

Performance of the STAR_ICMi macroinvertebrate index and implications for classification and biomonitoring of rivers

Daniel Spitale^{1,2,*}

¹ BioMonitoring Team, Via Stenico 2, 38095 Tre Ville, Italy

² Museo delle Scienze – MUSE, Limnology and Phycology Section, Corso del Lavoro e della Scienza 3, 38123 Trento, Italy

Abstract – Although biomonitoring is the core approach adopted by the European Union's Water Framework Directive (WFD), many biotic indices still lack a thorough analysis of their performance and uncertainty. The multihabitat sampling and the application of STAR_ICMi index on macroinvertebrates are the standard methods to assess the ecological status of rivers in Italy. Ever since the Italians' implementation, dates back to 2010, few studies have tested the index performance with different sampling efforts, and even rarer are those assessing index uncertainty. However, these are worthwhile topics to investigate because all the Environmental Agencies are applying this index with both ecological and economic consequences. Aims of this study were (i) to assess the effect of subsampling on the STAR_ICMi index, (ii) to propose a standard method to calculate the index precision, and (iii) to test several less time-consuming alternatives to census all the individuals in the sample. I showed that the index is strongly affected by subsampling, and unbiased comparisons of ecological status can only be done at the same sampling effort. The index precision, calculated by bootstrapping the observed abundance of taxa, was so low in some circumstances, to increase the risk of misclassification. Finally, I showed that to avoid counting all the individuals in a sample, it is possible to estimate the most abundant taxa using a rank-abundance model. With this less time-consuming method, the STAR_ICMi index is predicted with sufficient precision.

Keywords: sampling effort / rarefaction / index uncertainty / sampling variation / Italy

Résumé – Performance de l'indice macroinvertébrés STAR_ICMi et implications pour la classification et la biosurveillance des cours d'eau. Bien que le biomonitoring soit l'approche de base adoptée par la directive-cadre sur l'eau de l'Union européenne, de nombreux indices biotiques manquent encore d'une analyse approfondie de leur performance et de leur incertitude. L'échantillonnage multihabitat et l'application de l'indice STAR_ICMi sur les macroinvertébrés sont les méthodes standard pour évaluer l'état écologique des rivières en Italie. Depuis leur mise en œuvre qui remonte à 2010, peu d'études ont testé la performance de l'indice avec différents efforts d'échantillonnage, et encore plus rares sont ceux qui évaluent l'incertitude de l'indice. Cependant, ce sont des sujets intéressants à étudier parce que toutes les agences environnementales appliquent cet indice avec des conséquences écologiques et économiques. Les objectifs de cette étude étaient les suivants : (i) évaluer l'effet du sous-échantillonnage sur l'indice STAR_ICMi, (ii) proposer une méthode standard pour calculer la précision de l'indice et (iii) tester plusieurs alternatives moins coûteuses au recensement de tous les individus dans l'échantillon. L'indice est fortement influencé par le sous-échantillonnage et des comparaisons impartiales de l'état écologique ne peuvent être faites qu'au même effort d'échantillonnage. La précision de l'indice, calculée par la méthode de bootstrapping sur l'abondance observée de taxons, était très faible dans certaines circonstances, augmentant le risque de classification erronée. Finalement, pour éviter de compter tous les individus d'un échantillon, il est possible d'estimer les taxons les plus abondants en utilisant un modèle de rang-abondance. Avec cette méthode moins longue, l'indice STAR_ICMi est estimé avec une précision suffisante.

Mots clés : effort d'échantillonnage / rarefaction / incertitude d'indice / variation d'échantillonnage / Italie

* Corresponding author: spitale@biomonitoraggi.it

1 Introduction

Biological monitoring, or biomonitoring, is the use of biota to track changes in the environment (Friberg *et al.*, 2011). In freshwater habitats, biomonitoring is a key tool used by managers and environmental agencies to support their decisions. In European Union, the Water Framework Directive (WFD 2000/60/EC) is the main normative in the field of water monitoring and conservation. The WFD introduced profound innovations, changing the perspective of monitoring from a mere detection of pollution to a determination of the ecological status comparing the observed status against the reference condition. Even though this approach is ecologically sound, as any measurement of the ecological status, it is associated with an unknown source of uncertainty that would hamper a correct decision-making process by water managers (Friberg, 2014).

The benthic macroinvertebrates are indicators used throughout the world for evaluating the river quality for their capacity to assess multiple anthropogenic pressures. Member States of EU employ different sampling methods to collect macroinvertebrates in rivers (*e.g.* fixed time kick sampling, Surber sampler on fixed or not area), and they record abundance either using real estimates or classes (Bennett *et al.*, 2011). Although assessed in the early stages of the intercalibration process among Member States, the use of fixed number of individuals to calculate macroinvertebrate index was not widely adopted (Clarke *et al.*, 2006a). However, the effect of using different sampling efforts (=different number of individuals) has long been recognised as a serious pitfall in comparing samples (Gotelli and Colwell, 2001).

