The method matters: a guide for indicator aggregation in ecological assessments

Simone D. Langhans[1,*], Peter Reichert, Nele Schuwirth

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

* Corresponding author at:
Leibniz-Institute for Freshwater Ecology and Inland Fisheries
Müggelseedamm 310, 12587 Berlin, Germany
Tel.: +49 (0)30 64181 618
Fax.: +49 (0) 30 64181 750
E-mail address: langhans@igb-berlin.de

nele.schuwirth@eawag.ch (N. Schuwirth)
reichert@eawag.ch (P. Reichert)

[1] Present address: Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany

Abstract

Ecological assessment requires the integration of many physical, chemical, and/or biological quality elements. The choice of the aggregation method of such partial assessments into an overall assessment can considerably affect the assessment outcome – an issue that has been controversially discussed within the scientific community for the last decade. Current practice often considers only two different aggregation methods, the weighted arithmetic mean (additive aggregation) and the one-out, all-out method (minimum aggregation). However, both have important drawbacks. Additive aggregation compensates a bad status of one quality element by a number of elements featuring good status. Minimum aggregation can lead to overly pessimistic assessment results, since only the quality element in the worst status is considered. Here, we introduce a toolbox containing current and new aggregation methods, demonstrate and discuss their properties with simple, didactical examples, and suggest in which situations best to use them. Then, we illustrate the consequences of selected aggregation schemes for ecological river assessment with the case study of the Swiss Modular Concept of stream assessment (SMC), which we apply to ten river reaches in the Mönchaltdorfer Aa catchment in Switzerland. To be able to do so, we used multi-criteria decision analysis, i.e., multi-attribute value theory, to arrange the SMC quality elements into an objectives hierarchy, and to translate their individual assessments into value functions. Our case study revealed that choosing the most appropriate aggregation method particularly matters, if objectives with significantly different qualities are aggregated. We argue that redundant objectives (i.e., quality elements), often found at the lower levels of the objectives hierarchy, should best be aggregated additively allowing for compensation to increase the statistical significance of the results. Further, we suggest that complementary sub-objectives that often occur at higher levels may be optimally aggregated with a mixture of additive and minimum aggregation. Such a mixed method will allow some compensation, but nevertheless penalize for very bad states. Since here we compare commonly used aggregation methods with some which we believe have never been discussed in an assessment context before, our study concurrently informs ecological assessment in theory and in practice.

Key words: aggregation, multi-criteria decision analysis, multi-attribute value theory, one-out all-out, river assessment, Water Framework Directive

1. Introduction

Over the last few decades, freshwaters have suffered from a multitude of pressures resulting in poor ecosystem condition and a drastic decrease in biodiversity (Dudgeon et al., 2006; Vörösmarty et al., 2010). In an attempt to address these problems, a key management issue is to assess the ecological status of freshwater ecosystems, to identify the main perturbations responsible for the observed condition, and to put regulations and recommendations into place for ecosystem recovery (e.g., the EU Water Framework Directive (WFD); European Commission, 2000). Thereby, to provide a balanced view of the ecological status of freshwaters, evidence from multiple biological, chemical, physical and hydrological quality elements is usually combined (Moss et al., 2003).

Examples of such comprehensive ecological river assessment schemes are widespread including for instance the WFD´s Common Implementation Strategy for the classification of ecological status (European Commission, 2003), the Ecosystem Health Monitoring Program in South East Queensland, Australia (Bunn et al., 2010), the National Rivers and Streams Assessment (USEPA, 2013), and the Swiss Modular Concept for Stream Assessment (SMC; Bundi et al., 2000). Although these programmes differ in the selection of indicators and the spatial and temporal monitoring scheme, all of them integrate different quality elements to higher-level indices, and some even to an overall score (Bunn et al., 2010; European Commission, 2000). This integration is usually done by using the one-out, all-out rule (hereafter referred to as the minimum aggregation; European Commission, 2005; USEPA, 2013), by averaging (hereafter referred to as the arithmetic mean or additive aggregation; Barbour et al., 1999; Plafkin et al., 1989), or by using one or the other at different hierarchical aggregation levels (LAWA, 2002; Smith and Storey, 2001).

The logic behind the application of a minimum aggregation in ecological assessment is that a river should not reach a good ecological status if any of the quality elements measured fail. This precautionary principle might be an appropriate approach for serious impacts, such as for instance a toxic level of a hazardous substance. However, for less acute pressures (Moss, 2007) the minimum aggregation increases the likelihood that we report a lower quality than the actual ecological status (Hering et al., 2010; Sandin, 2005), which is referred to as the pessimism bias (Cunningham, 2012). This source of pessimism is amplified by the number of quality elements included (Heiskanen et al., 2004). The additive aggregation, on the other hand, implies that a low value of one quality element can be compensated by large values of other quality elements. Therefore, it poses the risk of overlooking an impact, which in fact would ask for a measure.

Although a range of alternative aggregation methods exists, ecological river assessments have rarely adopted other methods than minimum aggregation or averaging. The reason for this may be the lack of studies that quantify the consequences, such alternative aggregation methods may have on the quality evaluation of comprehensive river assessment schemes.

Quantifying the ecological state of a river calls for a framework that allows assessing different elements of the river ecosystem, and aggregating these assessments to an overall score. Multi-criteria decision analysis (MCDA), specifically multi-attribute value theory (MAVT) (Eisenführ et al., 2010; Keeney, 1982; Keeney and Raiffa, 1976), offers such a framework (Corsair et al., 2009; Klauer et al., 2006; Reichert et al., 2007). In this framework, a value function represents the degree of fulfilment of the overall- or sub-objectives on a scale from zero to unity as a function of objectively measurable system properties, the attributes. In the river assessment terminology, the term attributes refers to indicators or assets (Langhans et al., 2013). To facilitate the construction of such a value function, the overall objective (aka the goal of the assessment) is broken down hierarchically in complementary sub-objectives (often referred to as assessment endpoints or quality elements in the river assessment terminology) that make the higher-level objective more concrete. The value function for the overall objective is then constructed by formulating individual value functions for each lowest-level sub-objective, as a function of a small number of attributes, and aggregating the values at higher levels. This requires the specification of value functions for all lowest-level objectives and aggregation rules at all higher levels.

If the overall objective is to reach a good ecological state of the river, the corresponding value function reflects an ecological assessment score. In other words, it quantifies the degree to which the good ecological state of the river is reached (Langhans et al., 2013). Similarly, sub-objectives describe the state of sub-systems, such as for example the invertebrate community or water quality (Fig. 1). This makes it possible to use such a value function informally for deficit analysis, as it is mostly done with traditional river assessment procedures, or to include it into a formal decision support process in environmental management (Reichert, personal communication).

The main objective of this study was to provide a toolbox containing a mix of currently applied and new aggregation methods along with some guidance on which one best to select. Hence in the following, we introduce a sequence of four generic aggregation methods that span the spectrum from allowing for full compensation of poor assessments of sub-systems to no compensation at all. In addition to these four generic types, we establish a range of alternative methods to allow for a finer resolution of the adequate degree of compensation.

We then derive important properties of the aggregation methods, and investigate how they can affect classification outcomes. To do so, we compared hypothetical examples and a monitoring dataset from ten river reaches in Switzerland assessed according to the Swiss Modular Concept of stream assessment (SMC). Thereby, we used MAVT to arrange the different SMC-quality elements in an objectives hierarchy and to translate their individual assessments into value functions. User guidance for the different aggregation methods was developed considering the properties as well as the on-ground assessment outcomes.

2. Material and methods

Aggregation methods integrate the values (which are the degrees of fulfilment of sub-objectives in decision science), v_i, to an overall value, v, representing the degree of fulfilment of the higher-level objective. An aggregation method is defined as a function f: $v = f(v_1,v_2,\ldots,v_n)$ that specifies how the higher-level value is calculated from the n values at the lower level. If all the sub-objectives are fulfilled to the same degree, it seems reasonable to assume that the higher-level objective is fulfilled to the same degree. This leads to the following condition for the aggregation function f:

$$f(v_1 = v, v_2 = v, \ldots, v_n = v) = v \quad . \tag{1}$$

In this paper, we will only consider aggregation methods that fulfil this condition.

2.1 Basic aggregation methods

To start off, we considered four generic aggregation methods that are either widely applied in river assessment (the weighted arithmetic mean (eq. 3) and the minimum aggregation (eq. 6)), or are rarely considered, but belong to the three most prominent means (the weighted geometric mean (eq. 4) and the weighted harmonic mean (eq. 5)). Note that for the aggregation methods (3) to (5), we assumed that the weights are normalized to sum up to one:

$$\sum_{i=1}^{n} w_i = 1 \quad . \tag{2}$$

2.1.1 The weighted arithmetic mean (hereafter called additive aggregation)

For additive aggregation, the aggregated value is calculated as the sum of the n values, v_i, of the sub-objectives each of them multiplied with its weight, w_i:

$$f_{\text{add}}(v_1, \ldots, v_n) = \sum_{i=1}^{n} w_i v_i = w_1 v_1 + w_2 v_2 + \cdots + w_n v_n \quad . \tag{3}$$

If the weights are equal for all elements ($w_i = 1/n$), the result is identical to the (unweighted) arithmetic mean which is often referred to as unweighted averaging (Guitouni and Martel, 1998). In decision science, the weighted arithmetic mean is called additive aggregation, which is by far the most widely used aggregation function for multi-criteria decision support (Keeney and Raiffa 1983, Eisenführ et al. 2010).

2.1.2 The weighted geometric mean

The weighted geometric mean is calculated as the product of the *n* values, v_i, of the sub-objectives, each of them taken to the power of its weight, w_i:

$$f_{\mathrm{geo}}(v_1, \dots, v_n) = \prod_{i=1}^{n} {v_i}^{w_i} = v_1^{w_1} \cdot v_2^{w_2} \cdot \dots \cdot v_n^{w_n} \quad . \tag{4}$$

If all weights are equal, the weighted geometric mean is the same as the (unweighted) geometric mean. In economics, the weighted geometric mean is also known as the Cobb-Douglas function. It was originally introduced as a production function, but later also used as a value function which is often called utility function in economics (Varian, 2010).

2.1.3 The weighted harmonic mean

The weighted harmonic mean is calculated as the inverse of the sum of the inverse values, v_i, of the sub-objectives, each of these inverse values multiplied with its weight, w_i:

$$f_{\mathrm{harmo}}(v_1, \dots, v_n) = \begin{cases} \frac{1}{\sum_{i=1}^{n} \frac{w_i}{v_i}} = \frac{1}{\frac{w_1}{v_1} + \frac{w_2}{v_2} + \dots + \frac{w_n}{v_n}} & \text{if all } v_i > 0 \\ 0 & \text{if at least one } v_i = 0 \end{cases} \quad . \tag{5}$$

If all weights are equal, the weighted harmonic mean is the same as the (unweighted) harmonic mean.

2.1.4 The minimum aggregation

For the minimum aggregation the aggregated value, *v*, is calculated as the minimum of the values, v_i, of the sub-objectives:

$$f_{\mathrm{min}}(v_1, \dots, v_n) = \min(v_1, \dots, v_n) \quad . \tag{6}$$

The minimum aggregation method comes along with the potential of a pessimism bias, which is addressed in section 1.

2.2 Mixed aggregation methods

Calculating a weighted average of the results of two basic aggregation methods (eqs. 3–6) leads to mixed aggregation methods in the form of:

$$f_{m_1\, m_2}(v_1, \dots, v_n) = \alpha \cdot f_{m_1}(v_1, \dots, v_n) + (1 - \alpha) \cdot f_{m_2}(v_1, \dots, v_n) \quad . \tag{7}$$

Thereby, the methods m_1 and m_2 can be any of the basic aggregation methods introduced in section 2.1, and α (a number between 0 and 1) and 1-α are the weights of m_1 and m_2, respectively. This weight is an additional parameter that allows us to switch continuously from method m_1 ($\alpha = 1$) to method m_2 ($\alpha = 0$) covering all compromises in between.

2.3 Reverse and mixed – reverse aggregation methods

By using reverse scores $1 - v_i$ instead of the original scores v_i, we get another set of aggregation methods in the form of:

$$f_{\mathrm{rev}\,m}(v_1, \dots, v_n) = 1 - f_m(1 - v_1, \dots, 1 - v_n) \quad . \tag{8}$$

Here again, m refers to any of the aggregation methods. Note that the additive aggregation does not change when being reversed:

$$f_{\mathrm{rev\,add}}(v_1, \dots, v_n) = f_{\mathrm{add}}(v_1, \dots, v_n) \quad . \tag{9}$$

The reverse of the minimum aggregation is the maximum aggregation

$$f_{\mathrm{rev\,min}}(v_1, \dots, v_n) = f_{\mathrm{max}}(v_1, \dots, v_n) = \max(v_1, \dots, v_n) \quad . \tag{10}$$

Finally, we can also define mixed-reverse aggregation methods by combining the reverse aggregation methods based on eq. (7).

2.4 Properties of the aggregation methods

Here, we introduce general properties of the aggregation methods, which we will evaluate in section 3 and discuss in section 4.

2.4.1 Value properties

To demonstrate the effect of the different methods introduced above, we selected three examples in each of which we aggregated two quality elements. Each example represents a specific value combination v_1 and v_2 (both values ranging between 0 and 1). Example 1 aggregates $v_1 = 0.2$ and $v_2 = 0.8$, example 2 aggregates $v_1 = v_2 = 0.5$, and example 3 aggregates $v_1 = 0.8$ and $v_2 = 0.2$. A value of 0 indicates the worst quality state of an element considered in a river-assessment context, whereas a value of 1 stands for the best possible state. In addition, to visualize the behaviour of the value function for two aggregated values ($n = 2$), we can plot isolines or a response surface as a function of these values. Isolines are curves along which a function of two variables has a constant value.

