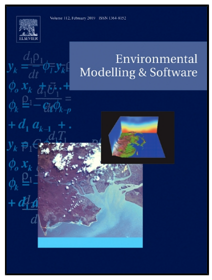
Developing composite indicators for ecological water quality assessment based on network interactions and expert judgment

Mao, Feng; Zhao, Xianfu; Ma, Peiming; Chi, Shiyun; Richards, Keith; Clark, Julian; Hannah, David M.; Krause, Stefan

DOI:
10.1016/j.envsoft.2019.01.011

Accepted Manuscript



Developing composite indicators for ecological water quality assessment based on network interactions and expert judgment

Feng Mao, Xianfu Zhao, Peiming Ma, Shiyun Chi, Keith Richards, Julian Clark, David Hannah, Stefan Krause

PII:	S1364-8152(17)30876-9
DOI:	10.1016/j.envsoft.2019.01.011
Reference:	ENSO 4375
To appear in:	*Environmental Modelling and Software*
Received Date:	22 August 2017
Accepted Date:	12 January 2019

Please cite this article as: Feng Mao, Xianfu Zhao, Peiming Ma, Shiyun Chi, Keith Richards, Julian Clark, David Hannah, Stefan Krause, Developing composite indicators for ecological water quality assessment based on network interactions and expert judgment, *Environmental Modelling and Software* (2019), doi: 10.1016/j.envsoft.2019.01.011

Environmental Modelling & Software

Developing composite indicators for ecological water quality assessment based on network interactions and expert judgment

Feng Mao[1], Xianfu Zhao[2], Peiming Ma[2], Shiyun Chi[2], Keith Richards[3], Julian Clark[1], David Hannah[1], Stefan Krause[1]

[1] School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

[2] Key Laboratory of Ecological Impacts of Hydraulic-Projects and Restoration of Aquatic Ecosystem of Ministry of Water Resources, Institute of Hydroecology, Ministry of Water Resources and Chinese Academy of Sciences, Wuhan, 430079, P. R. China

[3] Department of Geography, University of Cambridge, Downing Place, Cambridge, CB2 3EN, UK

Correspondence to: Feng Mao (f.mao@bham.ac.uk)

Highlights

- Networks are key features of ecosystems, providing useful information for decision-making.
- Interactions are seldom used in setting weights for ecological composite indicators.
- We develop a method to adjust indicator weights using ecological network information.
- We show an extensible and more justified indicator building strategy.

Abstract

Increasingly, composite indicators and multi-criteria approaches are applied in environmental assessment and decision-making, including the EU Water Framework Directive. For example, integrated evaluation of aquatic ecosystem conditions and functioning usually involves a group of criteria, such as biological organisms and communities, physicochemical and hydromorphological variables, which are measured individually and combined by a weighted linear function into an overall 'score'. We argue that the network interactions of evaluation components are useful information for expert judgments, which have not been sufficiently considered in existing multi-criteria combination strategies in environmental assessment and management. Built upon the Analytic Network Process and demonstrated with the Chishui River Basin in China, this paper introduces a network-based expert judgment approach to construct ecological water quality indicators, and to determine and adjust their variable weight settings with information of interaction networks. This approach has potential to construct composite indicators for a broad environmental context.

Keywords: Analytic Network Process; integrated river basin management; ecological network; composite indicator; ecological water quality assessment; multi-criteria decision-making

Software and data availability

The proposed approach is performed using R, which is an open-source programming language supported by the R Foundation for Statistical Computing. The latest version of this software can be accessed via https://www.r-project.org/. The calculation process of the approach, as well as the expert judgment data are presented in detail in the analytical example of this paper. To request these data in an excel file, or for additional information, please contact f.mao@bham.ac.uk.

1. Introduction

Integrated evaluation and composite indicators have been used widely in monitoring, management and multi-criteria decision-making (Hsu et al., 2013; Mizobuchi, 2014; Munda, 2005; Singh et al., 2012). These approaches help to summarise complex and multi-dimensional information, and facilitate communication among scientists, policy makers, regulators and the general public (OECD, 2008). Composite indicators have gained particular attention in the water sector, since there is an increasing tendency to use comprehensive ecologically-based methods for water quality evaluation and management to replace or complement conventional physicochemically-based approaches (Lumb et al., 2011). Successful large scale applications of ecological composite indicators include the European Union Water Framework Directive (EU WFD; European Commission, 2000), the Australian National River Health Programme (O'Connor et al., 1996), and the United States' National Rivers and Streams Assessment (Faustini et al., 2009). All these frameworks and programmes involve a process of combining a large group of elements, including biological components and indicators, into a single evaluation.

There are many strategies to combine evaluation components and build composite indicators, such as decision trees, the 'One-Out All-Out' (OOAO) principle, weighted-average models, and multivariate analysis, each having its own advantages and flaws (Borja et al., 2008; Boulton, 1999; Lücke and Johnson, 2009). A decision tree consists of a series of expert-judged *if-then* statements – *if* certain conditions are met, *then* corresponding outcomes are obtained. The OOAO principle can be an extreme example of a decision tree, in which the overall evaluation grade is determined by the lowest graded component. The weighted sum/average model combines variables using predetermined weights. Multivariate approaches, such as Principal Component Analysis (PCA), can also be used to allocate weights based on the covariance or correlation matrix of variables (see Esselman and Infante, 2014; Singh et al., 2012). However, these methods can be inefficient when the number of evaluation objects or components is large (decision tree), they may not be able to capture subtle changes in indicator values (OOAO), and the weighting sets may be difficult to interpret because they do not necessarily reflect the relevant importance of each component (multivariate methods, e.g. PCA). By comparison, expert judgment-based weighted sum/average models may be the most suitable option for their potential efficient, sensitive and easy-to-interpret characteristics.

Expert judgment strategies provide a valuable way to acquire information, that is effective and efficient to improve our understanding of complex problems, especially when empirical data is scarce or unavailable (Burgman et al., 2011; Martin et al., 2012). However, expert opinion must be used wisely to ensure accurate and reliable data are generated. This can be achieved through a number of approaches such as careful selection of experts, training and interacting with experts, optimising the questions for experts to elicit useful information, and consideration of what information to collect from experts (Martin et al., 2012; Sutherland and Burgman, 2015). This strategy has been used for composite indicator weight setting in many fields, such as in sustainable development (Hák et al., 2016; Rickels et al., 2016), resilience studies (Ostadtaghizadeh et al., 2015; Schipper and Langston, 2015) and generally assess the performance and status of systems (Mizobuchi, 2014; Molinos-Senante et al., 2014; Rogge, 2012). However, there are still concerns about the subjectivity and arbitrariness, and randomness of expert judgment methods, and on what information and knowledge the judgments can be based (Martin et al., 2012) – even though the weight for each component can be assigned by pooling experts' opinions on their relevant importance to the overall evaluation.

We argue that the current expert judgment-based weight setting practices seldom consider the networks of variables and their interactions, although these interactions are key features of ecosystems and can provide significant information about system behaviour. For example, the Ocean Health Index is a weighted average of ten categories, including food provision, natural products and biodiversity, which are also ten public goals for global ocean protection (Halpern et al., 2012). However, delivery or production of some goals may involve activities that have negative feedbacks on other goals. For example, overfishing and high level of food provision may risk biodiversity and iconic local species. These interactions and feedbacks are not assessed in the evaluation; and they are assumed to be neutral among goals (Halpern et al., 2012). Similarly, in integrated water quality evaluation of rivers and lakes, the biological, physicochemical and hydromorphological quality elements clearly all interact with each other. For example, hydromorphological types determine the ecological structure of macro-invertebrates (Vannote et al., 1980); and physicochemical status may influence the abundance of macrophytes (Lau and Lane, 2002), or other aquatic organisms (Monk et al., 2008; Webb et al., 2008).

This paper hypothesises that the interactions and networks of evaluation components and variables are long-neglected but useful information for expert judgments, especially in the field of water-dependent ecosystems and management where connections are complex and common. The Analytic Network Process (ANP) provides an opportunity to consider these interactions in evaluations. The ANP was introduced in the 1990s (Saaty, 1996, 1980) as a generalised form of the Analytic Hierarchy Process (AHP) method for Multiple-Criteria Decision Analysis. Different from the AHP which structures problems in a hierarchy, the ANP uses a network to describe much more complex inter-relationships of criteria. Herein, we aim to develop and test an ANP-based method integrate this additional network information into consideration of composite indicator construction and weight assignment.

In the following, the paper addresses three key questions about networks and interactions: Section 2 explains why interactions are worth considering, introduces the ANP approach, and proposes the methodology used in this study. Section 3 provides an analytical example to demonstrate how interactions can be considered in the weight assignment processes, with a sample application to the Chishui River Basin in the southwest of China. Section 4 analyses the results of the analytical example in Section 3, discusses the performance, features and potential applications of the proposed network-based approach, and explains how the addition information of network interactions introduces a risk-based thinking in the overall evaluation.

2. Neglected information and the Analytic Network Process (ANP)

2.1 Integrated evaluation and expert judgment

Integrated evaluation using composite indicators summarises a range of criteria by linear combination, with weighted sum/average form as shown in Figure 1. Expert opinion plays an active part in this process. It first helps to decide what components are included in the evaluation system. To calculate the overall score, at least two sets of information are required: weight values and performance scores for each component. The weights are usually decided by expert judgment. The performance scores describe the performance of evaluation components. In principle, a higher score suggests a better condition. The scores of evaluation components can be acquired by a series of methods such as monitoring, surveys, or even expert scoring when data collection is difficult. This paper discusses the information experts rely on when assigning weights.

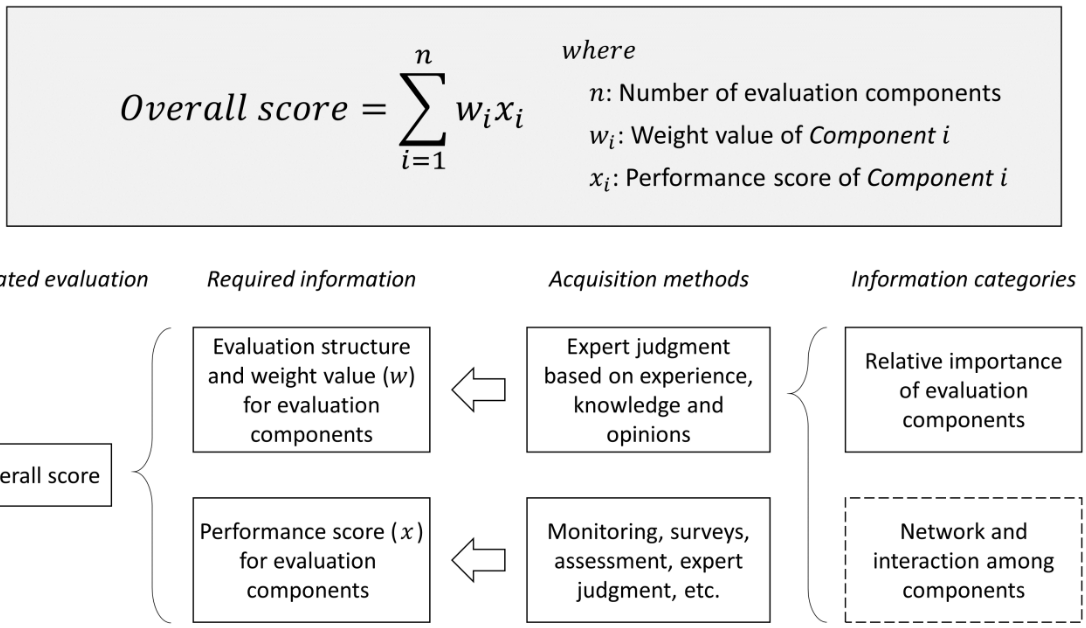


Figure 1. Typical information for integrated evaluation based on expert judgment. Two sets of information are required to obtain overall scores, including (1) weight value for each evaluation component, which is decided with expert judgment, experience, knowledge or opinion towards the relative importance of components; (2) standardised performance score for each evaluation component, which is acquired by monitoring, surveys, assessment or expert judgment. The interactions among components have not been sufficiently considered as a useful category of information for weight value determination. Top box: the weighted sum model for integrated evaluation. Bottom flow chart: informaton flow for generating overall scores.

