709:55–72.
- Jyväsjärvi, J., J. Nyblom, and H. Hämäläinen. 2010. Palaeolimnological validation of estimated reference values for a lake profundal macroinvertebrate metric (Benthic Quality Index). *Journal of Paleolimnology* 44:253–264.
- Jyväsjärvi, J., K. T. Tolonen, and H. Hämäläinen. 2009. Natural variation of profundal macroinvertebrate communities in boreal lakes is related to lake morphometry: implications for bioassessment. *Canadian Journal of Fisheries and Aquatic Sciences* 66:589–601.
- Kansanen, P. H., L. Paasivirta, and T. Väyrynen. 1990. Ordination analysis and bioindices based on zoobenthos communities used to assess pollution of a lake in southern Finland. *Hydrobiologia* 202:153–170.
- Lumley, T. 2004. The leaps package. R Project for Statistical Computing, Vienna, Austria (Available from: [cran.r-project.org / doc / packages / leaps.pdf](http://cran.r-project.org/doc/packages/leaps.pdf))
- Lyche-Solheim, A., C. K. Feld, S. Birk, G. Phillips, L. Carvalho, G. Morabito, U. Mischke, N. Willby, M. Søndergaard, S. Hellsten, A. Kolada, M. Mjelde, J. Böhmer, O. Miler, M. T. Pusch, C. Argillier, E. Jeppesen, T. L. Lauridsen, and S. Poikane. 2013. Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. *Hydrobiologia* 704:57–74.
- Maileht, K., T. Nöges, P. Nöges, I. Ott, U. Mischke, L. Carvalho, and B. Dudley. 2013. Water colour, phosphorus and alkalinity are the major determinants of the dominant phytoplankton species in European lakes. *Hydrobiologia* 704:115–126.
- Marchetto, A. 1994. Rescaling species optima estimated by weighted averaging. *Journal of Paleolimnology* 12:155–162.
- Mehner, T., K. Holmgren, T. L. Lauridsen, E. Jeppesen, and M. Diekmann. 2007. Lake depth and geographical position modify lake fish assemblages of the European ‘Central Plains’ ecoregion. *Freshwater Biology* 52:2285–2297.
- Meriläinen, J. J., J. Hynynen, A. Palomäki, K. Mäntykoski, and A. Witick. 2003. Environmental history of an urban lake: a palaeolimnological study of Lake Jyväsjärvi, Finland. *Journal of Paleolimnology* 30:387–406.
- Moss, B. 1980. *Ecology of freshwaters.* Blackwell Science Publications, London, UK.
- Mousavi, S. K. 2002. Boreal chironomid communities and their relations to environmental factors: the impact of lake depth, size and acidity. *Boreal Environment Research* 7:63–75.
- Oksanen, J., R. Kindt, P. Legendre, B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2008. *vegan: community ecology package.* Version 1.15-1. R Project for Statistical Computing, Vienna, Austria.
- Olden, J. D. 2003. A species-specific approach to modeling, biological communities and its potential for conservation. *Conservation Biology* 17:854–863.
- Ostermiller, J. D., and C. P. Hawkins. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. *Journal of the North American Benthological Society* 23:363–382.
- Phillips, G., A. Lyche-Solheim, B. Skjelbred, U. Mischke, S. Drakare, G. Free, M. Järvinen, C. de Hoyos, G. Morabito, S. Poikane, and L. Carvalho. 2013. A phytoplankton trophic index to assess the status of lakes for the Water Framework Directive. *Hydrobiologia* 704:75–95.
- Rask, M., K.-M. Vuori, H. Hämäläinen, M. Järvinen, S. Hellsten, H. Mykrä, L. Arvola, J. Ruuhijärvi, J. Jyväsjärvi, I. Kolari, M. Olin, E. Salonen, and P. Valkeajärvi. 2011. Ecological classification of large lakes in Finland: comparison of classification

- approaches using multiple quality elements. *Hydrobiologia* 660:37–47.
- Raunio, J., T. Ihaksi, A. Haapala, and T. Muotka. 2007. Within- and among-lake variation in benthic macroinvertebrate communities: comparison of profundal grab sampling and the chironomid pupal exuvial technique. *Journal of the North American Benthological Society* 26:708–718.
- Raunio, J., L. Paasivirta, and H. Hämäläinen. 2010. Assessing lake trophic status using spring-emerging chironomid pupal exuviae. *Fundamental and Applied Limnology* 176:61–71.
- Rossaro, B., A. Boggero, B. Lods-Crozet, G. Free, V. Lencioni, L. Marziali, and G. Wolfram. 2012. A benthic quality index for European alpine lakes. *Proceedings of the 18th International Symposium on Chironomidae. Fauna Norvegica* 31:95–107.
- Rossaro, B., L. Marziali, A.-C. Cardoso, A. Solimini, G. Free, and R. Giacchini. 2007. A biotic index using benthic macroinvertebrates for Italian lakes. *Ecological Indicators* 7:412–429.
- Smith, R. W., M. Bergen, S. B. Weisberg, D. Cadien, A. Dalkey, D. Montagne, J. K. Stull, and R. G. Velarde. 2001. Benthic response index for assessing infaunal communities on the Southern California mainland shelf. *Ecological Applications* 11:1073–1087.
- Stoddard, J., D. Larsen, C. Hawkins, R. K. Johnson, and R. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16:1267–1276.
- ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. *Advances in Ecological Research* 4:271–317.
- Thienemann, A. 1922. Die beiden Chironomusarten der Tiefenfauna der norddeutschen Seen. Ein hydrobiologisches Problem. *Archiv für Hydrobiologie* 13:609–646.
- Vejjola, H., J. J. Meriläinen, and V. Marttila. 1996. Sample size in the monitoring of benthic macrofauna in the profundal of lakes: evaluation of the precision of estimates. *Hydrobiologia* 322:301–315.
- Verbruggen, F., O. Heiri, J. J. Meriläinen, and A. F. Lotter. 2011. Subfossil chironomid assemblages in deep, stratified European lakes: relationships with temperature, trophic state and oxygen. *Freshwater Biology* 56:407–423.
- Wallach, D., and B. Goffinet. 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling* 44:299–306.
- Wiederholm, T. 1980. Use of benthos in lake monitoring. *Journal of the Water Pollution Control Federation* 52:537–547.