In Italy, the ecological status of rivers is assessed using diatoms, macrophytes, fish and macroinvertebrates, each one with specific methods and procedures (ISPRA, 2014). Concerning macroinvertebrates, the sampling method relies on the multihabitat approach (Hering *et al.*, 2003), and the ecological status is calculated with the STAR_ICMi index as ratio with the reference condition (Buffagni *et al.*, 2006; Bennett *et al.*, 2011). The STAR_ICMi (for an historical account of the index see Bo *et al.*, 2016) is a multimetric index composed of 6 metrics thought to represent different aspects of the macroinvertebrate community. Even though the performance of this index has been extensively tested across Europe especially for harmonisation purpose (*e.g.* Bennett *et al.*, 2011), a wide knowledge gap still exists concerning the accuracy of the metrics (but see Clarke *et al.*, 2006b; Clarke, 2013; Ramos-Merchante and Prenda, 2017).

All the assessments of the ecological status of water bodies using macroinvertebrate sampling are subjected to uncertainty and errors due to a range of factors including sample processing (subsampling, unfixed number of individuals), misidentification of organisms, different trained staff and so on (Clarke *et al.*, 2006a). The sources of uncertainty, which could generate variability in the STAR_ICMi index, are diverse and not simple to disentangle. Restricting the analysis to the fine scale, namely that one considered for the bioassessment, three main groups of causes could be recognised: ecological, errors and sampling-related causes (Clarke *et al.*, 2006a, 2006b). Quantification of these uncertainties is necessary to assign a confidence level in the classification of water bodies and to prescribe the correct management options in the case of deterioration or amelioration. Wrong classification of the ecological status has relevant

consequences in terms of expensive remedial measures but also as missed economic benefits.

In Italy, like in many MS of EU, macroinvertebrate indices are calculated examining a variable number of individuals (Bennett *et al.*, 2011). However, the use of unfixed abundance hinders the correct comparison of samples because of the inconsistent sampling effort. We urgently need additional studies addressing the effect of this unfixed abundance on bioassessment. The ecological status of many rivers could be wrongly assigned only as a result of this source of error. Moreover, the lack of a measure of uncertainty in the STAR_ICMi index is a serious drawback, which should be overcome (Friberg, 2014). An estimate of uncertainty can help the Environmental Agencies in interpreting situations in which the ecological status of rivers falls near boundary classes. If on one hand a careful attention to these topics seems desirable, it should be admitted that the application of the STAR_ICMi index is more demanding in terms of time, in respect to early indices like the IBE (the Italian version of the Extended Biotic Index, Ghetti, 1997). Therefore, it seems useful to devise alternative methods ensuring rigorous assessments in shorter time.

Aim of the study was threefold. (1) To test if and to what extent subsampling has an effect on the STAR_ICMi index and classification of the water bodies; (2) to suggest a method able to provide a measure of index precision, even on a single sample basis; and (3) to assess the effects of less time-consuming alternatives to the complete census of individuals in the samples.

2 Methods

2.1 The multihabitat approach

The multihabitat approach has been devised to provide a standardised assessment of the macroinvertebrates community in a water body. The method requires two steps: (1) visual quantification of the different abiotic substrates within a representative river reach, and (2) proportional assignments of samples according to the area occupied by substrates. For example, if in a reach the relative abundance of substrates is 80% megalithal (stones > 40 cm) and 20% mesolithal (20–40 cm), the multihabitat sampling requires to collect 8 subsamples on megalithal and 2 on mesolithal. Subsamples are taken disturbing an area of 0.1 m² and collecting macroinvertebrates with a hand net (mesh size 500 micron). In standard biomonitoring assessment, these subsamples are then integrated into a single sample representative of the river reach (ISPRA, 2014).

The 8 rivers used as example were localised in Trentino Alto Adige (Italy), in two hydro-ecoregions, Southern pre-Alps and Dolomites and Eastern Inner Alps. In contrast to what suggested in the official sampling protocol (ISPRA, 2014), samples were immediately fixed with ethanol (final concentration ~70%) and processed in laboratory. No attempts were made to sort the samples in the field. All the individuals were counted and identified at the taxonomic resolution of genus and family.

2.2 The STAR_ICMi index

The STAR_ICMi is a multimetric index which combines 6 metrics (Tab. 1): ASPT (Average Score Per Taxon), Log10_Sel_EPDT+1 (logarithm of the sum of selected

Table 1. The subindices used to compute the multimetric STAR_ICMi.

