

How to make river assessments comparable: A demonstration for hydromorphology



Simone D. Langhans*, Judit Lienert, Nele Schuwirth, Peter Reichert

Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland

ARTICLE INFO

Article history:

Received 8 October 2012

Received in revised form 12 March 2013

Accepted 20 March 2013

Keywords:

Ecological assessment

Comparability

Intercalibration

Bioassessment

River management

Multi-criteria decision analysis

ABSTRACT

River monitoring and assessment programs are important tools to quantify the condition of river ecosystems, identify deficits, and provide preliminary indication of how to improve them. But, they are limited in delivering comparable assessment results across national or transnational borders, aggregating site-specific assessments into broader scale assessments, and supporting river management decisions. We present a multi-criteria decision analysis approach for improving the comparability of ecological assessment methods of different origin and for combining these assessments into a joint procedure. The approach consists of seven consecutive steps. The most central ones concern the hierarchical allocation of ecological assessment endpoints, and the harmonization of the scoring procedure of attributes (ecological indicators or assets) to a common scale from 0 to 1. We demonstrate the approach integrating three programs developed to assess the hydromorphological river condition in Switzerland, Germany, and the USA. In our example, the integrated assessment produces comparable results for the whole range from natural to impacted rivers, while data continuity with original assessments was maintained. Our approach provides a common assessment standard due to the definition of the minimum amount of information required, is flexible regarding measurement and assessment endpoints, and bridges the gap between river quality assessment and management.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

In response to the poor condition of river ecosystems and the increasing risk of losing services that humans receive from them (Vörösmarty et al., 2010), water protection laws have been implemented globally (e.g., US Clean Water Act; European Commission, 2000; Swiss Water Law). These policies aim to evaluate the ecological status of freshwater ecosystems, to identify causes of poor river condition, and regulate the achievement of good river quality. To gather the necessary information, local, state, and national environmental agencies conduct ecological river assessments worldwide by applying a variety of monitoring programs (e.g., Bundi et al., 2000; Hughes et al., 2000; Verdonshot, 2000; Bunn et al., 2010). Such monitoring and assessment programs are valuable tools to document trends, detect deficits, and provide preliminary indications on how to improve the surveyed systems. However, as they often differ substantially in the choice of

ecological indices, the scales of interest, and how the indices are used to assess ecological conditions, the assessments are often not directly comparable (Raven et al., 2002; Feio et al., 2009; Cao and Hawkins, 2011; Birk et al., 2012a).

Comparability is a critical issue in the realm of bioassessment causing ongoing discussions on how to achieve it (e.g., Ghetti and Bonazzi, 1977; Diamond et al., 2012; Monaghan and Soares, 2012). We argue that comparability of ecological assessment in general is important due to several regulatory, cost, and management reasons. First, national and transnational legislation, e.g., the Water Framework Directive (Birk and Hering, 2006; Erba et al., 2009) increasingly requires comparability of river assessments used by member states (Solimini et al., 2009). In these cases, no elaborate and costly a posteriori intercalibration exercises (e.g., Heiskanen et al., 2004; Birk and Hering, 2006) would be needed if an assessment approach produced directly comparable results. Second, the integration of monitoring data collected with various programs could greatly strengthen local and state programs, reduce duplication of sampling effort, and provide databases for the development of indices if that data were comparable (e.g., Astin, 2006; Cao and Hawkins, 2011). Third, comparable assessments are essential for providing data continuity in long-term monitoring programs (Cao and Hawkins, 2011), and for aggregating site-specific assessments into broader scale assessments (e.g., Buffagni et al., 2007). Last, existing monitoring and assessment programs are often not

* Present address: Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany. Tel.: +49 (0)30 64181 618; fax: +49 (0)30 64181 750.

E-mail addresses: langhans@igb-berlin.de (S.D. Langhans), judit.lienert@eawag.ch (J. Lienert), nele.schuwirth@eawag.ch (N. Schuwirth), reichert@eawag.ch (P. Reichert).

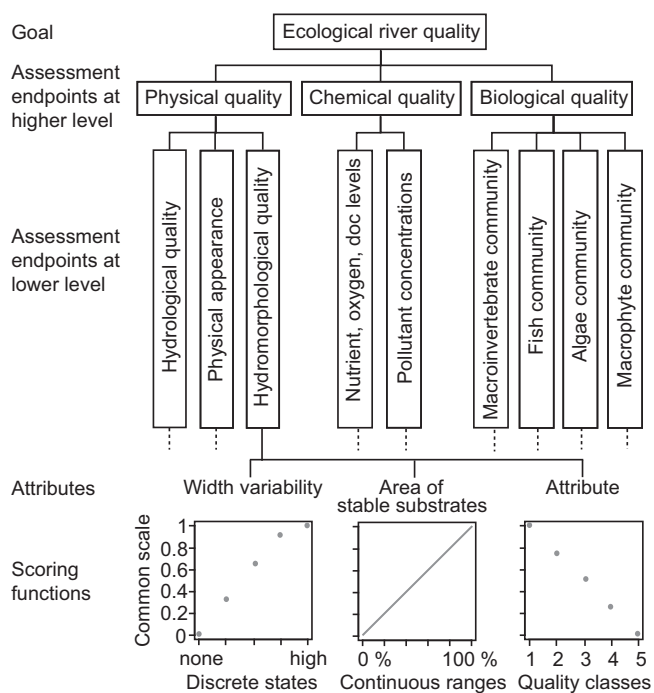


Fig. 1. Example of a possible assessment hierarchy with good ecological river quality as the main goal, the higher-level assessment endpoints, and the lower-level assessment endpoints with their corresponding attributes (indicators) at the lowest level, including examples of scoring functions to translate each attribute's individual measurement (e.g., discrete attribute states, continuous ranges (e.g., %), or quality classes) and corresponding scoring onto a scale between 0 and 1. Note that the endpoints are abbreviated by the field of assessment (e.g., the label *physical quality* means *good physical quality*).

incorporated into a conceptual management agenda (Beechie et al., 2010; but see Bunn et al., 2010). Data comparability, however, is a key prerequisite to efficiently plan for conservation or management measures.

Multi-criteria decision analysis (MCDA; specifically multi-attribute value theory, MAVT; see e.g., Keeney and Raiffa, 1976; Keeney, 1982; Clemen, 1996; Eisenführ et al., 2010) offers methods to standardize ecological assessments and integrate endpoints from different programs in an approach that can further be used to support river management decisions (Klauer et al., 2006; Reichert et al., 2007; Corsair et al., 2009). Thereby, objectives, which can be used synonymously to assessment endpoints in the common river assessment terminology (Table 1), are arranged hierarchically into different levels. The different levels culminate in the overall objective or goal (Table 1), for instance the good ecological quality of the river (Fig. 1). The lowest, most explicit level of endpoints is assessed with one or several attributes (often also called ecological indicators or assets, Table 1). To harmonize the assessments of the lowest level endpoints, the attribute-specific scorings are translated onto a common scale (from 0 to 1) and displayed as a mathematical function of the attributes: the common scale between 0 and 1 is given on the y-axis, and the different states (or levels) the attribute can adopt on the x-axis (e.g., discrete states, continuous ranges (e.g., %), or quality classes; Fig. 1). Such a function must be formulated as a measurable value function in the terminology of MAVT (Dyer and Sarin, 1979; Eisenführ et al., 2010), but we refer to it as the scoring function to adjust to a terminology more familiar in applied ecology and river management (Table 1). The higher level endpoints are, finally, assessed from bottom-up: the scores from the lower levels are mathematically aggregated to the next-higher level, etc.

In this study, we used MAVT methods to develop a new assessment approach that is capable of integrating ecological

assessments from different programs to produce comparable assessment results. First, we describe seven principle steps toward the new approach (Fig. 2). Each step is illustrated with a real world example – the integration of three existing hydromorphological assessment programs from the USA, Switzerland, and Germany. Second, we apply the new approach to assess the hydromorphological condition of a target river system in Switzerland. Third, we compare the original assessments of the three programs with the results obtained with the new approach to evaluate its performance. After a critical discussion, we conclude the study with the definition of six main advantages of our approach and its significance for river management.

2. Methods

2.1. Step 1: choose the assessment programs you want to compare or integrate

River assessment and monitoring programs are usually composed of a set of protocols dealing with different aspects of river quality (Bundi et al., 2000; Hawkins et al., 2008; Bunn et al., 2010). In the decision analysis terminology, these aspects correspond to sub-objectives that contribute to the overall objective of achieving a good ecological status of the river. With the new approach, biological, physico-chemical, and/or morphological aspects of river quality can be integrated depending on the main goal of the assessment. To demonstrate the approach, we combine three hydromorphological river assessments programs developed in Switzerland, Germany, and the United States (but other programs could have been chosen just as well).