2.4.2 Trade-off properties

The most widely applied technique to elicit aggregation rules is using trade-off questions (Eisenführ et al., 2010). Such a question could for example be: How much does the water quality need to improve to compensate for a given decrease in the quality of the fish biota? Hence, trade-offs are the core property to be discussed when aggregating degrees of fulfilment of ecological requirements.

Here, we define trade-offs in values *v* to be aggregated by the following implicit equation

$$f\left(v_1, \dots, v_j + \Delta v_j, \dots, v_n\right) = f\left(v_1, \dots, v_k + \Delta v_k(\Delta v_j), \dots, v_n\right) \qquad (11)$$

for the function

$$\Delta v_k(\Delta v_j) \quad . \qquad (12)$$

This function describes the change required in the argument v_k that leads to a state with the same value as the state in which the argument v_j was increased by Δv_j.

Trade-offs between two objectives can be estimated from the curvature of the isolines. Since each aggregation method is characterized by different trade-offs, isolines differ among aggregations methods. To demonstrate the difference in method-characteristic trade-offs, we quantified the change in either v_1 or v_2 that is required to get a gain in the aggregated value of 0.05 at nine different combinations of v_1 and v_2.

As we were primarily interested in trade-off ratios (i.e., the factor between Δv_j and Δv_k that lead to the same change in the aggregated value), rather than absolute trade-offs as formulated by the equations (11) and (12), we calculate the derivative of the function (12) and evaluate it at $\Delta v_j = 0$. This is an approximation to this factor approximately valid for small changes Δv_j and Δv_k. If this derivative is equal to unity, approximately the same change in v_k as in v_j is required to get a certain change in the aggregated value; if it is equal to two, we need to change v_k by twice as much as v_j to get the same change in the aggregated value. As the function (12) is defined implicitly by equation (11), we calculate its derivative by first taking the derivative of equation (11) with respect to Δv_j. Then we solve the resulting equation for the derivative $\partial \Delta v_k / \partial \Delta v_j$, and evaluate it at $\Delta v_j = 0$. This finally leads to the following expression for the trade-off ratio:

$$t_{k,j}(v_1, \dots, v_n, 0) = \left.\frac{\partial \Delta v_k}{\partial \Delta v_j}\right|_{\Delta v_j = 0} (v_1, \dots, v_n) = \frac{\frac{\partial f}{\partial v_j}(v_1 \dots, v_n)}{\frac{\partial f}{\partial v_k}(v_1, \dots, v_n)} \quad . \qquad (13)$$

2.5 Case study: Swiss Modular Concept of stream assessment (SMC)

To illustrate the consequences of the different aggregation methods on ecological quality assessments, we first harmonized and integrated the quality elements that are part of the Swiss

Modular Concept of stream assessment. To do so, we applied the MAVT-framework as described in Langhans *et al.* (2013).

The Swiss Modular Concept of stream assessment (SMC) (Bundi et al., 2000; http://www.modul-stufen-konzept.ch) has been introduced in 1994 to assess the fulfilment of the Swiss Water Protection Law and the Water Protection Ordinance in the early 90´s (Water protection act, 2013; Water protection ordinance, 2011). It comprises different methods that assess individual quality elements including river morphology (Hütte and Niederhauser, 1998), hydrology (Pfaundler et al., 2011), nutrient concentrations (Liechti, 2010), the physical appearance (Binderheim and Göggel, 2007), fish (Schager and Peter, 2004), macroinvertebrates (Stucki, 2010), and diatom communities (Hürlimann and Niederhauser, 2007). The assessment of each quality element is based on one to several attributes, which are aggregated with a minimum (the physical appearance), a mixed additive – minimum (hydrology), or an additive method (all remaining ones; see assessment protocols). Scores for each quality element are reported as one of the five quality classes bad, poor, moderate, good, and high.

To be able to integrate the different SMC-quality elements, we first translated the method-specific scorings into value functions with a common scale from 0 to 1. This common scale represents the degree of fulfillment of the corresponding objective (Langhans et al., 2013). We then arranged the individual quality elements hierarchically at seven levels. To culminate this objectives hierarchy into the main objective of a "good ecological state" of a river, three additional objectives which were not part of the original assessment scheme, were introduced: a "good physical state" aggregating the lower level endpoint "ecomorphology", "hydrology", and "physical appearance", a good "chemical state" based on the endpoint "nutrients", and a "good biological state" aggregating the objectives "fish", "macroinvertebrates", and "diatoms" (Fig. 1).

Being arranged in an objective hierarchy, quality scores of all the objectives can be calculated bottom-up. Thereby, to aggregate the scores at the various hierarchical levels, different aggregation methods may be used. The results can be visualized by colour coding as shown in Fig. 1.

2.5.1 Didactical aggregation examples

With the help of the didactical aggregation examples, we illustrate the extent of compensation possible when applying different aggregation methods for a single endpoint (i.e., for “nutrients”, Fig.1), and for the full assessment i.e., all endpoints of the objectives hierarchy.

For the single endpoint-example, we assigned a hypothetical high state, i.e., attribute levels corresponding to a maximum score of 1.0 to all but one attribute. The one attribute (dissolved organic carbon, DOC) we set to two slightly differing bad states, i.e., attribute levels of 12 mg/L and 11.99 mg/L corresponding to an assessment score of zero or slightly higher than zero, respectively. We then calculated the score of the endpoint “nutrients” applying the additive, the geometric mean, the mixed additive – geometric mean with $\alpha = 0.2$, and the additive – minimum aggregation method with $\alpha = 0.5$.

For the full assessment-example, we constructed three different hypothetical scenarios with a characteristic combination of values for the lowest level objectives: all third level objectives in a moderate state (the “even” scenario A), the third level objectives in either a good, a moderate or a poor state (the “middle” scenario B), and the states of the objectives ranging from bad to high (the “extreme” scenario C with the physical state in a bad, the chemical state in a high, and the biological state in a moderate state). We then calculated the full assessment for each of the three scenarios applying one of the five aggregation methods at a time: additive aggregation, the weighted geometric mean, minimum, the mixed additive – geometric mean and the mixed additive – minimum aggregation with $\alpha = 0.5$.

2.5.2 On-ground aggregation examples

To assess the potential influence of the chosen aggregation method on on-ground ecological river assessments, we evaluated the ecological state of ten river reaches in the Mönchaltorfer Aa catchment (northern Switzerland, Fig. 2), using different aggregation schemes. The Mönchaltorfer Aa catchment is located on the Swiss plateau 20 km south-east of Zurich, draining a total area of 51 km^2 to the Lake Greifen. Following the SMC-methods, environmental monitoring data in this catchment has been collected since the early 90s. For this study, we used a dataset from 2004 and 2005 that included quality information on fish, diatoms, and macroinvertebrate communities, nutrient concentrations, river morphology and the physical appearance. Data on the quality of river hydrology was not available, since the Swiss assessment protocol for hydrology was only published in 2011 and has not yet been applied in the study catchment.

Following the integrated assessment scheme described in section 2.3, we calculated the quality scores of the different objectives and the overall ecological state for the ten river reaches. Thereby, we applied two different aggregation schemes: First, the status quo SMC-aggregations including additive and minimum aggregation at the lower levels, and additivity to aggregate scores from the third level upwards. The combination of additive and minimum aggregation methods mirrors commonly used assessment schemes (Bundi et al., 2000; European Commission, 2005; LAWA, 2000; Smith and Storey, 2001). Second, we applied an alternative aggregation scheme including the status quo SMC-aggregation at the lower levels, and a mixed additive – minimum aggregation with $\alpha = 0.5$ (50% additive mixed with 50% minimum aggregation) at the three highest levels. We assumed equal weights except at the fourth level, where we attributed 10% to the physical appearance and 45% each to ecomorphology and hydrology. This combination seemed appropriate, since the physical appearance reports on deficits (e.g., litter or foam etc.) which do not necessarily impact river ecology.

Calculations and visualizations of the integrated assessments were done in R with the package ´utility´ (R Development Core Team, 2008; Reichert et al., 2013).

3. Results

3.1 Properties of the aggregation methods

3.1.1 Value properties

All basic methods introduced above aggregate equal values into the same aggregated value (property A in Tab. 1, and the circles ($v_1 = v_2 = 0.5$) in Figs. 3 A–D and 4 A–B). This corresponds to our basic requirement specified by eq. (1).

Besides this commonality, they all feature different combinations of the remaining five value properties (Tab. 1). For example, the additive aggregation allows assigning different weights to the values to aggregate, and improves the aggregated value even if other values than the worst one are increased. It does not, however, lead to an aggregated value of zero, if only one of the values to aggregate is zero (Tab. 1, properties B and C).

The basic methods from additive over geometric and harmonic to minimum aggregation show a decreasing possibility of compensating a poor value of one sub-objective by good values of other sub-objectives. This can best be shown with the trade-off ratios in Tab. 1 (property G). Whereas for additive aggregation this trade-off ratio only depends on the weights, it shows an increasingly strong dependence on the ratio of the values v_k/v_j for the geometric and the harmonic aggregation. For the minimum aggregation, no compensation is possible, since the overall value only increases if the worst value improves (the ratio of the values v_k/v_j would be infinite if v_j is the smallest value to aggregate, zero if v_k is the smallest, and undefined if one of the other values is the smallest). This ease of compensation for the additive aggregation method can also be identified through the values (Fig. 3 A): the square ($v_1 = 0.2$; $v_2 = 0.8$), the triangle ($v_1 = 0.8$; $v_2 = 0.2$) and the circle ($v_1 = v_2 = 0.5$) feature the same aggregated value.

The geometric mean, the harmonic mean and the minimum aggregation methods aggregate values to zero, if one of the quality objectives features a value of zero (Tab. 1, property C). Additionally, they all evaluate balanced value combinations (i.e., the circles in Fig. 3 B–D) better than more extreme ones with the same arithmetic mean (the squares and triangles; property F in Tab. 1). Aggregation schemes that mix additive with the three methods lose the former, but keep the latter property.

Reverse aggregation methods lead to aggregated values that are larger rather than smaller for “extreme” value combinations compared to balanced ones with the same mean (Fig. 4 D, and Supplementary Figs. S4 and S5 B–D, Figs. S6 and S7). For example, the geometric mean aggregates the three pairs of values $v_1 = 0.2$ and $v_2 = 0.8$ (square), $v_1 = v_2 = 0.5$ (circle), and $v_1 = 0.8$ and $v_2 = 0.2$ (triangle) into the aggregated values 0.4, 0.5, and 0.4, whereas the reverse geometric mean leads to the aggregated values 0.6, 0.5, and 0.6, respectively (Fig. 3 B and 4 D). This shows that dissimilar objectives (i.e., the value combinations for the squares and triangles) are better evaluated than objectives with similar values (circles), if all have the same arithmetic mean.

Properties of value aggregation for the remaining basic, mixed, reverse, and mixed – reverse methods are depicted in the Supplementary Figs. S1–S7 A–D and listed in Tab. 1 (only for basic, mixed and reverse aggregations).

3.1.2 Trade-off properties

The trade-off ratios in additive aggregation are independent of the values, i.e., they are the same irrespective of the position in the diagram (Figs. 3 E and 4 E, Tab. 1). However, they do depend on the weights. This means that the value with the lower weight (v_1 in the example in Fig. 4 E) needs to improve more than the value with the higher weight (v_2) to reach the same change in the aggregated value.

The trade-off ratios in the geometric aggregation depend on the weights (Figs. 3 F and 4 F, Tab. 1) and on the values i.e., on the position in the diagram. The more the aggregated value deviates from the diagonal where $v_1 = v_2$, the more the trade-off ratios deviate from the ones for additive aggregation (w_j/w_k) (Figs. 3 F and 4 F, Tab. 1). This indicates that, if the two values v_1 and v_2 are significantly different, the larger value needs to improve more than the smaller one to achieve the same change in the aggregated value. In other words, improving the worst value leads to a better assessment than improving a better value by the same amount, if their weights are the same. Thereby, owing to the fact that the gradient in the geometric mean is indefinite for a value of zero, considering a value that is only slightly higher than zero already leads to a significant increase in the aggregated value. The same is true for the harmonic mean aggregation (Fig. 3 G).

The tendency of achieving most, when improving the worst value, increases when moving from the geometric to the harmonic mean, and reaches its maximum in the minimum aggregation method (Fig. 3, Tab. 1). Contrarily, the reverse aggregation methods tend towards achieving more, when improving the better value (Fig. 4 H, Supplementary Figs. S4–S7).

Similar to the values, mixed aggregation methods exhibit trade-off ratios in between of the methods that were combined. For instance, the trade-offs of the mixed additive – geometric mean aggregation method depend on the weights and on the aggregated value, but to a lower degree than the geometric mean does (Supplementary Figs. S2 E, S3 E).

Trade-off ratios for the remaining basic, mixed, reverse and mixed – reverse aggregation methods with equal and different weights follow the same pattern as the cases described above (Supplementary Figs. S1–S7 E–H).

