Ecological Indicators 81 (2017) 491–502

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Review

When to use what: Methods for weighting and aggregating sustainability indicators

Xiaoyu Gan[a], Ignacio C. Fernandez[b], Jie Guo[c,*], Maxwell Wilson[d], Yuanyuan Zhao[e], Bingbing Zhou[b], Jianguo Wu[b,d,f]

[a] College of Architecture and Environment, Sichuan University, Chengdu, Sichuan Province, China
[b] School of Sustainability, Arizona State University, Tempe, AZ, USA
[c] College of Land Management, Nanjing Agricultural University, Nanjing, Jiangsu Province, China
[d] School of Life Sciences, Arizona State University, Tempe, AZ, USA
[e] School of Soil and Water Conservation, Beijing Forestry University, Beijing, China
[f] Center for Human-Environment System Sustainability (CHESS), State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Keywords:
Aggregation
Indicator framework
Sustainability indicators
Sustainability indices
Weighting

ABSTRACT

Context: Sustainability indices (SIs) have become increasingly important to sustainability research and practice. However, while the validity of SIs is heavily dependent on how their components are weighted and aggregated, the typology and applicability of the existing weighting and aggregation methods remain poorly understood.
Objectives: To close the knowledge gap regarding when to use which weighting and aggregation methods for constructing SIs, we review the most commonly used methods for weighting and aggregating SIs, discuss their benefits and drawbacks, and suggest a process-oriented approach for choosing appropriate weighting and aggregation methods depending on research objectives.
Methods: Our review synthesis was based on peer-reviewed journal articles, books, and reports by international organizations, governmental agencies, and research institutions. After carefully examining their principles, characteristics, and applications, we selected and classified the frequently used methods for indicator weighting and aggregation.
Results: We systematically discuss the benefits and drawbacks of nine weighting methods and three aggregation methods. We propose a four-step process for choosing the most suitable weighting and aggregation methods based on: research purposes, spatial and temporal scales, and sustainability perspectives.
Conclusions: In this research, we chose the most commonly used methods for weighting and aggregating SIs and analyzed the characteristics, strengths and weaknesses of each method. We found that choosing appropriate weighting and aggregation methods for a specific sustainability assessment project is an extremely important and challenging task. To meet this challenge, we propose a process-oriented approach for properly selecting methods according to the purpose, scale and sustainability concept. This approach can facilitate the proper selection of these methods in sustainability research and practice.

1. Introduction

Sustainability is the challenge of our time (Sachs, 2015). By seeking to achieve dynamic and simultaneous harmony among ecological subsystems (environmental sustainability), social subsystems (social sustainability), and economic subsystems (economic sustainability), sustainability is inherently complex, multi-dimensional, and embedded with trade-offs among multiple sustainability dimensions (Wu, 2013). However, as the public's desire for more sustainability grows stronger (Kates and Clark, 1999; Kates et al., 2001), so does the need to accurately assess the sustainability of our societies (Böhringer and Jochem, 2007), which is no easy task. To capture the complexity of sustainability, sustainability assessments often require the integration of multiple indicators to form composite indices (Özdemir et al., 2011; Wu and Wu, 2012). Thus, while developing sustainability indicators and indices (SIs) is a critical tool for assessing and ultimately attaining sustainability, the specifics of SI formulation can radically impact the measured sustainability of a system (Singh et al., 2009; Wilson and Wu, 2017).

The main procedures for building a sustainability index include

* Corresponding author at: College of Land Management, Nanjing Agricultural University, #1 Weigang Road, Nanjing 210095, Jiangsu Province, China.
E-mail address: guojie@njau.edu.cn (J. Guo).

http://dx.doi.org/10.1016/j.ecolind.2017.05.068
Received 8 February 2017; Received in revised form 9 May 2017; Accepted 29 May 2017

selecting appropriate sustainability indicators, weighting the selected indicators, and aggregating those indicators into a composite index (Meadows, 1998; Juwana et al., 2012). Disagreements on indicator selection are relatively easy to decipher, as existing guidelines, e.g., Bellagio Principles (Hardi and Zdan, 1997), or indicator frameworks, such as the Pressure-State-Response framework (OECD, 1993), can provide guidance for indicator selection (Wu and Wu, 2012). However, because the process of indicator integration is an inherently subjective procedure (Morse et al., 2001), selecting appropriate weighting and aggregation methods is challenging (Saisana and Tarantola, 2002; Wilson and Wu, 2017).

The weighting and aggregation of index components are critically important steps in any sustainability assessment. Weights of SIs reflect the relative importance of different dimensions in their contributions to the sustainability performance of a system, while aggregation essentially reflects the substitutability of different dimensions. The weighting and aggregation methods utilized in SI formulation define whether dimensions can compensate or substitute for each other. Whether complete, partial, or no substitution between environmental (or natural) and socioeconomic capital is legitimate underscores the two widely discussed sustainability perspectives: weak sustainability and strong sustainability (Daly et al., 1995; Markulev and Long, 2013). Weak sustainability allows for unlimited substitution between sustainability dimensions. Strong sustainability is a paradigm that views economic activities as part of the social domain, and both economic and social actions are constrained by the environment (Wu, 2013). Each perspective dictates a different set of criteria for indicator selection and fundamentally influences the final verdict of a sustainability assessment (Wu, 2013; Huang et al., 2015). Further, the weights of SIs not only reflect the relative importance of different dimensions in their contributions to overall sustainability but also symbolize the trade-off ratios among the dimensions if they are conceived as substitutable. Thus, it should come as no surprise that the inappropriate selection of weighting or aggregating methods can cause SIs to provide misleading information (Böhringer and Jochem, 2007). In this sense, one of the main challenges in developing and applying SIs is to know “when to use what.”

Informative reviews have been published on the strengths and weaknesses of commonly used sustainability indices (Böhringer and Jochem, 2007; Mayer, 2008; Singh et al., 2009; Mori and Christodoulou, 2012). These studies provided suggestions on how to choose appropriate sustainability indicators and indices for policy decisions and discussions of those SIs’ formulation and performance. Researchers have also proposed guidelines for constructing sustainability indices in various contexts, such as urban sustainability (Huang et al., 2015), industry sustainability (Singh et al., 2007), energy sustainability (Wang et al., 2009), and agricultural sustainability (Gómez-Limón and Sanchez-Fernandez, 2010). The main goal of the present study, therefore, is to provide a practical guide for the selection of weighting and aggregation methods in the formulation and application of SIs. We focus on three main questions: (1) What are the most commonly used methods for weighting and aggregation reported in the literature? (2) What are the strengths and weaknesses of these methods for measuring sustainability? (3) How can these methods be best utilized in SI development and applications? To address these questions, we systematically reviewed the main methods for weighting and aggregating SIs, identified the main advantages and challenges for applying these methods, and proposed a process-oriented approach to help researchers and practitioners select the most suitable weighting and aggregation methods for sustainability assessment using SIs in different contexts.

2. Methodology

2.1. Analytical framework

To ensure this review is as representative as possible, an analytical framework in which weighting and aggregation methods used for constructing SIs was needed. In our paper, the classification strategy of weighting and aggregation methods proposed by Nardo et al. (2005) and OECD (2008) was adopted.

Within this framework, methods for weighting indicators can be broadly categorized into three main groups: (1) equal weighting, (2) statistic-based weighting, and (3) public/expert opinion-based weighting. Equal weighting means that all the indicators are given the same weight. Statistic-based weighting derives weights from the statistical characteristics of the data (OECD, 2008). Unlike equal weighting and statistic-based weighting, public/expert opinion-based weighting relies on inputs from the participating public or experts, whose judgments ultimately determine the weights to be assigned to individual indicators (OECD, 2008). Thus, weights determined by public/expert opinion reflect the value judgments of the participants regarding different aspects of sustainability (e.g., relative importance, relative urgency, or substitution rates).

In contrast, aggregation methods integrate weighted components (e.g., indicators) into a single composite index. Different classification schemes for aggregation methods exist. In general, classification schemes include those based on the semantics of aggregation (Beliakov et al., 2007; Grabisch, 2009) and those based on the degree of permission of compensation (OECD, 2008). We adopt the latter classification scheme because it has a closer relation to the technical challenges of integrating weighted indicators based on sustainability concepts (Wilson and Wu, 2017). Widely used aggregation methods based on this classification scheme include additive aggregation methods (e.g., arithmetic), multiplicative aggregation methods (e.g., geometric) and non-compensatory aggregation methods (e.g., multi-criteria analysis).