2.1.1. Swiss modular concept for stream assessment (SMC)

In Switzerland, the SMC (Bundi et al., 2000; <http://www.modul-stufen-konzept.ch>) has been introduced to assess the fulfillment of the guidelines of the Water Protection Law of 1991 and the Water Protection Order of 1999. It consists of separate methods addressing different assessment fields. One of them assesses the morphological structure of rivers at two levels, a coarse overview survey at a regional scale and a more detailed survey at the scale of small sub-catchments. The goal of the overview survey comprises an assessment of the morphological quality of rivers in a wider region (BAFU, 1998). Only selected attributes are measured and evaluated to keep the sampling effort and costs low. The more detailed survey assesses the hydromorphological condition based on eleven attributes (BAFU, 2006). These belong to three categories suited to assess the condition of the riverbed structure, riparian zone, and longitudinal connectivity of river sections. For this study, we used the assessment based on information gathered in the more detailed survey. To facilitate the comparison of assessments with other programs, we averaged the quality of the riverbed structure, the riparian zone, and the longitudinal connectivity to an overall score (Supplementary Fig. 1A).

2.1.2. US rapid bioassessment protocol (RBP)

In the USA, the Clean Water Act established in 1977 led to the development of a wide range of assessment programs resulting in quantitative protocols e.g., by the U.S. Geological Survey (Fitzpatrick et al., 1998), the U.S. Forest Service/U.S. Bureau of Land Management (Gallo, 2002), or the U.S. Environmental Protection Agency (USEPA) (Stoddard et al., 2005, 2006). Besides these cost- and time-consuming assessments, less laborious rapid physical habitat protocols have been established: the qualitative habitat evaluation index of the Ohio Environmental Protection Agency (Ranking, 1989, 2006), USEPA's RBP (Plafkin et al., 1989; Barbour et al., 1999), or the more recent river visual assessment protocol of the Natural Resource Conservation Service (NRCS, 1998, 2009).

Table 1
Definition of different terminologies used in the realms of multi-attribute value theory (MAVT) and traditional river assessment. In this article, we used the river assessment terminology.

MAVT	River assessment
Overall objective: overall goal to be achieved.	Goal: target to meet, defined by the legislation, ecologists, and/or the general public (e.g., Boulton, 1999; Barbour et al., 2000).
Sub-objectives: each sub-objective covers an important aspect of the objective at the higher level; all sub-objectives associated with the same higher level objective cover all relevant aspects.	Assessment endpoint: explicit expression of the actual environmental values that are to be protected (USEPA, 1992, 1997, 1998).
Attribute: measurable system property to assess the degree of fulfillment of a sub-objective; all attributes together must make an assessment of the degree of fulfillment of all (sub-)objectives possible.	Attribute, indicator, asset: measurement endpoint to evaluate health of a system (economic, physical, biological, human) (Burger, 2006; see also Heink and Kowarik (2010) for further definitions).
Value function: description of the degree of fulfillment of the corresponding objective as a function of associated attributes on a common scale from 0 to 1.	Scoring function: description of the degree of fulfillment of the corresponding endpoint as a function of associated attributes (indicators, assets) on a common scale from 0 to 1.

Since all three qualitative indices are highly correlated (Hughes et al., 2010), we included the low-cost RBP. It can be applied to riffle-run prevalent rivers in moderate to high altitude with coarse bed sediments (i.e., high gradient rivers), or to glide-pool prevalent rivers in lowlands with fine bed sediments (i.e., low gradient rivers). Both approaches assess ten attributes, which are aggregated in one step to assess hydromorphology (Supplementary Fig. 1B).

2.1.3. Survey of the German working group of the federal states on water issues (LAWA)

Two standard methods are available for river habitat survey in Germany suggested by the German working group of the federal states on water issues (LAWA): a field survey for small to medium-sized rivers (LAWA, 2000) and an overview survey for larger rivers (LAWA, 2002; Weiss et al., 2008). To be comparable with the Swiss method, which was developed for small to medium sized rivers, we chose to work with the LAWA field survey (LAWA). The survey investigates 26 attributes, which are grouped into six main categories. These include development of the river course, longitudinal profile, riverbed structure, cross-sectional profile, bank structure, and riparian surroundings. The six categories are further

aggregated into valuations of the riverbed, riverbank, and surrounding landscape, which finally culminate in the valuation of the hydromorphological condition (Supplementary Fig. 1C).

2.2. Step 2: compile information from the chosen programs

The assessment approach presented here combines elements from existing programs to ensure assessment continuity and effectiveness, as monitoring data are often comprehensive and already available on a large spatial scale. Hence, to get an overview of the program-specific elements that will finally be integrated in the new approach, we disassembled the three programs and extracted information regarding attributes, assessment endpoints, scoring procedures and aggregation schemes (Table 2).

2.3. Step 3: standardize the scoring of the original endpoints to a common scale from 0 to 1

We transformed the program-specific scorings, i.e., the quality of an attribute or indicator which may be measured in any unit (e.g. %, m, or in classes) to a common assessment or value scale from

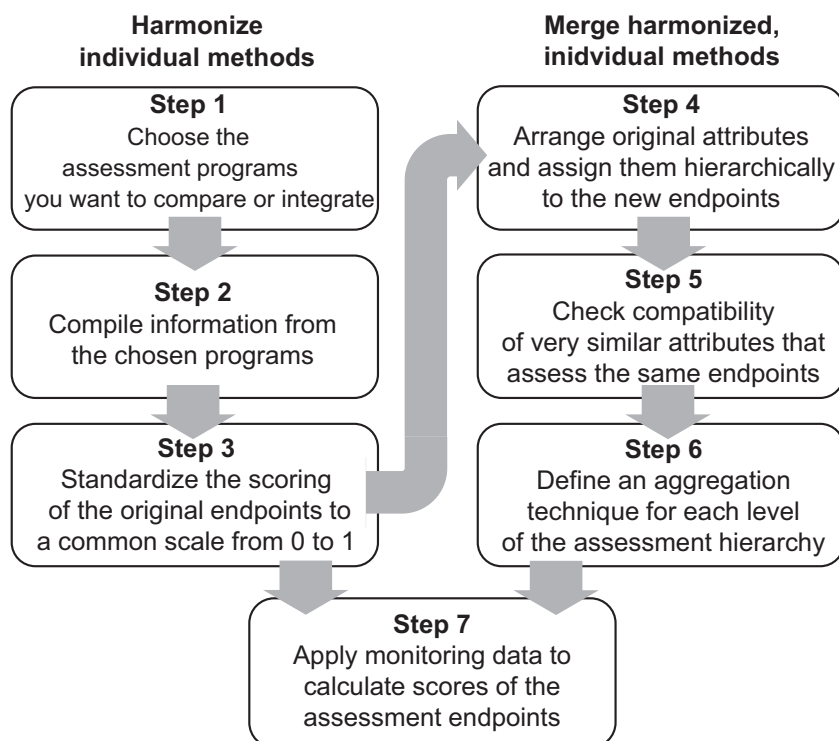


Fig. 2. The seven steps necessary to integrate existing assessment programs into the new approach to make their results comparable.

Table 2

Summarized information on existing river assessment programs that were integrated into the new approach developed in this paper: SMC (Switzerland), LAWA (Germany), and RBP (USA). LAWA's seven quality classes can be converted into five classes: 1 and 2 become quality class 1, and 6 and 7 become quality class 5.

Assessment characteristics	SMC	LAWA	RBP
Structure	Hierarchical	Hierarchical	Hierarchical
Number of hierarchical levels	3	4	2
Aggregation technique	Arithmetic mean and minimum	Arithmetic mean and minimum	Arithmetic mean
Pre-defined stream typology	No	Yes: 7 river types, 2 width classes	Yes: high and low order rivers
Scoring type	5 quality classes: 1 = high 2 = good 3 = moderate 4 = poor 5 = bad	7 quality classes: 1 = natural 2 = slightly modified 3 = moderately mod. 4 = considerably mod. 5 = heavily mod. 6 = very heavily mod. 7 = artificial	4 quality classes: 1 = high 2 = good 3 = fair 4 = poor
Reference condition	Near-natural	Near-natural (Leitbildzustand)	Optimal condition to support biology

0 to 1. We define this common scale in the sense that the same differences in scores of an indicator represent the same degree of improvement (Dyer and Sarin, 1979; Eisenführ et al., 2010). This means that an improvement from 0.1 to 0.3, or an improvement from 0.6 to 0.8 are of the same value. Additionally, we assume that the degree of improvement from one indicator level to the next or from one quality class to the next higher is the same, if an explicit statement of the meaning of discrete indicator levels or quality classes is missing in original protocols.