3.2 Didactical SMC-aggregation examples

The single endpoint-example revealed that aggregating an endpoint in the worst state (e.g., DOC of 12 mg/L) with six objectives in high states (TP, PO_4, NO_3, NO_2, NH_4, TOC) leads to a high or a bad nutrient status, when using additive aggregation or the geometric mean, respectively (Fig. 5 A and B, lower half of the boxes). Applying the mixed arithmetic – geometric mean aggregation with $\alpha = 0.2$ still leads to a bad nutrient state (while $\alpha = 0.5$ would lead to a moderate state, not shown), whereas the mixed additive – minimum aggregation with $\alpha = 0.5$ leads to a moderate state (Fig. 5 C and D lower half of the boxes). In this example, a very small improvement in the DOC level (from 12 to 11.99 mg/L, which is still within the measurement uncertainty) leads to results shown in the upper part of the boxes (Fig. 5). While there is no change in the additive aggregation and the mixed additive – minimum aggregation, the geometric mean shows a large increase in value changing from a bad ($v = 0$) to a poor state ($v = 0.34$). The mixed additive – geometric mean aggregation with $\alpha = 0.2$ leads to a change from a bad state ($v = 0.17$) to a moderate state ($v = 0.44$), which is an improvement of the aggregated value of 0.27. This is due to the very steep increase of the value response surface of the geometric aggregation (Supporting Fig. S8; the column with $\alpha = 0$ corresponds to purely geometric aggregation).

In the full assessment-example, the five different aggregations applied to the even scenario (i.e. in which all third level objectives have a moderate quality) resulted in very similar quality scores within the moderate quality class for the overall ecological state (v between 0.43 and 0.48) (Fig. 6). In the middle scenario (i.e., objectives are either in a poor, a moderate or a good state), the minimum aggregation and the mixed additive – minimum aggregation yielded to a poor ecological state (v = 0.24 and 0.35, respectively), whereas the remaining three, i.e., the geometric mean, the mixed additive – geometric mean and the additive aggregation resulted in a moderate state (v = 0.43, 0.45 and 0.47, respectively) (Fig. 6). The extreme scenario (i.e., objectives can take values between the bad and the high state) led to a large range of scores for the ecological state, depending on the aggregation method applied. The worst score was calculated with the minimum aggregation and the geometric mean (v = 0), increasing to a poor quality with the mixed additive – minimum (v = 0.24) and the additive – geometric mean aggregations (v = 0.35), and to a moderate quality with the additive aggregation (v = 0.48) (Fig. 6).

3.3 On-ground SMC-aggregation examples

When applying the additive aggregation method, the ecological state at the ten river reaches was either in a moderate or a good quality with the worst quality at reach 168 (v = 0.46), and the highest quality at reach 437 (v = 0.73). With the mixed additive – minimum aggregation, the ecological state at the reaches exhibited a poor, a moderate or a good quality being worst at reach 183 (v = 0.34) and best again at reach 437 (v = 0.67) (Figs. 1 and 7).

On average, the ecological quality calculated with the additive aggregation method was 0.10 scores higher than when using the mixed method (SD: ± 0.05). For three of the ten river reaches, this difference led to a change in the ecological quality class, from moderate to poor.

4. Discussion

Ecological river assessment often requires the integration of different quality elements including physical, chemical, and/or biological ones. The choice of the aggregation method of such partial assessments into an overall assessment can considerably affect the assessment outcome – an issue that has been controversially discussed within the scientific community (Caroni et al., 2013; Hering et al., 2010; Moss et al., 2003). Here, we assemble a toolbox of aggregation methods, demonstrate their properties, and provide guidelines on when best to use them. The toolbox contains some commonly used methods, such as the additive and the minimum aggregation (Bundi et al., 2000; European Commission, 2005; LAWA, 2000, 2002;

Smith and Storey, 2001). However, most of them are new and have, to our knowledge, never been described or discussed in a river-assessment context before.

4.1 Aggregation effects observed in the examples

When including multiple quality elements, which are likely arranged at different hierarchical levels (Fig. 1; Bundi et al., 2000; LAWA, 2000, 2002; Raven et al., 1998), the effect of the aggregation methods on the assessment result strongly depends on the range of values which are aggregated. However, multiple values that range within the same quality class aggregate into very similar values, independent of the method applied (Fig. 7, scenario A). This is due to the property established in eq. (1), which defines that the aggregation of values which are the same results in just that value. However, if the values differ (e.g., among several quality classes), the aggregation methods lead to different results. This effect is more pronounced the more the values differ from each other (Fig. 7, scenarios B and C).

For some aggregation methods, very small changes in the sub-objectives values, which may be within the measurement error, can lead to a large change in the aggregated value. This is the case for the geometric and harmonic mean, or mixtures that include one of these methods, if a value of one of the sub-objectives is very close to zero (Figs. 3 B–C, 4 B, 5, and Supplementary Figs. S1 B–C, S2 A–C, S3 A–C, and S8).

Minimum aggregation is usually applied to give a strong penalty for serious impacts to protect the ecosystem from the most dominant pressure or combination of pressures (Caroni et al., 2013; Smith and Storey, 2001; WG ECOSTAT, 2003). However, it also leads to the effect that an aggregated value only increases if the worst sub-objective is improved (Fig. 3 D and H). In other words, the aggregated value does not increase at all if a different sub-objective than the worst one is improved.

All aggregation methods discussed in this paper feature the property that equal values of sub-objectives lead to the same aggregated value, since we think that this is an essential requirement for value aggregation (see eq. 1). This criterion is, for example, not fulfilled for the multiplicative aggregation method (Keeney and Raiffa, 1976). Therefore, we have not included this particular aggregation method in our study, although it is frequently used to aggregate utilities in decision analysis (Keeney and Raiffa, 1976).

4.2 Guidelines for choosing an optimal aggregation scheme in ecological assessments

Searching for trade-offs that lead to indifference is an excellent technique to determine the adequate shape of aggregation methods (Eisenführ et al., 2010; Keeney and Raiffa, 1976).

Thereby, it is important that the values on each axis (i.e., between 0 and 1) are translated into corresponding attribute levels. This is because the attribute levels characterize the state of the underlying system, and the values are just mathematical representations of the preferences that may depend on the attributes in a non-linear way. Since preference structures may differ among river ecologists (who develop the assessment procedures), preferred aggregation methods may differ among them too. When choosing aggregation methods to formulate ecological assessment procedures, we are not interested in the subjective preferences of individuals. We want to find inter-subjective aggregation methods on which experts in this field may agree. To facilitate finding such a consensus, we suggest some selection criteria for generic situations.

4.2.1 Aggregation of redundant sub-objectives

It was argued (Caroni et al., 2013; Langhans et al., 2013) that in environmental assessments it may be advantageous to use redundant objectives in individual branches of the objectives hierarchies. There are two main reasons for this: First, redundant objectives increase the statistical significance of often highly uncertain measurements, which otherwise may have a large influence on classification outcomes (Caroni et al., 2013). This is for instance the case in the SMC, when the state of the macroinvertebrate community can be assessed with more than one of the three proposed indices (Fig. 1; macroindex, ibgn, and ibch; Agence de l´eau, 2000; Stucki, 2010). Second, allowing for redundant objectives makes the assessment more flexible, since it can still be done if data of one of the proposed quality elements is available, but is missing for the others.

When considering redundant sub-objectives, the chosen aggregation method should avoid a bias due to uncertain observations. If we assume that the data were taken with a symmetric, random observation error and there is no strong nonlinearity in the conversion to values, then the additive aggregation method would be our method of choice.

Redundant sub-objectives are often used at low hierarchical levels, while the level of complementarity of the objectives usually increases when climbing up the objectives hierarchy. Hence, the additive aggregation method is often the appropriate method at low hierarchical levels. To avoid having a too high weight of such redundant objectives, the overall weight (and other aggregation parameters) of the whole branch must be given independently of the number of assessed sub-objectives.

4.2.2 Aggregation of complementary sub-objectives

In ecological assessments, the quality elements that are aggregated into higher-level objectives are often complementary to each other. For example, a good biological state of a river reach may be described by the state of different communities such as diatoms, invertebrates, and/or fish (Bundi et al., 2000; Smith and Storey, 2001). Since a good ecological state should reflect a good state regarding all of the underlying aspects, we may want to avoid that a poor state in one of these aspects can be fully compensated by a good state in some others. This can be achieved by choosing the minimum aggregation, as it is recommended by the Water Framework Directive Classification Guidance (European Commission, 2005; WG ECOSTAT, 2003), the geometric or the harmonic mean aggregation, or one of the mixed forms. For all of these methods, we will reach a higher improvement of the aggregated value (e.g., for the ecological river quality), if we improve the quality of the endpoint with the lowest value instead of another one with a higher value. Thereby, the minimum aggregation reflects an improvement of the worst endpoint only. This may lead to undesired outcomes, e.g. when a river management measure improves other endpoints than the worst one, and this improvement is not reflected in the assessment of the rehabilitated river reach. Therefore, we do not recommend using a pure minimum aggregation.

4.2.3 Aggregation of mutually exclusive objectives

River assessments may include complementary objectives that are mutually exclusive. This is for example the case, if we consider improving the structural diversity of a river that has only few habitats left. In such a case, we may be neutral about which habitats to restore, since a better ecological state is already reached if we are restoring one habitat type. Pander and Geist (2010), for example, found high species richness in fish communities after implementing either of four different restoration measures. Such a mutual exclusivity could be represented with a reverse aggregation scheme.

Another example, in which a reverse aggregation scheme may be favourable, is the assessment of spatially distinct data. For example, one may want that the assessment of the ecological quality of a river consisting of its individual reaches leads to a higher result, when the river reaches are in high and bad qualities instead of moderate qualities only.

4.2.4 Aggregation of strongly conflicting objectives

River assessments are often part of whole river management strategies (European Commission, 2012; Moss, 2004). These strategies may include additional objectives to the good ecological state of the river, such as good ecosystem services, low costs, conformity with regulation or a robust design (Reichert, personal communication), which are often strongly conflicting. For example, it is usually not possible to improve the ecological state of a river while saving money.

In such a situation, it will not make sense to use minimum aggregation, as it is certainly an advantage either to use less money for the same ecological state or to improve the ecological state for the same amount of money. The elicitation of trade-offs between costs and the achieved ecological state is therefore particularly important for this aggregation problem. If these trade-offs do not strongly depend on the values an additive aggregation method would be appropriate. This choice would be consistent with cost-benefit analysis, which is based on additive aggregation, as the values are expressed in monetary units (and aggregated as a net monetary benefit through adding benefits and subtracting costs). Cost-benefit analysis is often applied at this level of decision support (Brouwer and Pearce, 2005; Hanley and Spash, 1993).

4.2.5 Summary of recommendations

From the discussion in the sections 4.2.1 to 4.2.4, we identified the necessity of aggregation methods whose trade-offs depend on the values to aggregate. Thereby, both types of changes may be relevant – a higher sensitivity to the smallest value (the more typical case; see section 4.2.2) and a higher sensitivity to the largest value (see section 4.2.3).

The extreme forms of these dependencies are represented by the minimum and the maximum aggregation methods. Both of the aggregation methods feature the often undesired property that only an increase in the worst (minimum) or the best value (maximum) leads to an improvement in the overall state. Hence, we may favour a weaker form of dependency such as with the geometric or the harmonic mean.

However, the geometric and the harmonic mean as well as the corresponding reverse techniques still have the undesired property, if one of the values to aggregate is zero (basic methods) or unity (reverse methods), respectively. Additionally, these methods are very sensitive to small changes in sub-objectives´ values, if one of these is close to zero (basic methods; see section 4.1) or close to unity (reverse methods).

Thus, it seems to be most reasonable to use a mixture of additive and minimum aggregation when dealing with complementary sub-objective (section 4.2.2), or a mixture of additive and maximum aggregation for the assessment of mutually exclusive sub-objectives (see section 4.2.3). These mixed methods span continuously between additive aggregation and one or the other of the extreme methods (minimum or maximum) depending on the weight of α. (see Supplementary Fig. S9 for the effect of different α.s in the mixed additive – minimum aggregation).

This property allows identifying the aggregation method and weighting parameters for which preferences are best approximated. In so doing, we find a weight α of the additive aggregation close to unity, when aggregating redundant or strongly conflicting objectives (see sections 4.2.1 and 4.2.4, respectively). Contrarily, we argue for significantly lower weights when aggregating complementary or mutually exclusive objectives (see sections 4.2.2 for mixed minimum aggregation, and 4.2.3 for mixed maximum aggregation).

4.3 Consequences for the case study

Considering the desired aggregation properties for ecological assessment discussed in section 4.2, we designed an optimal aggregation scheme for our case study. For the lower levels (four to seven; Fig. 1), where the objectives are (partly) redundant, the SMC-status quo method (i.e., additive aggregation) was a good choice. We also kept the status-quo method to aggregate the sub-objectives of the physical appearance, since this conforms with the water protection law in Switzerland (Water protection act, 2013; Water protection ordinance, 2011).

The higher levels (three and two; Fig. 1) feature complementary objectives, for which we do not want to allow for full compensation. However, we did want an increase in the aggregated value, if any of the sub-objectives are improved. Hence, following our reasoning in section 4.2.5, we chose a mixture of additive and minimum aggregation.

Since we also wanted different weights for the endpoints physical appearance (0.1), ecomorphology (0.45) and hydrology (0.45), we assigned the same weights for all sub-objectives except for those ones. Assuming equal weights for the physical, the chemical and the biological state (Fig. 1; second level), and for diatoms, invertebrates and fish (Fig. 1; third level) seemed appropriate, although different weights may be favourable too. For example, it could be argued that the biological state should receive a higher weight, since in contrast to the physical and the chemical state, it can be seen as an integrative indicator (European Commission, 2000).

In all cases, we set the weight of $\alpha = 0.5$ for additive versus minimum aggregation. Considering the present state of knowledge, including 50% additive aggregation seems to be a reasonable compromise: The mixture with the minimum aggregation allows for some compensation, but still considerably accounts for very bad impacts (Fig. S9).