2.2. Literature analysis

To evaluate what are the most commonly used methods for weighting and aggregation reported in the literature, we did a statistical analysis of published literature separate from the review discussed in the main text. To select the papers for this analysis, we followed the PRISMA flowchart (Liberati et al., 2009), shown in Fig. 1. We searched papers using the Web of Knowledge database by using the search by topic option with the search terms shown in Table 1. This search was done on April, 14th, 2017, resulting in 1319 publications, after removing duplicates. We added to these publications 98 documents that were considered relevant, but were not available on the Web of Knowledge database. We selected these publications based on the references from Nardo et al. (2005), Böhringer and Jochem (2007), Singh et al. (2009), and Huang et al. (2015). Titles and abstracts of these papers (n = 1417) were then screened to remove: (1) papers that were cited less than 30 times, (2)literature that was unrelated to sustainability assessments, and (3) papers on indicator sets instead of composite indicators. The remaining 230 articles were then assessed to remove articles presenting indices that did not provide specific weighting and aggregating methods or that provided duplicate indices without any modifications in the methods used for weighting or aggregation. A total of 90 SIs, including 96 weighting scheme variations and 90 aggregation scheme variations, were identified. As some SIs use different weighting/aggregation methods to integrate sub-indicators into the final composite indices, we counted each weighting/aggregation scheme as a separate index, and thus a total of 96 different SIs were used for the analysis.

2.3. Literature analysis results

Among the 96 SIs reviewed in our paper, 46.88% adopted equal weighting methods, 21.88% adopted statistical-based methods (principal component analysis, benefit of the doubt approach, regression analysis, unobserved component models), and 23.95% adopted

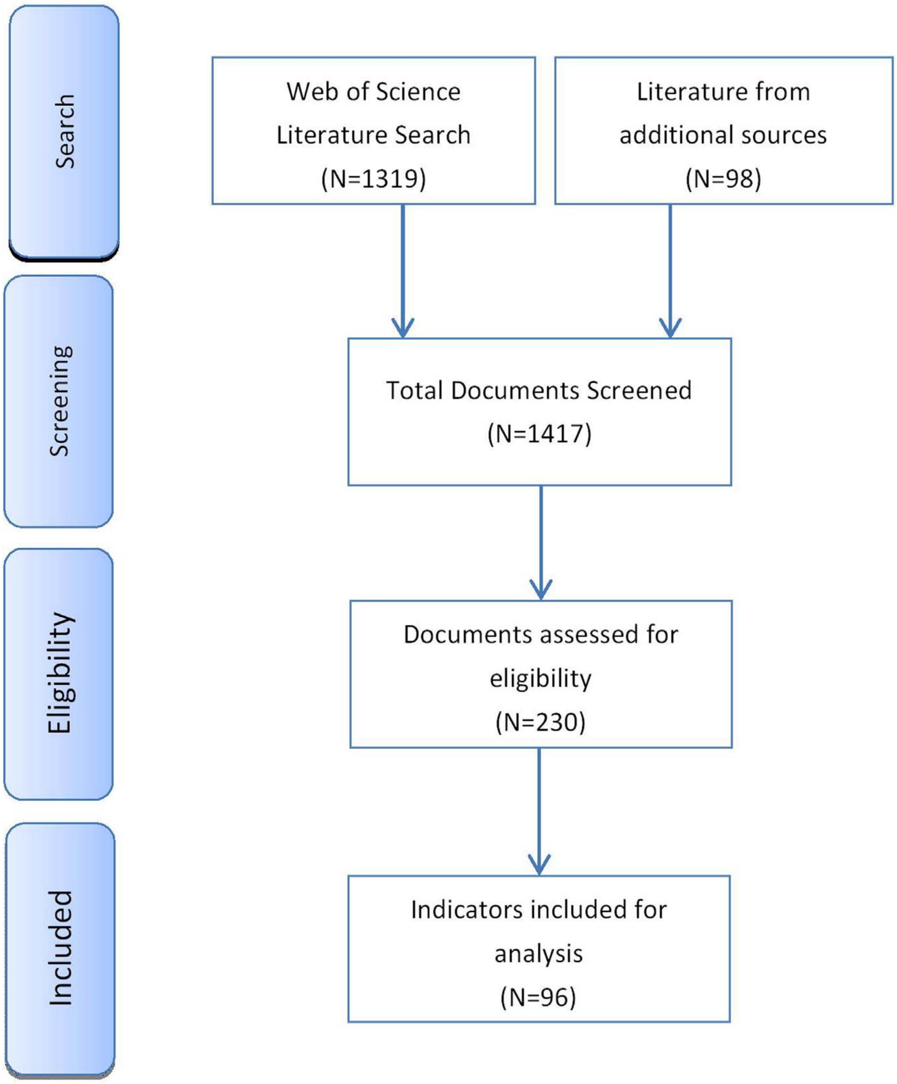

Fig. 1. Preferred reporting items for systematic reviews and meta-analysis flowchart (Liberatiet al., 2009).

Table 1
Keyword combinations used for the database searches.

Search no.	Search term
1	Sustainability indicators AND weighting
2	Sustainability indicators AND aggregation
3	Sustainability indicators AND weighting AND aggregation
4	Sustainability indices AND weighting
5	Sustainability indices AND aggregation
6	Sustainability indices AND weighting AND aggregation

participatory-based methods (budget allocation, public opinion, analytic hierarchy process, conjoint analysis) (Fig. 2). Regarding the aggregation methods, most SIs (86.46%) use an additive strategy (Fig. 2). In Sections 3 and 4 we review and discuss the 9 weighting and 3 aggregation methods analyzed in this section.

3. Commonly used methods for weighting

3.1. Equal weighting

Equal weighting can be used when all the indicators are considered equally important or when no statistical or empirical evidence supports a different scheme (Nardo et al., 2005). It is also recognized as the simplest strategy and can be replicated easily by others (Land, 2006). Several sustainability indices, such as the Living Planet Index (Loh et al., 1998; Loh et al., 2005), Human Development Index (UNDP, 1990), and Genuine Saving Index (WorldBank, 1999), have been built using an equal weighting strategy. Although simple and straightforward, the use of equal weighting has caused controversies, most of which focus on the validity and transparency of indices using this procedure (Table 2) (McClelland 1978; Gordon, 1995; Finnveden, 1999; Geniaux et al., 2009; Rowley et al., 2012; Mikulić et al., 2015).

3.2. Principal components analysis or factor analysis

Both principal components analysis (PCA) and factor analysis (FA) aim to reduce the dimensionality of the data without significant information loss using linear transformation techniques (Dunteman, 1989; DeCoster, 1998). Utilizing the correlation structure of the original data set, PCA extracts orthorhombic, or perpendicular, factors that highly correlated indicators are likely to share. Factors that account for the largest proportion of the variance are retained, while less informative factors are ignored (Smith, 2002). These retained factors are rotated so that each original indicator is loaded solely in one of the new principal factors. To be utilized as an SI, the new factors, which may include information from one or more indicators, are defined as different dimensions (or components) of sustainability. The weights of each dimension can then be calculated from the factor loadings, as these factor loadings express the ratio of the overall variance of the

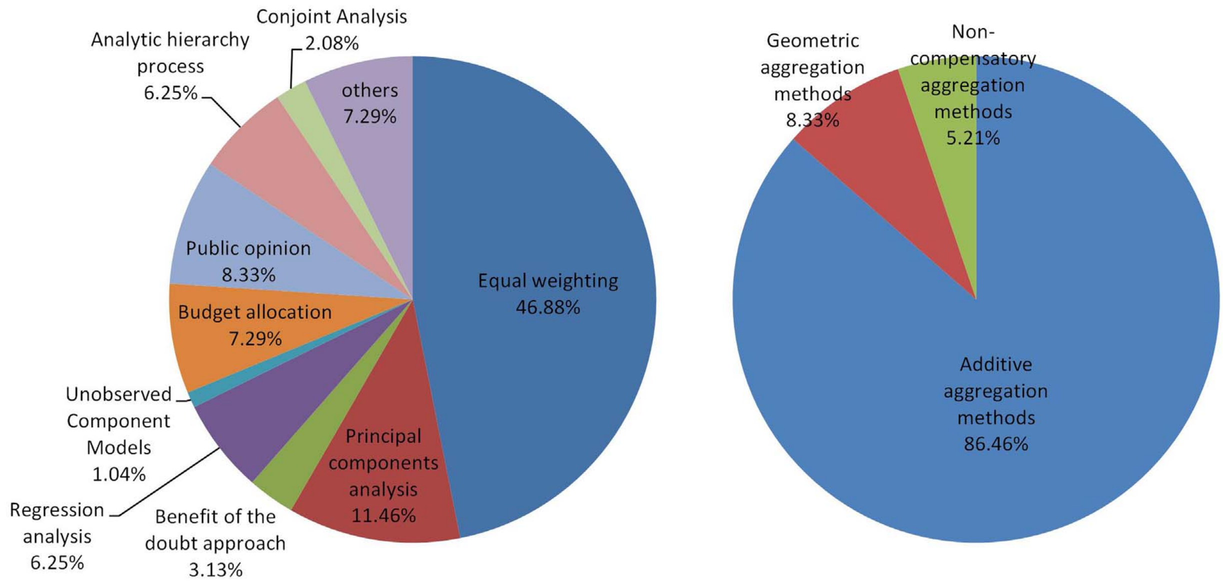

Fig. 2. Proportions of methods used for indicator weighting (left) and aggregation (right).

indicator explained by the factors (OECD, 2008; Riedler et al., 2015; Li et al., 2016; Vitunskiene and Dabkiene, 2016) (Table 2).