Indicator	Metric	Name	Taxa considered	Reference	Weight
Tolerance	Index	ASPT-2	Whole assemblage at family level	Armitage et al. (1983)	0.333
Habitat	Abundance	$\text{Log}_{10} (\text{Sel_EPTD}+1)$	Log_{10} (sum of Heptageniidae, Ephemeridae, Leptophlebiidae, Brachycentridae, Goeridae, Polycentropodidae, Limnephilidae, Odontoceridae, Dolichopodidae, Stratyomidae, Dixidae, Empididae, Athericidae e Nemouridae+1)	Buffagni et al. (2004)	0.266
	Abundance	1-GOLD	1-(Relative abundance of Gastropoda, Oligochaeta and Diptera)	Pinto et al. (2004)	0.067
Richness/diversity	Number of taxa	N° families	Sum of the families	Ofenböck et al. (2004)	0.167
	Number of taxa	N° EPT families	Sum of the families of Ephemeroptera, Plecoptera and Trichoptera	Ofenböck et al. (2004)	0.083
	Diversity index	Shannon–Wiener	Whole assemblage at family level		0.083

Ephemeroptera, Plecoptera, Diptera and Trichoptera taxa), 1-GOLD (where GOLD is the relative abundance of Gastropoda, Oligochaeta and Diptera), number of families, number of families in the EPT orders (=Ephemeroptera, Plecoptera and Trichoptera) and the Shannon–Wiener diversity index (H'). The selected indices include different parameters of benthic communities, such as taxa sensitivity, abundance and diversity. To calculate the index, each metric is combined according to a specific weight, and the final score is expressed as the ratio between the observed and the reference condition ([Tab. 1](#)) ([ISPRA, 2014](#)).

2.3 Data analyses

2.3.1 Rarefaction

In order to investigate the effect of subsampling on the STAR_ICMi index, I used rarefaction technique. Rarefaction consists in extracting, at random and without replacement, some individuals from the original community ([Gotelli and Colwell, 2001](#)). In this way, it is possible to study the index performance when a different number of individuals are counted and identified. The rarefaction curve was constructed by extracting the samples of different size (100 individuals, 200, 300, etc.), and the procedure was repeated 1000 times for each sample size. Thus, the rarefied community was composed of 1000 extractions of size 100 individuals, 1000 of 200 individuals, 1000 of 300 and so on until the total number of individuals (1600 in the example showed). On each one of these rarefied communities, the STAR_ICMi index was calculated to test its performance. Rarefaction was performed on all the 8 rivers, but results showed a similar common trend. Therefore, I presented only a single representative example. All the analyses were conducted using R ([R Development Core Team, 2016](#)).

2.3.2 Bootstrap

The index uncertainty was estimated bootstrapping the observed abundance of taxa ([Dixon, 1993](#); [Manly, 2006](#)). The

bootstrap distribution of the STAR_ICMi index was calculated by resampling the individuals of the original sample. For example, if the sample assemblage consists of 4 species for a total abundance of 100 individuals (80 individuals of sp1, 15 of sp2, 5 of sp3), the bootstrap algorithm extracts, at random and with replacement, 100 individuals. Because the extraction is random and with replacement, the occurrence and the abundance of the species change on each run of bootstrap. Repeating the procedure many times, it is possible to generate a family of bootstrap assemblages on which to calculate the STAR_ICMi index. The resampling procedure was repeated 10,000 times to generate the sampling distribution of the index. The confidence interval was calculated from the bootstrap distribution of the STAR_ICMi index using the percentile method (95%). The procedure accounts only for the stochastic sampling effect, and it represents only a fraction of the entire variability represented by other sources of variation, such as local heterogeneity (*e.g.* [Laini et al., 2014](#); [Ramos-Merchante and Prenda, 2017](#)). In other words, the procedure attempts to estimate the index precision, not the accuracy ([Dixon, 1993](#)). Elsewhere, this bootstrap sampling distribution would be called a “null distribution”, as the variability accounted for by the model is assumed to arise with no particular ecological process or mechanism: it represents a sampling variability generated by random selection from the species assemblage according to each species’ relative abundance ([Gotelli and Graves, 1996](#)).

2.4 Different input data

To evaluate possible means to save time in the enumeration of all the individuals in the sample, different data transformations were tested. The four methods, from the most to the less time demanding were: (1) counting all the individuals of taxa with abundance less than 10 and estimating the most abundant (>10 ind.) subsampling the original sample ([Buffagni and Erba, 2007](#)). However, subsampling introduces an additional source of uncertainty