The comparison of the results for the full assessment, calculated either with the fully additive or the optimized aggregation scheme showed that already a slight change in the aggregation scheme matters for on-ground river assessment. In our case, the value of the ecological state of all ten river reaches decreased with the optimized aggregation scheme, whereby three of them even changed into a worse quality class (Fig. 7). We argue that in contrast to the minimum aggregation, which may lead to a too pessimistic assessment (Hering et al., 2010; Sandin, 2005), the "pure" additive aggregation compensated too much for bad values.

5. Conclusions

There is no simple, universal solution for river assessment aggregation. We believe that this paper will considerably help towards choosing appropriate aggregation methods in future river assessment schemes. However, we still recommend eliciting the dependence of trade-offs between the values to be aggregated for each aggregation step individually. In this way, it can be checked which aggregation method best represents the preferences of the decision makers. The properties listed in Tab. 1 can further support this selection process.

Based on the properties of the basic aggregation methods and their combinations, we suggest that most preferences may reasonably well be described by a mixture of an additive aggregation scheme, with either minimum or maximum aggregation. These aggregation methods cover a wide range of shapes of value functions with only requiring a minimum number of parameters (i.e., weights, w_i, for additive aggregation of the values, v_i, and a single additional weight, α, which defines the proportion of additive relative to the minimum or maximum aggregation method). In addition, these mixed methods span between the two extremes of additive vs. minimum or maximum aggregation – methods that are often used in currently available ecological assessment protocols (Bundi et al., 2000; LAWA, 2000, 2002; Smith and Storey, 2001). Finally, they also miss undesired properties such as very steep slopes close to zero or unity.

Ideally, the decision on the optimal aggregation scheme should also consider the uncertainty in classification originating e.g. from field samples (Caroni et al., 2013). This is particularly important when extending objectives hierarchies to partly redundant sub-objectives, because this extension should not induce a bias into the assessment procedure. Visualizing the uncertainty of river assessments at all hierarchical levels can be done with the new R-package "utility" described in Reichert et al. (2013).

Considering that we describe commonly used aggregation methods, but also some which have never been discussed in a river assessment context before, our paper informs river assessment in theory and in practice.

Acknowledgements

We thank the cantonal Office of Waste, Water, Energy and Air, and the cantonal administration for fishery and hunting in Zurich for providing monitoring data of the Mönchaltdorfer Aa, and Rosi Siber (Eawag) for contributing Figure 2. Simone D. Langhans was supported by the Alexander von Humboldt-Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at

References

Agence de l'eau, 2000. Guide technique de l'indice biologique global normalisé.

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. U.S. Environmental Protection Agency, Office of Water, Washington D.C.

Binderheim, E., Göggel, W., 2007. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Äusserer Aspekt, Umwelt-Vollzug Nr. 0701. Bundesamt für Umwelt, Bern.

Brouwer, R., Pearce, D., 2005. Cost-benefit analysis and water resources management. Edgar Elgar Publishing, Cheltenham, UK.

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, 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.

Caroni, R., van de Bund, W., Clarke, R.T., Johnson, R.K., 2013. Combination of multiple biological quality elements into waterbody assessment of surface waters. Hydrobiologia 704, 437-451.

Corsair, H.J., Bassmann Ruch, J., Zheng, P.Q., Hobbs, B.F., Koonce, J.F., 2009. Multicriteria decision analysis of stream restoration: potential and examples. Group Decision and Negotiation 4, 387-417.

Cunningham, R., 2012. Glass half full or half empty? Why 2009 Water Framework Directive classification results are over-optimistic about the state of rivers despite the one-out, all-out rule. RSPB a million voices for nature.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Leveque, C., Naiman, R.J., Prieur-Richard, A.H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. 81, 163-182.

Eisenführ, F., Weber, M., Langer, T., 2010. Rational Decision Making. Springer-Verlag Berlin, Heidelberg.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy, Official Journal of the European Commission. European Commission.

European Commission, 2003. Water Framework Directive Common Implementation Strategy working group 2.A Ecological Status (ECOSTAT). Overall approach to the classification of ecological status and ecological potential. European Commission, p. 86 pp.

European Commission, 2005. Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance document n. 14. Guidance on the intercalibration process 2004-2006. Luxembourg: Office for Publications of the European Communities.

European Commission, 2012. River Basin Management Plans. Report from the commission to the European Parliament and the Council on the implementation of the Water Framework Directive (2000/60/EC), Brussels, Belgium. Last accessed 7.1.2014.

Guitouni, A., Martel, J.-M., 1998. Tentative guidelines to help choosing an appropriate MCDA method. European Journal of Operational Research 109, 501-521.

Hanley, N., Spash, C.L., 1993. Cost-benefit analysis and the environment. Edward Elgar Cheltenham.

Heiskanen, A.S., van de Bund, W., Cardoso, A.C., Noges, P., 2004. Towards good ecological status of surface waters in Europe - interpretation and harmonisation of the concept. Water Science and Technology 49, 169-177.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., Van De Bund, W., 2010. The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of the Total Environment 408, 4007-4019.

Hürlimann, J., Niederhauser, P., 2007. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Kieselalgen Stufe F (flächendeckend), Umwelt-Vollzug Nr. 0740. Bundesamt für Umwelt, Bern.

Hütte, M., Niederhauser, P., 1998. Methoden zur Untersuchung und Beurteilung der Fliessgewässer in der Schweiz. Ökomorphologie Stufe F (flächendeckend), Mitteilungen zum Gewässerschutz Nr. 27. Bundesamt für Umwelt, Wald, und Landschaft, Bern, Switzerland.

Keeney, R.L., 1982. Decision-analysis - an overview. Operations Research 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. Environment and Planning C: Government and Policy volumes, 235-256.

Langhans, S.D., Lienert, J., Schuwirth, N., Reichert, P., 2013. How to make river assessments comparable: A demonstration for hydromorphology. Ecol. Indic. 32, 264-275.

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.

Liechti, P., 2010. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Chemisch-physikalische Erhebungen, Nährstoffe, Umwelt-Vollzug Nr. 1005. Bundesamt für Umwelt, Bern.

Moss, B., 2007. Shallow lakes, the water framework directive and life. What should it all be about? Hydrobiologia 584, 381-394.

Moss, B., Stephen, D., Alvarez, C., Becares, E., Van de Bund, W., Collings, S.E., Van Donk, E., De Eyto, E., Feldmann, T., Fernandez-Alaez, C., Fernandez-Alaez, M., Franken, R.J.M., Garcia-Criado, F., Gross, E.M., Gyllstrom, M., Hansson, L.A., Irvine, K., Jarvalt, A., Jensen, J.P., Jeppesen, E., Kairesalo, T., Kornijow, R., Krause, T., Kunnap, H., Laas, A., Lille, E., Lorens, B., Luup, H., Miracle, M.R., Noges, P., Noges, T., Nykanen, M., Ott, I., Peczula, W., Peeters, E., Phillips, G., Romo, S., Russell, V., Salujoe, J., Scheffer, M., Siewertsen, K., Smal, H., Tesch, C., Timm, H., Tuvikene, L., Tonno, I., Virro, T., Vicente, E., Wilson, D., 2003. The determination of ecological status in shallow lakes - a tested system (ECOFRAME) for implementation of the European Water Framework Directive. Aquat Conserv 13, 507-549.

Moss, T., 2004. The governance of land use in river basins: prospects for overcoming problems of institutional interplay with the EU Water Framework Directive. Land Use Policy 21, 85-94.

Pander, J., Geist, J., 2010. Seasonal and spatial bank habitat use by fish in highly altered rivers - a comparison of four different restoration measures. Ecology of Freshwater Fish 19, 127-138.

Pfaundler, M., Dübendorfer, C., Zysset, A., 2011. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Hydrologie - Abflussregime Stufe F (flächendeckend), Umwelt-Vollzug Nr. 1107. Bundesamt für Umwelt, Bern.

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. Environmental Protection Agency, Office of Water, Washington.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project. org (accessed 1 April 2013), Vienna, Austria.

Raven, P.J., Holmes, N.T.H., Dawson, F.H., Everard, M., 1998. Quality assessment using River Habitat Survey data. Aquat. Conserv. 8, 477-499.

Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Sporri, 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., 2013. 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.

Sandin, L., 2005. Testing the EC water Framework Directive "one-out, all-out" rule - simulating different levels of assessment errors along a pollution gradient in Swedish streams. Verhandlungen Internationaler Vereinigung für Limnologie 29, 334-336.

Schager, E., Peter, A., 2004. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Fische Stufe F, Mitteilungen zum Gewässerschutz Nr. 44. Bundesamt für Umwelt, Bern.

Schuwirth, N., Stamm, C., Reichert, P., 2012. Incorporation of uncertainty in decision support to improve water quality, in: Seppelt, R., Voinov, A.A., Lange, S., Bankamp, D. (Eds.), International Environmental Modelling and Software Society (iEMSs), International

Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, Leipzig, Germany, pp. 1005-1012.

Smith, M.J., Storey, A.W., 2001. Design and Implementation of Baseline Monitoring (DIBM3): Developing an Ecosystem Health Monitoring Program for rivers and streams in South East Queensland. Report to the South East Queensland Regional Water Quality Management Strategy, Brisbane.

Stucki, P., 2010. Methoden zur Untersuchung und Beurteilung der Fliessgewässer. Makrozoobenthos Stufe F, Umwelt-Vollzug Nr. 1026. Bundesamt für Umwelt, Bern.

USEPA, 2013. National Rivers and Streams Assessment 2008-2009. A collaborative survey. Draft. U.S. Environmental Protection Agency, Washington.

Varian, H.R., 2010. Intermediate microeconomics: a modern approach. W.W. Norton and Company, New York.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467, 555-561.

Water protection act, 2013. Bundesgesetz über den Schutz der Gewässer (Gewässerschutzgesetz, GSchG) vom 24. Januar 1991. Bundesversammlung der Schweizerischen Eidgenossenschaft, Bern, Switzerland.

Water protection ordinance, 2011. Gewässerschutzverordnung (GSchV) vom 28. Oktober 1998, Bern, Switzerland.

WG ECOSTAT, 2003. Overall approach to the classification of ecological status and ecological potential. Version 4.

Figure captions

Figure 1. Objectives hierarchy suggested for the Swiss Modular Concept of stream assessment (SMC). Colours indicate the quality of the individual endpoints along the river reach 183 (Mönchaltdorfer Aa; Switzerland) assessed with the status quo SMC-aggregation at levels seven to three, and a mixed additive – minimum aggregation ($\alpha = 0.5$) for higher levels. Vertical black lines within the boxes indicate the position of the value within the respective quality class.

Figure 2. Case study sites along the Mönchaltorfer Aa catchment. Circles indicate the sampling locations and star signs the location of waste water treatment plants (WWTP). Figure adapted from Langhans *et al.* (2013).

Figure 3. Properties of the four basic aggregation methods additive, geometric mean, harmonic mean, and minimum, demonstrated with the aggregation of two values (v_1, v_2; $w_1 = w_2$). A–D) Isolines of aggregated values (shown at the intersection between the quality classes) and aggregated values for three selected argument pairs depicted as squares ($v_1 = 0.2$, $v_2 = 0.8$), circles ($v_1 = v_2 = 0.5$), and triangles ($v_1 = 0.8$, $v_2 = 0.2$). E–H) Trade-offs (horizontal and vertical lines) showing the change in value on the corresponding axis required to get a gain in the aggregated value of 0.05. Dashed lines indicate that no trade-off can be found within the range up to unity that leads to such a gain. See Figure 1 for colour coding.

Figure 4. Properties of selected aggregation methods. A) and E) show additive aggregation, B) and F) the geometric mean (all four with $w_1 = w_2/2$), C) and G) mixed additive – minimum aggregation, and D) and H) the reverse geometric mean (all four with $w_1 = w_2$). See Figure 3 for further explanation and Figure 1 for colour coding.

Figure 5. The SMC-quality element "nutrients" assessed with selected aggregation methods. A) Additive, B) geometric mean, C) mixed additive – geometric mean ($\alpha = 0.2$), and D) mixed additive – minimum aggregation ($\alpha = 0.5$). The upper and the lower part of the boxes show the consequences of one attribute being in its almost worst (DOC = 11.99 mg/L) or worst state (DOC = 12mg/L), respectively, while all other attributes are in their best state. White boxes indicate missing data. Vertical black lines within the boxes indicate the position of the value within the respective quality class. Equal weights are assumed.

Figure 6. A) Results for the objectives of the three highest levels of the SMC-objectives hierarchy. The assessments are calculated for three scenarios (A, B, and C) applying five different aggregation methods each: additive, geometric mean, minimum, mixed additive – geometric mean ($\alpha = 0.5$), and mixed additive – minimum ($\alpha = 0.5$). B) Colour-coded objectives hierarchy. The upper and lower part of the boxes show row 11 and 15, respectively. Equal weights are assumed except for ecomorphology and hydrology (0.45), and physical appearance (0.1).

Figure 7. Quality scores calculated for the objectives of the three highest levels of the SMC-objectives hierarchy of the ten river reaches along the Mönchaltdorfer Aa (107–489; Fig. 2). The upper part of the boxes shows additive and the lower half mixed additive – minimum aggregation ($\alpha = 0.5$). Equal weights are assumed except for ecomorphology and hydrology (0.45), and physical appearance (0.1).

Tables

Table 1. Summary of the key properties of the basic, the mixed and the reverse aggregation methods. Extended from Schuwirth et al. (2012).