We used two different procedures to harmonize the individual assessments: one for continuously measured attributes, and another for discretely measured ones. Whenever possible, we prefer to use continuously defined attributes (ideally with an assessment of the measurement uncertainty) as this avoids unnecessary inaccuracy due to rounding errors. The RBP-attribute *area of stable substrates*, for instance, is continuously measured. It can adopt four continuous attribute ranges from 0 to 10%, 10–30%, 30–50%, and 50–100% that are assessed with the RBP-quality classes *natural* (quality class 1), *good* (2), *fair* (3), and *bad* quality (4), respectively. To construct a scoring function for such continuous attributes, first, the continuous attribute ranges are mapped on the x-axis (Fig. 3A). Then the number of program-specific quality classes is represented as equally long intervals on the y-axis unless otherwise specified by the assessment program. In the case of the RBP-attribute *area of stable substrates*, the y-axis is divided into four intervals of 0.25 length (Fig. 3A). Thereby, the different intervals represent the four RBP-quality classes: 0–0.25 the *bad*, 0.25–0.5 the *fair*, 0.5–0.75 the *good*, and 0.75–1 the *high* one. Finally, the points at the class boundaries are connected by a piecewise linear function.

An example for a discretely measured attribute is the LAWA attribute *profile depth*. It can adopt five attribute states: *very flat*, *flat*, *moderately flat*, *deep*, and *very deep*. LAWA assigns five of the total seven possible quality classes to assess these categories as *natural* (or quality class 1), *slightly modified* (2), *considerably modified* (4), *very heavily modified* (6), and *artificial* (7). The quality classes *moderately modified* (quality class 3) and *heavily modified* (5) are not used in this case, as the attribute can only adopt five categories. To construct the scoring function for such a discrete attribute, we assume equal spacing of scores if the method description does not give specific hints for another interpretation. First, the attribute states are again mapped on the x-axis (Fig. 3B). For each of these attributes' states, we define a discrete score on the common scale between 0 and 1 (y-axis). We assume that the best state corresponds to 1, and the worst to 0. To define the scores for the remaining states, we divide the interval between 0 and 1 into a number of equally long intervals, depending on how many states remain. In our example of the LAWA attribute *profile depth*, the states *very flat*, *flat*, *moderately flat*, *deep*, and *very deep* are mapped on the x-axis (Fig. 3B). Then,

the state *very flat* is associated with 1, and *very deep* with 0. Scores for the remaining three states are calculated as 0.25 (*deep*), 0.5 (*moderately flat*), and 0.75 (*flat*). To finally check whether the constructed function corresponds with the original assessment of the attribute, we represent the program-specific quality classes on the common scale on the y-axis (as done for the continuous attribute). This results in seven equally long intervals (Fig. 3B). If the calculated, discrete scores lie in between the range of the corresponding quality classes, the scoring function can be accepted. E.g., the discrete score (0.25) calculated for the attribute state *deep* should be located within the interval representing its original quality class 6 (*very heavily modified*) etc.

The scoring of all SMC- and LAWA-attributes was standardized according to the discrete procedure (Supplementary Figs. 2 and 3). RBP-attributes were either discrete or continuous (Supplementary Fig. 4).

After translating all assessment endpoints that directly depend on the attributes, the original aggregation schemes can be used to finalize the harmonization of the individual assessment methods (e.g., SMC, LAWA, RBP). Although these methods now use a common scale, assessing a river reach with each of them (Fig. 2) may lead to different results. Finding the reasons for such differences, e.g. procedure-specific differences in attributes or in the definition of reference conditions, is very much facilitated when the methods are already harmonized.

2.4. Step 4: arrange original attributes and assign them hierarchically to the new endpoints

In addition to facilitate the comparison of individual assessment methods (see step 3), our approach can be used to merge harmonized methods into a single hierarchical structure of a joint assessment procedure. Structuring assessed river characteristics that culminate in the main goal good ecological river quality (Fig. 1) in a hierarchical way in the first place has several advantages: Hierarchies make it easier to (1) concretize assessment endpoints at lower levels, (2) facilitate the evaluation of their completeness, (3) increase the transparency of the assessed river characteristics, their aggregation structure and weights and (4) make deficiencies more obvious.

Together, all three programs resulted in a total of 49 attributes (ecological indicators). We grouped similar attributes into sets that could be associated with 16 lowest-level assessment endpoints. For example, the attributes *width variability* from the SMC and LAWA were associated with the endpoint *stream width variability*, and *sinuosity factor*, *channel sinuosity*, and *channel alteration* (from LAWA and RBP) with *sinuosity* (Fig. 4). These lowest level endpoints were then again hierarchically grouped and associated with higher-level endpoints.

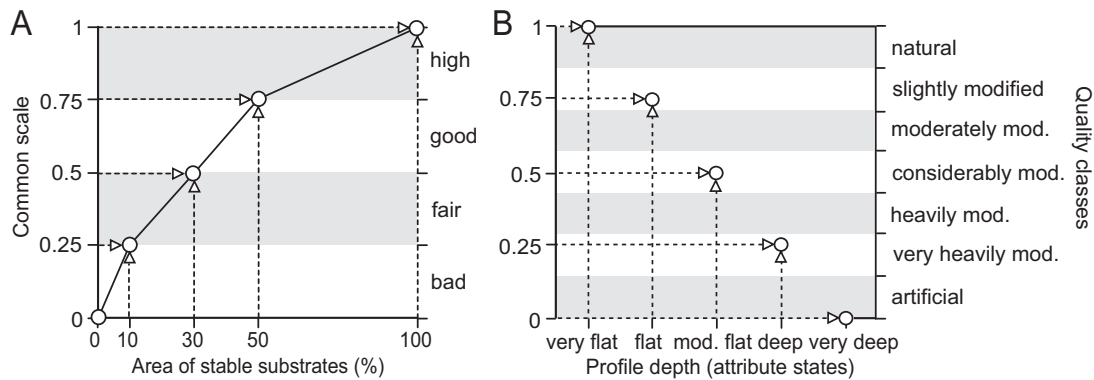


Fig. 3. Examples of scoring functions that are needed to transfer the individual attribute measures on the x-axis onto a common scale of 0–1, given on the y-axis. Two different approaches for continuous (A) and discrete (B) assessments are needed. (A) Standardization of the continuous assessment of the RBP-attribute *area of stable substrates*, and of (B) the discrete assessment of the LAWA-attribute *profile depth*. Intervals, according to the program-specific number of quality classes (four for RBP and seven for LAWA), are shown on the y-axis on the right-hand side.

In our example, we associated *stream width variability* with the higher-level endpoints *channel geometry*, *sinuosity*, *channel profile*, and *channel structures* (Fig. 4). This process finally led to a hierarchy with attributes, three levels of assessment endpoints, and a main goal. Due to the combination of the different programs, some endpoints were associated with more than one attribute. Since in practice one attribute should be sufficient to assess the corresponding endpoint, it will not be necessary to include monitoring data of all 49 attributes: Endpoints can either be assessed by implementing monitoring data for the attributes of just one program or by combining attributes from different programs. Depending on which attribute data are available, the weights of some hierarchical branches may need to be adjusted (explained in Section 2.6). In any case, while one attribute is sufficient, but may sometimes be a rather coarse measure, integrating monitoring data of more attributes solidifies the assessment results and decreases their uncertainty.

Although this data flexibility is beneficial, we have to ensure the quality of the assessment. Therefore, we identified a minimal, manageable subset of information needed to perform the assessment (Boulton, 1999). We defined all assessment endpoints at the first hierarchical level to be mandatory, i.e., *channel structure*, *flow features*, *longitudinal connectivity*, *river banks*, and *surrounding landscapes* (Fig. 4). Each of them represents an important aspect of river hydromorphology. At the next lower level, mandatory and optional endpoints were defined. We decided to define endpoints to be optional when they are partly covered by their partner endpoints at the same level. In this case, even if they are not included in the assessment we would not lose too much information. Finally, to maximize the flexibility of our approach, all endpoints at the third hierarchical level were defined to be optional (Fig. 4). However, at least one endpoint associated with a mandatory one at a higher level must be available to perform the assessment.