in the STAR_ICMi index that should be avoided. A possible way to predict the unknown abundance is to use a model of species abundance distribution (Magurran, 2004 and references therein). Among the many species of abundance distribution models, the geometric series was selected as one of the oldest, and most known models. When the species abundance distribution follows a geometric series, the relationship between the log abundance and the species ordered by rank is linear (Fattorini, 2005). Therefore, the most abundant taxa are estimated in two steps, (i) calculating the linear regression between the log abundance and rank using only the low abundant taxa and then (ii) forecasting the abundance of the most abundant taxa assuming their correct rank. In Appendix 1, a spreadsheet was provided to perform this calculation. See also Fattorini (2005) to an explanation on how to fit a geometric series using linear regression. (2) The second method thought to be less time demanding, is the use of rank, instead of the real abundance of taxa. Rank could be more quickly determined using subsamples. (3) For similar reasons, I tested the effect on the STAR_ICMi index of using classes of abundance (10 equal classes). The advantage of using classes is that they are attributed with more confidence than rank. (4) Finally, I evaluated the effect of using only the presence of taxa. Even though this last method is not consistent with the WFD, which requires quantitative assessment (European Union, 2000), this last option was tested for the sake of completeness.

3 Results

3.1 Subsampling effects

The effect of subsampling on the STAR_ICMi index was substantial in all the 8 rivers considered (Fig. 1 reported only one representative example, Arnò High). Half of the subindices were sensible to the number of individuals (logEPTD, number of families and EPT), whereas ASPT, GOLD and Shannon index were poorly affected. Even though the subindices have different weight in the STAR_ICMi (Tab. 1), the index was clearly dependent by the number of individuals. In the example, the ecological status inferred from the index changed from POOR (when only 100 individuals were counted), to MODERATE (between 200 and 800 individuals) and GOOD above. The final STAR_ICMi index with all the individuals counted was 0.826 in the example.

3.2 Index uncertainty

The precision of the STAR_ICMi index was different in the 8 rivers and ranges from 1.2% to 5.9% (coefficient of variation). Although quite precise that variation increased the risk of misclassification when the index was close to a boundary class. For example, in Pescara river, the index was 0.724 (0.71 was the boundary class between MODERATE and GOOD), and in the 16.5% of cases, the ecological status would fall into the lower class. The 95% confidence interval for the river was 0.699–0.747. The STAR_ICMi index distribution was not always symmetric and centred on the observed value as shown in the example (Fig. 2). In some rivers, the index distribution was skewed, as deducible observing the asymmetrical error bars in Figure 3.

3.3 Different input data

The method that estimates the most abundant taxa produced the less biased results (Fig. 3). The STAR_ICMi index, calculated on the estimated abundance, was close to the observed value and most of the times within the confidence interval calculated by bootstrap (Tab. 2). Using rank, classes and presence as input data produced results that were substantially different from the real ecological status calculated by counting all the individuals.

4 Discussion

The risk of misclassification can have serious ecological and economic consequences, especially when the water body is put in the moderate rather than in the good class. It has been demonstrated that the risk of misclassification is not negligible, and a measure of index uncertainty is essential for managers to take informed decisions. The approach used here to calculate index uncertainty is general, and as such could be applied also for other indices and taxonomic groups beyond the macro-invertebrates and the STAR_ICMi index.

The STAR_ICMi is a multimetric index affected by subsampling because three out of six subindices are sensible to the number of individuals counted. The three subindices affected by number are calculated as sums (logEPTD, number of families, EPT), whereas the three that are not have ratios in their equations (ASPT, GOLD, Shannon). The subindex logEPTD strictly depends on taxa abundance as it is calculated by summing the individuals belonging to several EPTD families (Tab. 1). As the number of individuals is added on, the number of taxa increases, and similarly does the probability to include the EPTD families. Similar arguments hold true for the subindices number of families and EPT. In contrast, the ASPT subindex reached an asymptote with about 400 individuals. The ASPT is calculated considering the score of taxa, and more precisely it is computed as the ratio between the sums of scores and the number of taxa (Armitage *et al.*, 1983). Thus, as individuals are added on, both the numerator and denominator increase, and the ratio changes little. Analogous explanations apply for the GOLD (Pinto *et al.*, 2004) and Shannon index (Shannon and Weaver, 1949). The cumulative sum of these effects determines the strong sensibility of the STAR_ICMi index to the number of individuals enumerated. Interestingly, the STAR_ICMi index did not show clear signs of levelling off in any of the 8 rivers examined. There is no general agreement on the minimum number of individuals needed to get the true ecological status. For example, Doberstein *et al.* (2000) showed that subsamples containing 100–300 individuals produced highly inaccurate results, while Ostermiller and Hawkins (2004) suggested to consider samples with more than 350 individuals. It is unlikely that more studies on the minimum sampling effort will produce general recommendations, as precision changes according to the type of stream and the indices used (Lorenz *et al.*, 2004). The most appropriate suggestion seems to define a target number, and to rarefy the communities to that common sampling effort. Only in this way, samples with different number of individuals can be correctly compared (Gotelli and Colwell, 2001). According also to other authors (e.g. Clarke *et al.*, 2006b), 700 individuals could be a