Property	Aggregation method*								
	add	geo	harmo	min	add–geo	add–harmo	geo–harmo	harmo–min	add–min
A) Equal values of sub-objectives lead to the same aggregated value (eq. 6)	yes	yes	yes	yes	yes	yes	yes	yes	yes
B) Objectives can have different weights	yes	yes	yes	no	yes	yes	yes	yes	yes
C) One sub-objective in the worst state (value = 0) leads to an overall value of 0	no	yes	yes	yes	no	no	yes	yes	no
D) The overall value is improved if any of the values is improved (not only if the worst is improved)	yes	yes**	yes**	no	yes	yes	yes**	yes**	yes
E) A low value of a sub-objective can be compensated by a high value of another sub-objective	yes	partly	partly	no	partly	partly	partly	partly	partly
F) Balanced assessment results are better evaluated than extreme ones with the same arithmetic mean	no	yes	yes	yes	yes	yes	yes	yes	yes
G) Trade-off ratios, $t_{k,j}$, according to eq. (13)	$\frac{w_j}{w_k}$	$\frac{w_j}{w_k}\frac{v_k}{v_j}$	$\frac{w_j}{w_k}\left(\frac{v_k}{v_j}\right)^2$	***	1)	2)	3)	***	***

* add = weighted arithmetic mean (eq. 1), geo = weighted geometric mean (eq. 2), harmo = weighted harmonic mean (eq. 3), min = minimum aggregation (eq. 4), add–geo, add–harmo, geo–harmo, harmo–min, and add–min according to eq. 7., rev geo, rev harmo, and max according to eqs. 8–10; ** if none of the other values is zero; *** no simple expression due to discontinuities

1) $\frac{w_j}{w_k}\ \frac{v_k}{v_j}\ \frac{\alpha v_j+(1-\alpha)f_{\mathrm{geo}}}{\alpha v_k+(1-\alpha)f_{\mathrm{geo}}}$ 2) $\frac{w_j}{w_k}\left(\frac{v_k}{v_j}\right)^2\ \frac{\alpha v_j^2+(1-\alpha)f_{\mathrm{harmo}}^2}{\alpha v_k^2+(1-\alpha)f_{\mathrm{harmo}}^2}$ 3) $\frac{w_j}{w_k}\left(\frac{v_k}{v_j}\right)^2\ \frac{\alpha v_j f_{\mathrm{geo}}+(1-\alpha)f_{\mathrm{harmo}}^2}{\alpha v_k f_{\mathrm{geo}}+(1-\alpha)f_{\mathrm{harmo}}^2}$

Table 1 continued.

Property	Aggregation method*		
	rev geo	rev harmo	max
A) Equal values of sub-objectives lead to the same aggregated value (eq. 6)	yes	yes	yes
B) Objectives can have different weights	yes	yes	no
C) One sub-objective in the worst state (v = 0) leads to an overall value of 0	no	no	no
One sub-objective in the best state (v = 1) leads to an overall value of 1	yes	yes	yes
D) The overall value is improved if any of the values is improved (not only if the worst is improved)	yes	yes	only if the best is improved
E) A low value of a sub-objective can be compensated by a high value of another sub-objective	partly	partly	no
F) Balanced assessment results are better evaluated than extreme ones with the same arithmetic mean	no, worse	no, worse	no, worse
G) Trade-offs, $t_{k,j}$, according to eq. (13)	$\frac{w_j}{w_k}\frac{1-v_k}{1-v_j}$	$\frac{w_j}{w_k}\left(\frac{1-v_k}{1-v_j}\right)^2$	****

**** depending on the order of the v_i

Figure 1

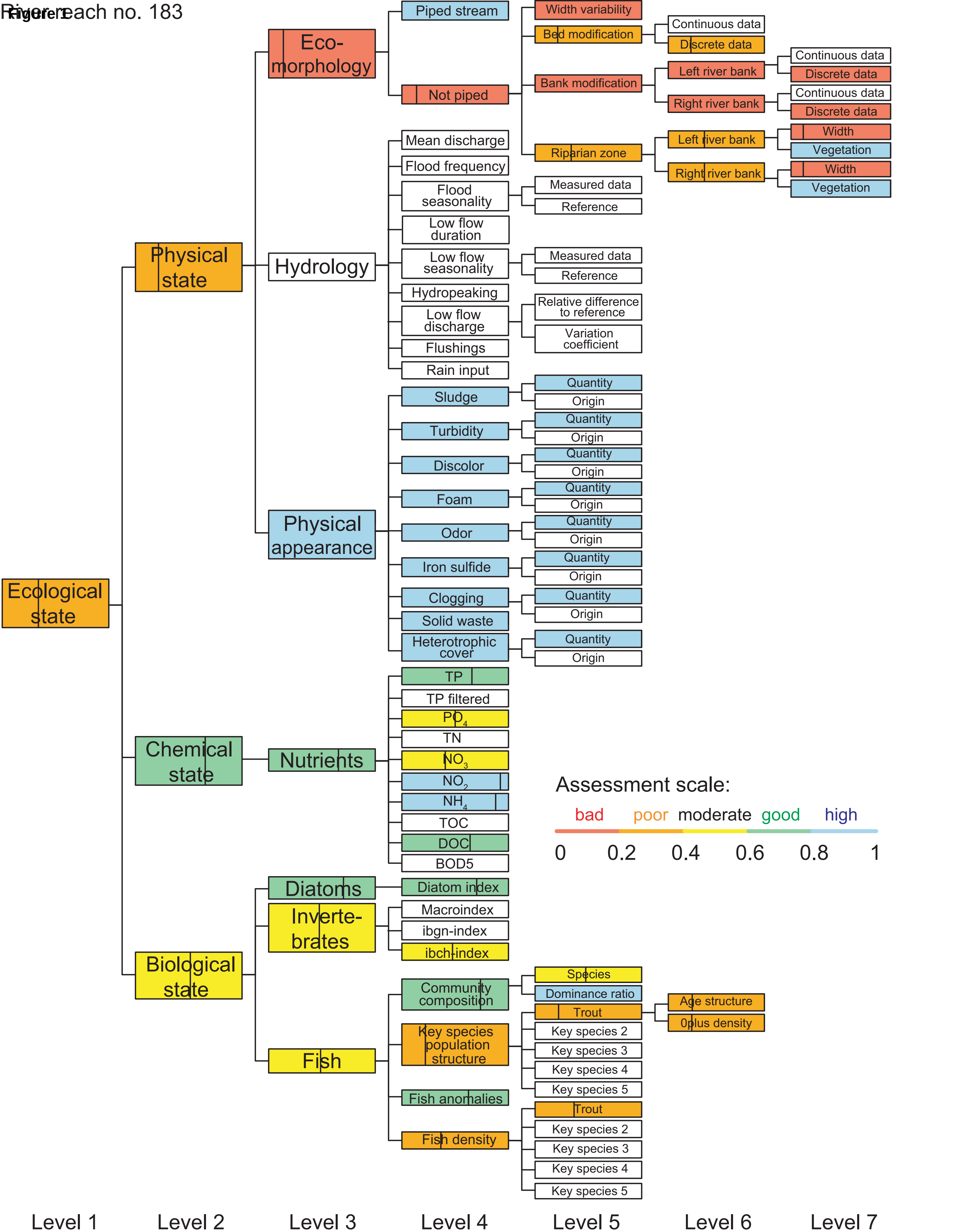

River reach no. 183
Ecological state
Physical state
Chemical state
Biological state
Eco-morphology
Hydrology
Physical appearance
Nutrients
Diatoms
Inverte-brates
Fish
Piped stream
Not piped
Width variability
Bed modification
Continuous data
Discrete data
Bank modification
Left river bank
Right river bank
Continuous data
Discrete data
Continuous data
Discrete data
Riparian zone
Left river bank
Right river bank
Width
Vegetation
Width
Vegetation
Mean discharge
Flood frequency
Flood seasonality
Measured data
Reference
Low flow duration
Low flow seasonality
Measured data
Reference
Hydropeaking
Low flow discharge
Relative difference to reference
Variation coefficient
Flushings
Rain input
Sludge
Turbidity
Discolor
Foam
Odor
Iron sulfide
Clogging
Solid waste
Heterotrophic cover
Quantity
Origin
TP
TP filtered
PO_4
TN
NO_3
NO_2
NH_4
TOC
DOC
BOD5
Diatom index
Macroindex
ibgn-index
ibch-index
Community composition
Key species population structure
Fish anomalies
Fish density
Species
Dominance ratio
Trout
Key species 2
Key species 3
Key species 4
Key species 5
Age structure
0plus density
Trout
Key species 2
Key species 3
Key species 4
Key species 5
Assessment scale:
bad
poor
moderate
good
high
0
0.2
0.4
0.6
0.8
1
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
Level 7

