Detecting Invertebrate Species Change in Running Waters: An Approach Based on the Sufficient Sample Size Principle

Terje Bongard[1],*, Ola H. Diserud[1], Odd Terje Sandlund[1] and Kaare Aagaard[2]

[1]*Norwegian Institute for Nature Research, P.O. Box 5685, 7485 Trondheim, Norway*
[2]*Museum of Natural History and Archaeology, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

Abstract: The EU Water Framework Directive (WFD) is in need of a practical method and index suitable for comparing the ecological status of invertebrate fauna in rivers throughout Europe. The uncertainties that characterize lotic sampling methods can be omitted by increasing sampling effort up to a species detection probability threshold. We describe a statistical model which provides the background for applying a sampling protocol aiming to record the total number of common and abundant species. Methods, sample standards and sorting time become less significant to the result as long as the total sample is sufficiently large to detect common species. An index called the Intercalibrated Benthic Invertebrate Biodiversity Index – IBIBI – is proposed based on Observed / Expected species ratios for EPT groups. The Ecological Quality Ratio thus obtained may then be adjusted to WFD's ecological status scale within each region or river type across Europe, which then can be compared and intercalibrated. Although offering a possible solution to the problem of intercalibration, further studies are needed.

INTRODUCTION

Functional freshwater ecosystems are essential parts of human life in many ways which need to be more recognized [1-3]. Benthic invertebrates are one of the biological quality elements used in assessing freshwater ecological status within the EU Water Framework Directive [4-6]. Several projects have, over the last decade, addressed issues associated with the practical implementation of the WFD (AQEM, STAR, REBECCA). Reviews and results of this work are published in Furse *et al.* [7]. For instance, the STAR project has addressed two main problems of comparison, not yet solved: a) The adoption of identical sampling techniques across Europe, and b) the harmonization of how the national assessment systems are classified [8-10]. The Geographical Intercalibration Groups (GIGs) aimed to intercalibrate national index data to compare regions and countries across Europe [10-15]. The complex problems of intercalibration, methodology and statistics are substantial, and the applicability and reliability of the results are still unclear [7, 8]. A useful index should be comparable both over time and space, and at the same time anchored to the local reference condition.

Here we propose an index that offers a solution to these problems. By increasing sample size and shift focus to species detection instead of abundances, the Observed - Expected ratio can be used to compare each locality to the

reference condition of its own region or river type [16, 17]. We present an index based on Ephemeroptera, Plecoptera and Trichoptera (EPT) that offers a feasible method of practical biomonitoring across regions in accordance with the five-scale ecological status classification of the WFD.

A river substrate is a micro-mosaic that complicates sampling on every level, and consequently renders all forms of investigations uncertain [18]. All known sampling techniques and methods have practical and statistical limitations. The two most commonly used methods for collecting invertebrates in rivers are the Surber and kick nets [19]. Both methods depend on variables such as sampling time, substrate type, clogging, water velocity, depth, and individual differences in sampling performance and effort. The Surber is limited to depths less than about 40 cm, and is highly susceptible to stony substrate. Traditionally, three to ten Surber samples have been collected from each locality. This corresponds to about 0.5-2 m^2 of substrate, depending on the size of the Surber sampler. Test results indicate that the number of samples should at least be doubled in order to obtain reliable data [19]. Even with large data sets and advanced statistical analyses, the results concerning species distributions and abundances are often dubious if each sample is too small [19-23]. Similarly, the quality of samples and the area sampled by the kick net method are not statistically reliable. Metrics and indices based on a standardized number of small samples from both these methods are therefore uncertain. These methodological concerns also became evident through a research and monitoring projects in Norway: The Atna River project has collected benthic data since 1986, and is one of Norway's most extensive national data sets on lotic invertebrates [24]. Atna is an undisturbed reference river

*Address correspondence to this author at the Norwegian Institute for Nature Research, P.O. Box 5885, 7485 Trondheim, Norway;
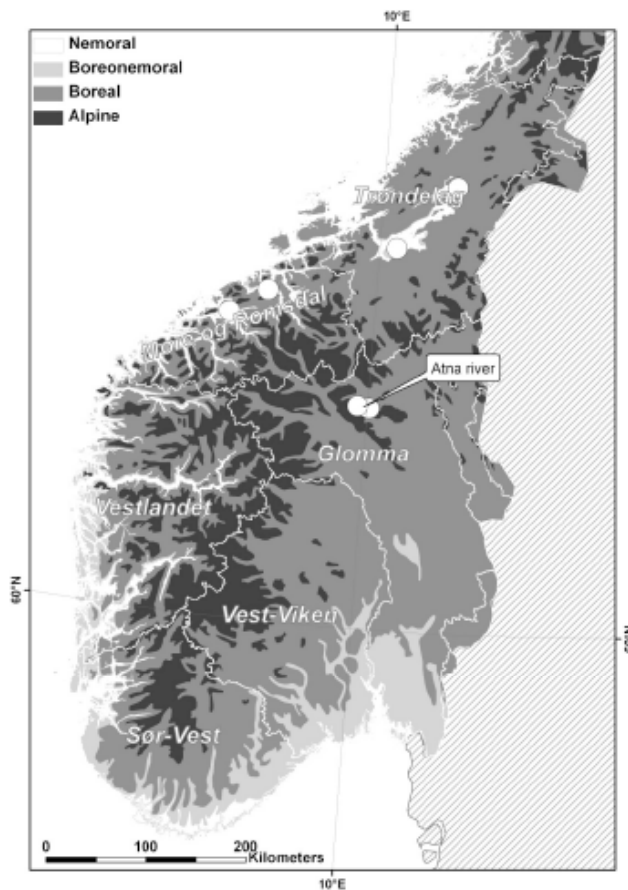Tel: +4798644786; Fax: +4773801401; E-mail: terje.bongard@nina.no

Fig. (1). The National River Basin Districts (RBD), and the main vegetation ecological regions in south Norway. The dominant mountain chains are parallel to the coast from south to north in Norway. The three main ecological zones; boreonemoral, boreal and arctic-alpine follow a similar pattern and are determined by both climate and altitude. We have applied the IBIBI to three localities in three different ecoregions or RBDs in Norway: Glomma RBD in the alpine and boreal zone, Trøndelag RBD in the boreal zone, and Møre and Romsdal RBD in the boreal zone. Localities where samples are taken are marked with circles (Map by Marc Daverdin).

situated in southern Norway (Fig. **1**). The method used for monitoring, mainly five Surber samples per location, was unsuitable for statistical documentation of either distribution or abundances, or time trends in the invertebrate community. Phenological variations from year to year are common for all species, and contribute to the uncertainties connected to the use of number of specimens as a reliable measure.