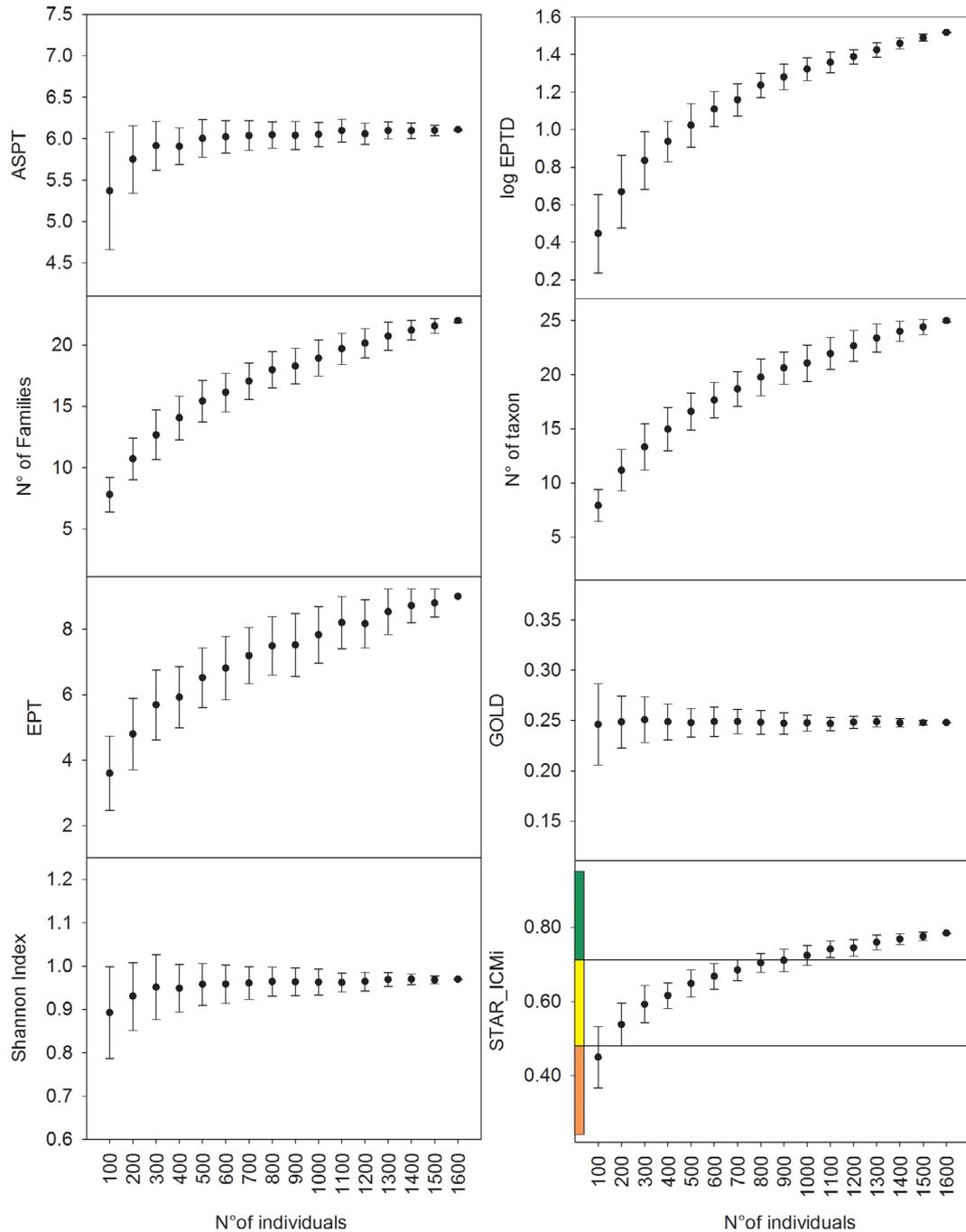


Fig. 1. The effect on the STAR_ICMi index, and relative subindices, of the incomplete counting of individuals in a sample. Error bars are standard deviation.

reasonable compromise between effort and precision. Hence, the best strategy would be to census at least, or more than 700 individuals, and when more, to rarefy the assemblage to the target.