The weight values are assigned according to experts' understanding of the relative importance of components to the overall evaluation, but do not typically reflect the interactions among components (Figure 1). This absence of interaction information may undermine the justification and performance of the composite indicators, firstly because the indicators are designed to assess ecological water quality comprehensively in integrated ways, which should include the interactions that are essential features of aquatic ecosystems. The other set of reasons is caused by the unavoidable inter-dependence among ecological variables. If all the selected variables are independent of each other, the combination of selected variables and their assigned weights reflect perfectly the experts' understanding of ecological water quality. However, if interactions exist but are not considered, the weights of the influencing variables are likely to be improperly estimated, because of their *de facto* higher relative importance in the evaluation system. If all the intricate interactions among the evaluation variables are considered, the initial weights assigned by experts should be adjusted accordingly. Thus, the challenge is to find a standardised and straightforward way to build the interaction network, and adjust the weight values with this additional information. We argue that the ANP method can be used to meet this challenge.

2.2 The ANP and interactions

As structured techniques for multi-criteria decision making, AHP and ANP methods are useful methods to find the best solution from alternatives based on multiple criteria (Saaty, 2008). AHP organises the elements (i.e. criteria and alternatives) in a hierarchy and compares their relative importance to the upper level through pairwise comparisons. ANP moves one step further by considering the interdependence and feedback between and within element clusters in a network structure, and by conducting more comprehensive pairwise comparisons among elements.

These two methods are widely used in environmental management and decision-making. Most studies use AHP/ANP techniques to select the best decisions or strategies among a group of candidates. For example, Bottero et al. (2011) used both AHP and ANP to help select the most sustainable water treatment technologies from among three candidates, while Toosi and Samani (2012) evaluated 10 water transfer projects in the Karun River, Iran, based on 30 influential factors using ANP. These methods have been incorporated into decision support systems to evaluate sites for sustainable cage aquaculture (Halide et al., 2009), to cope with sea level rise (Sahin and Mohamed, 2013), and to manage European coastal lagoons (Casini et al., 2015). In addition, AHP/ANP methods have also been used to establish weights for environmental indices (Herva and Roca, 2013; Singh et al., 2009). Examples include the development of the Environmental Quality Index (Bisset, 1988), the Environment Performance Index for Industries (Hermann et al., 2007), and the Composite Sustainability Performance Index (Singh et al., 2007).

As noted above, different from the AHP method, the ANP helps to integrate information about interactions between processes into the assignment of weights for composite indicators, where the interactions can be identified using existing analytical frameworks. Frameworks such as the SWOT (Strengths, Weaknesses, Opportunities and Threats) and BOCR (Benefits, Opportunities, Costs and Risks) methods can be used as starting points for building a network and hierarchy structure, and selecting influential variables given both positive and negative aspects of the alternatives (Bottero et al., 2011; Sipahi and Timor, 2010; Wijnmalen, 2007). The Driver-Pressure-State-Impact-Response (DPSIR) model is an additional tool frequently used in environmental management problems. Wolfslehner et al. (2005) used the DPSIR model and its variant frameworks in an ANP analysis to choose the best sustainable forest management strategies (Vacik et al., 2007; Wolfslehner et al., 2005; Wolfslehner and Vacik, 2008). However, these studies did not produce an index for environmental condition measurements, nor did they consider the ecosystem interactions. The ecological network also provides a potential basis for network-style analyses, and it refers to interactions among organism groups, such as predator-prey relationships, symbiotic cooperation or interactions amongst living organisms and their physical environment. In environmental assessment systems, these interactions may be represented by the interactions of variables and indices, which are metrics and measurements of the quality elements. For example, a high concentration of nitrate-N in freshwater may bring down the scores of certain sensitivity biological indices (Sandin and Hering, 2004); or abundant macrophytes in the river may increase the abundance of the macro-invertebrates with the ‘climber’ ecological trait, by providing suitable micro habitats (Cummins and Merritt, 2001).

2.3 Designing an ANP-based method to consider interaction networks

To determine the weight values of the components when considering interaction networks, this paper proposes a procedure of three mains steps, including the integration of an adapted ANP method.

- *Step 1 constructs the evaluation structure and interaction network*. This preparation step defines the hierarchical evaluation structure, and the interaction network (Figure 2). The evaluation structure explains how the overall evaluation is divided into evaluation clusters and components, while the network demonstrates how the evaluation components interact with each other, and with other key non-evaluation components within the system.
- *Step 2 determines the initial variable weights of different components with the hierarchical evaluation structure*. The overall evaluation is divided into categories, then components and variables. An initial weight is assigned to each component variable to represent its relative importance. In this step, the interaction network is not considered (Figure 2-a).
- *Step 3 adjusts the initial weights with the interaction network using the ANP method*. Pairwise comparisons are performed for each pair of nodes to determine the relative importance of variables (Figure 2-b). The results of pairwise comparisons are summarised in a form of a super-matrix and eventually transformed into the variable weights.

In the following sections, the detailed process of these three main steps is demonstrated in an analytical example, and the initial and adjusted weight sets are compared.

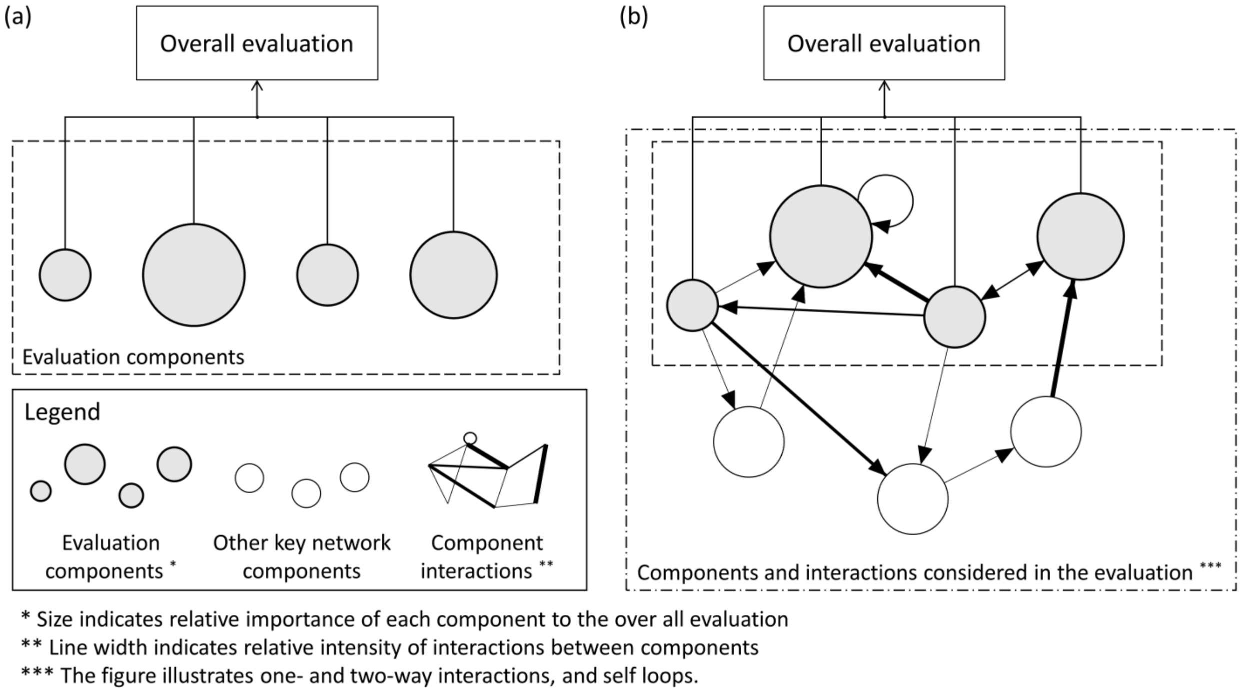


Figure 2. Two types of evaluation structures. (a) Evaluation structure without consideration of networks and interactions of components. Evaluation components are shown in the dashed box. (b) Evaluation structure with consideration of networks and interactions. The dashed box highlights the evaluations components as shown in panel a. The dash-dotted box contains all the evaluation components, the other key network components and the component interactions.

3. An analytical example

The Chishui River Basin in China was selected as the study area of the analytical example (Section 3.1), mainly because it has a unique environmental context and a comprehensive hydro-ecological dataset representing the entire river basin. Section 3.2 to 3.4 explain the application of the core concepts and the three-step procedure of the approach for this specific example. In Section 3.5, the initial and modified variable weights as well as the evaluations using both weighting systems are provided. It is followed by a sensitivity analysis of the approach in Section 3.6.

The calculations were performed in R 3.0.1 (RC Team, 2013) which offers free and flexible tools as alternatives to commercial decision-making support software such as Super Decisions (Saaty and William, 2000) or Expert Choice (Forman et al., 1983).

3.1 Study Area

The Chishui River Basin (27˚20' - 28˚50' N, 104˚45' - 106˚51' E) is located in the Southwest of China. With a catchment area of 21,000 km^2, and a main stream length of 425 km, the Chishui River lies between the Yungui Plateau and the Sichuan Basin. It is a tributary of the upper Yangtze River. The Basin has no dams or hydropower development in the mainstream and main tributaries, which is unique in the Yangtze river basin. According to China's National Surface Water Standard (SEPA, 2002), the Chishui River has good physicochemical water quality overall, but it does have a spatial variety of different water quality levels in different reaches and tributaries. The Chishui River Basin is famous for its liquor industry, especially the iconic Maotai brand, which is named after its place of production. Over 600 liquor factories are scattered within and around Maotai town, in the middle reaches of the Chishui River. Discharges from the liquor industry contribute nearly 80% of the pollution in Maotai town; and untreated domestic sewage is the other main source of water pollution in this region (Zhou, 2010).

Biological, physicochemical and hydromorphological data of 71 sampling reaches were obtained in the spring of 2012, covering the mainstream and 13 tributaries of the Chishui River. These data were collected under the River Basin Governance Research Network (RiBaGo, 2012) and the EU-China River Basin Management Programme (RBMP, 2012), which aimed to test the European-style ecologically-centred water quality management approaches in China, to support China to enhance Integrated Water Resources Management (IWRM) policies through learning from international experience, and to establish links with the EU WFD. The Chishui River was evaluated using this dataset and the two weighting systems (initial and ANP-adjusted). The evaluation outcomes generated from both weighting systems are compared below.