2.5. Step 5: check compatibility of very similar attributes that assess the same endpoints

Whenever very similar attributes from different programs are grouped to assess the same endpoint, we have to check whether they are compatible. To guarantee adequate intercalibration of the averaged assessments, they need to lead to similar results, i.e., have similar scoring functions. For instance, the endpoint *width variability* could be assessed with either the attribute *width variability* from the SMC or the one from LAWA, as the standardization of this endpoint in both protocols led to similar scoring functions. The same applied to the endpoint *structure building riverbed features*:

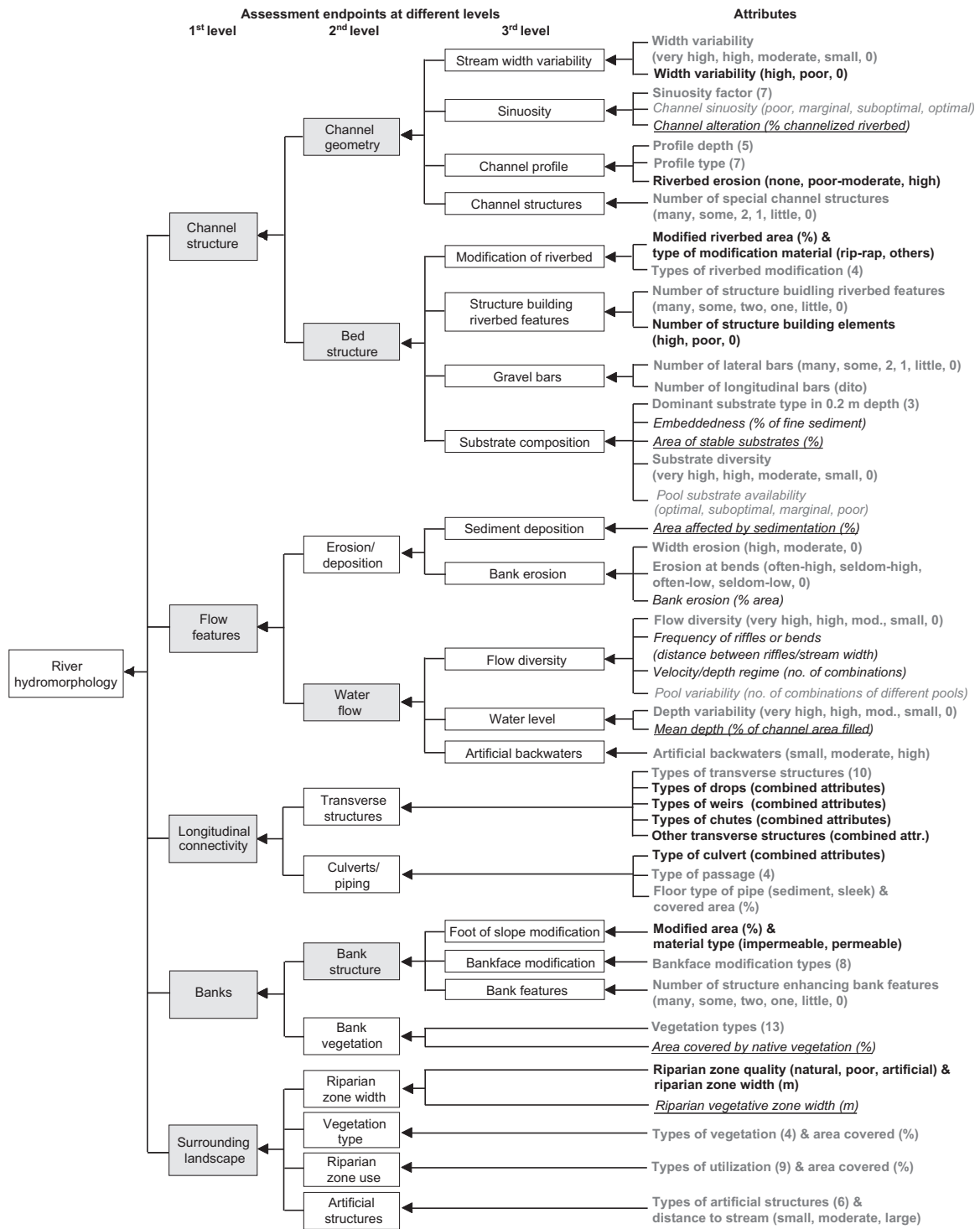
as the scoring functions of the LAWA attribute *number of structure building riverbed features* and the SMC attribute *number of structure building elements* were very similar, scores could be averaged. All other attributes arranged in groups were not similar enough to cause an intercalibration problem. Differences in scoring functions of very similar assessment endpoints from different programs may emerge, when the original assessments were calibrated according to different reference conditions. Most stream assessments compare a test stream with a natural reference stream. When natural or pristine reference streams are lacking, assessment strategies often use a desired, pursued, or near-natural condition (Verdonschot, 2000; Table 2).

2.6. Step 6: define an aggregation technique for each level of the assessment hierarchy

Assessments for lower level endpoints are merged hierarchically to scores of endpoints at the next higher level. In existing assessment programs, this aggregation step is often done by weighted or un-weighted averaging (arithmetic mean, additive aggregation). When translated into scoring functions, this is formulated as (e.g., Eisenführ et al., 2010)

$$v_{add} = \sum_{i=1}^n w_i v_i \quad (1)$$

where v_i is the score of the assessment endpoint i , w_i is the corresponding weight (all weights are normalized to sum up to 1), and v_{add} is the aggregated score of the endpoint at the next higher level. We illustrate this with the simple example of two endpoints and an un-weighted mean (both are equally important, i.e., weights $w_i = 0.5$): To evaluate the *bank vegetation* (Fig. 4), one endpoint might be assessed to have a rather bad value (e.g., *vegetation types* has a value of $v_1 = 0.2$), while the other is in a rather good state (e.g., *area covered by native vegetation* has a value of $v_2 = 0.7$). The aggregated assessment for *bank vegetation* then results in a medium value of $v_{add} = 0.45$ ($0.5 \times 0.2 + 0.5 \times 0.7$). This example immediately shows a possible drawback of additive aggregation, namely that a group of endpoints with high scores can compensate a low score of another endpoint unless this endpoint has a high weight. This property is useful if the aggregation serves primarily the purpose of averaging out assessment errors of similar endpoints. However, this property is undesired if the endpoints cover complementary aspects. To avoid this problem, some assessment programs use a minimum aggregation technique. Thereby, the score of the higher level endpoint equals the minimum of the scores of the lower endpoints (i.e., the score at the higher



Attributes from: SMC, LAWA-FS, RBP high gradient streams, RBP low gradient streams, RBP both

Fig. 4. Suggested hierarchy for the main goal *good river hydromorphology* with three levels of assessment endpoints and associated attributes (indicators). The construction of the hierarchy is based on three existing river assessment programs from Switzerland (SMC), the USA (RBP), and Germany (LAWA). Mandatory endpoints appear in gray. If an attribute is assessed discretely, the number of states the attribute can adopt appears in brackets.

level can never be better than the worst score at the next lower level of the hierarchy; in our example, the aggregated score could never be higher than the low score of 0.2 for *vegetation types*). This aggregation technique, however, leads to the undesirable result that an improvement of any endpoint except the one with the lowest score will not translate into an improved scoring. Thus, rivers are possibly assessed worse than appropriate. Additionally,

as attributes are usually assessed with a certain error, the minimum approach also tends to misclassification the more attributes are included (Heiskanen et al., 2004). Another aggregation possibility is the weighted geometric mean (Cobb–Douglas aggregation scheme; Cobb and Douglas, 1928; Varian, 2010) or mixtures of these three techniques as a compromise between additive and minimum aggregation.

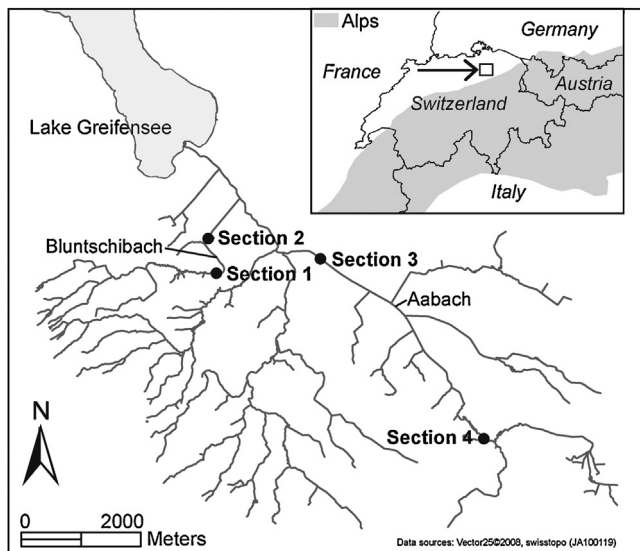


Fig. 5. Map of the four studied river sections in Switzerland.