Three main groups of causes could be recognised as sources of uncertainty in the STAR_ICMi index: (1) imputable to the ecological variability, (2) because of technical errors, and (3) due to sampling-related causes. To the first group belongs all the ecological processes which lead to the heterogeneous distribution of the macroinvertebrate assemblages in terms of species and density, such as habitat

preferences, competition, predation, dispersal, behavioural and so on (Taylor, 1984). Even though the multihabitat method requires a standard approach with fixed number of subsamples, fixed area and fixed proportion of substrates on a given stretch of river, the procedure will never provide two identical samples. These differences eventually lead to different results of the STAR_ICMi index, and sometimes to different ecological status. This variability can be accounted for adopting stratified sampling (Laini *et al.*, 2014), or hierarchical sampling (*e.g.* Li *et al.*, 2001). Similar methods can be adopted also to estimate uncertainty due to

errors (2), like identification of taxa, different trained staff and so on (Clarke *et al.*, 2006a). However, another source of variability exists, and this is related to the sampling process (3). Unless it would be possible to identify and count the

entire benthic community, we need some kinds of sampling to estimate the abundance of each taxon. As a result, the data used to calculate the index contain sampling error. If we have a jar with different coloured balls (the species), and we extract (the kick sample) several balls, results will not be the same in consecutive draws even replacing the balls into the jar. Uncertainty calculated by bootstrap represents the part of variability in the STAR_ICMi index that would be obtainable repeating the identical sampling survey. This form of variability is rarely included in studies of index uncertainty, but it is always present (see Dolph *et al.*, 2010 for a quantification of sampling variability with bootstrap). Importantly, it has been shown that this variability was not negligible, especially when the ecological status was near a boundary class. Therefore, it is suggested to present the STAR_ICMi index with the uncertainty at least due to the sampling variability (3) when other estimates (1, 2) are unavailable (as is usually the case in routine biomonitoring).

One of the critical aspects of routine application of the multihabitat method and STAR_ICMi application is that it is extremely time consuming (Bo *et al.*, 2016). Low budget and limited human resources put at risk the biomonitoring of many watercourses, which were previously surveyed with less demanding methods and procedures. Indeed, early works considered pro and cons of this new approach providing evidences about the size of sample, the minimum number of individuals, and the time requested by the different options (Vlek, 2004). However, still lacking was a work testing the effect of less time-consuming alternatives than the whole census of all the individuals in the sample. The most promising method, namely that one producing results close to

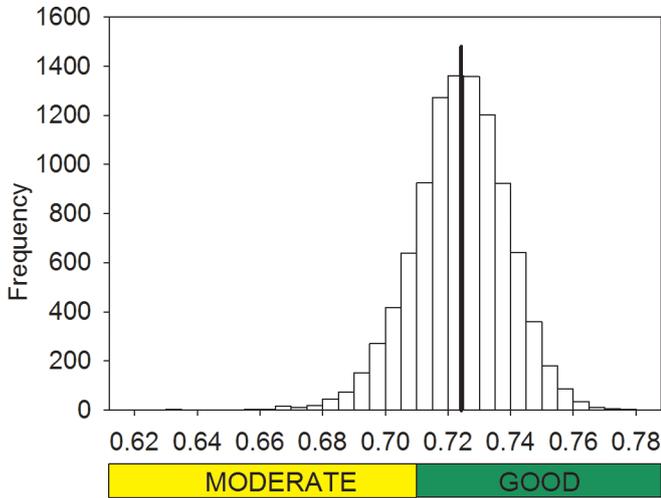


Fig. 2. Application of the bootstrap approach to estimate the confidence interval of the STAR_ICMi index in the Pescara river. The observed index was 0.724, and the 95% confidence interval was 0.699–0.747. The boundary between MODERATE and GOOD class is 0.71. In this example, the river would be classified as GOOD (0.724>0.71), but the confidence interval suggested that the probability of wrong classification was not negligible (in the 16.5% of cases the ecological status would fall into the MODERATE class).