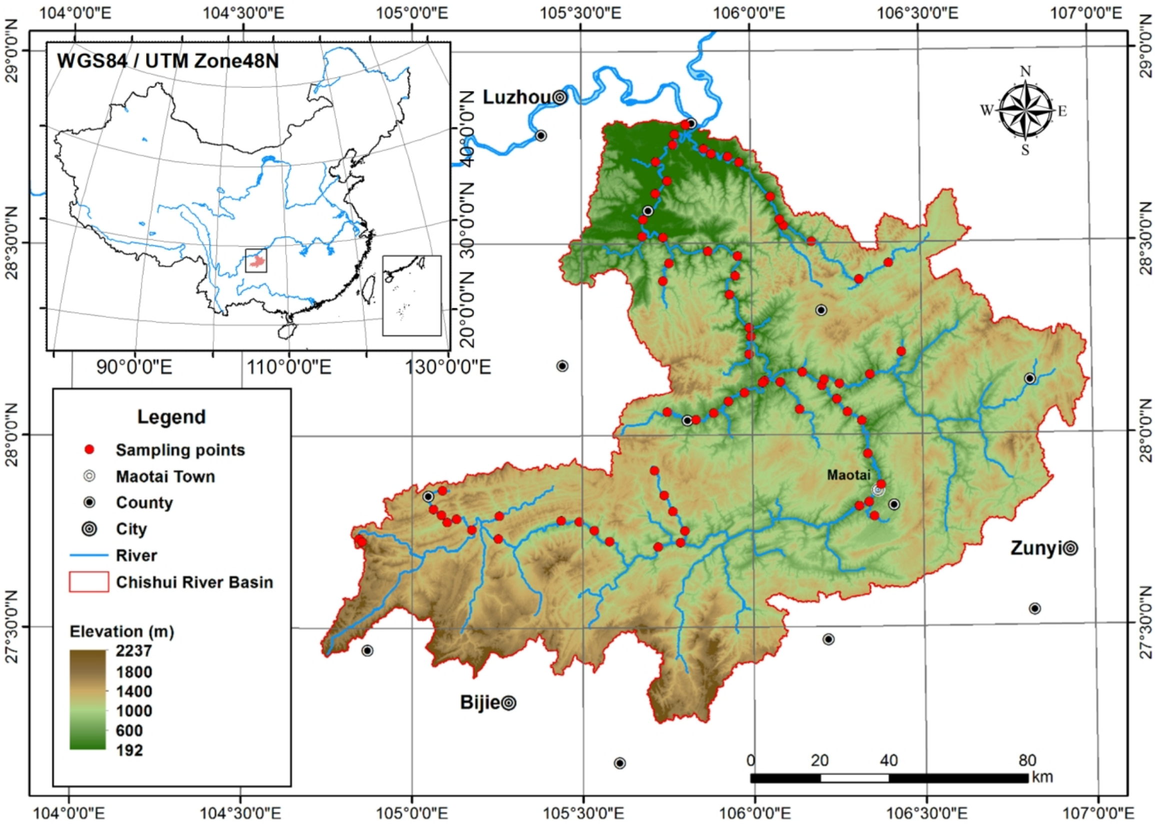


Figure 3. Location of the Chishui River Basin and the sampling reaches.

3.2 Step 1: Evaluation structure and interaction network construction

This section illustrates the two main parts of evaluation structure construction, including (1) choosing components and variables in an evaluation hierarchy, and (2) building a network of linked variables.

3.2.1 Evaluation of components and variables

This study applied initially a simplified ecological water quality evaluation system based on the WFD system and documents, with identifying the essential clusters, components, variables and interactions. According to the WFD, ecological water quality refers to a comprehensive assessment of the state of aquatic ecosystems (European Commission, 2000; Mao and Richards, 2012); and it is composed of three clusters of variables representing biological, physicochemical and hydromorphological conditions. In total, 7 variables commonly used in WFD practices are selected to represent the components in these three clusters. Macro-invertebrates and diatoms are selected as two biological groups to be assessed, with the Average Score Per Taxon (ASPT) and the Trophic Diatom Index (TDI) as the sensitivity metrics for invertebrates and diatoms respectively (UKTAG, 2008a, 2008b). The Shannon Diversity Index is used to measure biodiversity for both biological groups (Hering et al., 2006a). For the physicochemical cluster, an integrated chemical water quality index is selected as the only variable, which is built upon the China's current surface water quality evaluation (SEPA, 2002), and consistent with the physicochemical category in the WFD (UKTAG, 2009). This chemical water quality index is used to reflect the general state of eutrophication and pollution (i.e. dissolved oxygen concentration, pH and soluble reactive phosphorus concentration) in the water. The condition of the river banks and riparian zone, and the condition of the floodplain are selected and measured as hydromorphological components to indicate the naturalness of hydromorphology at different scales and in different river zones (CEN, 2010; NIEA, 2014). For example, concrete reinforced banks and agricultural floodplain have low condition scores in terms of hydromorphological naturalness. Each of the 7 variables was standardised and normalised to a range between 0 and 1 as an Ecological Quality Ratio (EQR), with larger values representing a better water quality condition or higher degree of naturalness (European Commission, 2000). Table 1 lists the sources of the 7 variables and their rescaling strategies.

Table 1. Calculation methods and rescaling strategies for the 7 variables.

Variable	Calculation method and rescaling strategy	Calculation reference
ASPT	ASPT value is calculated as below $$ASPT = \frac{\sum_{i=1}^{N} p_i \times BMWP_i}{\sum_{i=1}^{N} p_i}$$ where p_i denotes the presence absence of the i^{th} taxon (p=1 for presence; p=0 for absence), and *BMWP* is a sensitivity score ranging from 1 (least sensitive) to 10 (most sensitive). 10 is the largest possible ASPT value. EQR_{ASPT} = ASPT/10	UKTAG (2008b)
TDI	TDI value is calculated as below $$TDI = \frac{\sum_{i=1}^{N} a_i \times s_i}{\sum_{i=1}^{N} a_i} \times 25 - 25$$ where a_i denotes the proportional abundance of the i^{th} specimen and s_i is the nutrient sensitivity class (1–5) of the i^{th} specimen. The original TDI value ranges from 0 to 100, with lower value indicating lower trophic levels. EQR_{TDI} = (100-TDI)/100	Kelly et al. (2008) and UKTAG (2008a)
Shannon Diversity for invertebrates and diatoms	Shannon diversity is calculated as below, $$H' = -\sum_{i=1}^{S} p_i \ln p_i$$ where *H'* denotes Shannon diversity, and p_i denotes the proportional abundance of the i^{th} specimen out of the total number of species *S*. According to the distribution of Shannon diversity values, the reference value is set as 3. If $H' \geq 3$, $EQR_{Shannon} = 1$; if $H' < 3$, $EQR_{Shannon} = H'/3$.	Hauer and Lamberti (2007)
Chemical water quality	An EQR is allocated for each chemical water quality class. Class I – 1, Class II – 0.8, Class III – 0.6, Class IV – 0.4, Class V – 0.2, Worse than Class V – 0.	SEPA (2002)
Banks and riparian zone (BR) and Floodplain (FP)	According to CEN (2000), BR and FP are classified into five levels from 1 to 5 based on the naturalness of the hydromorphological condition. The largest possible score for these two variables is 5, representing severely modified condition. EQR_{BR} = (5-BR)/5, EQR_{Fp} = (5-FP)/5.	CEN (2010)

These variables are firstly organised in a hierarchy of four levels that are linked by arrows (see top half of Figure 4). Each arrow starts at a 'source' and ends with the 'target', indicating the influence of the source on a specific target. In the top half of Figure 4, the arrows point upwards and identify how the seven variables are integrated into an overall evaluation of ecological water quality, *via* components and clusters. At the top level of the hierarchy, two stages of the DPSIR model are considered; the *State*, representing the overall evaluation of the aquatic ecosystem, and the *Pressure*, representing the anthropogenic pressure imposed on the ecosystem, which is here, the condition of point source pollution (PSP).

3.2.2 Interactions among components and variables

Notwithstanding this hierarchical structure, the components and variables interact with each other, and the interactions and their magnitudes can be decided based on experience and previous studies. The bottom half of Figure 4 highlights the interactions among the variables. Biological communities are influenced by both physicochemical and hydromorphological conditions. Physicochemical conditions are influenced by hydromorphological conditions and by external point-source pollution. The hydromorphological condition, especially the floodplain modification, is on the one hand a component of ecological water quality evaluation, and on the other hand represents agricultural land-use and land-cover in this case, which implies non-point source pollution that impacts on both the physicochemical and biological condition.

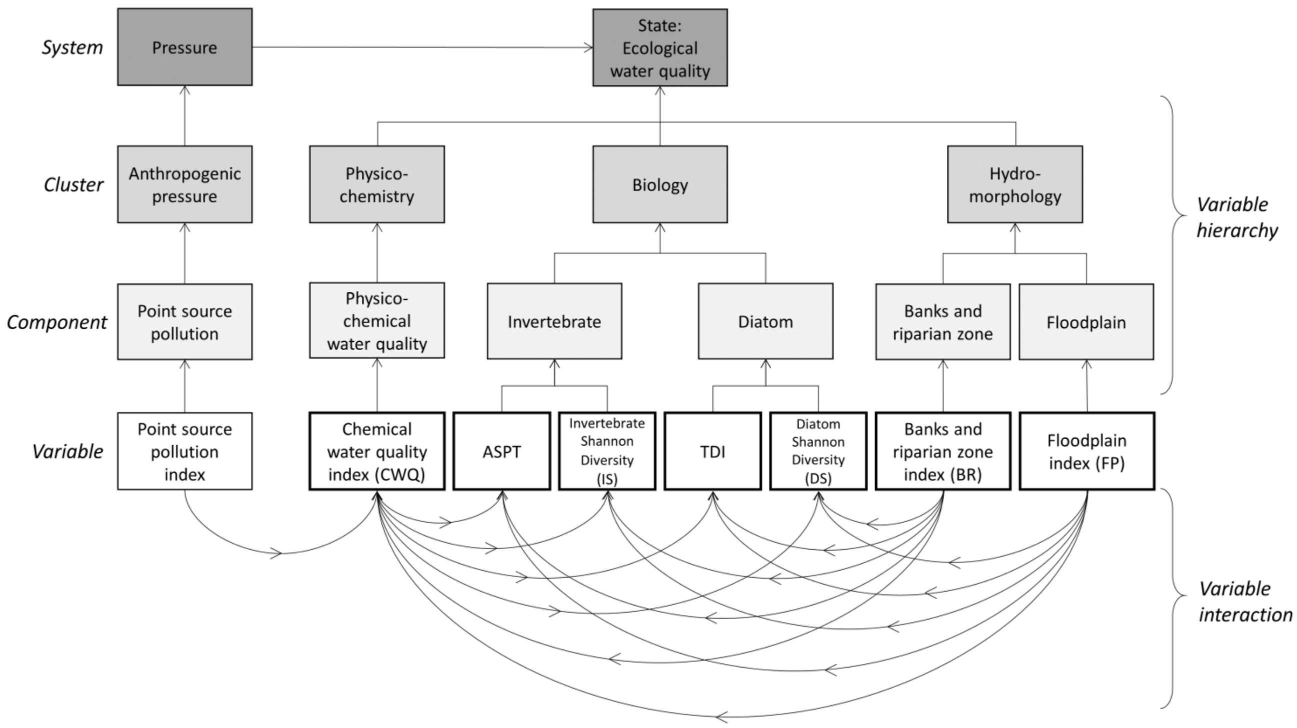


Figure 4. Structure of ecological water quality evaluation. The straight arrows denote the hierarchy of variables while the curved arrows denote interactions among variables. The seven bold-bordered white boxes denote the variables included in the composite indicator.

