Ecological Economics 114 (2015) 117–127

Contents lists available at ScienceDirect

Ecological Economics

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

Methodological and Ideological Options

Applications of aggregation theory to sustainability assessment☆

N. Pollesch [a,b,*], V.H. Dale [b]

[a] Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320, United States
[b] Center for BioEnergy Sustainability, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6035, United States

ARTICLE INFO

Article history:
Received 19 November 2014
Received in revised form 2 March 2015
Accepted 10 March 2015
Available online xxxx

Keywords:
Aggregation functions
Bioenergy sustainability
Compensatory functions
Distance to target
Indicators
Mathematical aggregation theory
Sustainability assessment
Uncertainty
Weak versus strong sustainability

ABSTRACT

In order to aid operations that promote sustainability goals, researchers and stakeholders use sustainability assessments. Although assessments take various forms, many utilize diverse sets of indicators numbering anywhere from two to over 2000. Indices, composite indicators, or aggregate values are used to simplify high dimensional and complex data sets and to clarify assessment results. Although the choice of aggregation function is a key component in the development of the assessment, there are few literature examples to guide appropriate aggregation function selection. This paper applies the mathematical study of aggregation functions to sustainability assessment in order to aid in providing criteria for aggregation function selection. Relevant mathematical properties of aggregation functions are presented and interpreted. Cases of these properties and their relation to previous sustainability assessment research are provided. Examples show that mathematical aggregation properties can be used to address the topics of compensatory behavior and weak versus strong sustainability, aggregation of data under varying units of measurements, multiple site multiple indicator aggregation, and the determination of error bounds in aggregate output for normalized and non-normalized indicator measures.

1. Introduction

A challenge for assessing sustainability is that it is not a single entity that can be readily measured. Instead sustainability is a combination of several aspects of the physical and biotic environment, social welfare, and economic wellbeing. Furthermore, it is an aspiration rather than a state. Its meaning is largely determined by contextual circumstances (Efroymson et al., 2013). Yet it is important to be able to measure, quantify, and discuss progress toward that goal.

Current sustainability assessment approaches often represent sustainability using multiple indicators, multiple variables, or multiple data points. At a minimum, the consensus is that sustainability needs to incorporate environmental, social, and economic conditions, which are referred to as the three pillars of sustainability (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008; Brundtland and World Commission on Environment and Development, 1987). In practice, sustainability indices can incorporate data from over 2600 indicator variables (The Living Planet Index, (McRae et al., 2012)). To add further complexity, each input variable often has an associated data set containing multiple observations. These large amounts of data about diverse components of sustainability are difficult to manage and nearly impossible to visualize without some sort of compression or reduction of dimensionality.

Aggregation functions are one method employed to accomplish this task of clarifying and simplifying data. Aggregation theory is the area of mathematics that explores the form and properties of such aggregation functions. In ecological economics the topic of aggregation comes up in regard to spatial aggregation (Su and Ang, 2010), valuation of ecosystem benefits (Tait et al., 2012; Lele and Srinivasan, 2013), calculation of conservation benefits (Winands et al., 2013), and combining information across sectors (Lenzen, 2007; Marin et al., 2012).

This study introduces basic properties, definitions, and theory related to the process of aggregation in order to aid in providing a rigorous mathematical baseline for further development of sustainability assessment techniques and methodologies. This paper deals with the conditions that must be met in order for information to be combined in an accurate, consistent, and overall robust manner. Five examples highlight some of the many relationships that can be derived between mathematical aggregation theory and sustainability assessment. These examples include mathematical interpretation of weak and strong sustainability, a proof that provides a simple bound for aggregate outputs under varying levels of relative error using the arithmetic mean, and two examples

☆ This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The United States Government and the publisher, by accepting the article for publication, acknowledge that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

* Corresponding author at: Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320, United States.
E-mail addresses: pollesch@math.utk.edu (N. Pollesch), dalevh@ornl.gov (V.H. Dale).

http://dx.doi.org/10.1016/j.ecolecon.2015.03.011
0921-8009/

of how grouping and aggregation can lead to inconsistent results depending on how aggregation takes place. The final section discusses multiple invariance properties with respect to the scale of measurability of the indicators to be aggregated and includes an example of how a simple change in measurement units can create inconsistent aggregate outputs. The 2004 paper by Ebert and Welsch, which provides a guideline for choosing aggregation functions, is interpreted and placed into the larger mathematical aggregation theoretic context.