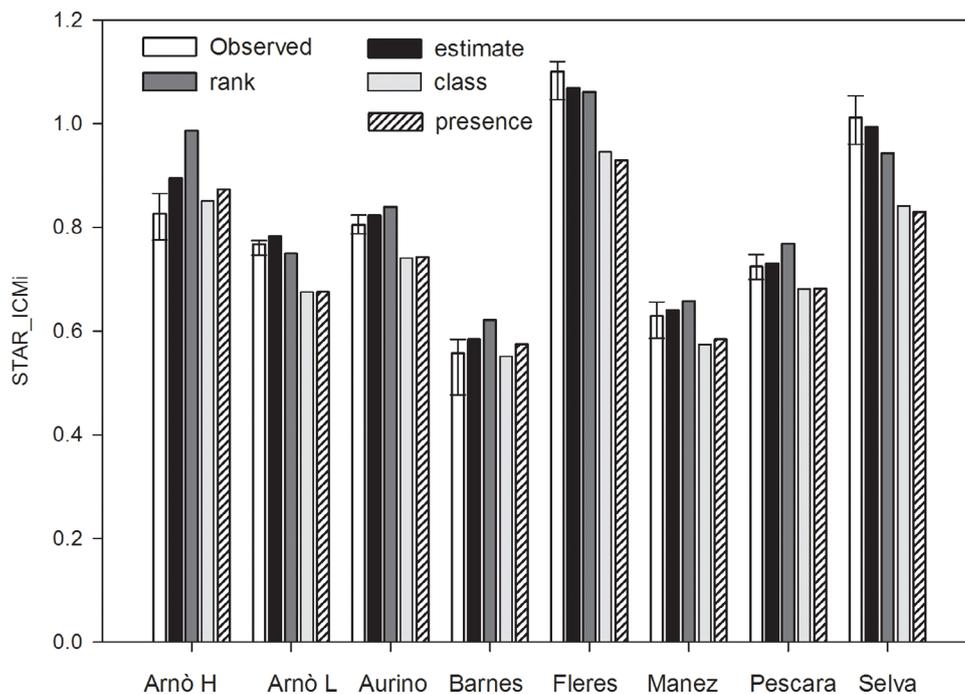


Fig. 3. Comparison of the effect of different input data on the STAR_ICMi index in 8 rivers. Observed is the real score counting all the individuals; the error bars on the observed value are calculated using the bootstrap approach and represent 95% confidence intervals. Estimate is the index value when the abundance of the most abundant taxa was estimated using the geometric series (see Appendix 1 for examples); rank, class and presence is the index value when the real input data were reduced to rank, classes and presence respectively.

Table 2. Effects of different input data on the STAR_ICMi index. Observed is the index score counting all the individuals; CI = confidence interval, the precision calculated using bootstrap. Estimate is the index score when the abundance of the most abundant taxa was estimated; rank, class and presence is the index score when the real input data were reduced to rank, classes and presence respectively.

River	Observed	CI	Estimate	Rank	Class	Presence
Arnò high	0.826	0.775–0.865	0.895	0.986	0.851	0.873
Arnò low	0.767	0.746–0.774	0.783	0.749	0.675	0.675
Aurino	0.804	0.787–0.823	0.822	0.840	0.741	0.742
Barnes	0.557	0.477–0.584	0.584	0.621	0.551	0.574
Fleres	1.100	1.046–1.119	1.068	1.061	0.945	0.928
Manez	0.628	0.585–0.656	0.640	0.658	0.574	0.583
Pescara	0.724	0.699–0.747	0.730	0.769	0.681	0.681
Selva	1.012	0.960–1.053	0.993	0.943	0.841	0.830

those observed by processing the whole sample, was to estimate the most abundant taxa using the rank-abundance model. The limit of this approach is that it assumes the geometric distribution in all the samples, which might not be the case in all the assemblages (Magurran, 2004).

5 Conclusions

Three main conclusions can be drawn from this study. First, because the STAR_ICMi index is affected by subsampling, it is suggested to sort at least 700 individuals, and when more to calculate the index rarefying the community to the target (700). In this way, sampling effort is consistent among surveys, and comparison is unbiased. As a medium-term perspective, the STAR_ICMi index might be modified in order to become less dependent on sampling effort. Second, index precision should always be presented along with the index score to quantify uncertainty in assigning the quality classes. Although the bootstrap approach does not account for all the sources of index variability, it is a valuable resource for the competent authorities to take informed decisions. Third, borrowing one of the oldest and most universal laws of ecology, I showed how to use a rank-abundance model for application purpose. The geometric series is an easy model, simple to compute, and for this it could be added to the practitioner's toolbox.

Supplementary Material

Appendix 1. Excel spreadsheet with two examples showing how to estimate the number of individuals of the most abundant taxa.

The Supplementary Material is available at <http://www.kmae-journal.org/10.1051/kmae/2017012/olm>.

References

Armitage PD, Moss D, Wright JF, Furse MT. 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Res* 17: 333–347.

Bennett C, Owen R, Birk S, et al. 2011. Bringing European river quality into line: an exercise to intercalibrate macro-invertebrate classification methods. *Hydrobiologia* 667: 31–48.

Bo T, Doretto A, Laini A, Bona F, Fenoglio S. 2016. Biomonitoring with macroinvertebrate communities in Italy: what happened to our past and what is the future? *J Limnol*. DOI: [10.4081/jlimnol.2016.1584](https://doi.org/10.4081/jlimnol.2016.1584).

Buffagni A, Erba S. 2007. Macroinvertebrati acquatici e Direttiva 2000/60/EC (W.F.D.) – Parte A. Metodo di campionamento per i fiumi guadabili. *IRSA-CNR Notiziario dei Metodi Analitici* 1: 1–27.

Buffagni A, Erba S, Cazzola M, Kemp JL. 2004. The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. *Hydrobiologia* 516: 313–329.

Buffagni A, Erba S, Cazzola M, Murray-Bligh J, Soszka H, Genoni P. 2006. The STAR common metrics approach to the WFD intercalibration process: full application for small, lowland rivers in three European countries. *Hydrobiologia* 566: 379–399.

Clarke RT, 2013. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). *Hydrobiologia* 704: 39–56.