3.3 Step 2: Initial weight determination

Step 2 calculates the initial weights without information about variable interactions. An AHP-style pairwise comparison approach was used, which compares the contributions of elements (i.e. clusters, components and variables) on their hierarchical levels to the upper level, and combines them into an overall score of ecological water quality. It is worth noting that the aim of this step is to assign initial weights that are later to be adjusted in the next step; and the pairwise comparison approach is not the only option. However, it was used to aid a comparison with the ANP approach in the next step.

A scoring system developed by Saaty (2008) can be used for pairwise comparison. Scores range from 1 to 9, representing increasing importance of the effect of one item on another. If the relative importance of two items is unclear or hard to compare, a score of 1 is applied reflecting an assumption of equal importance. A score of 9 represents the greatest difference in relative importance. If one activity has a number from 1 to 9 assigned to it when compared with other activity, then the latter activity has the reciprocal value when compared with the former one. Comparison matrices are built to calculate the weights. In this study, the value of the score denotes how much more influence or importance the column item has on the row item (Vacik et al., 2007; Wolfslehner and Vacik, 2008).

Table 2 demonstrates the process of weight calculation using pairwise comparisons at three levels of the evaluation hierarchy (i.e. cluster, component and variable), while the two variables of Component Invertebrate (i.e. ASPT and Diversity) are selected as examples. The top level shows a 3×3 comparison matrix of the three main clusters: biology, physicochemistry and hydromorphology. Following the biologically-centred idea of water quality assessment (European Commission, 2000), the biological group has a more important role than the physicochemical and hydromorphological groups in determining ecological water quality evaluation. So, the cells in the first row, second and third columns in the matrix of Table 2 have values of 5; accordingly, the cells in the second and third row, first column, have scores as the reciprocal of 5. Physicochemistry and hydromorphology have equal contributions to the overall evaluation, so the cells in the second row third column and third row second column have values of 1. A value of 1 is assigned for self-comparisons. The largest eigenvector of this 3×3 comparison matrix is produced, and is standardised by making the sum equal to 1. This standardised eigenvector is filled in the adjacent column as a priority weight. In the comparison matrices of (2) components in Cluster Biology and (3) variables in Component Invertebrate, all the cells are assigned by 1, according to the assumptions that the invertebrate and diatom have similar contributions to the biological cluster while the ASPT and Diversity contribute evenly to the component of invertebrates.

Table 2. Pairwise comparisons to calculate the weights for Invertebrate ASPT and Shannon Diversity. The priority weights are the standardised eigen vectors of the matrix of scores at each level in the hierarchy.

Comparison level with respect to overall evaluation		(1.1)	(1.2)	(1.3)	(2.1)	(2.2)	(3.1)	(3.2)	Priority weight
(1) Clusters in overall evaluation	(1.1) Biology	1	5	5					0.714
	(1.2) Physicochemistry	1/5	1	1					0.143
	(1.3) Hydromorphology	1/5	1	1					0.143
(2) Components in Cluster Biology	(2.1) Invertebrate				1	1			0.5
	(2.2) Diatom				1	1			0.5
(3) Variables in Comp. Invertebrate	(3.1) ASPT						1	1	0.5
	(3.2) Shannon Diversity						1	1	0.5

The initial weight of each variable derives from the combination of relevant priority weights across the three levels. For example, the initial weight of the invertebrate ASPT is the product of: (a) the priority weight of biology among the clusters (0.714), (b) the priority weight of invertebrate within the biological cluster (0.5), and (c) the priority weight of ASPT within the invertebrate component (0.5) (Equation 1; see Table 3).

$$\text{Weight}_{\text{ASPT}} = 0.714 \times 0.5 \times 0.5 = 0.179 \quad \textbf{(1)}$$

To summarise, pairwise comparisons are performed for all clusters, components and variables following a simple set of assumptions. First, the biological cluster is ‘strongly’ more important than physicochemical and hydromorphological clusters according to the WFD regulations. Second, the physicochemical and hydromorphological variable clusters make equal contributions to the overall evaluation. Third, in each cluster, all the variables have the same weight. The full set of initial weights is provided in Table 3.

Table 3. Initial weights for each variable.

Cluster	Variable	Abbreviation	Initial weights
Biology	Invertebrate ASPT	ASPT	0.179
	Invertebrate Shannon Diversity	IS	0.179
	Diatom TDI	TDI	0.179
	Diatom Shannon Diversity	DS	0.179
Physicochemistry	Chemical water quality	CWQ	0.143
Hydromorphology	Banks and Riparian zone	BR	0.071
	Floodplain	FP	0.071

3.4 Step 3: Weight adjustment to reflect interaction

This step adjusts the initial weights by considering the variable interactions. For each variable, the relative importance of all other influencing variables is compared by reviewing literature, with the result summarised in Table 4. Hering et al (2006b) compared the responses of organism groups to different anthropogenic stresses and found invertebrates and diatoms are both highly sensitive to chemical water condition and sensitive to land-use. However, invertebrate and diatom communities are more sensitive to large-scale land use than reach scale hydromorphology, and this sensitivity difference is more significant for diatoms (Feio et al., 2009; Hering et al., 2006b, 2006c, Johnson et al., 2006a, 2006b). ASPT and TDI as sensitivity metrics are designed to detect physicochemical conditions in the water (UKTAG, 2008a, 2008b). However, the hydromorphological condition has a smaller but non-neglectable influence on these indices (Dahm et al., 2013). The Shannon Diversity Index as a biodiversity index, usually has a weaker correlation relationship with the stresses than those sensitivity indices (Dahm et al., 2013; Hughes et al., 2009). However, degraded physical habitat and chemical conditions usually do lead to decreased biodiversity, and even to a loss of aquatic ecosystem function (Bona et al., 2008; Heatherly et al., 2007).

There is a significant relationship between land use and chemical water quality. For example, agricultural and impervious urban land release higher nitrogen and phosphorus levels into surface water as diffuse pollutants (Tong and Chen, 2002). The condition of the floodplain is believed to have a slightly greater influence on chemical water quality than that of the banks and riparian zone (Sliva and Williams, 2001). Banks and riparian zones can be regarded as buffers protecting water quality from non-point source pollution which mainly originates from agricultural land-use (Norris, 1993; Wasson et al., 2010). Point source water pollution refers to an identifiable source of pollution, such as wastewater discharges from factories and sewage treatment plants. Point source pollution directly changes the physicochemical condition, but indirectly influences biological communities via the changes of physicochemical condition.

Table 4. Relative importance of variables with respect to each other. The last column 'Importance score' indicates the importance of the effect of element A on element B. For example, the first row implies that Cluster Physicochemistry is more strongly important (5) than Cluster Hydromorphology with respect to ASPT.

With respect to variable	Comparison level	Element A	Element B	Importance score (A compares to B)
ASPT	Cluster	Physicochemistry	Hydromorphology	5
	Hydro. variable	Floodplain	Banks and Riparian zone	3
IS	Cluster	Physicochemistry	Hydromorphology	1
	Hydro. variable	Floodplain	Banks and Riparian zone	3
TDI	Cluster	Physicochemistry	Hydromorphology	5
	Hydro. variable	Floodplain	Banks and Riparian zone	5
DS	Cluster	Physicochemistry	Hydromorphology	1
	Hydro. variable	Floodplain	Banks and Riparian zone	5
CWQ	Cluster	Hydromorphology	Anthropogenic pressure	1
	Hydro. variable	Floodplain	Banks and Riparian zone	1/3

With the importance scores in Table 4, comparison matrices can be constructed similar to those in Table 2. Table 5 is an example demonstrating pairwise comparisons between and within clusters with respect to the invertebrate ASPT. This table reflects the fact that the physicochemistry is more influential than the hydromorphology in determining the ASPT levels, while within the hydromorphology clusters, the floodplain is slightly more important than the banks and riparian zone in relation to the ASPT values (Table 5).

Table 5. Pairwise comparisons to calculate the relative importance of variables to Invertebrate ASPT.

Comparison level with respect to Invertebrate ASPT		(1.1)	(1.2)	(1.3)	(1.4)	Priority weight
(1) Clusters	(1.1) Physicochemistry	1	5			0.833
	(1.2) Hydromorphology	1/5	1			0.167
(2) Variables and components in Cluster Hydromorphology	(2.1) Banks and Riparian zone			1	1/3	0.250
	(2.2) Floodplain			3	1	0.750

The comparison iterates until all links have been considered. A weighted super-matrix can be obtained from the priority weights of all the pairwise comparisons (Table 6). The super-matrix has two main parts. The first column represents the initial weights of all evaluation variables, while the remaining columns denote the relative importance of influencing factors in respect to certain variables, or the interaction among variables. The sum of each column is set to be 1. According to the calculated priority weights, the conditions of Banks and Riparian zone (BR) and Floodplain (FP) respectively contribute 0.042 (0.167×0.25) and 0.125 (0.167×0.75) of the total influence on ASPT (see Table 6 for full results).

Table 6. Weighted super-matrix. The sum of each column is 1. The first column denotes initial weights calculated by Step 2. The remaining columns denote the interaction among variables, which are used to adjust the initial weights. All diagonal cells yield zero, because no self-looping is identified.

		EWQ	ASPT	IS	TDI	DS	CWQ	BR	FP	PSP
Ecological Water Quality	EWQ	0	0	0	0	0	0	0	0	0
Biology	ASPT	***0.179***	0	0	0	0	0	0	0	0
	IS	***0.179***	0	0	0	0	0	0	0	0
	TDI	***0.179***	0	0	0	0	0	0	0	0
	DS	***0.179***	0	0	0	0	0	0	0	0
Physicochemistry	CWQ	***0.143***	0.833	0.500	0.833	0.500	0	0	0	0
Hydromorphology	BR	***0.071***	0.042	0.125	0.028	0.083	0.125	0	0	0
	FP	***0.071***	0.125	0.375	0.139	0.417	0.375	0	0	0
Point source pollution	PSP	0	0	0	0	0	0.500	0	0	0

Based on the *weighted super-matrix*, a *limit super-matrix* can be calculated as the resultant of the priority weights. The limit super-matrix is calculated by summing up the weighted super-matrix with raising exponent value k (Equation 2). The right side of Equation 2 has a form of the Cesaro summation, which converges to a limit (Saaty, 1996; Wolfslehner et al., 2005). The limited super-matrix is obtained when the convergence is achieved.

$$W_{limit} = \lim_{k\to\infty} \left(\frac{1}{N}\right) \sum_{k=1}^{N} W^k \quad (2)$$

where W_{limit} is the limit super-matrix, W is the weighted super-matrix, N denotes the step of iteration, and k denotes the exponent, which is arbitrarily selected (Wolfslehner et al., 2005). W^k is a matrix product of a k number of W.

Table 7 shows the calculated limit super-matrix, which becomes stable when k≥4. The bold italic digits in the first column represent the adjusted relative importance of each variable with respect to the overall ecological water quality evaluation. These values were then rescaled to sum to one, to remove the weight of the evaluation variable point source pollution.

Table 7. Limit super-matrix. The sum of each column is 1. The bold italic digits are the adjusted weights before standardisation.