2. Basic Properties of Aggregation Functions

The process of aggregation is ubiquitous in the sciences. However, the word aggregation can take on different meanings within different disciplines. The book *Aggregation Functions* (Grabisch et al., 2009) presents a comprehensive mathematical treatment of aggregation functions and their properties and is a unique resource within the mathematics literature. The definitions provided in Grabisch et al. (2009) are adopted here. Beginning with the formal definition for an aggregation function, the following section establishes the basic terms and properties used. For each set of properties presented, a mathematical definition is provided along with interpretations related to sustainability assessment to provide context.

2.1. Definition of an Aggregation Function

In general, an aggregate value is a single representative value for an arbitrarily long set of related values. An aggregation function is the mathematical operation that maps the input values to the representative output value or 'aggregate'. Formally, for some nonempty real interval $\mathbb{I}\subseteq\mathbb{R}$ containing the values to be aggregated, an *aggregation function* in $\mathbb{I}^n$ is a function:

$$A^{(n)}:\mathbb{I}^n\rightarrow\mathbb{I}$$

that

(i) is nondecreasing (in each variable)
(ii) fulfills the following boundary conditions:

$$\inf_{\mathbf{x}\in\mathbb{I}^n} A^{(n)}(x)=\inf\mathbb{I} \text{ and } \sup_{\mathbf{x}\in\mathbb{I}^n} A^{(n)}(x)=\sup\mathbb{I} \quad (1)$$

where n represents the number of variables in the argument of the function, that is, the number of values to be aggregated or the dimension of the input vector, x. In general, an aggregation function $A^{(n)}(x)$ is written as $A(x)$ with the number of variables in its argument suppressed. Also note that the domain associated with a given aggregation function often changes with assessment context.

As an interpretation, condition (i) states that if any input value increases, the aggregate output value cannot decrease. Condition (ii) dictates what must happen at the boundary values. For example, if a set of indicators are normalized to values between 0 and 1, then the nonempty interval is given by $\mathbb{I}=[0,1]$, and an aggregation function $A^{(n)}(x)$ must satisfy $A^{(n)}((0,...,0))=0$ and $A^{(n)}((1,...,1))=1$.

Table 1 gives some common aggregation functions and their definitions. The aggregation functions most frequently used in practice for sustainability assessment are the arithmetic and weighted arithmetic means (Singh et al., 2009; Böhringer and Jochem, 2007). Although the mathematical properties used to describe function behavior are numerous, certain properties of functions have particular importance to aggregation and are included here. The properties presented may help determine appropriate choices of aggregation functions given the sustainability indicator variables selected and the intended use within the assessment. Some of the mathematical definitions and properties presented, such as continuity, are familiar to mathematicians, while others, such as internality, conjunctivity, and disjunctivity as well as some of the grouping-based properties, are less familiar. However, within the context of sustainability assessment and aggregation theory, even familiar properties of functions can take on new meanings. The function property definitions in this paper follow the format of Grabisch et al. (2009), and interpretations relevant to sustainability assessment are provided when possible. Examples relating selected properties to sustainability assessment follow each set of properties provided.

2.2. Continuity Properties

Continuity relates closeness in the input variable(s) to closeness in the output variable(s) where closeness is defined using a specified norm. As such, continuity is important for understanding how the aggregation function performs with variable data or noise. Stronger and weaker forms of continuity exist. A strong form, Lipschitz continuity, allows for computing exact bounds in the output error of the aggregation function by knowing the error present in the input. An example of how the property of Lipschitz continuity of an aggregation function may be put to practical use in sustainability assessment is given next. Table 2 includes definitions for standard continuity and Lipschitz continuity for comparison and reference.