In order to obtain correct abundance numbers per area, or to detect species within acceptable fractions of a total, sampling effort must be increased beyond some threshold [21]. The problem of sample size is common in many forms of sampling from different environments and organism groups. A statistical model of this threshold is presented here.

The species distributions and abundances of Ephemeroptera, Plecoptera and Trichoptera (EPT) follow a general three-way pattern: Some species are always rare regardless of distribution pattern, some species may be locally common, but rare over large areas, and some are common over large areas and regions. A number of species are often com-
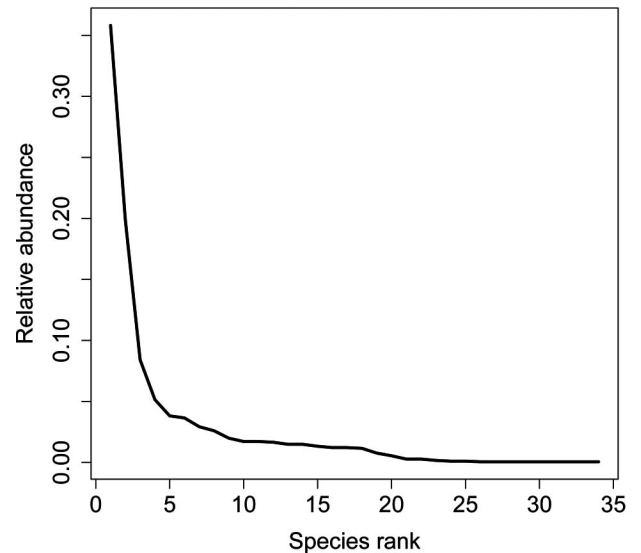

Fig. (2). The species rank – relative abundance relationship for our illustrative data set.

mon within a region even regardless of river type. On this basis, we suggest an index based on the EPT species predicted to be commonly present in pristine running waters of the same kind within a specific region, or in a specific river type within a region. Each locality may be given an Ecological Quality Ratio based on deviance from a predicted common species benchmark list from its own region (EQR is also called the Observed/Expected ratio). The EQR fraction for each locality may then be calibrated to the WFD classification system. The ecological status of every locality in Europe can, in this way, be characterized by the WFD scale. The intercalibration will consequently be a comparison of each locality's deviance from pristine conditions within the specific locality's own region. Species predictions may be based on species distributions from regions, or in different river types within regions (classified by variables like size, altitude etc). The topography, environment or size of these regions may of course vary considerably.

We call this index approach the Intercalibrated Benthic Invertebrate Biodiversity Index (IBIBI). The index is based on two pillars: A sufficiently large sample size and predictions of species presence.

MATERIALS AND METHODOLOGY

Statistical Basis for a Sufficient Sample Size

Ninety-one kick samples from the Atna River are used to illustrate a model of the statistics behind our index. The accumulated rank - relative abundance relationship is presented in Fig. (**2**). This community has 34 observed species where the most common species constitutes 35.8 % of the individuals in the population sample, the second most common species constitutes 19.9 % of the individuals etc., down to the 11 rarest species each of which constitutes less than 0.1 % of the individuals in the population sample. We define "common species" from this community as the species predicted to be present in any appropriately sized sample from an undisturbed river in this region. We further assume that the
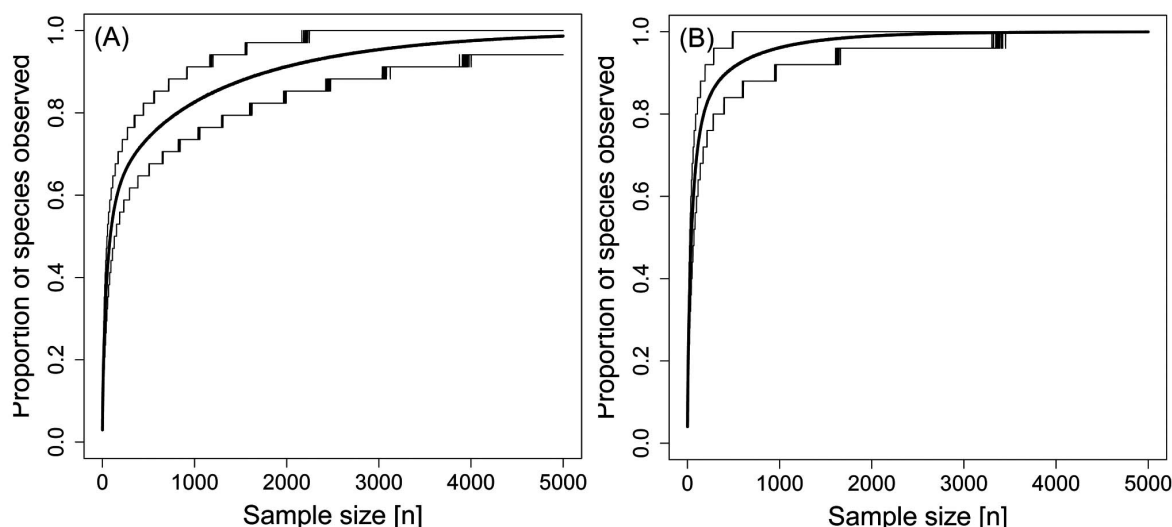
Fig. (3). The expected proportion of species observed as a function of sample size for all 34 species (**A**) and for the 25 most common species only (**B**), simulated from the relative species abundance distribution (solid lines). The step lines indicate 90% confidence belts, based on 10 000 simulated samples.

sampling can be described by the Poisson distribution, that is, so-called Poisson sampling [25]. If a species with abundance x is represented by X individuals in the sample, then conditionally on, x, X is Poisson distributed with expectation vx, where v is a measure of the sampling intensity, or the proportion of the total community being sampled.

This implies that each individual in the community has the same probability of being observed, regardless of species. We can then simulate the sampling procedure by assigning individuals to species multinomially from the relative species abundance distribution. Each different species abundance distribution will generate new species accumulation curves, and curves for sufficient sample size determination. A species' relative abundance is then defined as the probability of observing an individual of that species.