Because PCA/FA reduces the dimensionality of the original indicator set, these methods are most valuable when a large number of indicators need to be considered. Techniques based on PCA/FA can reduce the risk of double weighting, which may occur in the equal weighting method (Yeheyis et al., 2013). However, PCA/FA was originally developed to examine the relationships between variables or indicators, not determine weights (Hermans et al., 2008). Hence, two main drawbacks are evident when weighting SIs using PCA/FA. First, the real meaning of the dimensions extracted using these methods may be difficult to define, especially when seemingly unrelated indicators are grouped into the same dimension because of spurious correlations. Further, the optimal number of dimensions may fluctuate if different methods for principal components extraction are used (OECD, 2008), and unpredictable outcomes may also occur when weighting dimensions because weights are based on correlations instead of real-world links among assessed indicators (Hermans et al., 2008). For example, PCA/FA may assign lower weights to a crucial dimension simply because it is weakly correlated with other dimensions. As weights do not correspond to the relative importance of dimensions in the real world, weights derived through PCA/FA can be invalid when used for a sustainability assessment (Mikulić et al., 2015). The second disadvantage of this method is that it only works if a sufficient number of indicators are used and if they present a certain degree of correlation (OECD, 2008).

3.3. Benefit of the doubt approach

The benefit of the doubt approach (BOD) is an application of data envelopment analysis (DEA) (Nardo et al., 2005). BOD has the following key characteristics (see its mathematical formulation in Table 2): (1) The weights of indicators are revealed by the relative performance of a set of indicators, which means that indicators with more beneficial impacts on the unit's overall estimation of sustainability are assigned higher weights, and *vice versa*. Thus, the weights of indicators are retrieved endogenously and specifically for each unit to be assessed (e.g., a country, region, or locality) to optimize the composite measure using weighting summation (Shwartz et al., 2009). (2) Indices constructed by BOD are ratio based, where the composite measure of a particular unit is compared with a benchmark value, which is acquired endogenously through optimization. (3) The benchmark is based on the maximum weighted sum of a unit employing the same weight as the one to be assessed (Cherchye et al., 2007). Thus, if no unit has the highest score in all sub-indicators, the benchmark will be unit dependent, which means that no unique benchmark will exist for all the units under analysis.

An advantage of this method is that it can effectively integrate the processes of weighting, aggregation, and index formation. In addition, weights obtained using this method may help define trade-offs, mitigating some of the difficulties arising from linear aggregation methods (Nardo et al., 2005). These weights are selected to maximize the index for each unit, and therefore, potential complaints about unfair or harmful weighting schemes can be minimized (Cherchye et al., 2007). However, multiple potential results, the incomparability among them, and low transparency are the major disadvantages of this method (Nardo et al., 2005; Shwartz et al., 2009) (Table 2).

3.4. Regression analysis

Regression analysis is a multivariable technique, whose purpose is to assess relationships among a set of variables using statistical methods (Kleinbaum et al., 2013). Using regression analysis, weights can be determined by discerning the relationship between a set of indicators and a single output measure (Nardo et al., 2005) (Table 2). This method performs well when there are a large number of independent variables or indicators, and it can be used for updating or validating the applied set of weights (Nardo et al., 2005). However, some requirements of this method hinder its application for weighting indicators. First, multi-collinearity is not acceptable when building multiple linear regression models. This limitation is particularly problematic in sustainability assessments, where multi-collinearity is common (e.g., income is often positively associated with education and health indicators, but all three are independently relevant for measuring sustainability). Second, the basic objective of multiple linear regression analysis is to discern how the predictor variables explain the variation of the response variable. This condition assumes that the indicators selected can properly explain the variation of the response variable, which is not always the case. Therefore, it is extremely important to choose an appropriate dependent variable that can reflect the target and be explained by the indicators. Under these circumstances, there would be no need to calculate another index based on those weights to replace the "dependent variable" used in regression models (Muldur, 2001).

Table 2
Common methods for indicator weighting (primarily based on Nardo et al. (2005), OECD (2008), Hermans et al. (2008), and Mikulić et al. (2015)).

Method Name	Type	Examples	Formulas	Benefits	Drawbacks
Equal weighting	Equal weighting	Human Development Index (UNDP, 1990) Genuine Savings (WorldBank, 1999)	$\omega_i = \omega$, $i = 1, \ldots, m$, where ω_i is the weight of the i[th] indicator and ω a constant used as the weights for all the indicators	Simple, replicable and straightforward.	No insights into indicator relationships; risk of double weighting.
PCA/FA	Statistic-based	Environmental Sustainability Index (Sands and Podmore 2000) The 2006 European e-Business Readiness Index (Pennoni et al., 2006)	$\omega_i = r_j({l_{ij}}^2/E_j)$ $i = 1,\ldots,m$; $j = 1,\ldots,n$ where r_j is the proportion of the explained variance of factor j (or the intermediate composite j) in the data set, l_{ij} the factor loading of the i^{th} indicator on factor j and E_j the variance explained by the factor j	Reduces the risk of double weighting, classifying ungrouped indicators.	Dimensions of sustainability are unpredictable, and weights may differ from reality.
Benefit of the doubt approach (BOD)	Statistic-based	Meta-index of Sustainable Development (Cherchye and Kuosmanen 2004) Macro-economic performance evaluation (Melyn and Moesen 1991)	$\omega_c = \arg\max_{\omega_{c,i}} \frac{\sum_{i=1}^{m} \omega_{c,i} I_{c,i}}{\max_{y_{ij} \in \{studied\ units\}} \sum_{i=1}^{m} \omega_{c,i} I_{j,i}}$ $s.t. \sum_{i=1}^{m} \omega_{c,i} I_{j,i} \leq 1$, $\omega_{c,i} \geq 0$ $\forall\ i = 1, \ldots, m$; $\forall\ j = 1, \ldots, n$ where $\boldsymbol{\omega_c}$ is the weight vector of unit c, $\omega_{c,i}$ the weight of the i^{th} indicator of unit c, $I_{c,i}$ the normalized score of the i^{th} indicator of unit c, and $I_{j,i}$ the normalized score of the i^{th} indicator of the j^{th} unit	The processes of weighting, aggregation, and index construction are efficiently integrated. Weights are selected to maximize the index for each unit.	Results may not be comparable and lack transparency. A multiplicity of solutions exists.
Regression analysis (RA)	Statistic-based	National Innovative Capacity (Porter and Stern 2001)	$\omega_i = \beta_i$, $i = 1,\ldots,m$ where β_i is the regression coefficient of the i^{th} indicator	Results can be used for updating or validating weights.	Either multi-collinearity among indicators or an improper dependent variable may lead to poor results.
Unobserved component models (UCM)	Statistic-based	The aggregate governance indicators (Kaufmann et al., 1999)	$\omega_i = \frac{\delta_i^{-2}}{1 + \sum_{i=1}^{m} \delta_i^{-2}}$ $i = 1\ldots,m$ where δ_i is the variance of the i^{th} indicator	The processes of weighting, aggregation, and index construction are efficiently integrated. Statistical significance can be expressed when conducting comparisons.	Results are sensitive to outliers. Problems of identification may occur if indicators are highly correlated. Reliability and robustness of the model may be lost when adequate data are not available.
Budget allocation (BAL)	Public/Expert opinion-based	The Eco-indicator 99 (Goedkoop and Spriensma, 2001) Overall Health System Attainment (Murray et al., 2000)	–	Transparent and explicit.	Measuring urgency instead of importance; region-specific.
Public opinion (PO)	Public/Expert opinion-based	Concern about environmental problems Index (Parker, 1991)	–	Transparent and participatory.	Measuring concern instead of importance; region-specific.
Analytic hierarchy process (AHP)	Public/Expert opinion-based	Composite sustainability performance index (Singh et al., 2007) Index of Environmental Friendliness (Puolamaa et al., 1996)	$\mathbf{A}\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$ where $\mathbf{A}$ is the comparison matrix, λ the largest eigenvalue of $\mathbf{A}$, and $\boldsymbol{\omega}$ the weight vector as well as the eigenvector corresponding to λ	Has a hierarchical structure that is in line with the structure of sustainability frameworks. Simple and flexible. Providing consistent verification operation. Available for both quantitative and qualitative data.	Requirement of a high number of pairwise comparisons. Inconsistency and cognitive stress may exist if there are too many indicators in each cluster.
Conjoint analysis (CA)	Public/Expert opinion-based	Indicator of quality of life in the city of Istanbul (Ülengin et al., 2001)	$\omega_i = \frac{\partial P(I_1, \ldots, I_m)}{\partial I_i}$ where $P(I_1,\ldots,I_m)$ is the preference function defined by researchers and I_i the i^{th} indicator	Results can be easily used for making sustainability plans. Available for both quantitative and qualitative data.	Requires a large sample of respondents. Has complicated estimation process.