		EWQ	ASPT	IS	TDI	DS	CWQ	BR	FP	PSP
Ecological Water Quality	EWQ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Biology	ASPT	***0.065***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	IS	***0.065***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	TDI	***0.065***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	DS	***0.065***	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Physicochemistry	CWQ	***0.226***	0.455	0.333	0.455	0.333	0.000	0.000	0.000	0.000
Hydromorphology	BR	***0.091***	0.080	0.125	0.072	0.097	0.125	0.000	0.000	0.000
	FP	***0.235***	0.239	0.375	0.246	0.403	0.375	0.000	0.000	0.000
Point source pollution	PSP	0.187	0.227	0.167	0.227	0.167	0.500	0.000	0.000	0.000

3.5 Weight assignment results and river basin evaluation

After being adjusted by variable interactions, the weight of the biological cluster significantly decreases from 0.714 to 0.321, and correspondingly the weight of each biological variable drops from 0.179 to 0.080. The weight of chemical water quality nearly doubles from 0.143 to 0.278. The weight of the hydromorphological cluster also increases, from 0.143 to 0.402. However, the weight of the floodplain has a higher increase than that of the banks and riparian zone; the former weight increases from 0.071 to 0.112 while the latter weight increases from 0.071 to 0.290.

Two different evaluation scores of the Chishui River basin were calculated using the two sets of weights, and are compared in Figure 5 using an R plotting package *rivervis* (Mao et al., 2019). The figure shows a similar pattern and variation of initial and ANP-adjusted evaluation scores, indicating the water quality is impaired by towns and recovers along with the river. However, the adjusted scores seem slightly more sensitive in detecting water quality dynamics along the river, because of the larger ranges of their evaluation values. For example, downstream from Maotai on the main stream, as well as on tributaries such as the Zhaxi, Baisha and Gulin (see Figure 5), the overall evaluations with adjusted weights are reduced more significantly by the towns, and recover more rapidly downstream from towns. This is because of the contributions from hydromorphological and physicochemical factors – the hydromorphology around the towns is heavily modified, and the pollutant discharges are high, but the former improves and the latter are diluted downstream from the urban areas according to the sampling data.

The increased weights of the physical variables change the overall evaluations in different ways depending on the location of sampling reaches in the basin. Along the mainstream of the Chishui River, most sampling reaches have a higher score after weight adjustment, especially in reaches around the three major towns (Figure 5). In these areas, hydromorphological scores were better than the biological and physicochemical variables, because the valley morphology makes the river less easily accessible to, or useable by, local people. However, several tributaries show the opposite situation, and the initial indices have higher values. These tributaries include the Dacun, Gulin, Yanjin, Baisha, Zhaxi and Yuhe (Figure 5). In most midstream and upstream tributaries, the physicochemical and hydromorphological scores are relatively lower than the biological scores, since the settlements along the rivers are much smaller.

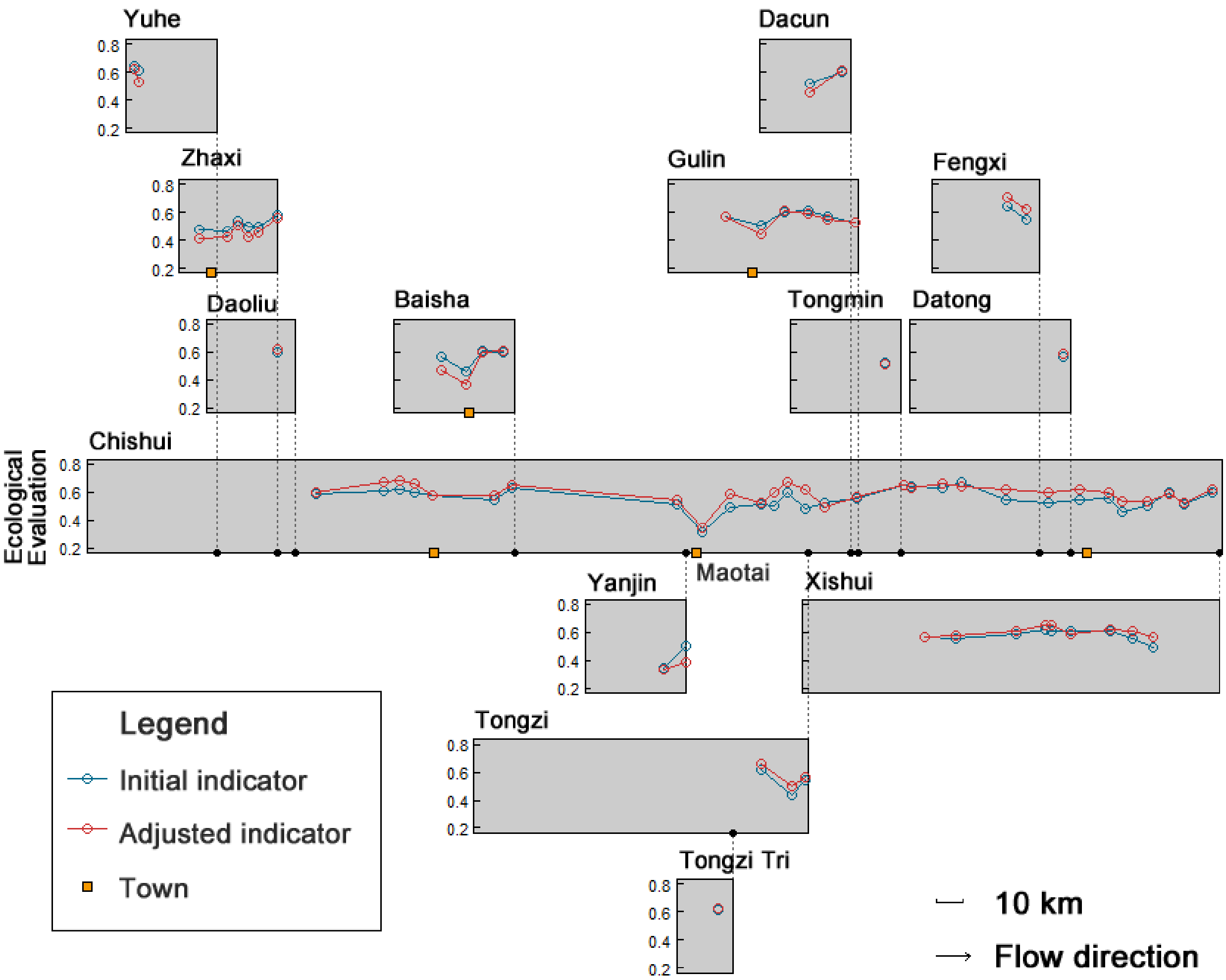


Figure 5. Comparison between initial and ANP-adjusted evaluation scores for the Chishui River Basin, based on ecological data from 2012. The blue circles denote initial scores calculated with the initial set of weights, and the red circles denote ANP-adjusted scores calculated with the adjusted set of weights. This figure is created with the rivervis R package (Mao et al., 2019).

3.6 Sensitivity analysis

Two sensitivity analyses were performed to test the robustness of the approach to the input perturbation. In the first analysis, we tried to determine the sensitivity of adjusted variable weights to changes in the assumed importance of interaction among variables. Twenty sets of experiments in five groups were performed (Table 8). Each experiment replaced one column in the weighted super-matrix with an alternative set of the relative importance of influencing variables (Table 6). Each group has four experiments with respect to the same variable to be influenced. Within each group, the first experiment represents an even contribution of the influencing variables, and the other three each represents a scenario that the influence is dominated by one influencing variable. For example, in Experiment 1, the CWQ, BR and FP have equal importance to ASPT, and this setting replaced the ASPT column in Table 6, and formed a new weighted super-matrix.

Table 8. Sensitivity analysis experiments for the variable weights. Variables not shown in this table have all-zero weights.

With respect to	ASPT				IS				TDI				DS				CWQ			
Experiment	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8	Exp 9	Exp 10	Exp 11	Exp 12	Exp 13	Exp 14	Exp 15	Exp 16	Exp 17	Exp 18	Exp 19	Exp 20
CWQ	***0.33***	***0.90***	***0.05***	***0.05***	***0.33***	***0.90***	***0.05***	***0.05***	***0.33***	***0.90***	***0.05***	***0.05***	***0.33***	***0.90***	***0.05***	***0.05***	0	0	0	0
BR	***0.33***	***0.05***	***0.90***	***0.05***	***0.33***	***0.05***	***0.90***	***0.05***	***0.33***	***0.05***	***0.90***	***0.05***	***0.33***	***0.05***	***0.90***	***0.05***	***0.33***	***0.90***	***0.05***	***0.05***
FP	***0.33***	***0.05***	***0.05***	***0.90***	***0.33***	***0.05***	***0.05***	***0.90***	***0.33***	***0.05***	***0.05***	***0.90***	***0.33***	***0.05***	***0.05***	***0.90***	***0.33***	***0.05***	***0.90***	***0.05***
PSP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***0.33***	***0.05***	***0.05***	***0.90***

In total, twenty weighted super-matrices were constructed and therefore twenty alternative sets of weights were calculated using the procedure described in Section 3.4. The results, together with the initial and adjusted weights, are shown in Figure 6 which illustrates that:

- Experiments 1-17 have similar results to the ANP-adjusted weight setting.
- Experiments 18-19 have significantly higher hydromorphological weights, but Experiment 18 has much higher BR weight and Experiment 19 has much higher FP weight.
- Experiment 20 has lower hydromorphological weights.

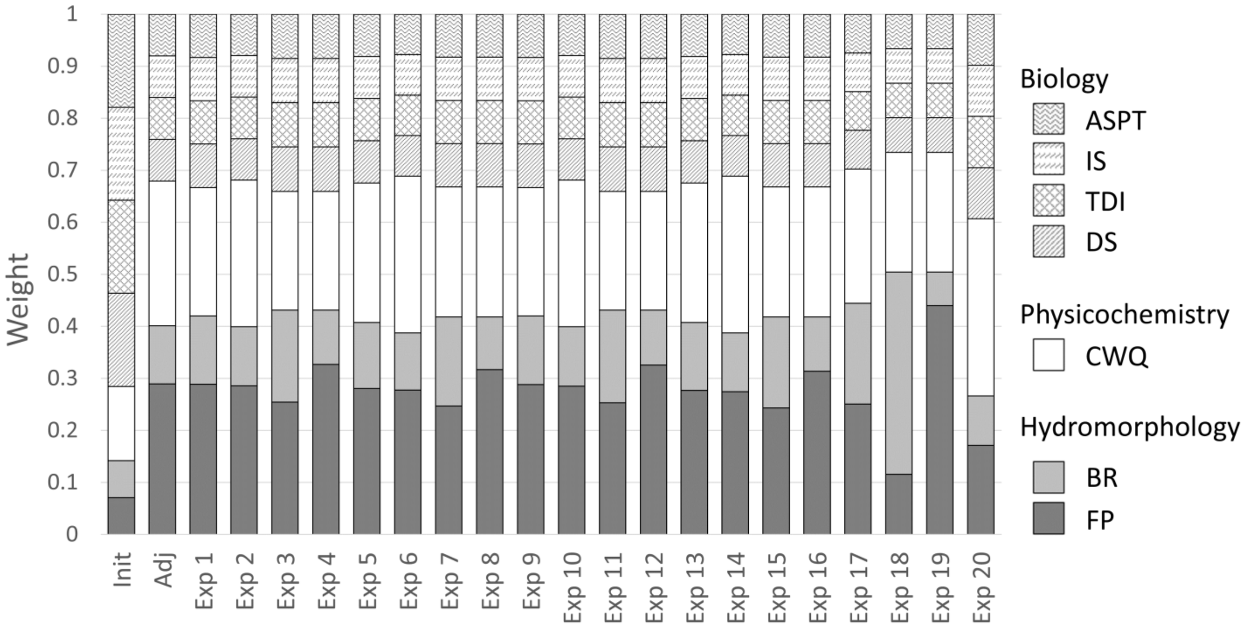


Figure 6. Sensitivity analysis of the variable weights to changes of interaction among variables. The initial, adjusted and 20 experimental weight settings are shown.