Figure 2

Figure 2

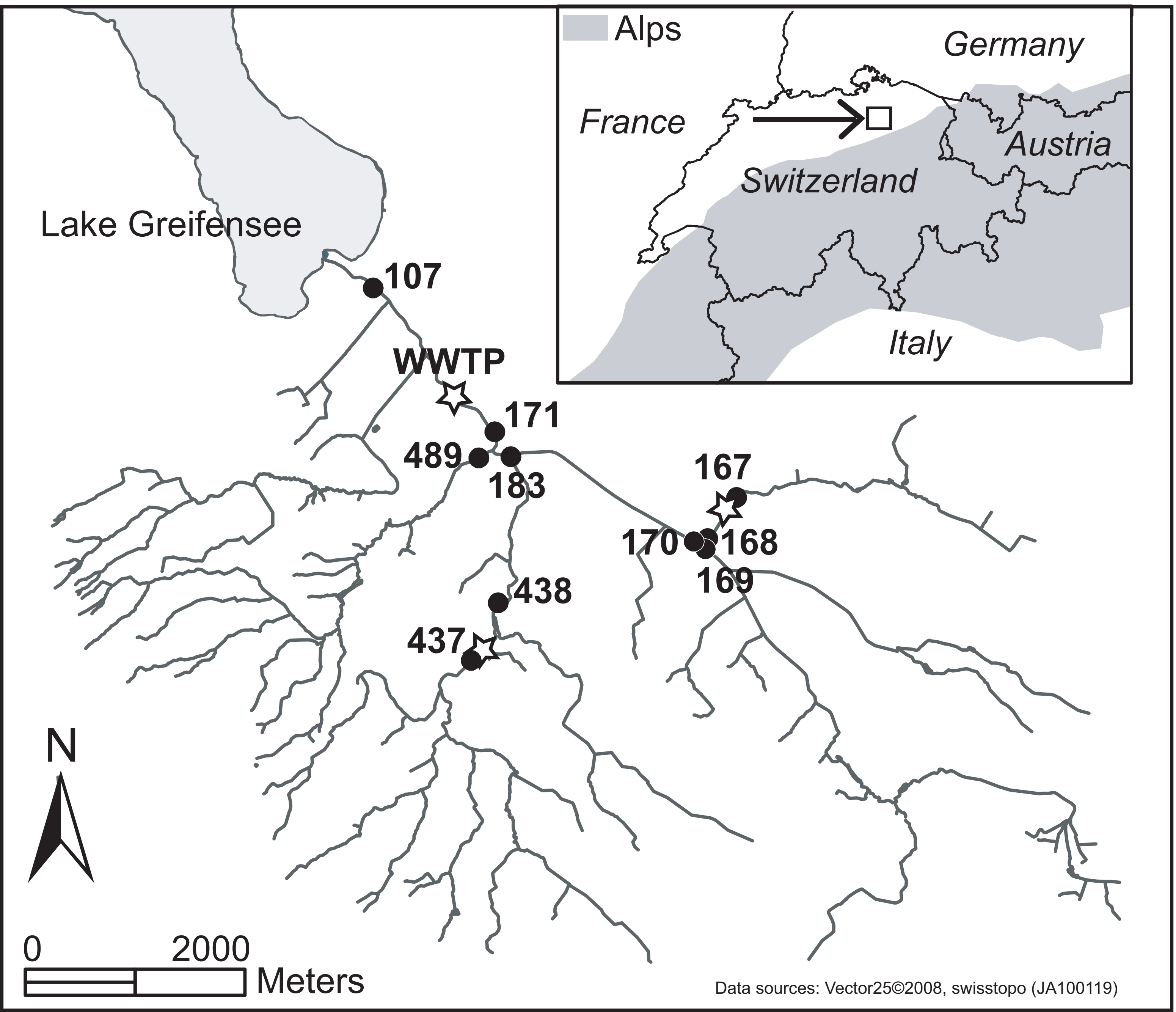


Figure 3

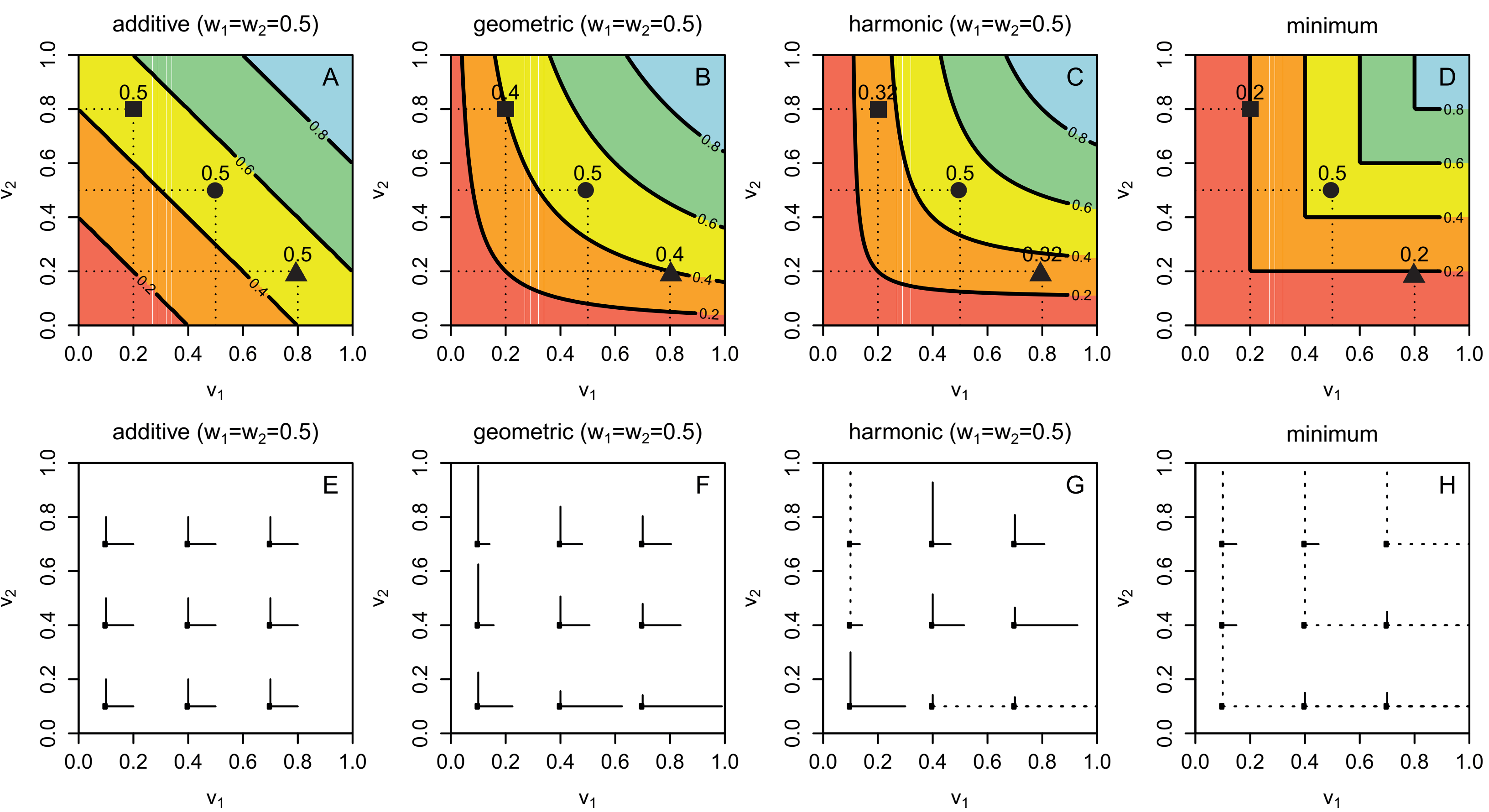
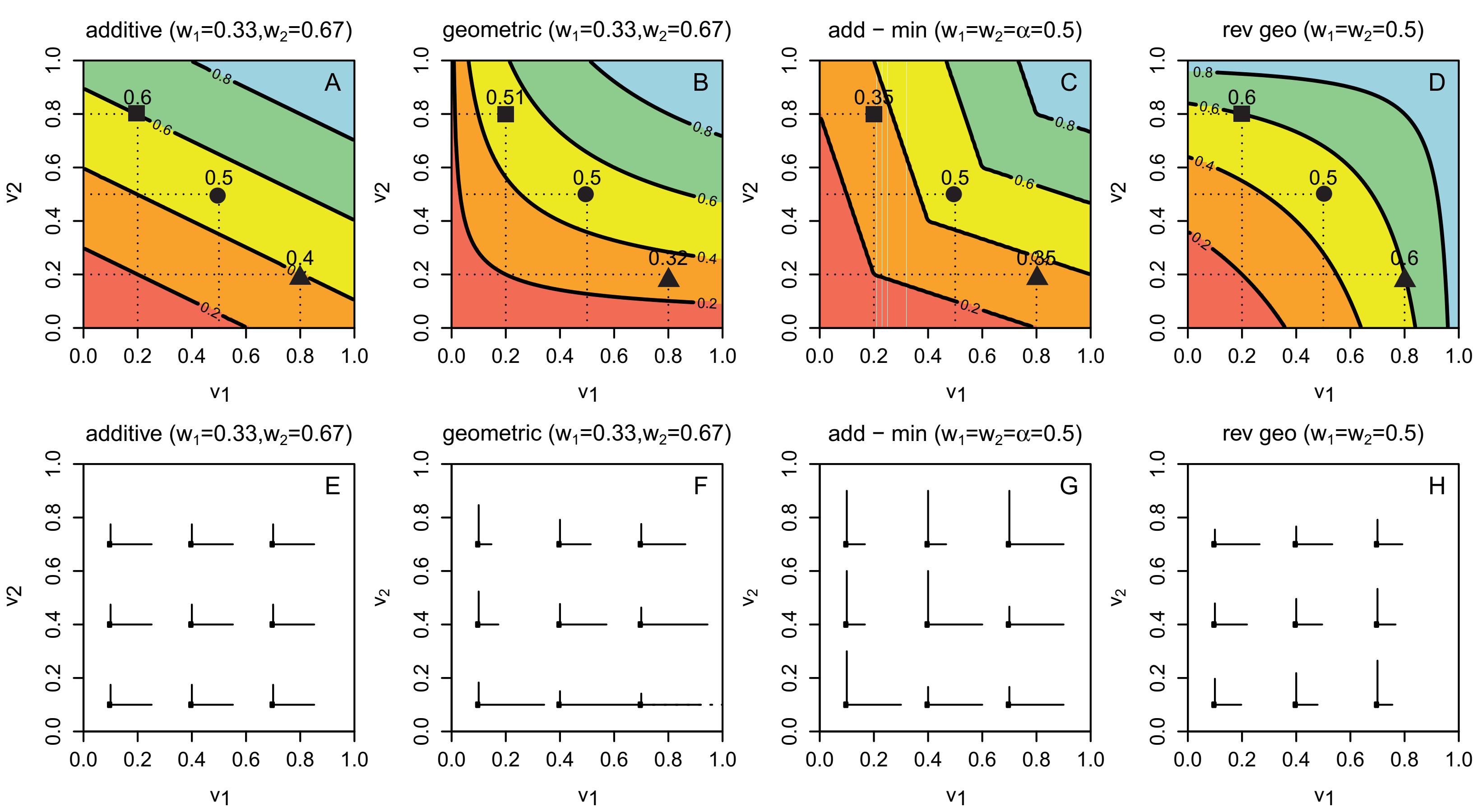

additive (w_1=w_2=0.5)
geometric (w_1=w_2=0.5)
harmonic (w_1=w_2=0.5)
minimum
A
B
C
D
0.5
0.4
0.32
0.2
0.2
0.4
0.6
0.8
E
F
G
H
v_1
v_2
0.0
0.2
0.4
0.6
0.8
1.0

Figure 4



additive (w_1=0.33,w_2=0.67)
geometric (w_1=0.33,w_2=0.67)
add − min (w_1=w_2=α=0.5)
rev geo (w_1=w_2=0.5)
A
B
C
D
E
F
G
H
0.6
0.5
0.4
0.51
0.32
0.35
v1
v2

Figure 5

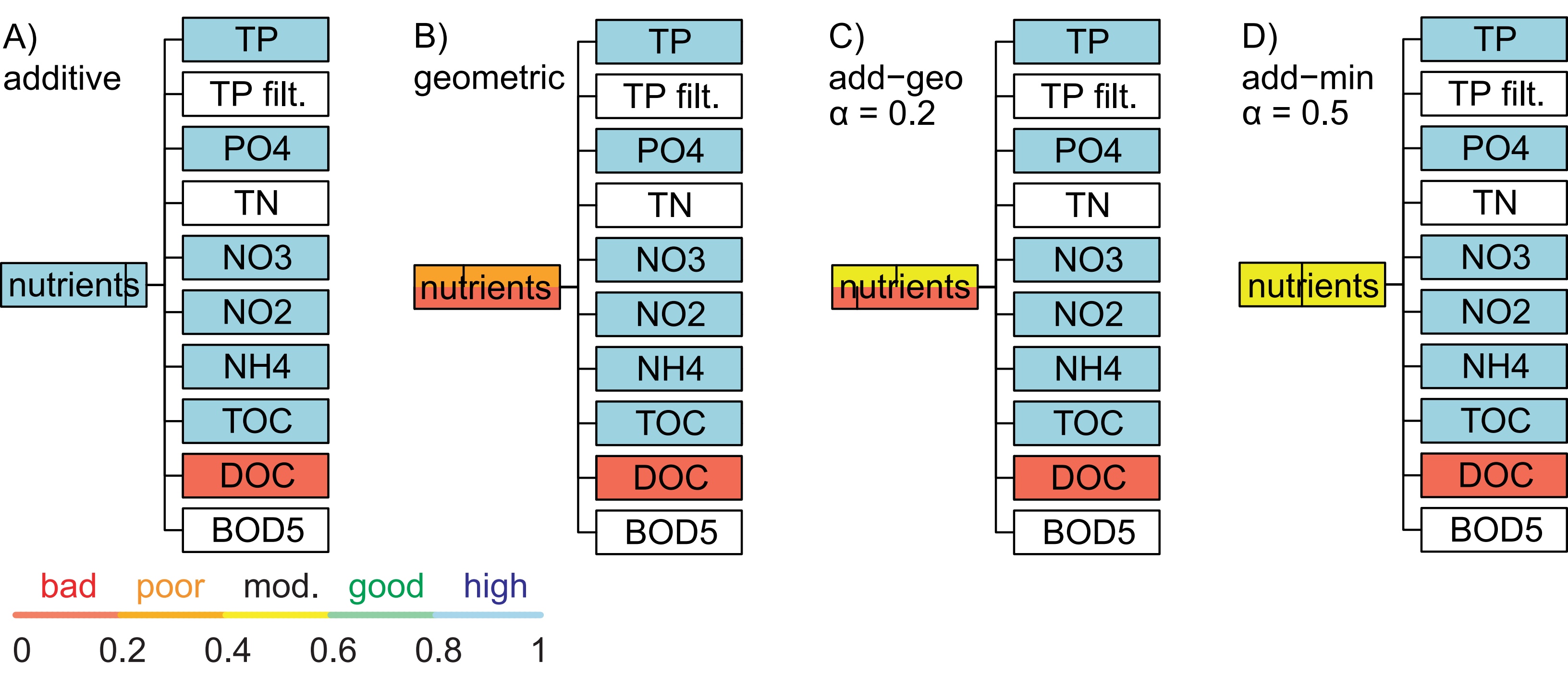

A)
additive
B)
geometric
C)
add−geo
α = 0.2
D)
add−min
α = 0.5
nutrients
TP
TP filt.
PO4
TN
NO3
NO2
NH4
TOC
DOC
BOD5
bad
poor
mod.
good
high
0
0.2
0.4
0.6
0.8
1

Figure 6

A)

	Ecological state	Physical state	Chemical state	Biological state	Ecomorphology	Hydrology	Physical appearance	Nutrients	Diatoms	Invertebrates	Fish
A add	.48	.45	.48	.50	.46	.43	.50	.48	.50	.52	.49
A geo	.48	.45	.48	.50	.46	.43	.50	.48	.50	.52	.49
A min	.43	.43	.48	.49	.46	.43	.50	.48	.50	.52	.49
A add geo	.48	.45	.48	.50	.46	.43	.50	.48	.50	.52	.49
A add min	.45	.44	.48	.50	.46	.43	.50	.48	.50	.52	.49
B add	.47	.30	.66	.46	.30	.24	.55	.66	.70	.27	.41
B geo	.43	.29	.66	.43	.30	.24	.55	.66	.70	.27	.41
B min	.24	.24	.66	.27	.30	.24	.55	.66	.70	.27	.41
B add geo	.45	.29	.66	.44	.30	.24	.55	.66	.70	.27	.41
B add min	.35	.27	.66	.36	.30	.24	.55	.66	.70	.27	.41
C add	.48	.08	.95	.42	.05	0	.55	.95	.70	.27	.29
C geo	0	0	.95	.38	.05	0	.55	.95	.70	.27	.29
C min	0	0	.95	.27	.05	0	.55	.95	.70	.27	.29
C add geo	.35	.04	.95	.40	.05	0	.55	.95	.70	.27	.29
C add min	.24	.04	.95	.34	.05	0	.55	.95	.70	.27	.29

Level 1 Level 2 Level 3

Three individual branches

B)