3.5. Unobserved component models

Unobserved component models (UCMs) are statistical tools pioneered in economics that have been used for constructing aggregate governance indicators (Kaufmann et al., 1999). The core assumption of this approach is that sustainability is difficult to observe directly or that an indicator is only an imperfect signal of an unobserved sustainability component. UCMs seek to isolate the informative signal of the unobserved sustainability component common to each indicator and develop the best possible index performance through an optimal combination of the available data. To be more concrete, the values of indicators generated by the common unobserved sustainability component are expressed as a linear function of the unobserved component plus a random error term. Based on a number of important assumptions regarding the error terms and unobserved component, and by estimating parameters of the linear functions, the unobserved sustainability component of the unit can be finally estimated. In addition, the weights of indicators can be retrieved using a series of decreasing functions of variances of indicators (Table 2) (Thomas, 2010; Kaufmann et al., 2011). This relation implies that when the precision of an indicator is lower, the weight of that indicator will also be lower.

The UCM method, like the BOD method, is an approach that combines the processes of weighting, aggregation, and index construction. The most interesting characteristic of UCMs is that they can provide interval estimates of a sustainability index instead of a specific value based on observed indicators. Therefore, statistical significance can be expressed when two different systems, or a single system in time, are compared to each other (Kaufmann et al., 2011). However, because outliers in the indicator set may lead to low weights for this indicator due to the decreasing function of variance of indicators, weights based on this method are sensitive to outliers. This method also requires enough data to keep the model reliable and robust. Lastly, when using this method, indicators cannot be highly correlated, and this method may perform poorly due to identification problems (Nardo et al., 2005).

3.6. Budget allocation

Budget allocation (BAL), or expert opinion (Saisana and Tarantola, 2002), is a participatory method wherein experts representing extensive knowledge and experience are joined together to distribute a budget of “n” points over a number of indicators (Nardo et al., 2005; OECD, 2008). Based on the experts’ experience, indicators judged to be more important are given a larger proportion of the budget. Indicator weights are then calculated according to the distribution of the points.

BAL has advantages of transparency and explicitness. However, the meanings of weights based on this method may be obscure or misleading, as weighting may measure the urgency or need for political intervention instead of importance (Nardo et al., 2005). As a result, the weights may not be transferable from one region to another because they may reflect specific local conditions. In addition, weighting a large number of indicators in a short period of time may lead to inconsistencies arising from the cognitive stress of experts (Saisana and Tarantola, 2002) (Table 2).

3.7. Public opinion

Weights based on public opinion can be obtained by public opinion polling, which is easy and inexpensive (Parker, 1991). Stakeholders can express their preferences on a public agenda in terms of “concern”(Van Haaster et al., 2017). Indicators receiving high concern are allocated relatively high weights and *vice versa.* This process is useful for multi-criteria decision processes and can make the process participative and transparent (Munda, 2004). However, weights based on this method are defined by measuring public concern rather than importance (Nardo et al., 2005), which may not reflect the real relevance of measured factors for sustainability. For example, people may pay much less attention to biodiversity protection than to air pollution, although the former is as important as the latter for sustainable development (Huang and Wang, 2001). Similar to BAL, this method also lacks transferability among different systems (Table 2).

3.8. Analytic hierarchy process

Analytic hierarchy process (AHP) is a structured technique for multiple criteria decision making based on pairwise comparisons of alternative elements. As implied by its name, the first step is to translate a complex problem into a hierarchical structure consisting of an overall goal (e.g., sustainability assessment), several criteria contributing to this goal (e.g., sustainability dimensions), and a number of attributes (e.g., indicators). The second step requires comparisons in a pairwise fashion of each cluster pertaining to the same level in the hierarchy. The comparisons are performed by experts, who answer two questions: First, which of these two elements is more important? Second, by how much?

A fundamental scale is used in making the comparisons, which consists of verbal judgments ranging from equal to extreme and corresponding numerical judgments often ranging from 1 to 9 (Saaty, 1990, 2008). The comparisons result in a series of reciprocal squared matrices. The third step is to calculate the relative weights of indicators from the comparison matrix using an eigenvector technique (Table 2). Some degree of inconsistency may occur due to careless errors or over stated judgments during the process of comparisons made by experts (Krajnc and Glavič, 2005). However, AHP tolerates these inconsistencies only if the consistency ratio, an index measuring the consistency of a matrix, does not exceed 0.10 (Nardo et al., 2005; Hermans et al., 2008).

AHP has been widely used as a multiple-criteria decision-making tool (Vaidya and Kumar, 2006; Arranz-López et al., 2017) and is a useful method for weighting sustainability indicators. AHP has the characteristics of a hierarchical structure, which is aligned with the structures of most sustainability frameworks and makes the process easy to comprehend for stakeholders (Singh et al., 2007; Hermans et al., 2008). Second, AHP is simple and flexible, which allows it to be combined with other techniques such as mathematical programming, entropy weight and DEA (Ho, 2008; Freeman and Chen, 2015). Third, unlike other participatory methods, AHP provides a consistent verification operation, which can be considered a feedback mechanism for experts or decision-makers to review and revise their judgments (Ho et al., 2010). Lastly, this method can be used with both qualitative and quantitative data (Nardo et al., 2005). Disadvantages of AHP include the high number of pairwise comparisons and the requirement for a parsimonious number of indicators in each analyzed cluster.

3.9. Conjoint analysis

Conjoint analysis (CA) is a statistical technique used for decomposing multivariate data and has been widely applied in marketing to reveal how individuals make trade-offs among different choices (Wind and Green, 2013). Because of its usefulness in participatory modeling and problem structuring, CA has been used for weighting sustainability indicators (Ülengin et al., 2001). The weighting process using CA is based on the concept of “decomposition”. Individuals are first given a set of alternative scenarios, and then each individual responds to the scenarios by sorting them according to his or her overall preferences. Afterwards, these preferences are decomposed based on a form of the preference function (e.g., a weighted sum) to find a set of “part worth” for the assessed attributes or indicators (Green and DeSarbo, 1978). Weights for the indicators can then be calculated if the parameters of the preference function are estimated (Table 2).

CA requires the respondents’ opinions about their overall preference of scenarios, not about each different attribute or indicator. Thus, unlike preference aggregation methods such as AHP, this preference-

Table 3
Common methods for indicator aggregation (based mainly on Munda and Nardo (2005), Beliakov et al. (2007), OECD (2008), and Pollesch and Dale (2015)).

Common methods for aggregation	Examples	Formulas	Benefits	Drawbacks
Additive aggregation	Environmental Performance Index (Esty et al., 2006) Well Being Index (Prescott-Allen 2001)	$SI = \omega_1 I_1 + \omega_2 I_2 + \ldots + \omega_m I_m = \sum_{i=1}^{m} \omega_i I_i$ where SI is the sustainability index, ω_i the weight of the i^{th} indicator, and I_i the normalized score of the i^{th} indicator.	Transparent and simple. Easy to execute sensitivity analysis and uncertainty quantification.	Rigorous prerequisites exist, such as mutually preferentially independence.
Geometric aggregation	Living Planet Index(Loh et al., 1998; Loh et al., 2005)	$SI = I_1^{\omega_1} I_2^{\omega_2} \ldots I_m^{\omega_m} = \prod_{i=1}^{m} I_i^{\omega_i}$ where SI is the sustainability index, ω_i the weight of the i^{th} indicator, and I_i the normalized score of the i^{th} indicator.	Transparent and simple. Can be used for all kinds of ratio-scale variables.	Rigorous prerequisites exist, such as mutually preferentially independence.
Non-compensatory aggregation methods	Index for "Social Multi-Criteria Evaluation"(Munda 2004)	Rank($Unit_i$) $s.\ t.\ \varphi_* = \max \sum e_{jk}$ $i = 1,\ldots,n$ where Rank ($Unit_i$) is the overall ranking of the n researched units, φ_* the corresponding score of the final ranking of the researched units, and e_{jk} the generic element of the outranking matrix.	No ad hoc restrictions.	Computational problems may be caused by the increasing number of units or indicators. Losing information on the intensity of sustainability.

disaggregation method first focuses on what respondents want and then decomposes those preferences (Ülengin et al., 2001). Disadvantages of CA include the requirement of a large sample of respondents and a large set of their preference data, potentially inconsistent results from different classes of respondents, and a relatively complicated estimation process.

4. Commonly used methods for aggregation

4.1. Additive aggregation methods

Additive aggregation methods employ functions that sum up the normalized values of sub-indicators to form a sustainability index. By far the most widespread additive method is the weighted arithmetic mean (Table 3). The continuity characteristic of the weighted arithmetic mean implies that the bound for the sustainability index can be precisely defined if the relative measurement error of a set of indicators is already known. This property can be used for sensitivity analysis and uncertainty quantification, both of which are important elements in sustainability assessment (Pollesch and Dale, 2015).