This simulation enables us to illustrate how the expected proportion of species observed will increase with the sample size for this community (solid line in Fig. **3A**). The dashed lines in Fig. (**3A**) are the 90 % confidence belts based on 10000 simulated samples, illustrating the effect of sampling uncertainty on the proportion of species observed. For a sample size of 1000 individuals we expect to observe 82.5 % of the species in our community. The sample size has to be close to 3000 individuals before we expect to observe more than 95 % of the species in this community. This number corresponds to about eight to ten minutes of kick sampling in an undisturbed location in the Atna area (i.e., the southeast boreal highland of Norway). In boreal central Norway, 250 kilometers north of Atna, a smaller sample of six to eight minutes is sufficient in order to obtain 95 % of the species (Table **1**).

This model indicates that insufficient sampling effort, for example five Surber samples or a three-minute kick sample, will provide uncertain information on the presence or absence of both common and uncommon species. The sample size must be very large before observations regarding the rarest species can be trusted as more than sampling randomness. In order to construct a robust and practical index, we therefore ignore the rarest species and focus our attention on

detecting the species predicted to be relatively common in an undisturbed community. In the IBIBI context, the term "relatively common species" refers to a species with probability larger than 0.90 for being observed in a sample of 3000 individuals in this example from Atna, corresponding to a relative species abundance larger than 0.0008. All species up to this level are omitted from calculation of the IBIBI idea, we only consider the 25 most common species and simulate samples multinomially from the truncated relative abundance distribution (Fig. **3B**). The proportion of predicted species observed will now have a steeper increase with sample size, the IBIBI will have a smaller sampling uncertainty since we are omitting the rarer species, and identification of individuals will be simpler for the field worker since recognition of every rare species is not required. This results in a more robust index, both with regard to sampling uncertainty and possible identification error.

In order to evaluate how well the IBIBI separates between an undisturbed community and one that has been affected by a disturbance, we need to know how the disturbance has altered the community composition, i.e. the species abundance and distribution for a given sampling intensity. Qu *et al.* [26] concluded that the response to disturbance for benthic invertebrate communities in streams was that the slopes of rank abundances became characteristically steeper, with decreasing species richness at the more polluted sites. To illustrate our approach, we therefore assume that the slope of the rank abundances for the undisturbed site becomes steeper after some disturbance. For example, assume that for an intermediate disturbance the abundance of species i is reduced by a factor so that the abundance of the most common species is only reduced by a factor 1/34, whereas the lowest ranked species (number 34) becomes extinct (reduced by a factor 34/34). We then obtain the rank – ln (abundance) relationship as given in Fig. (**4**) (dashed line), with the reference, the undisturbed rank – ln (abundance) relationship, as the solid line. The relative abundances of the more common species will thereby be increased, while the relative abundances of the less common species will be reduced, i.e. a trend towards dominance.
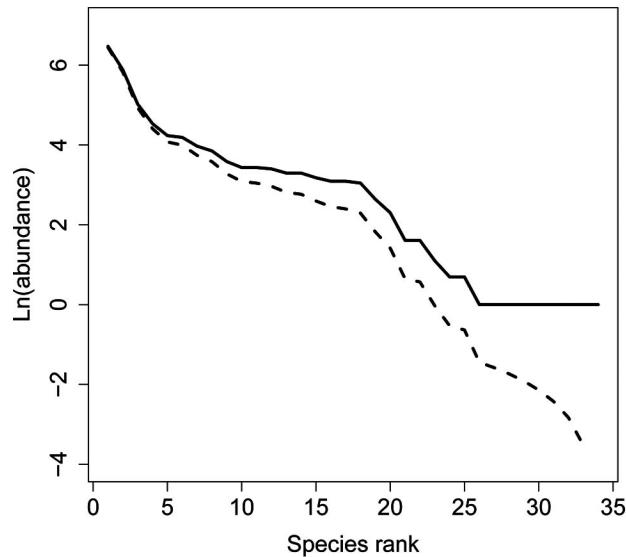
Fig. (4). The species rank – ln (abundance) relationship for the undisturbed community (solid line) and the disturbed community (dashed line).
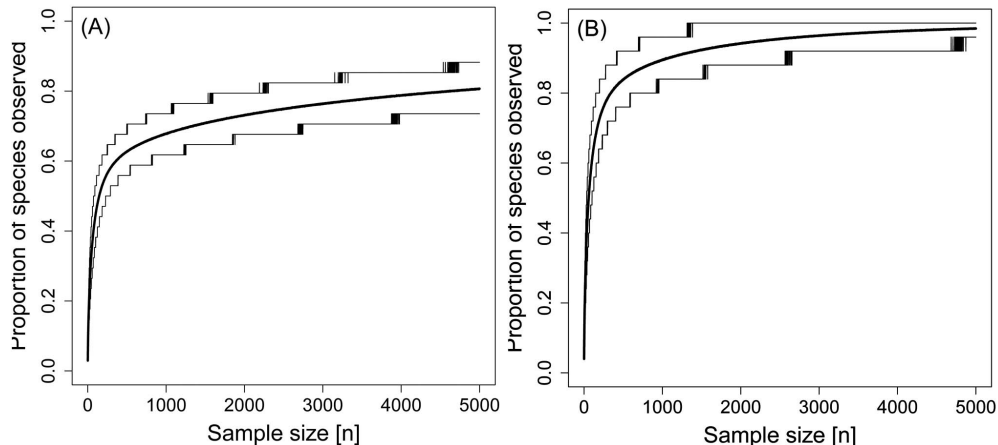
sample to the correct category. Therefore, we will need to include more of the less common species, e.g. those more affected by the disturbance, in the IBIBI. Increasing the list of expected species to 30 results in the curves shown in Fig. (**6B**). When we sample more than 1000 individuals, we have hardly any overlap between the confidence belts and can be fairly certain of classifying a sample in the right category; i.e., either the reference or the disturbed locations as defined by Fig. (**4**). All calculations and simulations were performed with the free statistical software R [27].

This gives a statistical background for applying a sampling protocol which aims to focus on the total number of common species. Methods, sample standards and sorting time become less significant to the result as long as the total sample is large enough and a sufficient number of individuals are sorted to species.