To map the aggregation schemes of the original programs, we applied additive aggregation with equal weights (Eq. (1); with $w_i = 1/n$; arithmetic mean) to all levels of the new, integral hierarchy. In case of missing data, only endpoints for which data were available were aggregated. However, for the future, we recommend applying a mixed technique that combines additive, minimum, and weighted geometric mean. In a recent river assessment study, most experts opted for such a compromise aggregation, which circumvents some of the problems described above. It is, of course, also possible to assign different weights to different attributes or endpoints if they are considered more important than others. Correct weighting procedures are heavily discussed in the MAVT literature and the elicitation of weights from experts is not trivial (see e.g., Poyhönen and Hämäläinen, 2001; Morton and Fasolo, 2009).

While the harmonization of the individual assessment methods (step 3) is done including the original, individual aggregation schemes, merging different harmonized methods (steps 4–6) may require compromising about aggregation structure and weighting schemes. Assessment results obtained with an individual method or the merged procedure on the same quality data may therefore differ. Such differences should be analyzed and weighting schemes adapted if required.

2.7. Step 7: apply monitoring data to calculate scores of the assessment endpoints

To evaluate the performance of the new approach, we sampled four hydromorphologically distinct 100 m long river sections along two lowland rivers within the Greifensee catchment, 20 km southeast of Zurich on the Swiss plateau (Fig. 5). The four sections comprised (1) a free flowing stretch along the Bluntschibach with a mostly natural (forested) riparian zone, (2) a channelized, clear-cut section bordered on each bank by agricultural land along the Bluntschibach, (3) a channelized section with a constrained riparian zone and little riparian cover along the Aabach, and (4) a free flowing stretch along the Aabach with near-natural riparian cover, but a constrained riparian zone. Along each section, all attributes from the SMC, LAWA, and RBP programs were surveyed according to the original protocols (lowland rivers' protocol for LAWA-attributes, low gradient rivers' protocol for RBP-attributes).

2.8. Evaluation of the new approach

Attribute data were used to calculate the hydromorphological condition of all four river sections applying (i) assessments from the three original programs (SMC, LAWA, RBP), (ii) the new approach with only SMC-, only LAWA-, or only RBP-attributes, and (iii) the new approach including all attributes from the three original programs. Comparing (i) and (ii) allowed us to evaluate whether, and if so how, the harmonization and integration of the different attributes and endpoints changed the results obtained with the original programs. In principle, a deviation from original assessments is undesirable as the new approach should provide results consistent with previous ones. Contrasting (i) and (iii) revealed the consequences and prospects of integrating the original programs into a single approach.

Assessments of the original programs (i) were calculated manually (Table 2) and reported as quality classes. To calculate the assessments with the new approach (ii and iii), we implemented all scoring functions, and the aggregation techniques in an R-package "utility" (Reichert et al., 2013; <http://www.r-project.org>). Results are reported as scores between 0 and 1, and corresponding color-coded quality classes: blue for scores between 1 and 0.8, green (0.8–0.6), yellow (0.6–0.4), orange (0.4–0.2), and red for scores between 0.2 and 0. The R-package can be downloaded at no charge, which should promote its application by practitioners.

3. Results

3.1. Results from the original programs

The hydromorphological condition of the four river sections, assessed with the original programs, covered the entire range from high (=best possible condition) to bad (=worst possible condition) with the SMC and RBP (quality classes 1–5 and 1–4, respectively) (Table 3A and B; see Supplementary Table 1 for complete assessment details). With the LAWA (quality classes 2–5), it covered the range from good (=second best condition) to bad (=worst possible condition) (Table 3C). Hence, the assessments were comparable except that the LAWA evaluated the condition of river section 1 to be good instead of high as the RBP and SMC did. The discrepancy in these evaluation patterns arose because LAWA does not use the width to evaluate the riparian zone, but only vegetation types, utilization types, and artificial structures within it. The only near-natural vegetation and the minor utilization along section 1, which both were not part of the original SMC and RBP programs, explained the worse assessment applying LAWA. However, the natural extension of the riparian zone led to high assessments with the SMC and RBP.

3.2. Results from the new approach including program-specific attributes separately

Eleven out of 12 assessments calculated with the new approach including only SMC-, only LAWA-, or only RBP-attributes were the same as calculated with the original programs (Table 3D–F). An exception was section 3. While the original SMC assessed it with a quality class of 3 (Table 3A), the new approach with only SMC-attributes attributed a class 2 (Table 3D). The reason for this was that, in the new approach, the high score of the endpoint *foot of slope modification* contributed to the overall hydromorphological condition with a weight of 0.25 (Supplementary Fig. 5), but only with 0.067 in the original SMC (Supplementary Fig. 1A). The higher weight came about because of the changed assessment structure in concert with the un-weighted, additive aggregation. Additionally to the discrepancy in endpoints' weights, the original SMC averaged

Table 3

Hydromorphological assessments of the four river sections calculated with the original programs (A) SMC, (B) LAWA, (C) RBP, and the new approach including (D–F) only specific attributes of one program each, and (G) all attributes. Assessments are given in quality classes. LAWA's seven quality classes are converted into five classes: 1 and 2 become quality class 1, and 6 and 7 become quality class 5 (see main text). Continuous scores on a common scale between 0 and 1, calculated with the new approach, are added in brackets.

River section	1	2	3	4	Quality classes: high to bad
(A) SMC	1	5	3	2	1–5
(B) LAWA	2	5	3	2	1–5
(C) RBP	1	4	3	2	1–4
(D) New approach, SMC-attributes	1 (0.91)	5 (0.11)	2 (0.69)	2 (0.80)	1 (1)–5 (0)
(E) New approach, LAWA-attributes	2 (0.63)	5 (0.17)	3 (0.46)	2 (0.70)	1 (1)–5 (0)
(F) New approach, RBP-attributes	1 (0.86)	4 (0.19)	3 (0.47)	2 (0.53)	1 (1)–4 (0)
(G) New approach, all attributes	2 (0.73)	5 (0.20)	3 (0.49)	2 (0.69)	1 (1)–5 (0)

foot of slope modification with three additional endpoints to the next higher level. These were of worse quality, and therewith diluted the high quality of the foot of slope along river section 3. Hence, this is an example of possible problems associated with the use of an additive aggregation technique for a larger number of attributes: it can result in a loss of information provided by one of the single attributes (be it better or worse) compared with the others.

3.3. Results from the new approach including all attributes

The new approach including all attributes from the different programs assessed the four river sections with scores from 0.20 to 0.73, where 0 is the worst and 1 the best-possible case (shown exemplarily for section 1 in Fig. 6, and for sections 2–4 in the Supplementary Fig. 6) and corresponding quality classes from good to bad (quality classes 2–5, where 1 is the worst- and 5 is the best-possible case, Table 3G). These results mirrored the assessments calculated with the original programs. An exception was river section 1, which was assessed one quality class worse with the new approach than with the original SMC and RBP (but corresponded to the LAWA assessment). This difference arose from the merging of attributes as explained in the previous paragraphs, resulting in a worse assessment of the riparian zone (compared to the SMC and RBP) and the longitudinal connectivity (compared to the SMC).

4. Discussion

Current river assessment efforts are extremely valuable. However, ongoing difficulties e.g., when making comprehensive assessments on large spatial scales, or informing river management to plan for ecosystem recovery indicate that these programs are often insufficient (but see Bunn et al., 2010). Along the multinational Danube River, for instance, river catchment management is coordinated, but mostly relies on quality data evaluated according to different national assessment methods (Birk et al., 2012b). Hence, the resulting quality classifications are not comparable among adjoining countries, and have to be harmonized in a costly and laborious intercalibration exercise (Heiskanen et al., 2004), before they can inform multinational river management projects.

Here, we present a new approach which accounts for these problems. We have developed a step-by-step guide on how to structure existing assessment programs in a hierarchical way, and standardize attribute-specific scoring judgements to a common scale. This allows harmonizing assessments from different programs, and integrating them into a single assessment. By following the seven steps developed in this study, we exemplarily harmonized and integrated three existing hydromorphological assessment programs from the USA, Germany, and Switzerland, and assessed four river sections with different morphological conditions. The application demonstrated that our approach is

practicable, effective, accurate, and effective, all of which are important characteristics of river assessment strategies (Boulton, 1999).

4.1. Practicability

Structuring assessment endpoints of the three original programs hierarchically before integrating them into the new approach enhanced the visualization of attributes, endpoints, aggregation structures, and weights of attributes and endpoints. This visualization helped to identify elements from existing programs, which we wanted to include or improve in the new approach. The original SMC program, for instance, does not culminate in a quantification of the overall hydromorphological condition. Rather, *riverbed structure*, *riparian zone*, and *longitudinal connectivity* are only assessed individually (BAFU, 2006; Supplementary Fig. 1). However, with the new approach it was possible to aggregate these individual assessments into an overall score. It can be very useful to have such an aggregated assessment measure, e.g., to communicate assessment results to non-experts. For instance, the Australian Ecosystem Health Monitoring Program produces, among more detailed assessment results, single quality scores for each river catchment (EHMP, 2008). These scores are presented to policy makers in a public event providing transparent reporting to the public about regional river conditions (Bunn et al., 2010). As Fig. 6 demonstrates, our approach enables an aggregation into such an overall score without losing information regarding the condition of endpoints at lower hierarchical levels. Information at these levels is crucial to help interpreting probable causes of deficits knowing that some attributes respond more closely to different impacts than others (Bunn et al., 2010).