However, two important features of additive aggregation must be noted. The first is connected to preferential independence. SIs must be mutually preferentially independent when using linear additive aggregation methods. This means that the contributions of all indicators can be added together to yield a total value, implying that no synergy or conflict exists among different indicators, an assumption that seems unrealistic in many situations (Chen and Pu, 2004; Nardo et al., 2005). Second, weights used in additive methods are substitution rates instead of importance coefficients because the intrinsic nature of additive methods implies a compensatory logic. Thus, additive methods should not be used when interactions between indicators are substantial.

4.2. Geometric aggregation methods

Geometric aggregation methods utilize multiplicative instead of additive functions. The most widespread geometric aggregation function is the weighted geometric mean (Table 3).

Unlike additive aggregation methods, which are fully compensatory, geometric mean-based methods only allow compensability between indicators within certain limitations. This requirement exists because of the "geometric-arithmetic means inequality" (Beliakov et al., 2007; Bullen, 2013), which limits the ability of indicators with very low scores to be fully compensated for by indicators with high scores. Simultaneously, significant marginal effects maybe measured using geometric methods when increasing the values of indicators with relatively low absolute values (OECD, 2008).

The limitations of geometric aggregation methods also must be noted. First, geometric aggregation methods are not fully non-compensatory techniques, and thus they allow for trade-offs among indicators, because geometric methods, like additive methods, have the characteristic of being preferentially dependent (Keeney, 1973, 1974; OECD, 2008). Furthermore, with geometric aggregation methods, sensitivity analyses and uncertainty quantifications cannot be analyzed using measurement errors of indicators (Calvo et al., 2002; Beliakov et al., 2007).

4.3. Non-compensatory aggregation methods

As both additive and geometric aggregations imply that compensation among the sub-components of sustainability is acceptable, the use of these methods to aggregate indicators is often contentious, especially when taking the perspective of strong sustainability into account (Rowley et al., 2012; Pollesch and Dale, 2015). When substitution between sub-components is deemed unacceptable, non-compensatory aggregation methods become important. These methods are apparently based on two points of view: the properties of aggregation

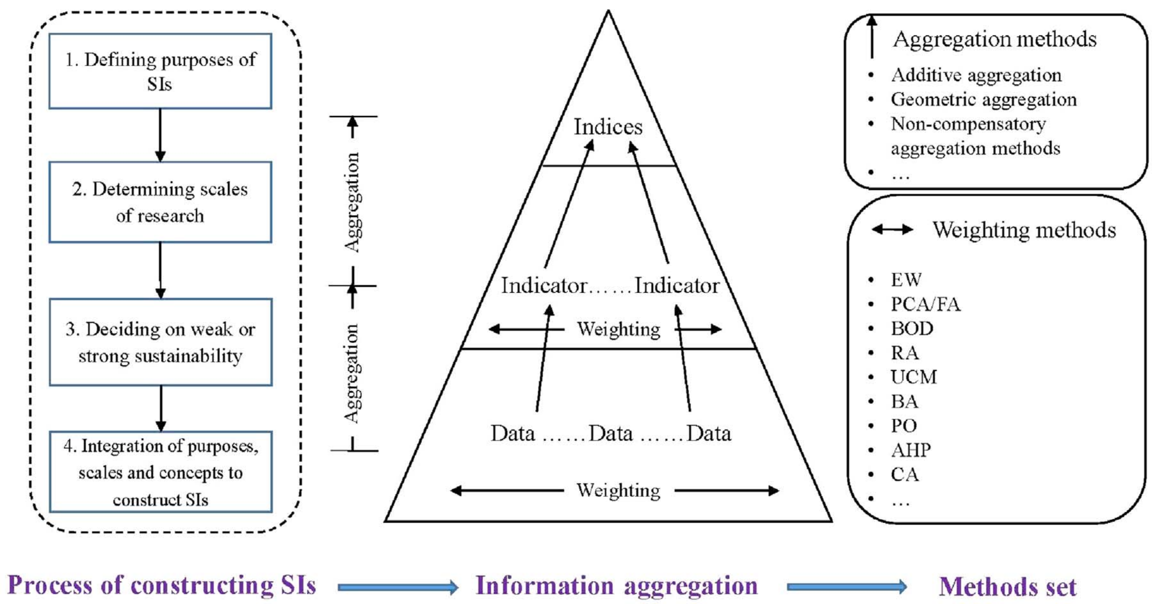

Fig. 3. An illustration of the relationships among the key steps in constructing sustainability indices and selecting weighting and aggregation methods.

functions (Pollesch and Dale, 2015) and the perspective of multi-criteria decision making (MCDM) (Guitouni and Martel, 1998; Munda, 2005).

The literature on aggregation function properties deals with the use of extreme elements as input variables and particularly focuses on conjunctivity and disjunctivity (Beliakov et al., 2007). Neither conjunctive nor disjunctive aggregation functions permit the existence of compensation because the output of the conjunctive function must be bounded above by the min(x) function, whereas the output of the disjunctive function must be bounded below by the max(x) function (Pollesch and Dale, 2015).

For example, SIs based on conjunctive function depend on the lowest value among the indicators. This suggests that sustainability follows the Law of the Minimum, or Liebig's Barrel, and that sustainable development is limited by the dimension with the poorest performance, no matter how well other indicators perform. Therefore, an aggregation function may satisfy strong sustainability if the function is conjunctive (Pollesch and Dale, 2015). Aggregation methods that use conjunctive functions can be easily understood and operated; however, these methods lose the information of indicators whose values are not extreme, and thus, the index may be of limited use for decision making.

In contrast, MCDM is a major class of research that deals with decision making in the presence of a number of decision criteria. The aim of MCDM is to elicit clear subjective preferences (Munda, 2003; Pohekar and Ramachandran, 2004). From this perspective, every sustainability assessment is in fact a multi-criteria problem (Munda, 2005; OECD, 2008). Different classes of MCDM methods exist, almost all of which use decision-maker preferences to reach their conclusions (Guitouni and Martel, 1998). Among these is the non-compensatory multi-criteria approach, which has been used for aggregating sustainability indicators (Nardo et al., 2005).

The non-compensatory multi-criteria approach is based on decision-maker preferences and is centered around the fact that a general objective of most indices is to create rankings (Munda and Nardo, 2005). Therefore, the core of this method is to construct a ranking algorithm that is more consistent than the linear aggregation rule (Munda, 2005). Two procedures are used to calculate the index: (1) units are compared pairwise according to the whole set of sub-indicators to construct a ranking matrix; (2) units are ranked in a complete pre-order according to the ranking matrix (Table 2) (Munda and Nardo, 2005).

The output of this method is a rank rather than a concrete output value for each unit. No compensation is allowed among indicators in the method, and thus, all the weights reflect the relative importance of each indicator instead of a trade-off ratio. Furthermore, there are no restrictions on the type of variables or indicators that can be used, which means that both quantitative and qualitative data can be used. Two possible drawbacks of this method are computational limitations associated with the increasing number of units or indicators and the loss of information on the intensity of sustainability (Munda and Nardo, 2005).

5. A process-oriented approach for answering "when to use what"

Weighting and aggregation are key steps in building SIs (Jollands, 2005; Nardo et al., 2005), as they are the processes by which sustainability information is transferred from variables to sub-indicators and then from sub-indicators to indices (Fig. 3). Although various methods for weighting and aggregating exist (see Sections 2–4), each method has unique strengths and weaknesses (Tables 2 and 3). Despite the critical importance of weighting and aggregation for SI performance, the literature lacks information that specifically guides researchers and practitioners on "when to use what" weighting and aggregation method. If weighting and aggregation methods are not properly selected, the results from SIs will not successfully represent what the particular SI seeks to measure.

To solve the aforementioned question, it is necessary to consider the characteristics of sustainability science. As noted by Kates (2012), sustainability science is place-based and use-inspired fundamental research. Thus, purpose orientation, scale dependence, and concept definition should be fundamental features underlying the construction of SIs (Meadows, 1998). It has been widely accepted that the steps for constructing SIs include defining the policy goals, selecting indicators based on a framework, selecting suitable weighting and aggregation techniques, and checking for robustness and sensitivity (McGranahan, 1971; Booysen, 2002; Jollands, 2005). These steps are not separated and are in fact closely related to each other (Nardo et al., 2005). Therefore, in the process of indicator weighting and aggregation, as in other steps, it is extremely important to take into account research purposes, scales and sustainability concepts (Singh et al., 2009).

In this context, the following questions are critical for the selection of weighting and aggregation methods for building SIs: (1) How should appropriate weighting and aggregation methods be selected according to the explicit purpose of SIs, as the purposes of sustainability are so diverse that no single method can meet all the requirements (Singh

Table 4
A process-oriented approach for answering "when to use what.".