RESULTS

Sampling Protocol

The statistical model described here indicates that a more extensive sampling of invertebrate communities in running

Fig. (5). The proportion of species observed as a function of sample size for all 34 species (**A**) and for the 25 most common species only (**B**), simulated from the relative species abundance distribution for the disturbed community (solid lines). The thin lines indicate the 90% confidence belts, based on 10 000 simulated samples.

From the relative species abundance distribution for the disturbed community, we can simulate new samples to see how the proportion of expected species (all 34) now increases with sample size (Fig. **5A**). The less common species will become more difficult to observe, and we would need a very large sample ($n > 4300$) before we could expect to record more than 27 species. We will never observe all species since the rarest species is assumed to have gone locally extinct. If we now apply the IBIBI approach and consider the 25 most common species from the disturbed community, we obtain the curve in Fig. (**5B**).

Comparing the curves from Figs. (**3B** and **5B**) (merged in Fig. **6A**), we see a large overlap in the confidence belts, indicating that the sampling uncertainty will still make it hard to separate between this disturbed community (dashed lines) and the undisturbed reference condition (solid lines). The rarest of the selected 25 species is still "too common", i.e. the reduction in relative abundance after the disturbance has not been sufficiently large to ensure that we can classify the

water is necessary to provide satisfactory species distribution and abundance data, especially in order to reveal smaller environmental perturbations. The most convenient sampling method is to use many kick net samples in order to overcome the limitations of sample size. Preferably, each sample may last from one to three minutes to avoid net clogging. The use of 500 µm mesh size gives less debris in the samples, less clogging of the net, and samples that are easier to sort. Smaller instars may be lost, but each locality should be sampled at least four times during the ice-free season, both to cover the different species phenologies and to detect species with larger instars present at the next sampling date. This yearly sampling procedure is not, per se, included in the statistical modelling. Standard equipment for sorting invertebrate samples may be used, like pipettes and pincers. The samples may be sorted live, or conserved and sorted under a microscope. The main point is that samples of a sufficient size should be sorted until new species become increasingly far between and species detection declines, i.e. the curve
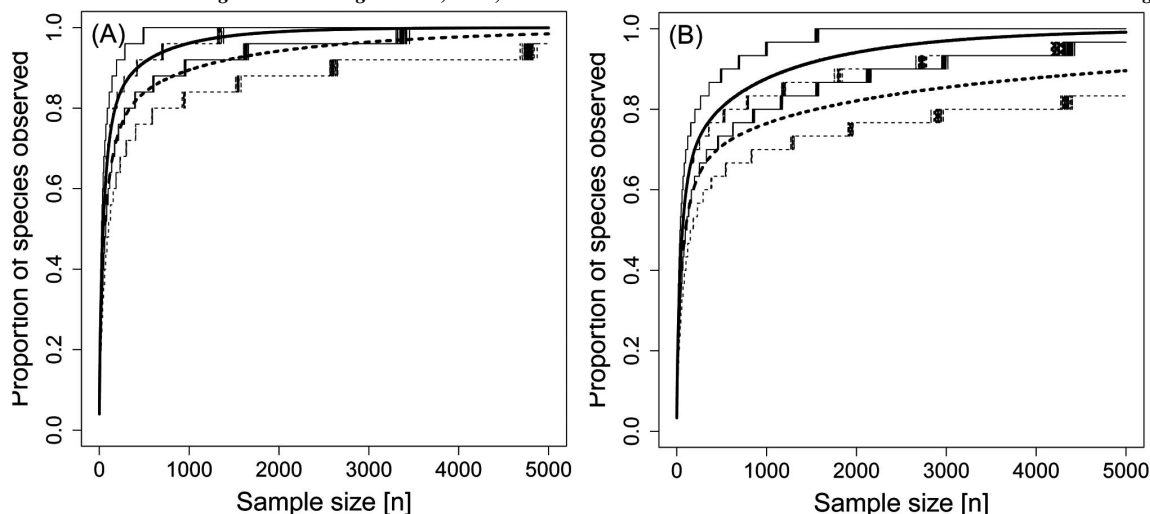
Fig. (6). The proportion of species observed as a function of sample size for the 25 most common species (**A**) and for the 30 most common species (**B**). The estimated mean and confidence belts for the simulations from the undisturbed community are given as solid lines (fat and thin lines, respectively) and results for the simulations from the disturbed community are given as dashed lines (fat and thin lines, respectively).

close in on the upper asymptote (cf. Figs. **3**, **5** and **6**). This leads to different sample sizes and different lengths of sorting for each locality. Sampling time may vary substantially according to the substrate, amount of organic debris, and diversity present. The statistical model indicate that sampling methods and their weaknesses become less important, as long as the sample size from a locality is larger than the threshold of statistic uncertainty for detecting predicted species (cf. Figs. **3**, **5** and **6**). As an illustration; if a predicted species has a relative abundance of 0.001, it will require a sample size of around 3000 to detect this species with a probability greater than 0.95.

By using this method over the years, we have therefore experienced that required sampling times were of different duration in surveys from the example regions from Norway, shown in Table **1**. Different regions or river types have different abundances, leading to different sample sizes required in order to detect predicted species. The statistical detection model outlined above shows that standardization of sample size is not necessary for a locality. The standard is in itself constituted by a sufficiently large sample followed by a sufficient effort invested in sorting. To secure that species detection has reached the upper, right part of the curve, we propose sorting for about half an hour beyond the last registration of a predicted species (see IBIBI index examples of predictions below).

Some specimens of each species and groups are collected during sorting, and preserved in 96 % ethanol for identification in the laboratory. Although conservation followed by laboratory sorting is preferable, field sorting substantially lowers project costs. Samples may preferably be taken indoors and sorted. A lamp with magnifying glass has proven useful for sorting. The procedure requires attention to new species, morphologically diverging specimens and new groups in the samples in order to register as many taxonomic groups as possible. Consequently, genera and families normally represented by sibling species which are difficult to distinguish in the field should be sampled more extensively for later identification. One example is the family Baetidae

(Ephemeroptera). In Norway, a total of 14 species of this family have been recorded and about half of them are rather similar in morphology and behavior. Therefore, a considerable number of specimens from this family should be sampled for later identification in the laboratory.