The hierarchical structure also helped comparing weights of original endpoints with weights assigned to endpoints in the new approach. A transparent communication of these weights is especially important in the new approach, as its flexible character allows including new or excluding unwanted attributes and endpoints. This process may modify some of the weights if they are not defined per se (see Section 4.2 for further explanations). Flexible assessment strategies will become more important in the future when more recent environmental problems, such as invasive species (Hermoso and Clavero, 2012) need to be added as additional endpoints to existing river assessment. Further, they will facilitate replacing original attributes with more complex ones which will soon be accessible due to recent technological advances (Boulton, 1999). In any case, whenever elements within an assessment change, the endpoints' weights have to be adjusted accordingly. This can quite easily be done when endpoints and their weights are displayed and structured hierarchically as suggested in our approach.

Finally, aggregating the endpoints into several hierarchical levels prevented the 'dilution' of the impact of single attributes or endpoints in the new approach. In contrast, the original RBP averages ten endpoints into the overall hydromorphological

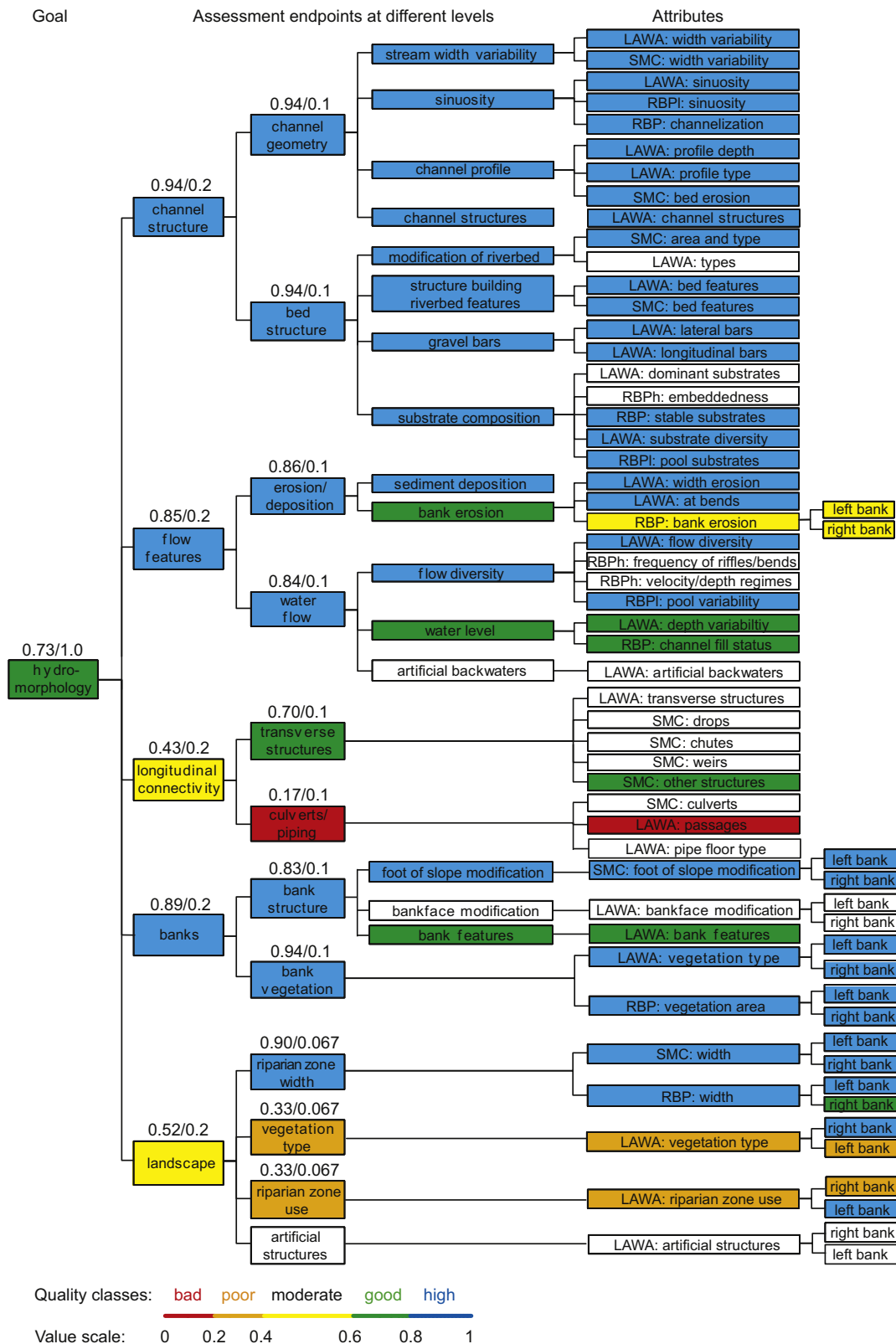


Fig. 6. Quality assessment calculated with the new approach including all attributes from the three original assessment programs (RBP, SMC, and LAWa) exemplarily shown for river section 1. The numbers above the boxes show 1) the scores on a common scale from 0 (worst possible condition) to 1 (best-possible condition), calculated with the new approach, and 2) the weights for the higher level endpoints. Scores were further separated into colored quality classes. RBP_L = RBP-attributes for low gradient streams, RBP_H = RBP-attributes for high gradient streams, RBP = RBP-attributes for both.

condition (Plafkin et al., 1989; Barbour et al., 1999). This property is useful to offset assessment errors of similar endpoints (Schuwirth et al., 2012). However, if some of these endpoints are complementary, it may be more reasonable to use a multi-level approach assigning complementary endpoints to separate, higher level endpoints.

4.2. Accuracy and effectiveness

The new approach produced assessment results that reflected river conditions quantified with the original programs. This is important to provide data continuity when switching between the original and the new assessment strategy. Causes for deviations in the results could be explained by the changes made in the assessment structures (from the programs' original structures to the integrated one), and consequential changes in attributes' weights. The three original assessments average scores of the lower level endpoints to higher level endpoints or to the main goal without specifying weights for single endpoints explicitly (Bundi et al., 2000; Plafkin et al., 1989; LAWA, 2000, 2002; Table 2). This strategy yields to attributes' or endpoints' weights which are not defined per se, but depend on the number of aggregated elements and hierarchical levels (Poyhönen and Hämäläinen, 2001). For instance, each of ten averaged attributes receive half the weight compared to a case where only five attributes are averaged to the final score, if they are located on the same level. When more levels are included, weights are assigned top-down from the main goal (which has a weight of 1) to the attributes: In this case the weights of the endpoints that quantify the main goal are defined by the goal's weight (1) divided by the number of endpoints. Each of these weights is then divided by the number of the next lower level endpoints and so forth (Supplementary Fig. 1A–C). To mirror original weights in the integrated assessment, we could therefore adjust the hierarchical assessment structure. If this does not lead to the desired effect, we could also use weighted averaging instead of just average scores, and assign specific weights to individual elements.

Combining the three original programs into an integral assessment led to an assessment that covered a wide range of morphological river characteristics, some of which were very similar, others were complementary. In theory, such a strategy should lead to more precise assessment results as it considers all indicators of a wide range of programs. Additionally, due to the averaging of similar indicators from several programs, it leads to more robust results. But how much the assessment accuracy increases, and whether river assessment in general would benefit from it, has to be investigated in the future. In any case with the new approach, practitioners have more flexibility in the choice of attributes to evaluate. Moreover, organizations with higher budgets, which collect more data, can use all data for their assessment. Finally, due to the fact that the integral assessment approach is built on original programs it is highly effective: attributes can further be collected, processed, and assessed according to the original protocols.