Methods			Purposes		Scales				Sustainability perspectives	
			Assessing State/ Prediction	Comparison	Spatial scales		Temporal scales		Weak sustainability	Strong sustainability
					Coarse	Fine	Long	Short		
Weighting	EW		REC	OK	REC	REC	REC	REC	REC	NO
	PCA		REC	OK	REC	OK	OK	REC	REC	NO
	BOD		REC	NO	REC	OK	OK	REC	REC	NO
	RA		REC	OK	REC	OK	OK	REC	REC	NO
	UCM		REC	OK	REC	OK	OK	REC	REC	NO
	BAL		REC	OK	NO	REC	REC	REC	OK	OK
	PO		REC	OK	NO	REC	REC	REC	OK	OK
	AHP		REC	OK	NO	REC	REC	REC	REC	NO
	CA		REC	OK	NO	REC	REC	REC	REC	NO
Aggregation	Additive aggregation		REC	OK	REC	REC	REC	REC	REC	NO
	Geometric aggregation		REC	OK	REC	REC	REC	REC	REC	NO
	Non-compensatory aggregation methods	Conjunctive/ disjunctive functions	REC	OK	REC	REC	REC	REC	NO	NO
		MCDM	NO	REC	REC	REC	REC	REC	NO	NO
	Combined methods		REC	OK	REC	REC	REC	REC	REC	REC

Note: REC, OK, and NO denote that a method is recommended, applicable, and not to be used in a given situation, respectively, assuming that each method is used alone.

et al., 2009)? (2) How should weighting and aggregation methods that consider differences of spatial and temporal scales be chosen, as proper weighting and aggregation methods may vary from national to local scales between and long- and short-term perspectives (Farrell and Hart, 1998; Mayer, 2008)? (3) How should suitable weighting and aggregation methods be employed in accordance with sustainability concepts (weak or strong), as no single aggregation method can capture these concepts thoroughly (Wilson and Wu, 2017)? (4) How should the purposes, scales and sustainability concepts be systematically integrated to generate SIs that effectively measure what they are intended to measure?

Deciding which weighting and aggregation methods are best fit for particular SIs should be framed as a multi-step decision. Herein, we provide a process-oriented approach that explicitly aims to answer this question by focusing on what particular methods are more suitable for use based on the purposes, scales, and sustainability concepts underlying the structuring of the SIs. Under this approach, the process of selecting the most suitable weighting and aggregation techniques from the commonly used methods set can be completed by following four main steps: (1) defining the purposes of SIs; (2) determining spatial and temporal scales of analysis; (3) deciding on sustainability concepts (weak or strong sustainability); and (4) integrating the purposes, scales and concepts to construct SIs (Fig. 3).

5.1. Confirming the purposes of the SI

Weighting and aggregation methods should be carefully chosen according to what the SI is meant to measure. Generally, the main purposes of SIs include assessing development states in relation to goals and targets; comparisons across time periods, locations or situations; and anticipating future conditions and trends (Saisana and Tarantola, 2002; Singh et al., 2009). Each of these purposes may be best achieved by different combinations of weighting and aggregation methods.

Most weighting and aggregation methods, except for BOD and preference-based methods, result in concrete SI values. Methods providing concrete values possess information integrity, accuracy, and high sensitivity to change, and hence, they can be used for state monitoring and projection (Cairns et al., 1993). However, these SIs do not necessarily measure whether the resulting value is good or bad, especially when no reference or target is available (Warhurst, 2002). Conversely, outputs taking the form of ranks or relative values based on preferences are easy to understand, provide a relative measure on how good or bad the results are, and can promote action in favor of sustainability development (Gordon, 1996). Nevertheless, information intensity may be lost during the processes of preference-based methods (Munda, 2012). Therefore, preference-based methods are good choices for handling comparisons. This does not mean that weighting and aggregation methods that produce concrete values of SIs cannot be used for comparison. However, due to their data-dependent structures and complex formulations, they are less efficient for building indicators based on rankings. Therefore, weighting and aggregation methods resulting in concrete values are best used for state monitoring and projections in time. In contrast, weighting and aggregation methods based on preferences are effective when SIs are used for comparison (Table 4).

5.2. Determining scales of research

Sustainability assessments must be conducted at specific spatial and temporal scales. It is therefore essential to select weighting and aggregation methods that are consistent with the scale at which the assessment will be conducted (Table 4). Weighting methods based on public participation are more powerful at finer scales than at coarser scales not only because of the high cost for organizing stakeholders but also because opinions on controversial issues are spatially auto correlated (Van de Kerk and Manuel, 2008). As a result, weights based on participatory methods at local scales cannot be directly applied at larger scales as they likely reflect local conditions. However, weighting methods based on statistical methods are more effective at coarse spatial scales and in politically bounded systems (Mayer, 2008).

The temporal scale of the sustainability assessment also affects which weighting and aggregation method should be selected. Statistic-based methods rely heavily on data completeness and integrity. However, both data availability and inter-operator variability become more challenging problems as study periods extend. Therefore, statistic-based methods may be more appropriate for assessments covering short timescales, while participatory or equal weighting methods are preferable for long timescales, when data completeness cannot be guaranteed. In addition, public opinions on sustainability may change substantially over long periods of time.

5.3. Deciding on weak or strong sustainability

The type of sustainability an SI measures is directly defined by the weighting and aggregation method it utilizes (Wilson and Wu, 2017). On one hand, weak sustainability allows for substitution, and therefore, compensatory weighting and aggregation methods can be used for assessing sustainability under a weak sustainability paradigm. On the other hand, according to the definition of strong sustainability, some types of social and environmental capital are critical and cannot be substituted for by economic capital. Nevertheless, substitutability among different kinds of capital is still allowed as long as a system exists within the constraints of its environmental and social structures (Wu, 2013). Therefore, indices representing strong sustainability must take into account non-compensability as well as threshold values for each indicator, above which substitutability cannot be allowed (Ekins et al., 2003; Mori and Christodoulou, 2012). This implies that combined aggregation methods should be employed for constructing strong SIs. For example, Díaz-Balteiro and Romero (2004) introduced a function:

$$SI = (1-\lambda)[\min_i(\omega_i I_i)] + \lambda \sum_{i=1}^{m} \omega_i I_i, \tag{1}$$

where $\lambda \in [0,1]$ is a compensation parameter, and ω_i is the weight for indicator I_i. Formula (1) represents a combination of internal and conjunctive functions that is controlled by the parameter λ. When λ equals 1, the function becomes a weighted arithmetic mean, while when λ equals 0, the function can be used to represent a non-compensatory aggregation function. Therefore, the parameter λ determines which type of sustainability is being measured.

In terms of weighting, equally weighted methods cannot be used to measure strong sustainability (Huang et al., 2015; Wilson and Wu, 2017). Similarly, SIs developed through statistic-based methods (e.g., PCA/FA) cannot be considered strong sustainability indices, as the weights based on those methods represent the statistical characteristics of data and not the critical limits to substitution. Participatory weighting methods, in contrast, allow respondents to convey their own preferences on weights, which may or may not conform to the ideals of strong sustainability (Table 4).

Therefore, to conduct strong sustainability assessments, appropriate weighting and aggregation methods should (1) explicitly define the relationships among types of capital, (2) use proper non-equal weights, (3) combine both compensational and non-compensational aggregation functions and (4) consider utilizing adjustment parameters or threshold values to reflect the limit substitutions among dimensions.

5.4. Integration of purposes, scales and concepts to construct SIs

By this point, the SI designer should have defined the purposes of SI, the scales at which the SI will be used, and what type of sustainability the SI seeks to measure. The final step in SI formulation is to integrate research purposes, scales and concepts to construct SIs to assess whether the SI is correctly structured to meet its goals. In other words, it is necessary to confirm the selection of the weighting and aggregation methods in a specific context.

Although we have treated each of the previous steps as separate, they are intrinsically connected. Defining the purpose of an SI is a prerequisite for determining the scale at which it will be measured and applied. Furthermore, in a landscape in which development and protection must occur simultaneously, the scale at which the SI is to be applied is directly connected to the type of sustainability to be measured (Wu, 2013). Therefore, the usefulness of an SI is defined by the weighting and aggregation method utilized at each level.

Let us consider a specific scenario in which a sustainability index will be constructed to assess a natural reserve. The purpose of the index is to evaluate the sustainability status, while the research scale is finer. Given that the major function of a natural reserve is to protect ecosystems, the concept of strong sustainability should be employed for the sustainability assessment. Therefore, according to the process we have proposed, the participatory weighting methods (e.g., PO and BAL) are better choices under this specific situation. In addition, aggregation methods that combine compensatory and non-compensatory methods are more appropriate in this case.

6. Conclusions

Sustainability is defined to measure how far we are from our targets or how well aligned we are to a desired development path. SIs are fundamental for measuring current levels of sustainability, gauging whether implemented strategies are effectively accomplishing their objectives and designing strategies that target sustainability issues that are most relevant to sustainable development. The tenets of sustainability require that the process of SI construction be transparent and that their results are easily communicable and interpretable in order to be embraced by decision-makers and the non-expert community (Munda and Nardo, 2005).