Clarke RT, Davy-Bowker J, Sandin L, Friberg N, Johnson RK, Bis B. 2006a. Estimates and comparisons of the effects of sampling variation using 'national' macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. *Hydrobiologia* 566: 477–503.

Clarke RT, Lorenz A, Sandin L, et al. 2006b. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. *Hydrobiologia* 566: 441–459.

Dixon PM. 1993. The bootstrap and the jackknife: describing the precision of ecological indices. In: Scheiner S, Gurevitch J, eds. Design and analysis of ecological experiments. New York: Chapman and Hall, pp. 210–318.

Doberstein CP, Karr JR, Conquest LL. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Fresh Biol* 44: 355–371.

Dolph CL, Sheshukov AY, Chizinski CJ, Vondracek B, Wilson B. 2010. The Index of biological integrity and the bootstrap: can random sampling error affect stream impairment decisions? *Ecol Indic* 10: 527–537.

European Union. 2000. Directive 2000/60/EC. Establishing a framework for community action in the field of water policy. Luxembourg: European Commission PE-CONS 3639/1/100 Rev 1.

Fattorini S. 2005. A simple method to fit geometric series and broken stick models in community ecology and island biogeography. *Acta Oecol* 28: 199–205.

- Friberg N. 2014. Impacts and indicators of change in lotic ecosystems. *WIREs Water* 1: 513–531.
- Friberg N, Bonada N, Bradley DC, *et al.* 2011. Biomonitoring of human impacts in freshwater ecosystems: the good, the bad and the ugly. *Adv Ecol Res* 44: 1–68.
- Ghetti PF. 1997. Manuale di applicazione. Indice Biotico Esteso (I.B.E.). I macroinvertebrati nel controllo della qualità degli ambienti di acque correnti. Trento: Provincia Autonoma di Trento, 222 p.
- Gotelli NJ, Colwell RK. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4: 379–391.
- Gotelli NJ, Graves GR. 1996. Null models in ecology. Washington, D. C.: Smithsonian Institution Press, 368 p.
- Hering D, Buffagni A, Moog O, *et al.* 2003. The development of a system to assess the ecological quality of streams based on macroinvertebrates – design of the sampling programme within the AQEM project. *Int Rev Hydrobiol* 88: 345–361.
- ISPRA. 2014. Linee Guida per la valutazione della componente macrobentonica fluviale ai sensi del DM 260/2010. ISPRA Manuali e Linee Guida 107/2014. Available from: http://www.isprambiente.gov.it/files/pubblicazioni/manuali-lineeguida/MLG_107_14.pdf.
- Laini A, Vorti A, Bolpagni R, Viaroli P. 2014. Small-scale variability of benthic macroinvertebrates distribution and its effects on biological monitoring. *Ann Limnol - Int J Lim* 50: 211–216.
- Li J, Herlihy A, Gerth W, *et al.* 2001. Variability in stream macroinvertebrates at multiple spatial scales. *Fresh Biol* 46: 87–97.
- Lorenz A, Kirchner L, Hering D. 2004. ‘Electronic subsampling’ of macrobenthic samples: how many individuals are needed for a valid assessment result? *Hydrobiologia* 516: 299–312.
- Magurran AE. 2004. Measuring biological diversity. Oxford: Blackwell Publishing, 256 p.
- Manly BF. 2006. Randomization, bootstrap and Monte Carlo methods in biology. Boca Raton: Chapman and Hall/CRC, 480 p.
- Ofenböck T, Moog O, Gerritsen J, Barbour M. 2004. A stressor specific multimetric approach for monitoring running waters in Austria using benthic macroinvertebrates. *Hydrobiologia* 516: 251–268.
- Ostermiller JD, Hawkins CP. 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. *J North Am Benthological Soc* 23: 363–382.
- Pinto P, Rosado J, Morais M, Antunes I. 2004. Assessment methodology for southern siliceous basins in Portugal. *Hydrobiologia* 516: 191–214.
- R Development Core Team. 2016. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: <http://www.R-project.org/>.
- Ramos-Merchante A, Prenda J. 2017. Macroinvertebrate taxa richness uncertainty and kick sampling in the establishment of Mediterranean rivers ecological status. *Ecol Indic* 72: 1–12.
- Shannon CE, Weaver W. 1949. The mathematical theory of communication. Urbana: University of Illinois Press, 144 p.
- Taylor LR. 1984. Assessing and interpreting the spatial distributions of insect populations. *Annu Rev Entomol* 29: 321–357.
- Vlek HE. 2004. Comparison of (cost) effectiveness between various macroinvertebrate field and laboratory protocols. European Commission, STAR (Standardisation of river classifications), Deliverable N1, 78 p.

Cite this article as: Spitale D. 2017. Performance of the STAR_ICMi macroinvertebrate index and implications for classification and biomonitoring of rivers. *Knowl. Manag. Aquat. Ecosyst.*, 418, 20.