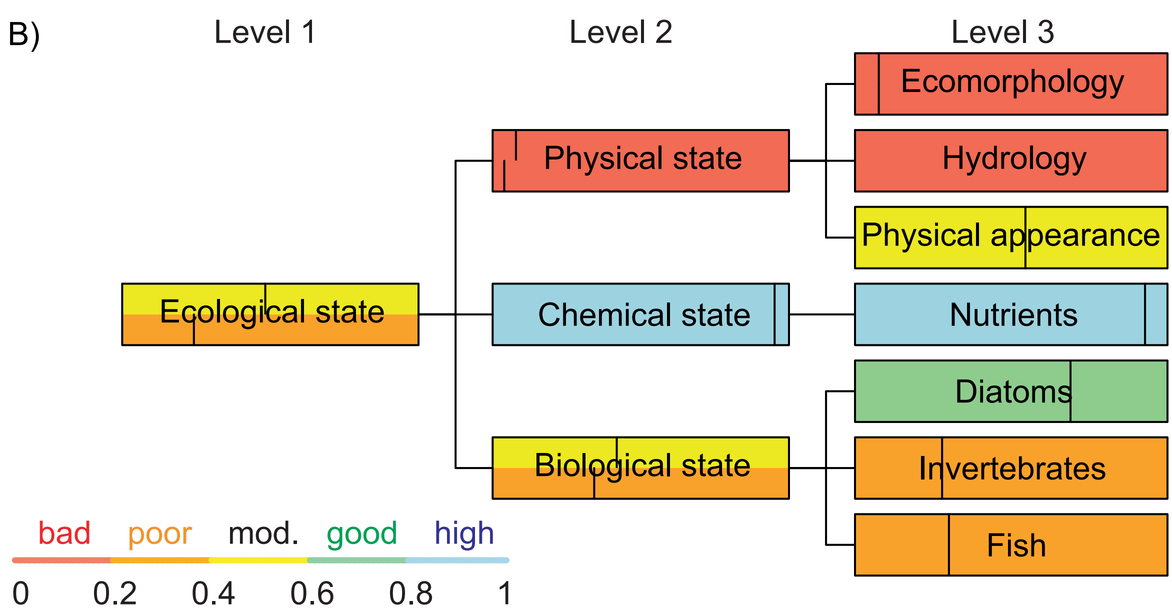


Figure 7

Reach no.	Ecological state	Physical state	Chemical state	Biological state	Ecomorphology	Hydrology	Physical appearance	Nutrients	Diatoms	Invertebrates	Fish
107	.55 .45	.38 .38	.70 .70	.56 .50	.37 .37		.42 .42	.70 .70	.65 .65	.57 .57	.45 .45
167	.58 .38	.23 .20	.82 .82	.69 .63	.18 .18		.42 .42	.82 .82	.81 .81	.57 .57	
168	.46 .36	.33 .29	.51 .51	.53 .51	.25 .25		.70 .70	.51 .51	.48 .48	.57 .57	
169	.57 .50	.51 .45	.51 .51	.70 .69	.40 .40		1 1	.51 .51	.73 .73	.68 .68	
170	.55 .48	.48 .44	.47 .47	.69 .66	.40 .40		.87 .87	.47 .47	.70 .70	.62 .62	.74 .74
171	.58 .46	.42 .38	.72 .72	.61 .51	.34 .34		.81 .81	.72 .72	.68 .68	.73 .73	.42 .42
183	.50 .34	.29 .22	.65 .65	.55 .51	.14 .14		1 1	.65 .65	.70 .70	.48 .48	.49 .49
437	.73 .67	.90 .89	.62 .62	.68 .65	.89 .89		.94 .94	.62 .62	.73 .73	.62 .62	
438	.59 .57	.60 .57	.56 .56	.61 .59	.54 .54		.87 .87	.56 .56	.64 .64	.57 .57	
489	.71 .62	.58 .54	.82 .82	.74 .71	.51 .51		.94 .94	.82 .82	.80 .80	.68 .68	

Level 1 Level 2 Level 3

Three individual branches

bad poor mod. good high

0 0.2 0.4 0.6 0.8 1

The method matters: indicator aggregation in ecological river assessment

Simone D. Langhans[1], Peter Reichert, Nele Schuwirth
Eawag: Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland
[1] Current address: Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany

Appendix A. Supplementary data

Figure S1. Properties of the basic aggregation methods additive, geometric mean, harmonic mean, and minimum demonstrated with the aggregation of two values (v_1 and v_2) assuming the weight of v_2 being twice as high as the one for v_1 (w_1 = 0.33, w_2 = 0.67). Panels A–D show isolines of aggregated values, and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines. The lines indicate the change in value on the corresponding axis that is required to get a gain in the aggregated value of 0.05. Dashed lines indicate that no trade-off can be found within the range up to unity that leads to such a gain.

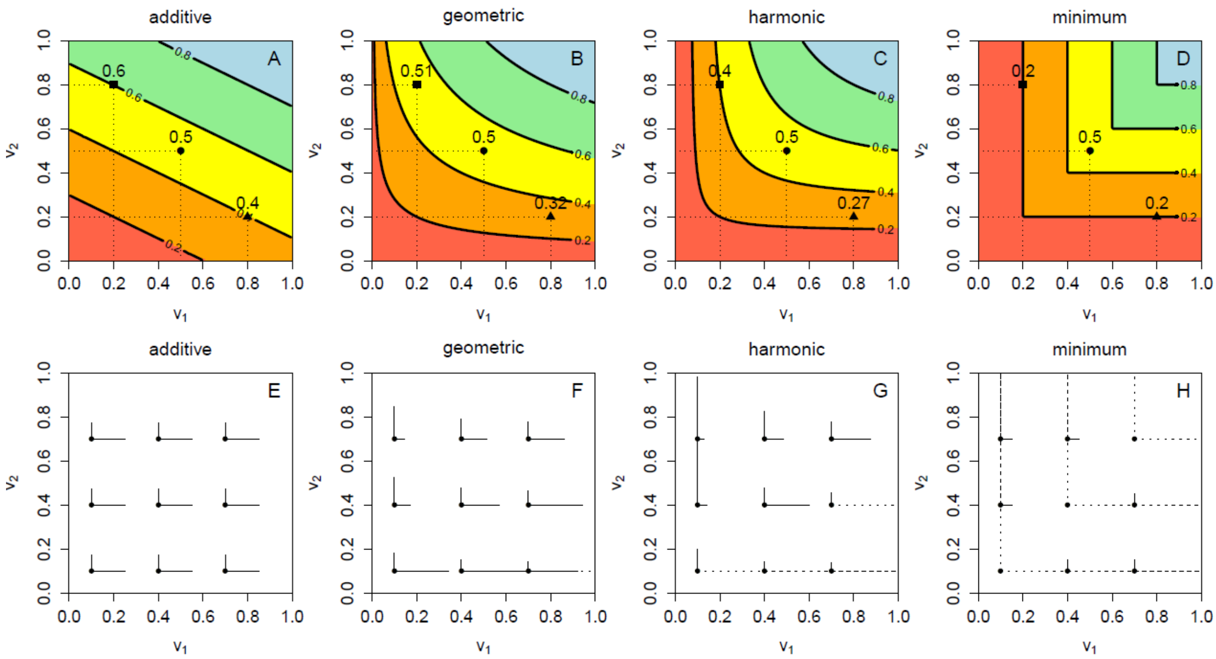


Figure S2. Properties of the mixed aggregation methods demonstrated with the aggregation of two values (v_1 and v_2) assuming equal weights for both values ($w_1 = w_2 = \alpha = 0.5$). Panels A–D show isolines of aggregated values and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

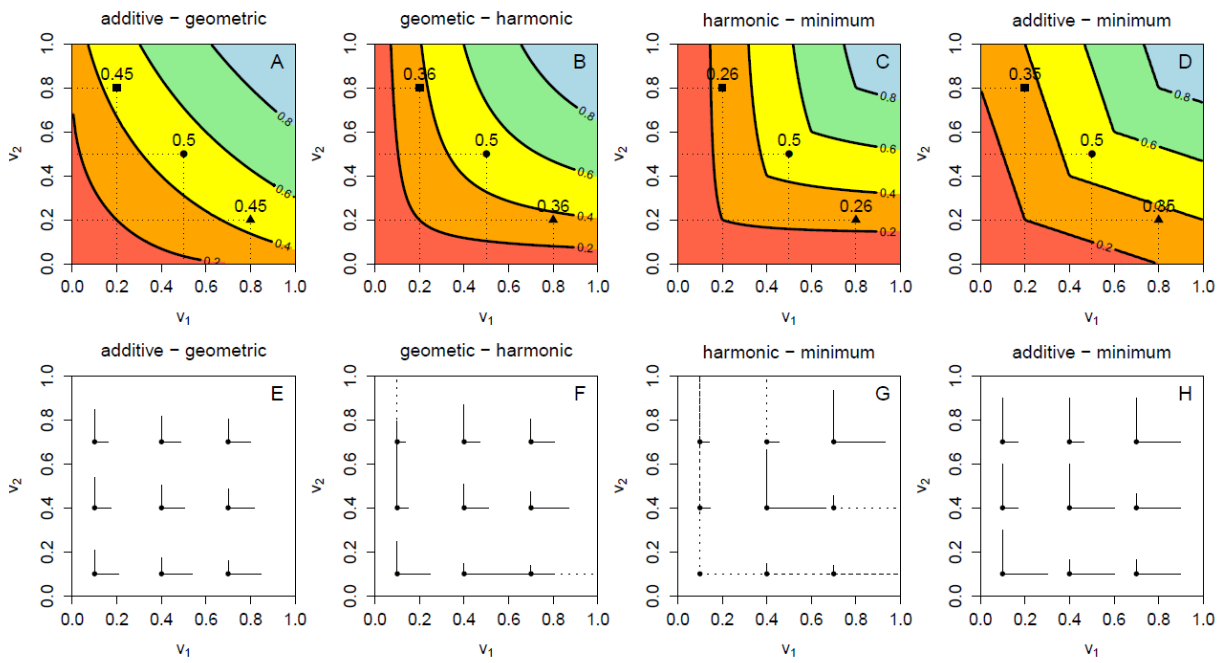


Figure S3. Properties of the mixed aggregation methods demonstrated with the aggregation of two values (v_1 and v_2) assuming the weight of v_2 being twice as high as the one for v_1 (w_1 = 0.33, w_2 = 0.67, α = 0.5). Isolines for the mixed harmonic mean–minimum and the additive–minimum aggregation methods are the same for equal and different weights (see Fig. S2). Panels A–D show isolines of aggregated values and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

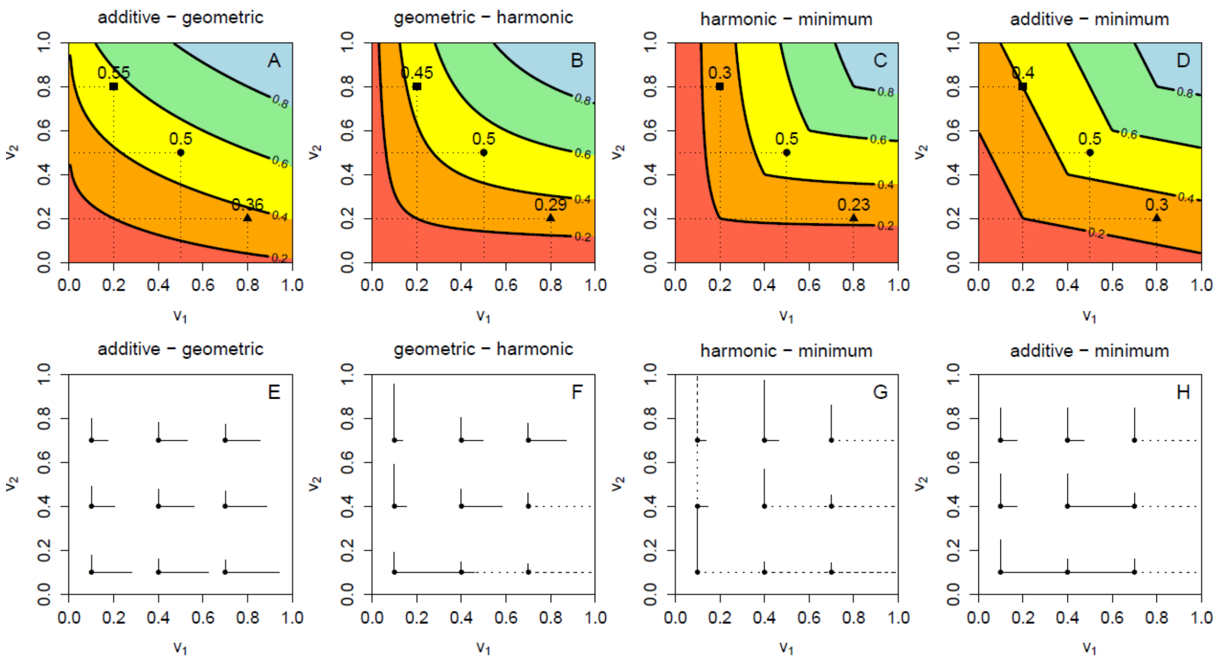


Figure S4. Properties of the reverse aggregation methods demonstrated with the aggregation of two values (v_1 and v_2) assuming equal weights of both values ($w_1 = w_2 = 0.5$). Isolines for the additive and the reverse additive aggregation methods are the same (see Fig. 1). Reversing the minimum aggregation leads to the maximum aggregation method. Panels A–D show isolines of aggregated values, and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

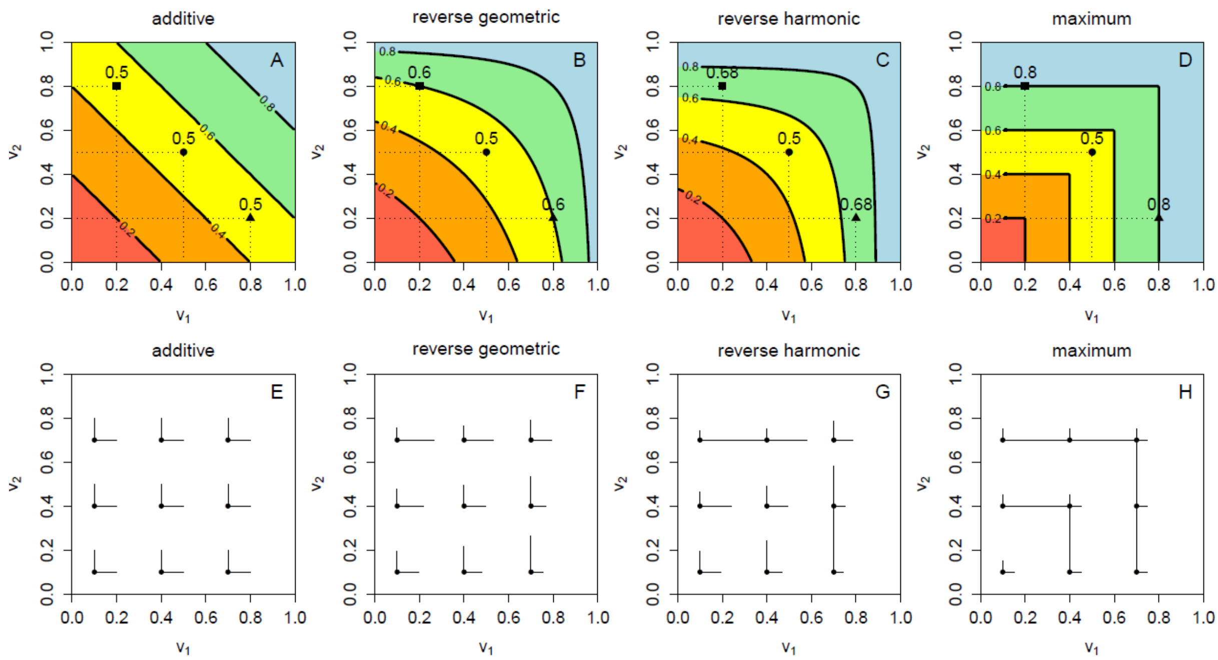


Figure S5. Properties of the reverse aggregation methods demonstrated with the aggregation of two values (v_1 and v_2) assuming the weight of v_2 being twice as high as the one for v_1 (w_1 = 0.33, w_2 = 0.67). Isolines for the maximum aggregation are the same for equal and different weights (see Fig. S4). Panels A–D show isolines of aggregated values, and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