However, measuring sustainability is a challenging task. SIs aim to measure outputs of complex systems and integrate multiple elements related to specific purposes, scales and concepts of sustainability used in their formulation (Saisana and Tarantola, 2002). Weighting and aggregation is an important step. There are various weighting and aggregation methods for compositing SIs. Weighting methods can be categorized into equal weighting, statistical-based methods, and participatory methods. Aggregation methods can be categorized into additive, geometric and non-compensatory methods. Different weighting and aggregation methods symbolize different substitutabilities for different dimensions of SIs.

Considering that each method has its own strengths, weaknesses, and applicable situations, it is important to know "when to use what." Nevertheless, it remains unclear which weighting and aggregation methods are more suitable for building SIs in different situations. This study has sought to close this knowledge gap. Because SIs can be developed to measure different types of sustainability (weak to strong) at different spatial and temporal scales, it is evident that the "one-size-fits-all" approach for weighting and aggregation is inappropriate. Instead, we propose a four-step process for choosing the most suitable weighting and aggregation methods:

(1) Clearly describe the purpose of developing or using SIs;
(2) Determine the particular spatial and temporal scales at which the SIs are to be applied;
(3) Be explicit about the specific type of sustainability that the SIs are used to assess; and
(4) Conduct a comprehensive evaluation of the built SIs based on the previous three factors.

We acknowledge that for building effective SIs, it is necessary to not only select appropriate weighting and aggregation methods but also address several other challenges, including data availability and the particular socio-ecological context in which the SIs will be implemented. Nevertheless, weighting and aggregation methods are critical to the final SI measurement. Therefore, it is our hope that this process-oriented approach will help researchers and practitioners select the appropriate weighting and aggregation methods for their goals and improve our capacity to measure sustainability in the long term.

Acknowledgments

This research was supported and funded by the National Natural Science Foundation of China(Nos. 71303119 and 51108284). We are particularly indebted to the Landscape Ecology and Sustainability Science Laboratory (LESL), Arizona State University, the United States, which provided substantial contributions to discussions of and modifications to the manuscript.

References

Özdemir, E.D., Härdtlein, M., Jenssen, T., Zech, D., Eltrop, L., 2011. A confusion of tongues or the art of aggregating indicators—Reflections on four projective methodologies on sustainability measurement. Renew. Sustain. Energy Rev. 15 (5),

2385–2396.
Ülengin, B., Ülengin, F., Güvenç, Ü., 2001. A multidimensional approach to urban quality of life: the case of Istanbul. Eur. J. Oper. Res. 130 (2), 361–374.
Arranz-López, A., Soria-Lara, J.A., López-Escolano, C., Campos, Á.P., 2017. Retail Mobility EnvironmentsA methodological framework for integrating retail activity and non-motorised accessibility in Zaragoza, Spain. J. Transp. Geogr. 58, 92–103.
Böhringer, C., Jochem, P.E., 2007. Measuring the immeasurable—a survey of sustainability indices. Ecol. Econ. 63 (1), 1–8.
Beliakov, G., Pradera, A., Calvo, T., 2007. Aggregation Functions: A Guide for Practitioners. Springer.
Booysen, F., 2002. An overview and evaluation of composite indices of development. Soc. Indic. Res. 59, 115–151.
Bullen, P.S., 2013. Handbook of Means and Their Inequalities. Springer Science & Business Media.
Cairns Jr., J., McCormick, P.V., Niederlehner, B., 1993. A proposed framework for developing indicators of ecosystem health. Hydrobiologia 263 (1), 1–44.
Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R., 2002. Aggregation operators: properties, classes and construction methods. Aggregation Operators. Springerpp. 3–104.
Chen, L., Pu, P., 2004. Survey of Preference Elicitation Methods. Ecole Politechnique Federale de Lausanne (EPFL), Lausanne, Switzerland.
Cherchye, L., Kuosmanen, T., 2004. Benchmarking Sustainable Development: A Synthetic Meta-index Approach, Research Paper. UNU-WIDER, United Nations University (UNU).
Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007. An introduction to 'benefit of the doubt' composite indicators. Soc. Indic. Res. 82 (1), 111–145.
Díaz-Balteiro, L., Romero, C., 2004. In search of a natural systems sustainability index. Ecol. Econ. 49 (3), 401–405.
Daly, H., Jacobs, M., Skolimowski, H., 1995. Discussion of beckerman's critique of sustainable developemnt. Environ. Values 4 (1), 49–70.
DeCoster, J., 1998. Overview of Factor Analysis. (Accessed 22 December 2015). http://www.stat-help.com/factor.pdf.
Dunteman, G.H., 1989. Principal Components Analysis. Sage.
Ekins, P., Simon, S., Deutsch, L., Folke, C., De Groot, R., 2003. A framework for the practical application of the concepts of critical natural capital and strong sustainability. Ecol. Econ. 44 (2), 165–185.
Esty, D.C., Levy, M.A., Srebotnjak, T., de Sherbinin, A., Kim, C.H., Anderson, B., 2006. Pilot 2006 Environmental Performance Index. Yale Center for Environmental Law & Policy, New Have.
Farrell, A., Hart, M., 1998. What does sustainability really mean?: The search for useful indicators. Environ.: Sci. Policy Sustain. Dev. 40 (9), 4–31.
Finnveden, G., 1999. A Critical Review of Operational Valuation/weighting Methods for Life Cycle Assessment. Swedish Environmental Protection Agency, Stockholm.
Freeman, J., Chen, T., 2015. Green supplier selection using an AHP-Entropy-TOPSIS framework. Supply Chain Manage. 20 (3), 327–340.
Gómez-Limón, J.A., Sanchez-Fernandez, G., 2010. Empirical evaluation of agricultural sustainability using composite indicators. Ecol. Econ. 69 (5), 1062–1075.
Geniaux, G., Bellon, S., Deverre, C., Powell, B., 2009. Sustainable Development Indicator Frameworks and Initiatives. SEAMLESS.
Goedkoop, M., Spriensma, R., 2001. The Eco-indicator99: A Damage Oriented Method for Life Cycle Impact Assessment: Methodology Report.
Gordon, D., 1995. Census based deprivation indices: their weighting and validation. J. Epidemiol. Community Health 49 (Suppl. 2), S39–S44.
Gordon, M., 1996. Problems and fundamentals of sustainable development indicators. Sustain. Dev. 4, 1–11.
Grabisch, M., 2009. Aggregation Functions. Cambridge University Press.
Green, P.E., DeSarbo, W.S., 1978. Additive decomposition of perceptions data via conjoint analysis. J. Consum. Res. 58–65.
Guitouni, A., Martel, J.-M., 1998. Tentative guidelines to help choosing an appropriate MCDA method. Eur. J. Oper. Res. 109 (2), 501–521.
Hardi, P., Zdan, T., 1997. Assessing Sustainable Development: Principles in Practice. International Institute for Sustainable Development, Winnipeg, Manitoba.
Hermans, E., Van den Bossche, F., Wets, G., 2008. Combining road safety information in a performance index. Accid. Anal. Prev. 40 (4), 1337–1344.
Ho, W., Xu, X., Dey, P.K., 2010. Multi-criteria decision making approaches for supplier evaluation and selection: a literature review. Eur. J. Oper. Res. 202 (1), 16–24.
Ho, W., 2008. Integrated analytic hierarchy process and its applications–A literature review. Eur. J. Oper. Res. 186 (1), 211–228.
Huang, F., Wang, Y., 2001. Difficulties and counter-measures for biodiversity conservation. Biodiv. Sci. 9 (4), 399–406.
Huang, L., Wu, J., Yan, L., 2015. Defining and measuring urban sustainability: a review of indicators. Landsc. Ecol. 30, 1175–1193.
Jollands, N., 2005. How to aggregate sustainable development indicators: a proposed framework and its application. Int. J. Agric. Resour. Governance Ecol. 5 (1), 18–34.
Juwana, I., Muttil, N., Perera, B., 2012. Indicator-based water sustainability assessment—A review. Sci. Total Environ. 438, 357–371.
Kates, R., Clark, W., 1999. Our Common Journey: a Transition Toward Sustainability. National Academy Press, Washington Dc.
Kates, R., Clark, W.C., Hall, J.M., Jaeger, C., Lowe, I., McCarthy, J.J., Schellnhuber, H.J., Bolin, B., Dickson, N.M., Faucheux, S., 2001. Sustainability science. Science 641–642.
Kates, R.W., 2012. From the unity of nature to sustainability science: ideas and practice. Sustainability Science: The Emerging Paradigm and the Urban Environment. Springerpp. 3–19.
Kaufmann, D., Kraay, A., Zoido, P., 1999. Aggregating Governance Indicators. World Bank Policy Research Working Paper.
Kaufmann, D., Kraay, A., Mastruzzi, M., 2011. The worldwide governance indicators: methodology and analytical issues. Hague J. Rule Law 3 (02), 220–246.
Keeney, R.L., 1973. A decision analysis with multiple objectives: the Mexico city airport. Bell J. Econ. Manage. Sci. 101–117.
Keeney, R.L., 1974. Multiplicative utility functions. Oper. Res. 22 (1), 22–34.
Kleinbaum, D., Kupper, L., Nizam, A., Rosenberg, E., 2013. Applied Regression Analysis and Other Multivariable Methods. Cengage Learning.
Krajnc, D., Glavič, P., 2005. A model for integrated assessment of sustainable developmentResources. Conserv. Recycl. 43 (2), 189–208.
Land, K., 2006. The Foundation for Child Development Child and Youth Well-Being Index (CWI), 1975–2004, with Projections for 2005. Duke University Durham, NC.
Li, Y., Shi, X., Yao, L., 2016. Evaluating energy security of resource-poor economies:A modified principle component analysis approach. Energy Econ. 58, 211–221.
Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement forreportingsystematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6 (7).
Loh, J., Randers, J., MacGillivray, A., Kapos, V., Jenkins, M., Groombridge, B., 1998. Living Planet Report, 1998. WWF, Gland, Switzerland.
Loh, J., Green, R.E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., Randers, J., 2005. The Living Planet Index: using species population time series to track trends in biodiversity. Phil. Trans. R. Soc. B : Biol. Sci. 360 (1454), 289–295.
Markulev, A., Long, A., 2013. On Sustainability: an Economic Approach. Productivity Commission Staff Research, Canberra.
Mayer, A.L., 2008. Strengths and weaknesses of common sustainability indices for multidimensional systems. Environ. Int. 34 (2), 277–291.
McClelland, G., 1978. Equal Versus Differential Weighting for Multiattribute Decisions: There Are No Free Lunches. DTIC Document.
McGranahan, D., 1971. Development indicators and development models. J. Dev. Stud. 91–102.
Meadows, D.H., 1998. Indicators and Information Systems for Sustainable Development. Sustainability Institute, Hartland Four Corners, Vermont.
Melyn, W., Moesen, W., 1991. Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information Is Available.
Mikulić, J., Kožić, I., Krešić, D., 2015. Weighting indicators of tourism sustainability: A critical note. Ecol. Indic. 48, 312–314.
Mori, K., Christodoulou, A., 2012. Review of sustainability indices and indicators: towards a new City Sustainability Index (CSI). Environ. Impact Assess. Rev. 32 (1), 94–106.
Morse, S., McNamara, N., Acholo, M., Okwoli, B., 2001. Sustainability indicators: the problem of integration. Sust. Dev. 9 (1), 1–15.
Muldur, U., 2001. Technical Annex on Structural Indicators. Two Composite Indicators to Assess the Progress of Member States in Their Transition Towards a Knowledge Based Economy. Directorate-General for Research and Innovation, Brussel.
Munda, G., Nardo, M., 2005. Non-compensatory Composite Indicators for Ranking Countries: A Defensible Setting. EUR Report. EUR (21833).
Munda, G., 2003. Multicriteria Assessment. International Society for Ecological Econimics. Internet Encyclopaedia of Ecological Economicspp. 10.
Munda, G., 2004. Social multi-criteria evaluation: methodological foundations and operational consequences. Eur. J. Oper. Res. 158 (3), 662–677.
Munda, G., 2005. Measuring sustainability: a multi-criterion frameworkEnvironment. Dev. Sustain. 7 (1), 117–134.
Munda, G., 2012. Intensity of preference and related uncertainty in non-compensatory aggregation rules. Theory Decis. 73 (4), 649–669.
Murray, C.J., Lauer, J., Tandon, A., Frenk, J., 2000. Overall Health System Achievement for 191 Countries. Global Programme on Evidence for Health Policy, Discussion Paper Series. World Health Organization, Geneva.
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., 2005. Tools for Composite Indicators Building. European Comission, EUR 21682 EN. Institute for the Protection and Security of the Citizen, Ispra, Italy.
OECD, 1993. Core Set of Indicators for Environmental Performance Reviews. OECD, Paris.
OECD, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD publishing.
Parker, J., 1991. Environmental Reporting and Environmental Indices. PhD Dissertation. University of Cambridge, Cambridge, UK.
Pennoni, F., Tarantola, S., Latvala, A., 2006. The 2005 European e-business Readiness Index. 2006. pp. 1–54.
Pohekar, S., Ramachandran, M., 2004. Application of multi-criteria decision making to sustainable energy planning—a review. Renew. Sustain. Energy Rev. 8 (4), 365–381.
Pollesch, N., Dale, V., 2015. Applications of aggregation theory to sustainability assessment. Ecol. Econ. 114, 117–127.
Porter, M.E., Stern, S., 2001. National Innovative Capacity. The Global Competitiveness Report. 2002. pp. 102–118.
Prescott-Allen, R., 2001. The Wellbeing of Nations. Island Press, Washington DC.
Puolamaa, M., Kaplas, M., Reinikainen, T., 1996. Index of Environmental Friendliness: a Methodological Study, Official Statistics of Finland.
Riedler, B., Pernkopf, L., Strasser, T., Lang, S., Smith, G., 2015. A composite indicator for assessing habitat quality of riparian forests derived from Earth observation data. Int. J. Appl. Earth Observ. Geoinf. 37, 114–123.
Rowley, H.V., Peters, G.M., Lundie, S., Moore, S.J., 2012. Aggregating sustainability indicators: beyond the weighted sum. J. Environ. Manage. 111, 24–33.
Saaty, T.L., 1990. How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. 48 (1), 9–26.
Saaty, T.L., 2008. Decision making with the analytic hierarchy process. Int. J. Serv. Sci. 1, 83–98.
Sachs, J.D., 2015. The Age of Sustainable Development. Columbia University Press, New