2.3. Example: Lipschitz Continuity and Error Estimation in the Arithmetic Mean

Error estimation and uncertainty quantification through the aggregation process may be approached by utilizing a variety of techniques. Certain aggregation functions have properties that allow one to provide

Table 1
Example aggregation functions.

Function name	Formula	Assumptions/notes
Arithmetic mean	$A(x):=\frac{1}{n}\sum_{i=1}^{n}x_i$	$A:\mathbb{I}^n\rightarrow\mathbb{I}, x:\in\mathbb{I}$
Weighted arithmetic mean	$A(x):=\sum_{i=1}^{n}w_ix_i$	$A:\mathbb{I}^n\rightarrow\mathbb{I}, x:\in\mathbb{I}(w_1,\ldots,w_n)\in[0,1]^n\sum_{i=1}^{n}w_i=1$
Ordered weighted average	[a]$A(x):=\sum_{i=1}^{n}w_ix_{(i)}$	$A:\mathbb{I}^n\rightarrow\mathbb{I}, x:\in\mathbb{I}$ $(w_1,\ldots,w_n)\in[0,1]^n\sum_{i=1}^{n}w_i=1$
Geometric mean	$A(x):=(\prod_{i=1}^{n}x_i)^{1/n}$	$A:\mathbb{I}^n\rightarrow\mathbb{I}, x:\in\mathbb{I}$ [b]If $n>1$ then $\mathbb{I}\subseteq(0,\infty)$
Weighted geometric mean	$A(x):=\prod_{i=1}^{n}x_i^{w_i}$	$A:\mathbb{I}^n\rightarrow\mathbb{I}, x:\in\mathbb{I}$ $(w_1,\ldots,w_n)\in[0,1]^n\sum_{i=1}^{n}w_i=1$ If $n>1$ then $\mathbb{I}\subseteq(0,\infty)$
Minimum	$A(x):=min\{x_1,\ldots,x_n\}$ (or $OS_1(x):=x_{(1)}$)	Also written $\mathrm{Min}(x)=\bigwedge_{i=1}^{n}x_i$ and OS_1 is the 1st order statistic
Maximum	$A(x):=max\{x_1,\ldots,x_n\}$ (or $OS_n(x):=x_{(n)}$)	Also written $\mathrm{Max}(x)=\bigvee_{i=1}^{n}x_i$ and OS_n is the nth order statistic

[a] $x_{(i)}$ represents the ith lowest coordinate of x, s.t. $x_{(1)}\leq\cdots\leq x_{(k)}\leq\cdots\leq x_{(n)}$.
[b] The geometric means are not aggregation functions on every domain, specifically, for $n>1$ then $\mathbb{I}$ must satisfy $\mathbb{I}\subseteq(0,\infty)$.

Table 2
Continuity properties.

Property	Definition	Interpretation/Notes
Standard continuity	F: $\mathbb{I} \rightarrow \mathbb{R}$, F is *continuous* at $x' \in \mathbb{I}$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{I} \lvert x - x' \rvert < \delta \Rightarrow \lvert F(x) - F(x') \rvert < \epsilon$	In essence, continuous functions have the property that *small* changes in input values ($\lvert x - x' \rvert < \delta$) will result in *small* changes in the output values ($\lvert F(x) - F(x') \rvert < \varepsilon$).
Lipschitz Continuity	$F : \mathbb{I}^n \rightarrow \mathbb{R}$, F is *Lipschitz continuous* (with respect to the norm, $\lVert \cdot \rVert$)[a] with *Lipschitz constant c* if $\lvert F(x) - F(x') \rvert \leq c \lVert x - x' \rVert$ for all $x, x' \in \mathbb{I}^n$	With Lipschitz continuity, knowledge about variation in input values ($\lVert x - x' \rVert$) can be used to give an exact bound for the variation in output values. This differs from the definition of standard continuity and is a stronger property. All Lipschitz continuous functions obey (standard) continuity, but not all continuous functions are Lipschitz continuous.