The second analysis was to test the sensitivity of the overall evaluation scores to the change of weights. The evaluation scores for each sampling reach were calculated using all the 20 experimental weight settings from the first sensitivity analysis. The percent deviations (%) were derived by comparing the experimental scores to the initial scores as well as the ANP-adjusted scores of each sampling reach. The result is visualised in heat maps in Figure 7.

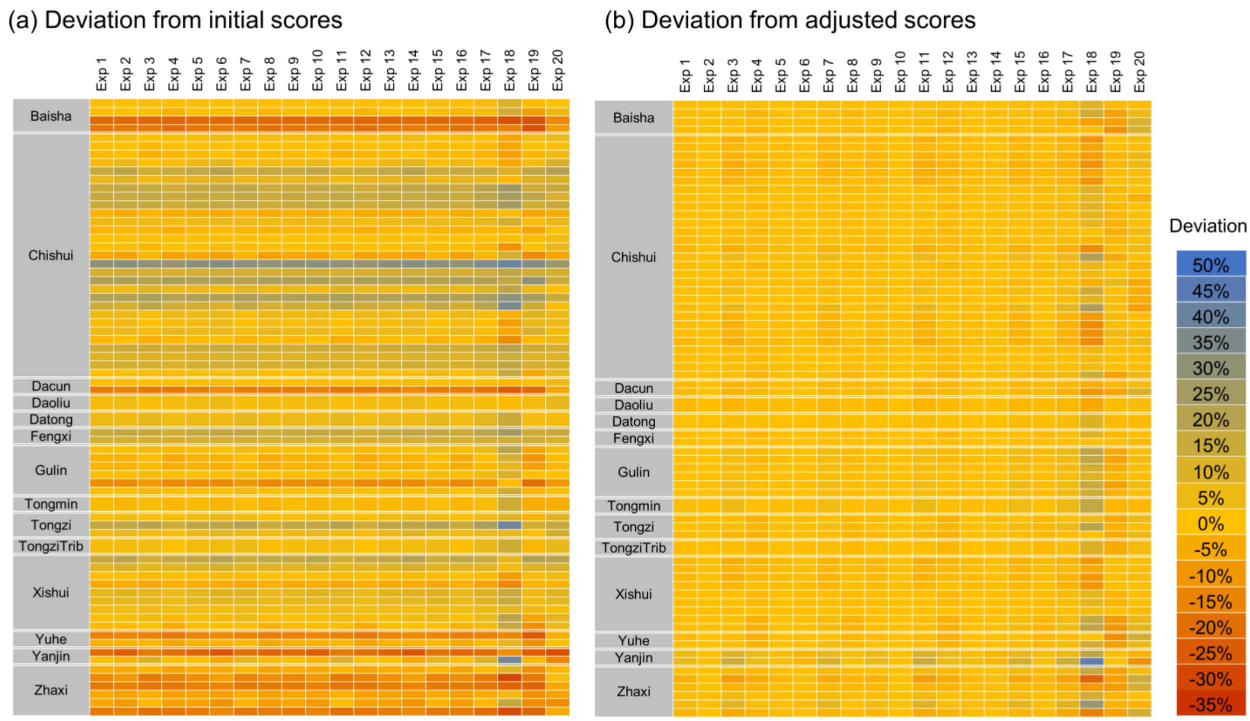


Figure 7. Sensitivity analysis of the evaluation scores to changes of weights. The 20 experimental weight settings are used to calculate experimental performance scores for each sampling sites. The experimental evaluation scores for each sampling site are compared with the initial (a) and adjusted (b) score, and the percent deviations (%) are derived. In each catchment, sampling sites are sorted from the river mouth (top) to the source (bottom).

According to Figure 7, nearly all experiments make almost the same adjustment to the initial scores (Figure 7-a), which is presented in Figure 5; most of the evaluation scores in all experiments have very similar values to the ANP-adjusted scores, with the percent deviation between -5% and 5% (Figure 7-b). This result shows the high robustness of the proposed method. However, Experiment 18-19 have different patterns of deviation from the rest experiments. For example, in Figure 7-b, Experiment 18 has a wider range of deviation, with some scores having much higher or lower estimations than the ANP-adjusted scores while Experiment 19 shows an opposite direction of deviation in many reaches compared to Experiment 18. It is because that these two experiments have much higher hydromorphological weights, and their estimated scores are largely influenced by the performance of BR and FP respectively. In addition, between these two variables, BR has a higher variety of performance across the Chishui River Basin. For example, in the downstream of the Chishui, Xishui and Zhaxi River and the midstream of the Chishui River, BR has lower scores due to the modification in and around towns. On the contrary, BR has a better performance in the Fengxi, Gulin Rivers and the upstream of the Tongzi River because of their near natural bank and riparian conditions. The changes of the overall evaluation scores in Experiment 18 are determined by the performance of BR due to its high weight. However, the weight of the highly variable BR is significantly reduced in Experiment 19, which leads to a different direction of changes in these reaches.

4. Discussion

This section discusses the performance and features of the proposed ANP-based approach, by comparing the initial and adjusted weight settings, as well as the overall evaluations calculated with both weight settings. From the comparisons and analyses in relation to three aspects: (1) weight settings, (2) overall evaluation, and (3) information considered in expert judgment, this section shows how the additional information of component interactions and networks is incorporated into the elicitation, evaluation, and decision-making processes, and making the processes more objective and justifiable.

4.1 Weight settings

Calculating the limited super-matrix using matrix multiplication (Equation 2) can be understood as a process to obtain the accumulated influence of the network on each variable. The changes in the weights for variables or variable groups are brought about by their interactions following one principle – more influential variables have a higher weight increase, and less influential variables a lower one. In this example, physicochemical and hydromorphological groups influence biological variables but the four biological variables do not influence each other. This results in a general increase in the weight of abiotic variables. Additive effects can be observed. For example, the importance of the hydromorphology increases most significantly because it directly affects biological conditions, and also has considerable effects on chemical water quality that also influence biological communities. This is evident in the sensitivity analysis (Section 3.6); changing the relative importance of hydromorphological variables for chemical water quality dramatically alters their weights in the outcome (Experiments 18-20).

Many factors potentially affect the outcomes of adjusting weights, but the selection of variables and the ecosystem types may be two particularly significant ones, and these may require special attention when using this method to combine evaluation components and variables. In the current study, the biological cluster has four variables: two of which are physicochemical-sensitivity indicators (ASPT and TDI) and two of which are diversity indicators (Shannon diversity). If the biological cluster in the scoring network is dominated by sensitivity variables, it receives greater influence from the physicochemical cluster; and therefore the weight of the physicochemical cluster increases relative to the current arrangement. An additional important factor is the location within different ecosystem types. Ecological variables may interact differently, and therefore have different weight settings, in different river types (García et al., 2014; Montgomery, 1999; Schmutz et al., 2000). For example, the ASPT scores may be more sensitive to organic pollution in midstream reaches than in other river types; while the riparian condition, such as the existence of buffer zones, is more effective in protecting rivers from chemical contamination in higher order streams (Anbumozhi et al., 2005). It therefore implies that type-specific weighting systems have a great potential if they don't bring in too much additional complexity and redundancy to the assessment (García et al., 2014; Jungwirth et al., 2002). In this condition, it can offer a considerable improvement over fixed methods of variable combination, and clearly shows how the whole system of assessment can be integrated in a more sophisticated way.

4.2 Overall evaluations and a risk-based interpretation

As stated, constructing composite indicators can be regarded as a process of multiple-criteria decision making. Conflicting criteria are typical, and actually a common phenomenon in this process (Huang et al., 2011). For example, how can the overall status of a river ecosystem be evaluated if the performance of macroinvertebrates is high but the performance of hydromorphology is low. Most of the time, there is not a single optimal representation of the system. Instead, this process reflects understanding of the ecosystem and the preferences among different settings. Compared to the initial indicator, the proposed approach provides a new strategy to incorporate additional ecological information (e.g. ecological networks and accumulated influence of interaction) in the decision-making process in an attempt to improve the treatment of such conflicting criteria.

There can be many ways to interpret the connections between the two indicators and the confliction of component performance. In this sub-section, we offer a 'risk-based' explanation: the *initial indicator* structure defined by experts gives a meaning to the ecological water quality; the *adjusted indicator value,* dependent on potentially dynamic interactions, refers to possible short-term changes of the *initial indicator*, and may imply environmental risks.

As discussed above, indicator variables do not always have the same performance. If the more influential variables underperform the variables that are less influential but more dependent, there are two interlinked consequences. First, a negative outlook of the initial score in the future. It is because the poorer performance of the more influential variables may bring down the dependent variables in long-term (see Section 4.1 for detailed explanation). Second, a lower ANP-adjusted score than the initial score. It is because the more influential variables are likely to gain weights through ANP adjustments. Vice versa, outperformed more influential variables suggest a positive future expectation of the initial score and a higher adjusted score. This implies that the mismatches between the initial and adjusted scores can be understood as signals of likely future tendencies, or risk of having a decreased overall performance – the risk is low if the adjusted score is higher than the initial one; whereas, the risk is high if the adjusted score is lower than the initial one.

In the Chishui example, physicochemical and hydromorphological variables were more influential, but performed less well than the biological one in many tributary sampling reaches. Although this does not necessarily predict that the biological condition will be impaired by unsatisfactory abiotic factors, this mismatch does imply a negative outlook of the biological components and consequently the overall evaluation. Therefore, in these reaches, the adjusted indicators have low values. On the contrary, in some mainstream reaches, the physical factors outperform the biological measures, which suggests there is little evidence of environmental risks that potentially bring down scores for the biological elements, which results in a higher adjusted indicator value. It is worth noting that the accuracy of this risk-based thinking depends on accurate considerations of component interactions. Incomplete or incorrect interactions may make the changes associated with the variables unreliable.

4.3 Interaction as additional information

In the analytical example, there are two calculation steps, two types of information, two weighting sets and two overall evaluations. *Step 2* determines the initial variable weights based on the hierarchy of composite indicators, reflecting experts' subjective understanding of the 'composition' of ecological water quality, how that composition is defined, and the preferences for variable selection. The Chishui example defined higher weights for the biological cluster; within which macro-invertebrates and diatoms had equal weights, and were assessed with sensitivity and diversity metrics. These settings were based on the WFD's ecologically-centred philosophy (European Commission, 2000), and on multi-metric index practices (Hering et al., 2006a). *Step 3* is a process to adjust the initial weights and the composite indicators with considering important ecological mechanisms and variable interactions. These interrelationships and interactions provide ecological meaning to the weight determination process by explaining why certain evaluation components and variables are more important than others. Although the construction of the interaction network is inevitably limited by knowledge and experience, the additional information makes the weight determination process less subjective, less arbitrary and more justifiable. Therefore, the adjustment changes the assessment of water quality and its evaluation, to a combination of not only the constitutive components, but also their interactions. This can potentially be applied to the existing composite indicators where expert opinions are substantially used and variable interactions are common, such as integrated environmental evaluations (see Halpern et al., 2012; Wiegand et al., 2010), sustainable development indicators (Hák et al., 2016), resilience assessment (Schipper and Langston, 2015), and socio-economic performance measurement (Bandura, 2008).