York.
Saisana, M., Tarantola, S., 2002. State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development. Citeseer.
Sands, G.R., Podmore, T.H., 2000. A generalized environmental sustainability index for agricultural systemsAgriculture. Ecosys. Environ. 79 (1), 29–41.
Shwartz, M., Burgess, J.F., Berlowitz, D., 2009. Benefit-of-the-doubt approaches for calculating a composite measure of quality. Health Serv. Outcomes Res. Methodol. 9 (4), 234–251.
Singh, R.K., Murty, H., Gupta, S., Dikshit, A., 2007. Development of composite sustainability performance index for steel industry. Ecol. Indic. 7 (3), 565–588.
Singh, R.K., Murty, H., Gupta, S., Dikshit, A., 2009. An overview of sustainability assessment methodologies. Ecol. Indic. 9 (2), 189–212.
Smith, L.I., 2002. A Tutorial on Principal Components Analysis. Cornell University, USA.
Thomas, M.A., 2010. What do the worldwide governance indicators measure & quest. Eur. J. Dev. Res. 22 (1), 31–54.
UNDP, 1990. Human Development Report 1990. Oxford University Press, New York.
Vaidya, O.S., Kumar, S., 2006. Analytic hierarchy process: an overview of applications. Eur. J. Oper. Res. 169 (1), 1–29.
Van Haaster, B., Ciroth, A., Fontes, J., Wood, R., Ramirez, A., 2017. Development of a methodological framework for social life-cycle assessment of novel technologies. Int. J. Life Cycle Assess. 22 (3), 423–440.
Van de Kerk, G., Manuel, A.R., 2008. A comprehensive index for a sustainable society: the SSI—the sustainable society index. Ecol. Econ. 66 (2), 228–242.
Vitunskiene, V., Dabkiene, V., 2016. Framework for assessing the farm relative sustainability: a Lithuanian case study. Agricultural Economics 62 (3).
Wang, J., Jing, Y., Zhang, C., Zhao, J., 2009. Review on multi-criteria decision analysis aid in sustainable energy decision-making. Renew. Sustain. Energy Rev. 13 (9), 2263–2278.
Warhurst, A., 2002. Sustainability Indicators and Sustainability Performance Management. Mining, Minerals and Sustainable Development [MMSD] Project Report. pp. 43.
Wilson, M.C., Wu, J., 2017. The problems of weak sustainability and associated indicators. Int. J. Sustain. Dev. World Ecol. 24 (1), 44–51.
Wind, Y., Green, P.E., 2013. Marketing Research and Modeling: Progress and Prospects: a Tribute to Paul E. Green Springer Science & Business Media.
WorldBank, 1999. World Development Indicators 1999. World Bank, Washington, DC.
Wu, J., Wu, T., 2012. Sustainability indicators and indices: an overview. Handbook of Sustainable Management. Imperial College Press, London, pp. 65–86.
Wu, J., 2013. Landscape sustainability science: ecosystem services and human well-being in changing landscapes. Landscape Ecol. 28 (6), 999–1023.
Yeheyis, M., Hewage, K., Alam, M.S., Eskicioglu, C., Sadiq, R., 2013. An overview of construction and demolition waste management in Canada: a lifecycle analysis approach to sustainability. Clean Technol. Environ. Policy 15 (1), 81–91.