The abundance of each species, genera or group is subsampled from each tray, using a black grid marked on the bottom of the tray. Abundances are recorded as numbers per one-minute sample for each species or group. Abundance data are not crucial for the IBIBI, but might be valuable in detecting impacts through discrepancies from expected numbers [28]. For example, in central Norway, undisturbed rivers normally have about 200-400 *Baetis rhodani* per 1-minute kick sample. This particular species is known to be sensitive to acidic conditions, and can be almost or completely absent in rivers with a pH below about 5,5 [29]. On the other hand, under eutrophic conditions, abundances of *B. rhodani* may reach thousands in a one-minute sample. Such conditions might require a more extensive survey, including water chemistry, in order to reveal possible alterations from a pristine state. In a typical sample, the common species are usually easily identified, while a relatively small number of species occupy a disproportionate amount of a researcher's time [30].

We have focused on the EPT species for the index, but in principle, any invertebrate group may be used. A species list collected with statistical significance is nevertheless the ultimate basis to reveal environmental impacts.

Sorting is performed beyond a level at which new species become rare, and the upper right of Figs. (**2** and **5**) is reached. The required sampling effort, both in terms of number of individual kick net samples and sorting time, is therefore adjusted to each locality. For example, a locality in boreal central Norway normally requires about three to five kick net samples from one to three minutes in duration, followed by about one to three hours of picking. Inter-annual species variability is less important, as we have focused on the common species, predicted to be in detectable numbers every year. This EPT species list forms the basis for

Table 1. IBIBI Benchmarks: Common Species Predicted to be Present in all Medium Sized, Oligotrophic, Stony Substrate Pristine Rivers in three Ecoregions in Norway (See Map in Fig. 1). The Predicted Lists are Based on Four Sampling Times, from Spring to Autumn. Suggested Kick Sampling Times are Given in Parentheses. Expected Abundances are Indicated for each Species in a One-Minute Kick Sample: * Usually Few (Below 10); ** Usually Common (10 to 50); *** Often Dominant in Numbers (More Than 50)

Species	Boreal Central (8 Minutes) (26 Species)	Southeast Boreal Highland (10 Minutes) (19 Species)	Northwest (8 Minutes) (16 Species)
Ephemeroptera:			
Baetis muticus	**	**	
B. rhodani	***	***	***
B. scambus	**	**	
Heptagenia spp.	**	*	*
Ameletus inopinatus	**		
Ephemerella aroni	**	*	*
E. mucronata	*		
Plecoptera:			
Diura nanseni	**	*	*
Amphinemura spp.	**	**	*
Nemoura cinerea	*	*	
Leuctra spp.	**	**	**
Taeniopteryx nebulosa	*/**	*	*
Isoperla obscura	*/**		
Isoperla grammatica	*/**		
Capnia spp.	**	**	*
Siphonoperla burmeisteri	*/**		*
Brachyptera risi	*/**		*
Protonemura meyeri	*/**		
Trichoptera:			
Rhyacophila nubila	*	*	*
Hydroptila spp.		*	
Polycentropus flavomaculatus	**	*	*
Apatania spp.	*	*	*
Potamophylax spp.	*		
Arctopsyche ladogensis	*	*	
Sericostoma personatum	*	*	*
Lepidostoma hirtum	*	*	*
Hydropsyche spp.	*/**	*	*

Table 2. IBIBI classification of Ecological Status for Running Water in Three Ecoregions in Norway Based on Number of Predicted EPT Species Present in Pristine Conditions (Ref. Table 1). Suggested Scaling is Based on Expert Opinion in Accordance with the WFD System for ECOLOGICAL status

Ecological Status	Boreal Central (26 Species)	Southeast Boreal Highland (19 Species)	Northwest (16 Species)
High (Abundances as expected, reference)	23 or more	16 or more	14 or more
Good (Slight alterations in abundances might be found)	18-22	12-15	10-13
Moderate (Alterations in abundances found)	13-17	8-11	7-9
Poor (Significant alterations in abundances found)	8-12	5-7	4-6
Bad (Large alterations in abundances found)	7 or less	4 or less	3 or less

Table 3. Example of IBIBI use on Data Sets from the Southeast Boreal Highland Norway, Locality Solbakken, Atna, 2005 (61^0 45' N, 10^0 45' E). Numbers are Calculated from Subsampling of 8-12 Minutes Kick Samples. Ecological Status is Characterised from Table 2

DATE	04.07.2005	25.07.2005	18.08.2005	20.10.2005
Lymnaea peregra	4	3	5	
Oligochaeta	2	3	2	1
Acari	10	10	10	5
Ephemeroptera				
Baetis muticus	15			
Baetis rhodani	135		45	200
Baetis scambus	25	3	10	
Baetis subalpinus	10	120	10	
Heptagenia dalecarlica	5		5	5
Afghanurus joernensis	3	15	5	
Ephemerella aroni	8	25	10	10
Plecoptera				
Diura nanseni				5
Taeniopteryx nebulosa				1
Amphinemura borealis	I			
Capnia atra				1
Leuctra digitata		1	I	I
Leuctra fusca		2	2	
Elmidae				
Elmis aenea	2	2	2	2
Trichoptera				
Rhyacophila nubila	5	4	3	5
Hydroptila spp.			5	5
Polycentropus flavomaculatus	3	1	2	
Hydropsyche nevae			3	
Hydropsyche pellucidula	3	2	5	2
Arctopsyche ladogensis		5		3
Micrasema sp.		1	2	3
Lepidostoma hirtum	5	5	8	
Apatania stigmatella	1		2	
Sericostoma personatum	3		3	

Table 3. cont…..

DATE	04.07.2005	25.07.2005	18.08.2005	20.10.2005
Athripsodes sp.	1	1	1	
Diptera	2			2
Tipulidae	2	2		1
Simulidae	50	10	20	2
Chironomidae	30	10	15	10
Number per 1-minute sample	323	225	172	263
EPT in total	15	13	18	12
EPT from Table 1	12	9	12	11
EPT for indexing	18			
Ecological status IBIBI (from Table 2)	HIGH (16 or more)			

the index proposal. The IBIBI considers the presence or absence of a set of predicted common species. A common species still missing after extensive sampling and sorting are likely to indicate environmental perturbation of some kind.