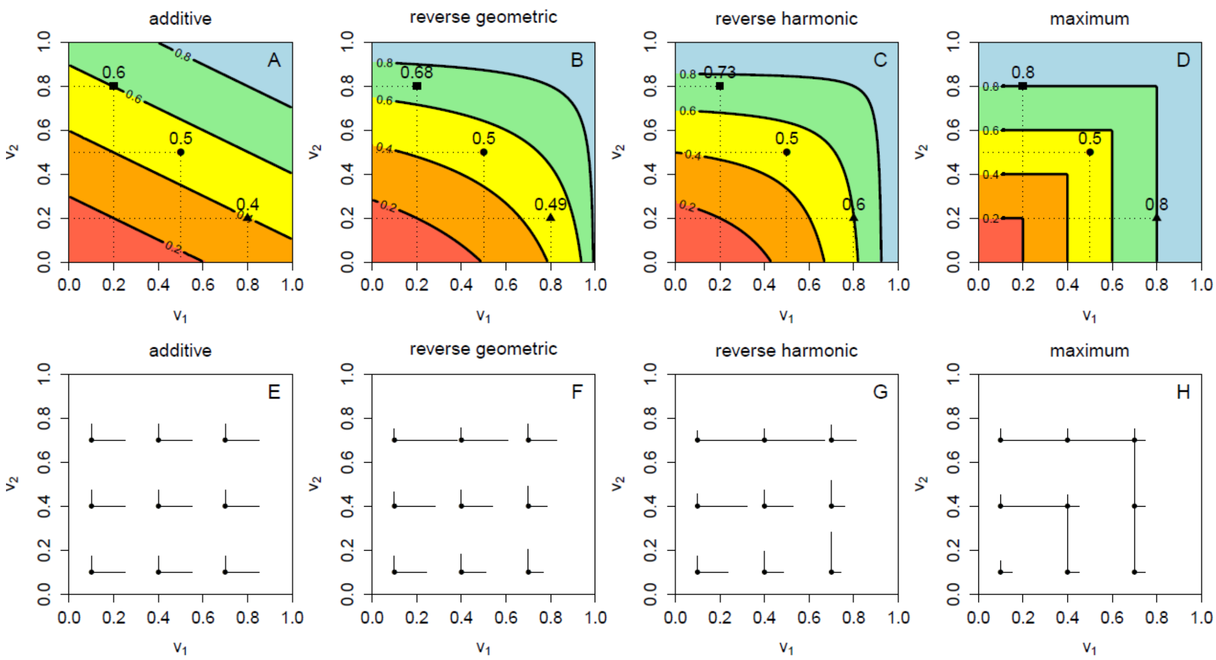


Figure S6. Properties of the mixed–reverse aggregation methods including the mixed additive–reverse geometric mean, the mixed reverse geometric mean–reverse harmonic mean, the mixed reverse harmonic mean–maximum, and the additive–maximum aggregation (from left to right), demonstrated with the aggregation of two values (v_1 and v_2) assuming equal weights of both values ($w_1 = w_2 = \alpha = 0.5$). Panels A–D show isolines of aggregated values, and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

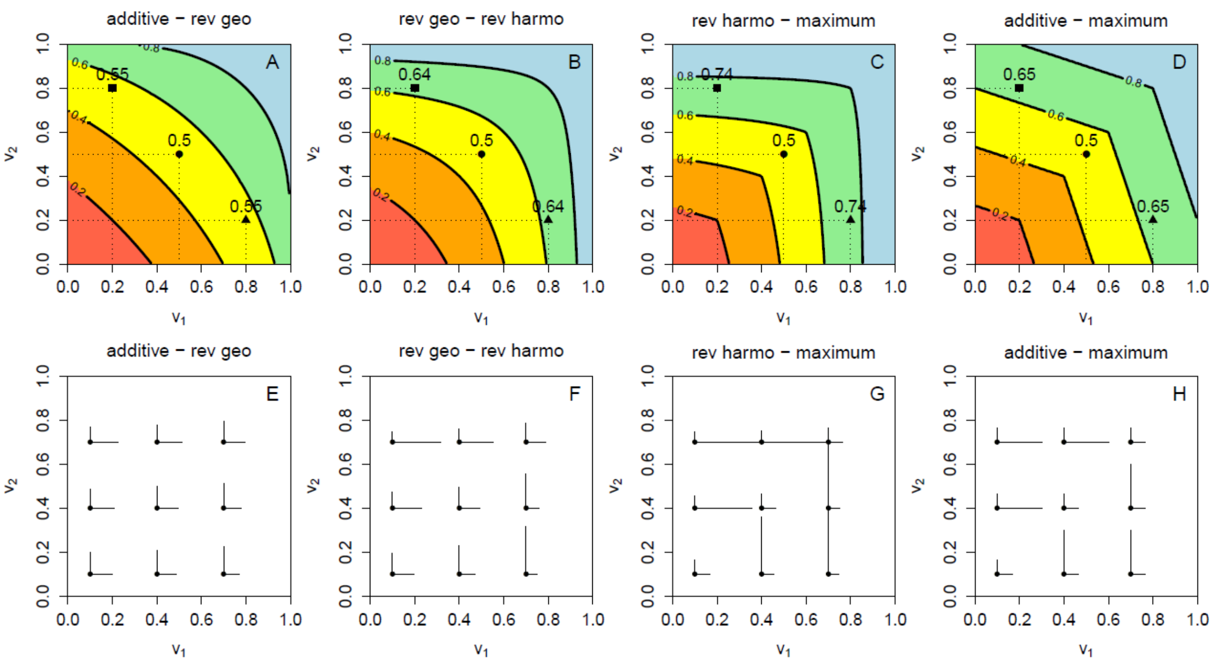


Figure S7. Properties of the mixed–reverse aggregation methods demonstrated with the aggregation of two values (v_1 and v_2) assuming the weight of v_2 being twice as high as the one for v_1 (w_1 = 0.33, w_2 = 0.67, α = 0.5). Isolines for the reverse harmonic mean–maximum aggregation (rev harmo–maximum) and the mixed additive–maximum aggregation with equal and different weights are the same (see Fig. S6). Panels A–D show isolines of aggregated values, and aggregated values for three selected argument pairs depicted as squares, circles, and triangles. Panels E–H visualize trade-offs with horizontal and vertical lines (see Fig. S1 for further explanation).

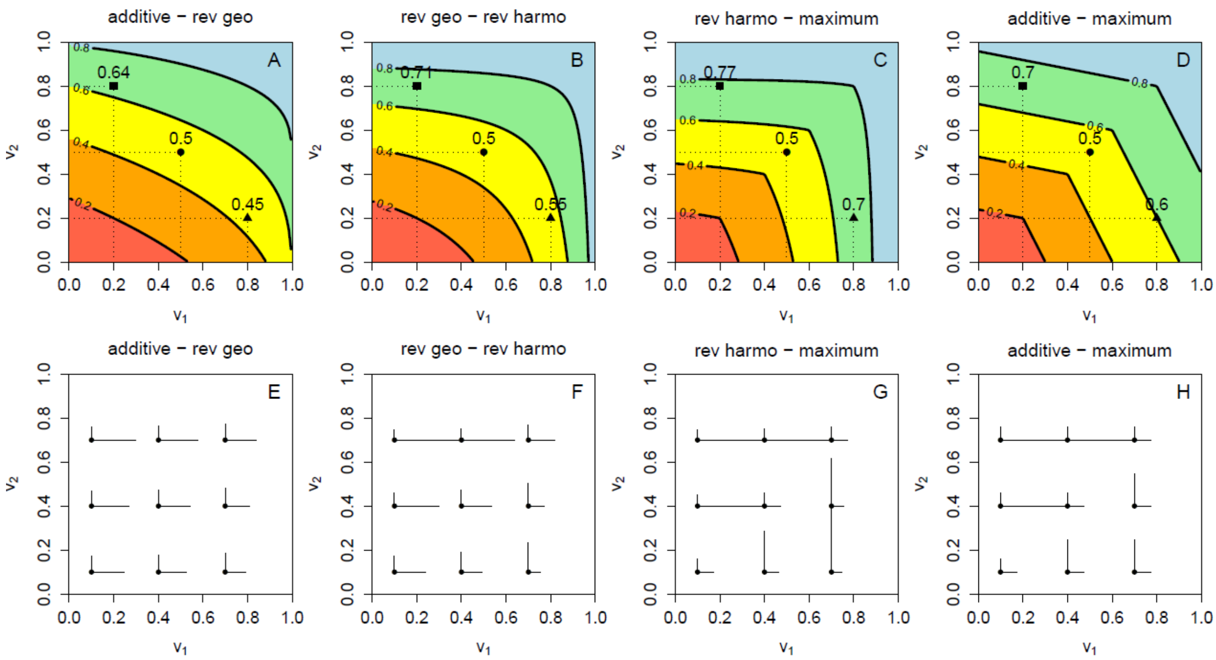


Figure S8. 3D Plots (first row panels), isolines (middle row panels) and trade-offs (third row panels) of the mixed additive–geometric mean aggregation method (add–geo) demonstrated with the aggregation of two values (v_1 and v_2) assuming equal weights ($v_1 = v_2 = 0.5$). The parameter α, which defines the weight of the additive aggregation, increases from left to right: $\alpha = 0$, $\alpha = 0.2$, $\alpha = 0.4$, $\alpha = 0.6$, $\alpha = 0.8$, and $\alpha = 1$.

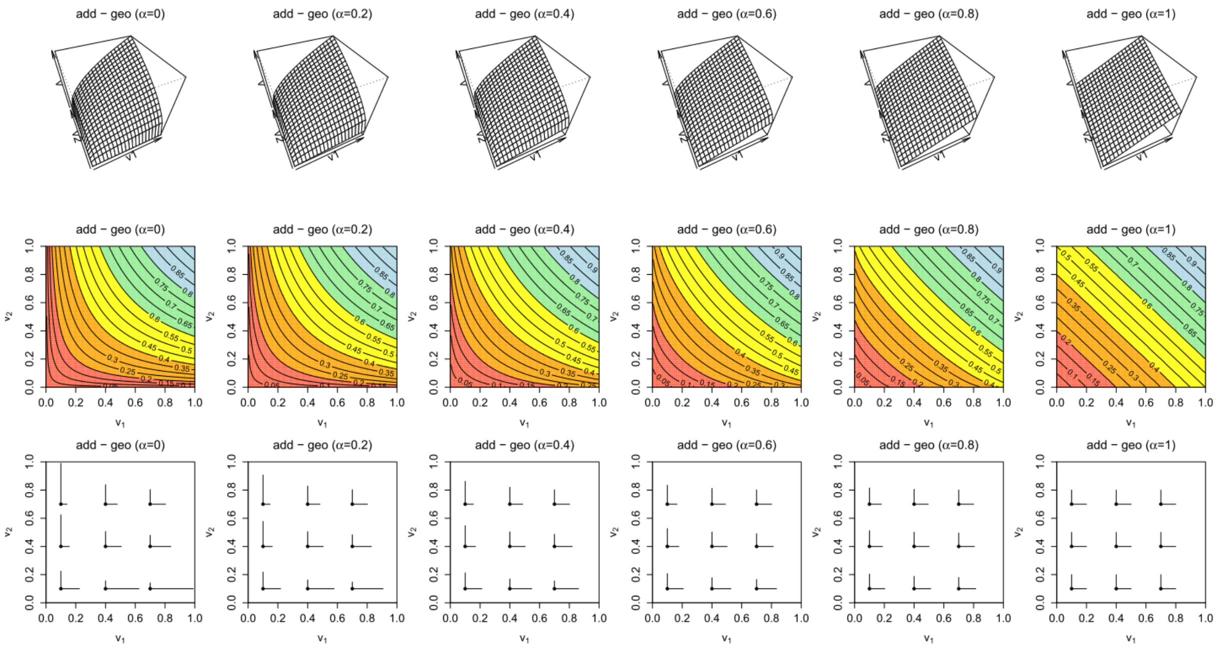


Figure S9. 3D Plots (first row panels), isolines (middle row panels) and trade-offs (third row panels) of the mixed additive–minimum aggregation method (add–min) demonstrated with the aggregation of two values (v_1 and v_2) assuming equal weights ($v_1 = v_2 = 0.5$). The parameter α, which defines the weight of the additive aggregation, increases from left to right: $\alpha = 0$, $\alpha = 0.2$, $\alpha = 0.4$, $\alpha = 0.6$, $\alpha = 0.8$, and $\alpha = 1$.