4.3. Limits and challenges

Despite the approach's transparency and high flexibility described above, there are some limits and challenges that should be considered. First, only protocols that calibrate the assessment of their endpoints to the same benchmark, typically sites in natural or least-disturbed conditions (Hawkins et al., 2010) can be merged directly. If programs use different reference conditions for the best (or worst) state, the scoring functions need to be harmonized by adjusting the range of the y-axis of the scoring function. Second, merging a lot of programs can make the approach confusingly complex. The optimal number, however, can not be defined per se but may depend on the complexity of the single programs

comprising the number of attributes, endpoints, hierarchical levels and the variety of applied aggregation schemes. Third, the elicitation of new scoring functions can be very time-consuming (Schuwirth et al., 2012). Such elicitations become necessary if we want to include new attributes or assessments for river types that are not already described in the original protocol (Verdonschot, 2006). However, this difficulty is not specific to our approach: adjusting original assessment programs to new river types or including new attributes is difficult and time-consuming in any case. Fourth, our approach can only be as accurate and detailed as the description of the original protocol it is based on. If for instance it does not clearly define whether the steps among different quality classes signify an equal improvement in quality, we need assumptions to harmonize the assessments. Further, assessments using multimetric indices (Stoddard et al., 2008) or modeled expected conditions that are compared to observed ones (e.g., RIV-PACS (Clarke et al., 2003), AusRivAS (Parsons et al., 2002, 2003)) could be harmonized, but the multiple information the models or indices aggregate are difficult to represent in detail with our approach. Finally, as so far most assessment programs apply discrete quality classes, working with continuous assessment scores may be unfamiliar at first.

5. Conclusions and implications for river management

From the experience gained through our case study, we synthesized six main advantages of the new approach:

1. Harmonizing original assessments facilitates the direct comparison of individual assessment methods and allows their combination into a single procedure that produces comparable results.
2. The development of a hierarchical assessment structure increases the transparency of the assessed river characteristics and facilitates the visualization of deficiencies.
3. The possibility of defining mandatory and optional endpoints prevents assessments based on inappropriately sparse data sets, at the same time promoting flexibility if more data become available. Including more information improves the significance of the results.
4. Assessments in the form of continuous scores on a common scale between 0 and 1 facilitate the direct comparison of different results. Additional color-coded quality classes simplify a quick identification of deficits, and facilitate communicating results for different purposes (e.g., action implementation versus information) or to different parties (e.g., experts versus politicians).
5. Any aspect of river quality assessment can be harmonized and integrated: e.g., macroinvertebrate indices, water quality measurements, or a combination of them, depending on the main goal of the assessment. When more programs and therewith more attributes are considered, the new approach becomes more holistic, but also more complex (e.g., the objectives hierarchy). However, because of the structured approach and the visualization, the gist of the results can easily be captured.
6. Using existing assessment programs as a source for input data is wise: it is cost-effective, and data are often comprehensive and available on a large spatial scale.

Besides river assessment, river management decisions such as rehabilitation prioritization involve an assessment process, namely the evaluation of the predicted outcomes of measures (Prato, 2003; Reichert et al., 2007; Steel et al., 2008; Hermoso and Clavero, 2012). To facilitate decision making in these cases, river quality assessment should be an integral component of river management (Nichols and Williams, 2006; Bunn et al., 2010). Our approach offers

such an integration: instead of measured attribute data, predictions of how attributes respond to potential rehabilitation measures can be used to calculate the expected future river condition on the continuous, common scale between 0 and 1. The alternative measures can then be prioritized according to this score. The management decision is supported by this ranking, but at least as much by the insights gained through the structured results of the assessment procedure.

Acknowledgements

This research was supported by a Discretionary Fund from the Swiss Federal Institute of Aquatic Science and Technology (Eawag). We thank Jacqueline Schlosser for help in the field, and Daniel Hering, Chris Robinson, Bernd Klauer, Peter Pollard, and two anonymous reviewers who provided helpful comments on earlier versions of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecolind.2013.03.027>.

References

- Astin, L., 2006. Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River Basin, USA. *Ecol. Indic.* 6, 664–685.
- Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B., 1999. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton Benthic Macroinvertebrates and Fish*, 2nd ed. EPA 841-B-99-002, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.
- BAFU, 1998. *Methoden zur Untersuchung und Beurteilung der Fließgewässer in der Schweiz. Ökomorphologie Stufe F (flächendeckend)*. Umwelt-Vollzug, Bundesamt für Umwelt, Wald, und Landschaft, Bern.
- BAFU, 2006. *Methoden zur Untersuchung und Beurteilung der Fließgewässer in der Schweiz. Ökomorphologie Stufe S (systembezogen)* Entwurf vom Juli 2006. Umwelt-Vollzug, Bundesamt für Umwelt, Bern.
- Barbour, M.T., Swietlik, W.F., Jackson, S.K., Courtemanch, D.L., Davies, S.P., Yoder, C.O., 2000. Measuring the attainment of biological integrity in the USA: a critical element of ecological integrity. *Hydrobiologia* 422/423, 453–464.
- Beechie, T.J., Sear, D.A., Olden, J.D., Press, G.R., Buffington, J.M., Moir, H., Roni, P., Pollock, M.M., 2010. Process-based principles for restoring river ecosystems. *Bioscience* 60, 209–222.
- Birk, S., Hering, D., 2006. Direct comparison of assessment methods using benthic macroinvertebrates: a contribution to the EU Water framework directive intercalibration exercise. *Hydrobiologia* 566, 401–415.
- Birk, S., van Kouwen, L., Willby, N., 2012a. Harmonising the bioassessment of large rivers in the absence of near-natural reference conditions – a case study of the Danube River. *Freshw. Biol.* 57, 1716–1732.
- Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W.V., Zampoukas, N., Hering, D., 2012b. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. *Ecol. Indic.* 18, 31–41.
- Boulton, A.J., 1999. An overview of river health assessment: philosophies, practice, problems and prognosis. *Freshw. Biol.* 41, 469–479.
- Buffagni, A., Erba, S., Furse, M.T., 2007. A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. *Environ. Sci. Policy* 10, 709–724.
- Bundi, U., Peter, A., Frutiger, A., Hütte, M., Liechti, P., Sieber, U., 2000. Scientific base and modular concept for comprehensive assessment of streams in Switzerland. *Hydrobiologia* 422/423, 477–487.
- Bunn, S.E., Abal, E.G., Smith, M.J., Choy, S.C., Fellows, C.S., Harch, B.D., Kennard, M.J., Sheldon, F., 2010. Integration of science and monitoring of river ecosystem health to guide investments in catchment protection and rehabilitation. *Freshw. Biol.* 55, 223–240.
- Burger, J., 2006. Bioindicators: a review of their use in the environmental literature 1970–2005. *Environ. Bioindicat.* 1, 136–144.
- Cao, Y., Hawkins, C.P., 2011. The comparability of bioassessments: a review of conceptual and methodological issues. *J. N. Am. Benthol. Soc.* 30, 680–701.
- Clarke, R.T., Wright, J.F., Furse, M.T., 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecol. Model.* 160, 219–223.
- Clemen, R.T., 1996. *Making Hard Decisions*, 2nd ed. PWS-Kent, Boston.
- Cobb, C.W., Douglas, P.H., 1928. A theory of production. *Am. Econ. Rev.* 18, 139–165.
- Corsair, H.J., Bassman Ruch, J., Zheng, P.Q., Hobbs, B.F., Koonce, J.F., 2009. Multi-criteria decision analysis of stream restoration: potential and examples. *Group Decis. Negot.* 18, 387–417.
- Diamond, J., Stribling, J.R., Huff, L., Gilliam, J., 2012. An approach for determining bioassessment performance and comparability. *Environ. Monit. Assess.* 184, 2247–2260.
- Dyer, J.S., Sarin, R.K., 1979. Measurable value functions. *Oper. Res.* 27 (4), 810–822.
- EHMP, 2008. Report Card 2008 for the Waterways and Catchments of South East Queensland, Ecosystem Health Monitoring Program, South East Queensland Healthy Waterways Partnership.
- Eisenführ, F., Weber, M., Langer, T., 2010. *Rational Decision Making*. Springer-Verlag, Berlin/Heidelberg.
- Erba, S., Furse, M.T., Balestrini, R., Christodoulides, A., Ofenböck, T., van der Bund, W., Wasson, J.-G., Buffagni, A., 2009. The validation of common European class boundaries for river benthic macroinvertebrates to facilitate the intercalibration process of the Water Framework Directive. *Hydrobiologia* 633, 17–31.
- European Commission, 2000. Directive 2000/60/EC of the European Council and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, Official Journal of the European Commission, L3271–L3272.
- Feio, J.M., Almeida, S.F.P., Craveiroand, S.C., Calado, A.J., 2009. A comparison between biotic indices and predictive models in stream water quality assessment based on benthic diatom communities. *Ecol. Indic.* 9, 497–507.
- Fitzpatrick, F.A., Waite, I.R., D'Arconte, P.J., Meador, M.R., Maupin, M.A., Gurtz, M.E., 1998. Revised methods for characterizing stream habitat in the National water-quality assessment program, Water-Resources Investigations Report 98-4052, U.S. Geological Survey, Raleigh, North Carolina.
- Gallo, K., 2002. *Aquatic and Riparian Effectiveness Monitoring Program For the Northwest Forest Plan*. U.S. Forest Service, Corvallis, Oregon.
- Ghetti, P.F., Bonazzi, G., 1977. A comparison between various criteria for the interpretation of biological data in the analysis of the quality of running waters. *Water Res.* 11, 819–831.
- Hawkins, C.P., Paulsen, S.G., Sickle, J.V., Yuan, L.L., 2008. Regional assessments of stream ecological condition: scientific challenges associated with the USA's national Wadeable Stream Assessment. *J. N. Am. Benthol. Soc.* 27, 805–807.
- Hawkins, C.P., Olson, J.R., Hill, R.A., 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. *J. N. Am. Benthol. Soc.* 29, 312–343.
- Heink, U., Kowarik, I., 2010. What are indicators? On the definition of indicators in ecology and environmental planning. *Ecol. Indic.* 10, 584–593.
- Heiskanen, A.-S., van den Bund, W., Cardoso, A.C., Nöges, P., 2004. Towards good ecological status of surface waters in Europe – interpretation and harmonisation of the concept. *Water Sci. Technol.* 49, 169–177.
- Hermoso, V., Clavero, M., 2012. Revisiting ecological integrity 30 years later: non-native species and the misdiagnosis of freshwater ecosystem health. *Fish Fish.*, DOI: 10.1111/j.1467-2979.2012.00471.x.
- Hughes, R.M., Paulsen, S.G., Stoddard, J.L., 2000. EMAP-Surface Waters: a multi-assemblage, probability survey of ecological integrity in the U. S. A. *Hydrobiologia* 422/423, 429–443.
- Hughes, R.M., Herlihy, A.T., Kaufmann, P.R., 2010. An evaluation of qualitative indexes of physical habitat applied to agricultural streams in ten U.S. states. *J. Am. Water Resour. Assoc.* 46, 792–806.
- Keeney, R.L., 1982. Decision analysis: an overview. *Oper. Res.* 30, 803–838.
- Keeney, R.L., Raiffa, H., 1976. *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge University Press, Cambridge, United Kingdom.
- Klauer, B., Drechsler, M., Messner, F., 2006. Multicriteria analysis under uncertainty with IANUS – method and empirical results. *Environ. Plan. C* 24, 235–256.
- LAWA, 2000. *Gewässerstrukturgütekartierung in der Bundesrepublik Deutschland – Verfahren für kleine und mittelgroße Fließgewässer*, Empfehlung, Länderarbeitsgemeinschaft Wasser.
- LAWA, 2002. *Gewässerstrukturgütekartierung in der Bundesrepublik Deutschland – Übersichtsverfahren*, Empfehlungen Oberirdische Gewässer, Länderarbeitsgemeinschaft Wasser.
- Monaghan, K.A., Soares, A.M.V.M., 2012. Bringing new knowledge to an old problem: building a biotic index from lotic macroinvertebrate traits. *Ecol. Indic.* 20, 213–220.
- Morton, A., Fasolo, B., 2009. Behavioural decision theory for multi-criteria decision analysis: a guided tour. *J. Oper. Res. Soc.* 60, 268–275.
- Nichols, J.D., Williams, B.K., 2006. Monitoring for conservation. *Trends Ecol Evol* 21, 668–673.
- NRCS (Natural Resources Conservation Service), 1998. *Stream visual assessment protocol*, Technical Note 99-1, Natural Resources Conservation Service, Washington, D.C.
- NRCS (Natural Resources Conservation Service), 2009. *Stream Visual Assessment Protocol 2*, Natural Resources Conservation Service, Washington, D.C.
- Parsons, M., Ransom, G., Thoms, M., Norris, R.H., 2002. *Australian River Assessment System: AusRivAS Physical and Chemical Assessment Module*, Monitoring River Health Initiative Technical Report no 23, Commonwealth of Australia and University of Canberra, Canberra.
- Parsons, M., Thoms, M.C., Norris, R.H., 2003. Development of a standardized approach to river habitat assessment in Australia. *Environ. Monit. Assess.* 98, 109–130.
- Poyhönen, M., Hämäläinen, R.P., 2001. On the convergence of multiattribute weighting methods. *Eur. J. Oper. Res.* 129, 569–585.
- Plafkin, J.L., Barbour, M.T., Porter, K.D., Gross, S.K., Hughes, R.M., 1989. *Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish*. EPA/444/4-89-001. U.S. Environmental Protection Agency, Office of Water, Washington.