Examples of IBIBI Index Regions

Kick net and Surber samples from four decades are registered in the Norwegian Vanninfo database, consisting of about 3500 invertebrate samples [31]. Based on these data, we have applied the IBIBI approach to three localities in three different ecological regions or National River Basin Districts (RBDs) in Norway: Glomma RBD in the alpine and boreal zone, and Møre and Romsdal RBD and Trøndelag RBD in the boreal zone. The locality examples are: From Glomma RBD; the Atna River, from Møre and Romsdal RBD locality; the Batnfjord River, and Trøndelag RBD localities are from the Trondheim and Steinkjer municipalities (Fig. **1**). The most commonly distributed EPT species expected to be found in all pristine rivers of medium size, oligotrophic and with a stony substrate in these regions are presented in Table **1**. About 300-500 specimens are predicted to be present in a one-minute kick sample from boreal central Norway. Hence, the expected abundance categories in Table **1** can be translated into three relative abundance classes. A species normally represented by few individuals in a sample is assigned one asterisk (<10), giving an approximate relative species abundance of 0.01 when using intermediate range values (5 out of 400 ~ 0.01). Species expected to be present at 10 to 50 individuals per minute kick sample are given two asterisks. Finally, the dominant species with normally more than 50 individuals per minute are given three asterisks and relative species abundances 0.20 (~ 80/400). From this relative species abundance distribution we can apply the same approach as described above.

We suggest, based on our own experience, expert opinion boundaries from "High ecological status" to "Bad ecological status" for three regions in Norway, see Table **2**. In boreal central Norway, 26 species are so common that they should be present in all undisturbed localities (Table **1**). We suggest that the limit for "High" should be 23 species (Table **2**). Between 18 and 22 recorded species corresponds to "Good", and between 13 and 17 species, corresponds to "Moderate".

With less than 50% of expected species present, the ecological status levels are "Poor" and "Bad". The expected species numbers for southeast boreal highland Norway to northwest Norway are similarly adapted to the WFD scale (Table **2**). The northwest Norway ecoregion has substantially lower species diversity than boreal central Norway, some 200 km to the east. Conspicuous alterations from expected species abundances should be taken into the final consideration, particularly if results are intermediate between classes. The boundaries we suggest have not yet been discussed with other experts in Norway, but they can easily be changed on the background of better knowledge and experience, also as experience is gained through time. Species data already collected will not lose its value by these changes.

One can also adapt the IBIBI to only one sampling per year by lowering the benchmark, i.e. by adjusting expected species number to sampling date. Table **3** shows, as an example, the IBIBI applied to data from one southeast boreal highland locality Solbakken, in the Atna river.

DISCUSSION

Macroinvertebrates like EPT species are widely used in monitoring water quality as they exhibit a relatively wide range of responses to chemical and physical water quality stress [32, 33]. EPT species usually have yearlong life cycles, which make them suitable for revealing different kinds of long term impacts. Biotic indices are in general designed to measure how an ecosystem deviates from a pristine state. An index should also ideally define the size of perturbations. The effects of acidity, inorganic or organic pollution, or hydromorphological changes on biodiversity should be described by the index, thereby making it a tool to operationalise the WFD scale.

Different macroinvertebrate taxa have different habitat preferences and various levels of perturbation tolerance. These qualities are used in many indices, included those which are preliminary selected as standards also for Norway: The Biological Monitoring Working Party (BMWP) score, and the derived average score per taxon (ASPT) [34-36]. These indices are mostly based on the presence/absence and/or range values of selected taxa with specific ecological demands. These indices lack universal application as they are limited by the distribution of the species that constitute and

ultimately define the index benchmark. In Norway, as in Scandinavia and other parts of Europe, the selection of species should be different than in Great Britain. The problems of comparison and intercalibration soon become evident. We therefore propose an index based on an EQR comparison with the high ecological status ("reference condition"), documented or expected, for each area or region.

The IBIBI given in Table **1** form the reference benchmark for medium rivers with stony substrate in three different ecoregions. In practice, the benchmark EPT species lists will be a trade-off between variables like ecoregion size, river types, altitude and geography. In Table **2** the levels of deviation from the reference condition leading to the various ecological status classes of the WFD are suggested. The boundaries between quality classes, i.e. the number of species deviating from expected pristine conditions, are based on the WFD description. The terms "High", "Good", "Moderate", "Poor" and "Bad" represent qualitative values which are difficult to define exactly by numbers. The WFD status classes are not exact descriptions of nature, but rather quality assessments of different EQRs. Extensive information on how different degrees of perturbations affect the various species is required if the goal is to set numerical boundaries. Table **2** is therefore based on expert opinion. Increasing knowledge from sampling of disturbed localities can be used to refine and calibrate the boundaries between the ecological status classes.

Bonada *et al.* [21] identify twelve criteria for an ideal biomonitoring index for running water invertebrates. We here briefly discuss each criterion in relation to the statistical model, the sampling protocol and the IBIBI index we propose:

Derived from Sound Theoretical Concepts in Biology

Our proposal is a very simple EQR (O/E) approach that is, in all aspects, firmly based in the concepts of biology. Species occurrences and abundances are the ultimate measures of biodiversity, and also the ultimate measures of any impact on an ecosystem. The simplicity strengthens the IBIBI approach, inasmuch as it is based on species presence or absence, collected in a way that reduces the significance of methodological problems.

A Priori Predictive

A species list is the basis for assessing community tolerance for all forms of perturbations. The calibration of the IBIBI to the five status classes (i.e. the number of species determining each status class) may change with time, but the data quality will not. A similar approach has been proposed by botanists by looking at species subsets as indicators of ecological status [37]. However, plant species are more patchily distributed over larger areas. Freshwater invertebrates normally have more continuous species distributions, and are therefore more a priori predictive. Consequently, they are more suitable for EQR approaches.

Potential of Assessing Biological Functions

Biological functions are linked to species. Hence, species detection will constitute the basis of assessing biological functions. Functional groups are likewise constituted by species. Resh and Rosenberg [18] have stated that species is the most reliable level of taxonomy when it comes to indicator organisms.

Potential to Discriminate Overall Human Impact (i.e., to Identify Anthropogenic Disturbance)

Perturbations of ecosystems are generally the outcome of human activities. Ecosystem deviations resulting from other reasons, like nature disasters, are rare exceptions. These are normally easily identified as such.

Potential to Discriminate Different Types of Human Impact (i.e., to Identify Specific Types of Anthropogenic Disturbance)

All forms of influences on an ecosystem alter the distributions and abundances of species [38]. Each species might react differently to specific impacts, for example the response of *Baetis* spp. to acidity. Substantial deviances of species composition and abundances over time are rarely registered in the absence of any human impact. Regardless of perturbation or form of pollution, whether inorganic or organic, chemical or biological, deviation from expected biodiversity is the most unambiguous and ultimate measure of impact. Once any impact is established, the next step will be to analyse specific species reactions, and to which forms of impact and concentration each species responds to [39-41].