The involvement of the variable network extends the scope of possible questions to be used in expert elicitation processes when determining variable weights. In addition to asking directly (I) how much each variable contributes to the overall evaluation, new questions include (II) whether one variable has an influence on others, (III) if one variable is influenced by others, what is the relative importance of the influencing variables for the influenced one. The new questions are also compatible with those expert elicitation methods that are used to converge and summarise the opinions from multiple experts (e.g. Delphi method, Huang et al., 2011). These additional questions imply two potential strategies. The first strategy is similar to the analytic example in this manuscript – the answers to Question II and III are used to adjust or calibrate the weighting system established by answering Question I. The second strategy is to avoid Question I but presume that all variables have equal initial weight before asking Question II and III. More importantly, deciding the initial contribution of each variable to the overall evaluation may sometimes inevitably reflect some degrees of inter-variable interactions and consequently introduce double-counting into the follow-up network-based process. Therefore, how to choose from these two strategies in different expert elicitation scenarios requires further research.

The ANP-based approach has been proven to be both flexible and extensible in defining the weighting set, which will develop through usage of the network approach. The example demonstrated in Section 3 described a basic network, but the network can be expanded easily for the purpose of full WFD application in a more sophisticated analysis. More complicated networks require more information in establishing the super-matrix, and there is accordingly a balance between practical utility and theoretical complexity. Some possibilities of network expansion are suggested here. First, the network may include more complete human-nature interactions covering a larger number of more specific variables. For example, including the Response component of the DPSIR model in the network may offer a chance to consider environmental institutions and management as positive anthropogenic influences on ecosystems. Research in socio-ecological and socio-hydrological networks may also provide a solid basis for variable selection and interaction construction (Kumar, 2015; Lane, 2014; Mao et al., 2017; McCluney et al., 2014). Second, more detailed interrelationships, such as two-way interactions, feedbacks and self-looping, can be considered (Sivapalan and Blöschl, 2015). For example, diatoms as well as macrophytes are important producers of dissolved oxygen in water, contributing one of the key components of physicochemical water quality. As a result, diatoms and macrophytes are not only influenced by, but also exert influence on physicochemical conditions. Third, spatial and geographical networks can be included in the analysis. The Chishui example, for instance, calculated a performance score for each of the 71 sampling reaches. However, there may be a need to combine these scores into an overall evaluation of the whole basin. Simple averaging can be a strategy but considering the river network provides an alternative way to consider the different relative importance of each sampling reach, which can be built upon the upstream-downstream relations of rivers. For example, upstream reaches in a tributary have influence on its downstream, but may not affect the reaches in other tributaries. Comparison matrices can be established by analysing the spatial connections of all sampling reaches within a river basin.

5. Conclusion

This paper introduces a new, more flexible and extensible strategy to determine variable weights for ecological water quality assessment, on the basis of the ANP approach. Our approach provides a novel contribution as it integrates the long-neglected but useful information about networks and interactions of ecological components into composite indicators, to assess ecological water quality more holistically. Our network-based approach improves the objectivity of expert-based strategies in determining variable weights: the importance of more influential elements is raised by the adjustment process. This ANP strategy introduces a risk-based thinking in ecological water quality assessment, by introducing the additional information of network interactions, allowing comparison of the initial and adjusted index values, and interpreting the difference in relation to the factors influencing the adjustment. These risk-based indications help in preparing the most appropriate programme of measures to improve the overall quality sustainably. The proposed approach has the potential to be applied in other multi-criteria decision-making fields beyond water, such as environmental management in a broader sense, and indicators for sustainable development.

Acknowledgement

This research was developed within the 'River Basin Governance Research Network' project, which is supported by the ESRC grant [reference: RES-810-21-0071]. It was also supported by the National Natural Science Foundation of China [Nos. 51279113] and the Special Funds for Public Industry Research Projects of the National Ministry of Water Resources [Nos. 201501030]. We also thank Wei Chen and Sheng Chen who processed the biological samples.

References

Anbumozhi, V., Radhakrishnan, J., Yamaji, E., 2005. Impact of riparian buffer zones on water quality and associated management considerations. Ecol. Eng. 24, 517–523. doi:10.1016/j.ecoleng.2004.01.007

Bandura, R., 2008. A Survey of Composite Indices Measuring Country Performance: 2008 Update, A UNDP/ODS Working Paper.

Bisset, R., 1988. Developments in EIA methods, in: Wathern, P. (Ed.), Environmental Impact Assessment: Theory and Practice. Routledge, London, pp. 47–61.

Bona, F., Falasco, E., Fenoglio, S., 2008. Response of macroinvertebrate and diatom communities to human-induced physical alteration in mountain streams. River Res. Appl. 1081, 1068–1081. doi:10.1002/rra

Borja, A., Bricker, S.B., Dauer, D.M., Demetriades, N.T., Ferreira, J.G., Forbes, A.T., Hutchings, P., Jia, X.P., Kenchington, R., Marques, J.C., Zhu, C., 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Mar. Pollut. Bull. 56, 1519–1537.

Bottero, M., Comino, E., Riggio, V., 2011. Application of the Analytic Hierarchy Process and the Analytic Network Process for the assessment of different wastewater treatment systems. Environ. Model. Softw. 26, 1211–1224. doi:10.1016/j.envsoft.2011.04.002

Boulton, A.J., 1999. An overview of river health assessment: philosophies, practice, problems and prognosis. Freshw. Biol. 41, 469–479. doi:10.1046/j.1365-2427.1999.00443.x

Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L., Twardy, C., 2011. Expert status and performance. PLoS One 6, 1–7. doi:10.1371/journal.pone.0022998

Casini, M., Mocenni, C., Paoletti, S., Pranzo, M., 2015. Decision support system development for integrated management of european coastal lagoons. Environ. Model. Softw. 64, 47–57. doi:10.1016/j.envsoft.2014.11.008

CEN, 2010. Water quality — Guidance standard on determining the degree of modification of river hydromorphology.

Cummins, K.W., Merritt, R.W., 2001. Application of Invertebrate Functional Groups to Wetland Ecosystem Function and Biomonitoring, in: Rader, R.B., Batzer, D.P., Wissinger, S.A. (Eds.), Bioassessment and Management of North American Freshwater Wetlands. New York, pp. 85–112.

Dahm, V., Hering, D., Nemitz, D., Graf, W., Schmidt-Kloiber, A., Leitner, P., Melcher, A., Feld, C.K., 2013. Effects of physico-chemistry, land use and hydromorphology on three riverine organism groups: a comparative analysis with monitoring data from Germany and Austria. Hydrobiologia 704, 389–415. doi:10.1007/s10750-012-1431-3

Esselman, P.C., Infante, D.M., 2014. An index of Cumulative Disturbance to River Fish Habitats of the Conterminous United States from Landscape Anthropogenic Activities. Ecol. Restor. 29, 133–151. doi:10.1353/ecr.2011.0011

European Commission, 2000. Directive 200/60/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 Communities. Official Journal of the European Communities.

Faustini, J.M., Kaufmann, P.R., Herlihy, A.T., Paulsen, S.G., 2009. Assessing Stream Ecosystem Condition in the United States. Eos, Trans. Am. Geophys. Union 90, 309–310.

Feio, M.J., Almeida, S.F.P., Craveiro, 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. doi:10.1016/j.ecolind.2008.07.001

Forman, E., Saaty, T., Selly, M., Waldron, R., 1983. Expert Choice.

García, L., Pardo, I., Delgado, C., 2014. Macroinvertebrate indicators of ecological status in Mediterranean temporary stream types of the Balearic Islands. Ecol. Indic. 45, 650–663. doi:10.1016/j.ecolind.2014.05.030

Hák, T., Janoušková, S., Moldan, B., 2016. Sustainable Development Goals: A need for relevant indicators. Ecol. Indic. 60, 565–573. doi:10.1016/j.ecolind.2015.08.003

Halide, H., Stigebrandt, A., Rehbein, M., McKinnon, A.D., 2009. Developing a decision support system for sustainable cage aquaculture. Environ. Model. Softw. 24, 694–702. doi:10.1016/j.envsoft.2008.10.013

Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., Katona, S.K., Kleisner, K., Lester, S.E., O'Leary, J., Ranelletti, M., Rosenberg, A. a, Scarborough, C., Selig, E.R., Best, B.D., Brumbaugh, D.R., Chapin, F.S., Crowder, L.B., Daly, K.L., Doney, S.C., Elfes, C., Fogarty, M.J., Gaines, S.D., Jacobsen, K.I., Karrer, L.B., Leslie, H.M., Neeley, E., Pauly, D., Polasky, S., Ris, B., St Martin, K., Stone, G.S., Sumaila, U.R., Zeller, D., 2012. An index to assess the health and benefits of the global ocean. Nature 488, 615–20. doi:10.1038/nature11397

Hauer, F.R., Lamberti, G.A., 2007. Methods in Stream Ecology, Second Edi. ed. Academic Press.

Heatherly, T., Whiles, M.R., Royer, T. V, David, M.B., 2007. Relationships between water quality, habitat quality, and macroinvertebrate assemblages in Illinois streams. J. Environ. Qual. 36, 1653–60. doi:10.2134/jeq2006.0521

Hering, D., Feld, C.K., Moog, O., Ofenböck, T., 2006a. Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566, 311–324. doi:10.1007/s10750-006-0087-2

Hering, D., Johnson, R.K., Buffagni, A., 2006b. Linking organism groups – major results and conclusions from the STAR project. Hydrobiologia 566, 109–113. doi:10.1007/s10750-006-0098-z

Hering, D., Johnson, R.K., Kramm, S., Schmutz, S., Szoszkiewicz, K., Verdonschot, P.F.M., 2006c. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. Freshw. Biol. 51, 1757–1785. doi:10.1111/j.1365-2427.2006.01610.x

Hermann, B.G., Kroeze, C., Jawjit, W., 2007. Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. J. Clean. Prod. 15, 1787–1796. doi:10.1016/j.jclepro.2006.04.004

Herva, M., Roca, E., 2013. Review of combined approaches and multi-criteria analysis for corporate environmental evaluation. J. Clean. Prod. 39, 355–371. doi:10.1016/j.jclepro.2012.07.058

Hsu, A., Lloyd, A., Emerson, J.W., 2013. What progress have we made since Rio? Results from the 2012

Environmental Performance Index (EPI) and Pilot Trend EPI. Environ. Sci. Policy 33, 171–185. doi:10.1016/j.envsci.2013.05.011

Huang, I.B., Keisler, J., Linkov, I., 2011. Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. Sci. Total Environ. 409, 3578–94. doi:10.1016/j.scitotenv.2011.06.022

Hughes, S.J., Santos, J.M., Ferreira, M.T., Caraça, R., Mendes, A.M., 2009. Ecological assessment of an intermittent Mediterranean river using community structure and function: evaluating the role of different organism groups. Freshw. Biol. 54, 2383–2400. doi:10.1111/j.1365-2427.2009.02253.x

Johnson, R.K., Hering, D., Furse, M.T., Clarke, R.T., 2006a. Detection of ecological change using multiple organism groups: metrics and uncertainty. Hydrobiologia 566, 115–137. doi:10.1007/s10750-006-0101-8

Johnson, R.K., Hering, D., Furse, M.T., Verdonschot, P.F.M., 2006b. Indicators of ecological change: comparison of the early response of four organism groups to stress gradients. Hydrobiologia 566, 139–152. doi:10.1007/s10750-006-0100-9

Jungwirth, M., Muhar, S., Schmutz, S., 2002. Re-estabilishing and assessing ecological integrity in riverine landscapes. Freshw. Biol. 47, 867–887. doi:10.1046/j.1365-2427.2002.00914.x

Kelly, M., Juggins, S., Guthrie, R., Pritchard, S., Jamieson, J., Rippey, B., Hirst, H., Yallop, M., 2008. Assessment of ecological status in U.K. rivers using diatoms. Freshw. Biol. 403–422. doi:10.1111/j.1365-2427.2007.01903.x

Kumar, P., 2015. Hydrocomplexity: Addressing water security and emergent environmental risks. Water Resour. Res. 51, 5827–5838. doi:10.1002/2015WR017342

Lane, S.N., 2014. Acting, predicting and intervening in a socio-hydrological world. Hydrol. Earth Syst. Sci. 18, 927–952. doi:10.5194/hess-18-927-2014

Lau, S.S.S., Lane, S.N., 2002. Nutrient and grazing factors in relation to phytoplankton level in a eutrophic shallow lake: the effect of low macrophyte abundance. Water Res. 36, 3593–601.