- Prato, T., 2003. Multiple-attribute evaluation of ecosystem management for the Missouri River system. *Ecol. Econ.* 45, 297–309.
- Ranking, E.T., 1989. The Qualitative Habitat Evaluation Index (QHEI), Rationale, Methods, and Application. Ohio EPA, Columbus, Ohio.
- Ranking, E.T., 2006. Methods for assessing habitat in flowing waters: using the qualitative habitat evaluation index (QHEI), Technical Report EAS/2006-06-01, Ohio EPA, Groveport, Ohio.
- Raven, P.J., Holmes, N.T.H., Charrier, P., Dawson, F.H., Naura, M., Boon, P.J., 2002. Towards a harmonized approach for hydromorphological assessment of rivers in Europe: a qualitative comparison of three survey methods. *Aquat. Conserv.* 12, 405–424.
- Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Spörri, C., Tockner, K., Truffer, B., 2007. Concepts of decision support for river rehabilitation. *Environ. Modell. Softw.* 22, 188–201.
- Reichert, P., Schuwirth, N., Langhans, S.D. Constructing, evaluating and visualizing value and utility functions for decision support. *Environ. Modell. Softw.*, <http://dx.doi.org/10.1016/j.envsoft.2013.01.017>, in press.
- Schuwirth, N., Reichert, P., Lienert, J., 2012. Methodological aspects of multi-criteria decision analysis for policy support: a case study on pharmaceutical removal from hospital wastewater. *Eur. J. Oper. Res.* 220, 472–483.
- Solimini, A.G., Ptacnik, R., Cardoso, A.C., 2009. Towards holistic assessment of the functioning of ecosystems under the Water Framework Directive. *TrAC* 28, 143–149.
- Steel, E.A., Fullerton, A., Caras, Y., Sheer, M.B., Olson, P., Jensen, D., Burke, J., Maher, M., McElhany, P., 2008. A spatially explicit decision support system for watershed-scale management of salmon. *Ecol. Soc.* 13, 50–81.
- Stoddard, J.L., Peck, D.V., Paulsen, S.G., Van Sickle, J., Hawkins, C.P., Herlihy, A.T., Hughes, R.M., Kaufmann, P.R., Larsen, D.P., Lomnický, G., Olsen, A.R., Peterson, S.A., Ringold, P.L., Whittier, T.R., 2005. An ecological assessment of western streams and rivers. EPA 620/R-05/005. U.S. Environmental Protection Agency, Washington, D.C.
- Stoddard, J.L., Herlihy, A.T., Hill, B.H., Hughes, R.M., Kaufmann, P.R., Klemm, D.J., Lazorchak, J.M., McCormick, F.H., Peck, D.V., Paulsen, S.G., Olsen, A.R., Larsen, D.P., Van Sickle, J., Whittier, T.R., 2006. Mid-Atlantic Integrated Assessment (MAIA): State of the Flowing Waters Report. EPA/620/R-06/001. U.S. Environmental Protection Agency, Washington, D.C.
- Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., Tarquinio, E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. *J. N. Am. Benthol. Soc.* 27, 878–891.
- USEPA, 1992. Framework for Ecological Risk Assessment, EPA/630/R-92/001, Risk Assessment Forum, Washington, D.C.
- USEPA, 1997. Interim Final, Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments, EPA 540/R-97/006, Office of Solid Waste and Emergency Response, Washington, D.C.
- USEPA, 1998. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F, Risk Assessment Forum, Washington, D.C.
- Varian, H.R., 2010. *Intermediate Microeconomics: A Modern Approach*. W.W. Norton and Company, New York.
- Verdonschot, P.F.M., 2000. Integrated ecological assessment methods as a basis for sustainable catchment management. *Hydrobiologia* 422/423, 389–411.
- Verdonschot, P.F.M., 2006. Evaluation of the use of Water Framework Directive typology descriptors, reference sites and spatial scale in macroinvertebrate stream typology. *Hydrobiologia* 566, 39–58.
- Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Reidy Liermann, C., Davies, P.M., 2010. Global threats to human water security and river biodiversity. *Nature* 467, 555–561.
- Weiss, A., Mauskova, M., Matschullat, J., 2008. Hydromorphological assessment within the EU-Water Framework Directive – trans-boundary cooperation and application to different water basins. *Hydrobiologia* 603, 53–72.