Low Costs for Sampling and Sorting (Field Approaches) or for Standardized Experimentation (Laboratory Approaches)

Hypothetically, it is possible to sample all species present, including the rare ones, in any specific location [37]. However, for all practical purposes, surveys have to strike a balance between available resources for sampling and the need for reliable knowledge of the species present at the location. The sampling protocol we propose substantially lower laboratory time and costs, and trade the quest for accurate numbers with an increase in field effort to register species. This trade-off provides data that meet the requirements for an index that may operationalise the WFD scale within present budget levels.

Simple Sampling Protocol

The strength of the IBIBI is the simplicity of both the sampling and the analysis of the results. It can be used for practical biomonitoring within acceptable cost limits. By avoiding sorting and counting of large numbers of common species, resources are, as a consequence, redirected to red-listed species. Bonada *et al.* [21] conclude (quote): *"....a tool that could use a simple sampling protocol (i.e. least as possible standardized techniques, such as 10 min of random kick sampling, collected once at any time of the year with a 0.5 mm mesh-sized net, Criterion 7) would be advantageous for practical reasons in routine biomonitoring programs".* We believe that once a year is not enough to detect species composition, and suggest four samples per season, but a project's economy often allows for only one sampling per year. The index can, as mentioned, be calibrated to one sampling per year, by reducing the predicted species list and recalibrate the WFD ecological classification scale.

Low Cost for Taxa Identifications (no Specialists in Taxonomy Required)

The most common species are normally the simplest to identify, while the rarer and more difficult species may be left to the experts. Local staff may do the sampling and picking and send the more problematic specimens to an expert laboratory for identification.

Large-Scale Applicability (Across Ecoregions or Biogeographic Provinces)

The IBIBI offers a solution to the problem of intercalibration. National experts can produce preliminary prediction lists of common species for their region and river types. Ecological reference conditions are, for many areas in Europe, lost forever. Sufficient species information might nevertheless be extrapolated from adjacent regions and historical material. Any changes in knowledge of species responses to perturbations will easily lead to re-evaluation of the WFD class boundaries. Shifts in abundances of species are common, but the data quality from a statistically sound sampling procedure will, nevertheless, not be lost.

Reliable Indication of Changes in Overall Human Impact

The robustness of the species registration, which is an effect of the sampling protocol, provides an ultimate and reliable indication of any deviances from a pristine ecosystem state.

Reliable Indication of Changes in Different Types of Human Impact

Different impacts affect different species in different ways. This knowledge can be used to describe and address different forms of influences on an ecosystem. For instance, the family Baetidae is sensitive to acidification. Another example from Norway is the fact that many rivers and streams are oligotrophic, and organic pollution often increase both species distributions and abundances. Both these examples of different impacts will be registered by our approach.

Human Impact Indication on a Linear Scale

Statistically valid samples are necessary to provide comparable data for the future. The species lists with relative abundances collected with excess sampling and sorting will not lose their value over time. By increasing the knowledge of each species and its preferences, historical data of species deviances over the years sampled with the IBIBI method will constitute a background for evaluating human impact on a linear scale over time.

Extreme localities to which species predictions do not apply are occasionally found. They could be identified as deviations from the expected values and described as such. WFD Annex II, no. 1.4 lists a number of situations that an index should be able to intercept, like point pollution, diffuse pollution, regulations, and drainage and morphological disturbances. These are all forms of stress that alter species composition, and they will therefore be uncovered by the IBIBI approach.

An argument for why some species are common might be that they are the ones least susceptible to various forms of environmental impact. Consequently, they are only expected to disappear in response to stronger environmental impacts. We believe this argument is overemphasized. For example, the most common mayfly in Norway, *Baetis rhodani*, is one of the species most sensitive to acid rain. In fact, of the 27 species suggested in our index, 14 are considered to be sensitive to acidity, according to Raddum's acidity index [29]. Similar differences in sensitivity to various impacts might be present for each species. EPT species in general are sensitive and delicate organisms that react quickly to environmental alterations. A minority of EPT species are tolerant to environmental impacts, regardless of commonness. Regardless of differences in species sensitivity, a general decline in biodiversity is the ultimate and most robust measure of any form of perturbation.

Comparisons between undisturbed, intermediate and heavily polluted sites can be used to tune the scale in accordance with the WFD definitions. With increasing species knowledge, species prediction lists could easily include more of the rarer species. It is possible to merge ecoregions that have a similar fauna in order to establish shorter species lists valid for larger areas. With relatively little effort, an initial version of ecological reference conditions for every region in Norway can be presented. We believe that this also is possible for most European countries.

CONCLUSION

The relationship between sample size and the number of species recorded is a well known statistical principle in biology. Although further studies are needed, comparing each area, region or ecotone to its own biodiversity benchmark, offers an interesting approach to intercalibration. The intercalibration problems within the WFD could find a solution along these lines. An increase in field sampling in order to close in on the threshold of finding new species might be a way of comparing localities across Europe. A more accurate species number is after all the ultimate measure of biodiversity. We propose a simple index based on a benchmark of common species expected to be present in a location. The EQR (Observed/Expected ratio) may easily be calibrated to the WFD ecological status scale. Each region, ecotone or river type may establish its own scale that actually can be intercalibrated with every other ecosystem. By the very fact that the species data from each locality is unambiguous, the WFD index scale descriptions (high-good-moderate-poor-bad) may continually be adjusted for each region as knowledge of expected species distributions and abundances increase.

REFERENCES

[1] Arthington AH, Naiman RJ, McClain ME, Nilsson C. Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities. Freshw Biol 2010; 55(1): 1-16.

[2] MEA. Millennium Ecosystem Assessment Synthesis Report. Island Press: Washington DC 2005.

[3] Ormerod SJ, Dobson M, Hildrew AG, Townsend CR. Multiple stressors in freshwater ecosystems. Freshw Biol 2010; 55: 1-4.

[4]   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. Off J Eur Commun 2000; 327: 72.

[5]   Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Rivers and lakes - Typology, reference conditions and classification systems. 2003; 10: 94.

[6]   Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance on the intercalibration process 2004-2006. 2005; 14: 31.