Lücke, J.D., Johnson, R.K., 2009. Detection of ecological change in stream macroinvertebrate assemblages using single metric, multimetric or multivariate approaches. Ecol. Indic. 9, 659–669. doi:10.1016/j.ecolind.2008.08.005

Lumb, A., Sharma, T.C., Bibeault, J.-F., 2011. A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. Water Qual. Expo. Heal. 3, 11–24. doi:10.1007/s12403-011-0040-0

Mao, F., Clark, J., Karpouzoglou, T., Dewulf, A., Buytaert, W., Hannah, D., 2017. HESS Opinions: A conceptual framework for assessing socio-hydrological resilience under change. Hydrol. Earth Syst. Sci. 21, 3655–3670. doi:10.5194/hess-21-3655-2017

Mao, F., Richards, K., 2012. Irreversible river water quality and the concept of the reference condition. Area 44, 423–431. doi:10.1111/j.1475-4762.2012.01124.x

Mao, F., Richards, K.S., Toland, M., Shi, Y., Hannah, D.M., Krause, S., 2019. Computers and Geosciences Rivervis : A tool for visualising river ecosystems 123, 59–64. doi:10.1016/j.cageo.2018.11.007

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., Mcbride, M., Mengersen, K., 2012. Eliciting Expert Knowledge in Conservation Science. Conserv. Biol. 26, 29–38. doi:10.1111/j.1523-1739.2011.01806.x

McCluney, K.E., Poff, N.L., Palmer, M.A., Thorp, J.H., Poole, G.C., Williams, B.S., Williams, M.R., Baron, J.S., 2014. Riverine macrosystems ecology: Sensitivity, resistance, and resilience of whole river basins with human alterations. Front. Ecol. Environ. 12, 48–58. doi:10.1890/120367

Mizobuchi, H., 2014. Measuring World Better Life Frontier: A Composite Indicator for OECD Better Life Index. Soc. Indic. Res. 118, 987–1007. doi:10.1007/s11205-013-0457-x

Molinos-Senante, M., Go´mez, T., Garrido-Baserba, M., Caballero, R., Sala-Garrido, R., 2014. Assessing the sustainability of small wastewater treatment systems: A composite indicator approach. Sci. Total Environ. 497–498, 607–617. doi:10.1016/j.scitotenv.2014.08.026

Monk, W.A., Wood, P.J., Hannah, D.M., Wilson, D.A., 2008. Macroinvertebrate community response to inter-annual and regional river flow regime dynamics. River Res. Appl. 24, 988–1001. doi:10.1002/rra.1120

Montgomery, D.R., 1999. Process domains and the river continuum. J. Am. Water Resour. Assoc. 35, 397–410. doi:doi.org/10.1111/j.1752-1688.1999.tb03598.x

Munda, G., 2005. "Measuring Sustainability": A Multi-Criterion Framework. Environ. Dev. Sustain. 7, 117–134. doi:10.1007/s10668-003-4713-0

NIEA, 2014. River Hydromorphology Assessment Technique (RHAT) training manual. Lisburn.

Norris, V., 1993. The use of buffer zones to protect water quality: a review. Water Resour. Manag. 257–272.

O'Connor, N.A., Lloyd, L.N., Moore, S.J., 1996. Evaluation of the National River Health Program (Report for Department of Environment, Sports & Territories and Land & Water Resources Corporation), Lloydia (Cincinnati).

OECD, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. Methodology 3, 162. doi:10.1787/9789264043466-en

Ostadtaghizadeh, A., Ardalan, A., Paton, D., Jabbari, H., Khankeh, H.R., 2015. Community Disaster Resilience: a Systematic Review on Assessment Models and Tools. PLoS Curr. Disasters 1–14. doi:10.1371/currents.dis.f224ef8efbdfcf1d508dd0de4d8210ed

RBMP, 2012. EU-China River Basin Management Programme [WWW Document]. URL http://eeas.europa.eu/archives/delegations/china/documents/eu_china/development_cooperation/rbmp_6083_20120627.pdf (accessed 3.13.17).

RC Team, 2013. R: A language and environment for statistical computing. R Found. Stat. Comput.

RiBaGo, 2012. River Basin Governance Research and Network (RiBaGo) [WWW Document]. URL http://www.geog.cam.ac.uk/research/projects/ribago/ (accessed 3.13.17).

Rickels, W., Dovern, J., Hoffmann, J., Quaas, M.F., Schmidt, J.O., Visbeck, M., 2016. Indicators for monitoring sustainable development goals: An application to oceanic development in the European Union. Earth's Futur. 4, 252–267. doi:10.1002/2016EF000353

Rogge, N., 2012. Undesirable specialization in the construction of composite policy indicators: The environmental performance index. Ecol. Indic. 23, 143–154. doi:10.1016/j.ecolind.2012.03.020

Saaty, T., 1996. Decision making with dependence and feedback: The analytic network process. RWS.

Saaty, T., 1980. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw, New York.

Saaty, T., William, A., 2000. Super Decisions software.

Saaty, T.L., 2008. The Analytic Hierarchy and Analytic Network Measurement Processes: Applications to Decisions under Risk. Eur. J. Pure Appl. Math. 1.

Sahin, O., Mohamed, S., 2013. A spatial temporal decision framework for adaptation to sea level rise. Environ. Model. Softw. 46, 129–141. doi:DOI 10.1016/j.envsoft.2013.03.004

Sandin, L., Hering, D., 2004. Comparing macroinvertebrate indices to detect organic pollution across Europe: a contribution to the EC Water Framework Directive intercalibration. Hydrobiologia 516, 55–68. doi:10.1023/B:HYDR.0000025258.63416.11

Schipper, E.L.F., Langston, L., 2015. A comparative overview of resilience measurement frameworks analysing indicators and approaches. Overseas Dev. Inst. - Work. Pap. 422 30. doi:10.13140/RG.2.1.2430.0882

Schmutz, S., Kaufmann, M., Vogel, B., Jungwirth, M., Muhar, S., 2000. A multi-level concept for fish-based, river-type-specific assessment of ecological integrity. Hydrobiologia 422/423, 279–289. doi:10.1023/A:1017038820390

SEPA, 2002. Environmental Quality Standards for Surface Water.

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, a. K., 2012. An overview of sustainability assessment methodologies. Ecol. Indic. 15, 281–299. doi:10.1016/j.ecolind.2011.01.007

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, a. K., 2009. An overview of sustainability assessment methodologies. Ecol. Indic. 9, 189–212. doi:10.1016/j.ecolind.2008.05.011

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, a. K., 2007. Development of composite sustainability performance index for steel industry. Ecol. Indic. 7, 565–588. doi:10.1016/j.ecolind.2006.06.004

Sipahi, S., Timor, M., 2010. The analytic hierarchy process and analytic network process: an overview of applications. Manag. Decis. 48, 775–808. doi:10.1108/00251741011043920

Sivapalan, M., Blöschl, G., 2015. Time scale interactions and the coevolution of humans and water. Water Resour. Res. 51, 6988–7022. doi:10.1002/2015WR017896

Sliva, L., Williams, D.D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res. 35, 3462–72.

Sutherland, W.J., Burgman, M., 2015. Policy advice: Use experts wisely. Nature 526, 317–318. doi:10.1038/526317a

Tong, S.T.Y., Chen, W., 2002. Modeling the relationship between land use and surface water quality. J. Environ. Manage. 66, 377–393. doi:10.1006/jema.2002.0593

Toosi, S.L.R., Samani, J.M. V, 2012. Evaluating Water Transfer Projects Using Analytic Network Process (ANP). Water Resour. Manag. 26, 1999–2014. doi:10.1007/s11269-012-9995-2

UKTAG, 2009. UK Technical Advisory Group on the Water Framework Directive: Recommendations on Surface Water Classification Schemes for the purposes of the Water Framework Directive, Framework.

UKTAG, 2008a. UKTAG rivers assessment methods: macrophytes and phytobenthos: phytobenthos - diatom assessment for river ecological status (DARES) 1–19.

UKTAG, 2008b. UKTAG river assessment methods benthic invertebrate fauna: River invertebrate classification tool (RICT). Edinburgh.

Vacik, H., Wolfslehner, B., Seidl, R., Lexer, M.J., 2007. Integrating the DPSIR approach and the analytic network process for the assessment of forest management strategies, in: Reynolds, K.M., Thomson, A.J., Köhl, M., Shannon, M.A., Ray, D., Rennolls, K. (Eds.), Sustainable Forestry: From Monitoring and Modelling to Knowledge Management and Policy Science. CABI.

Vannote, R., Minshall, G.W., Cummins, K.W., Sedell, J.R., Cushing, C.E., 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37, 130–137.

Wasson, J.-G., Villeneuve, B., Iital, A., Murray-Bligh, J., Dobiasova, M., Bacikova, S., Timm, H., Pella, H., Mengin, N., Chandesris, A., 2010. Large-scale relationships between basin and riparian land cover and the ecological status of European rivers. Freshw. Biol. 55, 1465–1482. doi:10.1111/j.1365-2427.2010.02443.x

Webb, B.W., Hannah, D.M., Moore, R.D., Brown, L.E., Nobilis, F., 2008. Recent advances in stream and river temperature research. Hydrol. Process. 22, 902–918. doi:10.1002/hyp.6994

Wiegand, J., Raffaelli, D., Smart, J.C.R., White, P.C.L., 2010. Assessment of temporal trends in ecosystem health using an holistic indicator. J. Environ. Manage. 91, 1446–55. doi:10.1016/j.jenvman.2010.02.004

Wijnmalen, D.J.D., 2007. Analysis of benefits, opportunities, costs, and risks (BOCR) with the AHP–ANP: A critical validation. Math. Comput. Model. 46, 892–905. doi:10.1016/j.mcm.2007.03.020

Wolfslehner, B., Vacik, H., 2008. Evaluating sustainable forest management strategies with the Analytic Network Process in a Pressure-State-Response framework. J. Environ. Manage. 88, 1–10. doi:10.1016/j.jenvman.2007.01.027

Wolfslehner, B., Vacik, H., Lexer, M.J., 2005. Application of the analytic network process in multi-criteria analysis of sustainable forest management. For. Ecol. Manage. 207, 157–170. doi:10.1016/j.foreco.2004.10.025

Zhou, F., 2010. The study about the contents of heavy metals and their influencing factors in the middle of Chishui river. Guizhou University.