[7]   Furse MT, Hering D, Brabec K, Buffagni A, Sandin L, Verdonschot PFM, Eds. The Ecological Status of European Rivers - Evaluation and Intercalibration of assessment methods. Developments in Hydrobiology Dordrecht, The Netherlands: Springer 2006.

[8]   Buffagni A, Furse M. Intercalibration and comparison - major results and conclusions from the STAR project. Hydrobiologia 2006; 566: 357-64.

[9]   Buffagni A, Erba S, Furse MT. A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. Environ Sci Policy 2007; 10(7-8): 709-24.

[10]   Erba S, Furse MT, Balestrini R, *et al*. The validation of common European class boundaries for river benthic macroinvertebrates to facilitate the intercalibration process of the Water Framework Directive. Hydrobiologia 2009; 633(1): 17-31.

[11]   Official Journal of the European Union Commission decision of 30 October 2008 establishing, pursuant to Directive 2000/60/ EC of the European Parliament and of the Council, the values of the Member State monitoring system classifications as a result of the intercalibration exercise. 2008; 332: 20-44.

[12]   McGarrigle ML, Lucey J. Intercalibration of ecological status of rivers in Ireland for the purpose of the Water Framework Directive. Biol Environ-Proc R Irish Acad 2009; 109B(3): 237-46.

[13]   Bennett C, Owen R, Birk S, *et al*. Bringing European river quality into line: an exercise to intercalibrate macro-invertebrate classification methods. Hydrobiol 2011; 667; 1: 31-48.

[14]   Murray-Bligh J, Buffagni A, Cazzola M, Birk S, Vlek H. Procedure for Calculating the ICMi Index: Step-By-Step Instructions For Central-Baltic river GIG 2006;Version 2.6.

[15]   Bonanno G, Lo Giudice R. Application of two quality indices as monitoring and management tools of rivers. case study: The Imera Meridionale River, Italy. Environ Manag 2010; 45(4): 856-67.

[16]   Hawkins CP, Norris RH, Hogue JN, Feminella JW. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol Appl 2000; 10(5): 1456-77.

[17]   Aroviita J, Mykra H, Hamalainen H. River bioassessment and the preservation of threatened species: Towards acceptable biological quality criteria. Ecol Indic 2010; 10(4): 789-95.

[18]   Resh VH, Rosenberg DM. Freshwater biomonitoring and benthic macroinvertebrates. New York: Chapman and Hall 1993; p. 488.

[19]   Stark JD. Performance of the Macroinvertebrate Community Index - Effects of Sampling Method, Sample Replication, Water Depth, Current Velocity, and Substratum on Index Values. NZ Mar Freshw Res 1993; 27(4): 463-78.

[20]   Doberstein CP, Karr JR, Conquest LL. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. Freshw Biol 2000; 44(2):355-71.

[21]   Bonada N, Prat N, Resh VH, Statzner B. Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. Ann Rev of Entomol 2006; 51: 495-523.

[22]   Diserud OH, Aagaard K. Testing for changes in community structure based on repeated sampling. Ecology 2002; 83(8): 2271-7.

[23]   Vallania A, Corigliano MD. The effect of regulation caused by a dam on the distribution of the functional feeding groups of the benthos in the sub basin of the Grande River (San Luis, Argentina). Environ Monit Assess 2007; 124(1-3): 201-9.

[24]   Aagaard K, Solem JO, Bongard T, Hanssen O. Studies of aquatic insects in the Atna river 1987-2002. In: Sandlund OT, Aagaard. K, Eds. The Atna river: Studies in an Alpine-Boreal Watershed. Hydrobiologia 2004; 521: 87-105.

[25]   Fisher RA, Corbet AS, Williams CB. The relation between the number of species and the number of individuals in a random sample from an animal population. J Anim Ecol 1943; 12: 42-58.

[26]   Qu XD, Song MY, Park YS, Oh YN, Chon TS. Species abundance patterns of benthic macroinvertebrate communities in polluted streams. Ann Limnol-Int J Lim 2008; 44(2): 119-33.

[27]   R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 2008.

[28]   Donohue I, McGarrigle ML, Mills P. Linking catchment characteristics and water chemistry with the ecological status of Irish rivers. Water Res 2006; 40(1): 91-8.

[29]   Raddum GG, Fjellheim A. Acid precipitation: Biological monitoring of streams and lakes. Sci Total Environ 1990; 96: 57-66.

[30]   Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. Phil Trans R Soc Lond B 1994; 345: 101-18.

[31]   Aagaard K, Dolmen D. Limnofauna Norvegica: Tapir forlag, 1996; p. 309.

[32]   Simon TP, Ed. Biological response signatures: indicator patterns using aquatic communities. Boca Raton, FL: CRC Press 2003; p. 576.

[33]   Karr JR, Chu EW. Restoring life in running waters: better biological monitoring. Washington: Island Press 1999; p.206.

[34]   Armitage PD, Moss, D.,Wright, J.F., Furse, M.T. The performance of a new biological water quality score system based on macroinvertebrates over av wide range of unpolluted running-water sites. Water Res 1983; 17(3): 333-47.

[35]   Walley WJ, Hawkes HA. A computer based reappraisal of the Biological Monitoring Working Party scores using data from the 1990 river quality survey of England and Wales. Water Res 1996; 30: 2086-94.

[36]   Walley WJ, Hawkes HA. A computer based development of the Biological Monitoring Working Party score system incorporating abundance ratio, site type and indicator value. Water Res 1997; 31: 201-10.

[37]   Vellend M, Lilley PL, Starzomski BM. Using subsets of species in biodiversity surveys. J App Ecol 2008; 45: 161-9.

[38]   Dobiesz NE, Hecky RE, Johnson TB, *et al*. Metrics of ecosystem status for large aquatic systems - A global comparison. J Great Lakes Res 2010; 36(1): 123-38.

[39]   den Brink PJ, Blake N, Brock TCM, Maltby L. Predictive value of species sensitivity distributions for effects of herbicides in freshwater ecosystems. Hum Ecol Risk Assess 2006; 12(4): 645-74.

[40]   Von Der Ohe PC, Liess M. Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. Environ Toxicol Chem 2004; 23(1): 150-6.

[41]   Rubach MN, Baird DJ, Van den Brink PJ. A new method for ranking mode-specific sensitivity of freshwater arthropods to insecticides and its relationship to biological traits. Environ Toxicol Chem 2010; 29(2): 476-87.