[a] A *norm* is used to convey a sense of distance between variables. With regards to indicator variable input, if x is the vector of *true* input values (without measurement error), and if x' represents the actual indicator variable measurements (with measurement error), then the norm of two inputs, $\lVert x - x' \rVert$, represents how different the true and measured values are (or how much error is present in the actual measurements).

exact bounds in output error depending on the input error (e.g., the arithmetic mean and its Lipschitz continuity).

Consider the following example: Let $x = (x_1, x_2,..., x_n)$, $x_i \in [0, 1]\ \forall\ i$ be a vector whose components are a set of n indicators to be aggregated using the arithmetic mean, $A(x) := \frac{1}{n}\sum_{i=1}^{n} x_i$. Assume that there is variability in the measures of each of the components, x_i, and that each indicator has a maximum relative error equal to ϵ for some $\epsilon > 0$. Let $\overline{x}_i$ be the best estimate of indicator x_i. Let $\hat{x}_i = \overline{x}_i + \overline{x}_i * \epsilon$, be the upper bound of measures of indicator i. Let $\check{x}_i = \overline{x}_i - \overline{x}_i * \epsilon$ be the lower bound in the measurement of indicator i. Let $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$ and $\check{\mathbf{x}} = (\check{x}_1, \ldots, \check{x}_n)$. It follows that $\lVert \hat{\mathbf{x}} - \check{\mathbf{x}} \rVert_1$ is the largest distance (using the L_1 norm[1]) between any two vectors of indicator measures. Further, this difference simplifies as follows:

$$\begin{aligned} \lVert \hat{\boldsymbol{x}} - \check{\boldsymbol{x}} \rVert_1 &= \\ &= \sum_{i=1}^{n} \lvert \hat{x}_i - \check{x}_i \rvert \\ &= \sum_{i=1}^{n} \lvert (\overline{x}_i + \overline{x}_i * \epsilon) - (\overline{x}_i - \overline{x}_i * \epsilon) \rvert \\ &= \sum_{i=1}^{n} 2\lvert \epsilon \rvert * \lvert \overline{x}_i \rvert \\ &= 2\lvert \epsilon \rvert * \lVert \overline{\boldsymbol{x}} \rVert_1. \end{aligned}$$

Using the Lipschitz continuity of the arithmetic mean (Grabisch et al., 2009) with Lipschitz constant $\frac{1}{n}$, the following bound must hold for the largest error in our aggregate value:

$$\begin{aligned} \lvert A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}}) \rvert &\leq \frac{1}{n} \lVert \hat{\boldsymbol{x}} - \check{\boldsymbol{x}} \rVert_1 \\ &= \frac{2\lvert \epsilon \rvert}{n} \lVert \overline{\boldsymbol{x}} \rVert_1. \end{aligned} \tag{2}$$

Inequality (2) gives a simple bound to the output in the aggregate value of a set of indicator variables with the same relative error, ϵ, using the arithmetic mean as the aggregation function. A relevant simplification comes when one considers measures that have been normalized to fall between the values of 0 and 1. In this case, $\lVert \hat{\boldsymbol{x}} \rVert_1 \leq n$, and thus $\lvert A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}}) \rvert \leq 2\lvert \epsilon \rvert$.

Since it is not generally expected that each of the indicators will have the same relative error in measurement, a more realistic example may include relative errors, ϵ_i, for each indicator variable x_i. To treat this case in which each indicator has its own maximum relative error, let $\hat{x}_i = \overline{x}_i + \overline{x}_i * \epsilon_i$ be the upper bound of measures of indicator i, let $\check{x}_i = \overline{x}_i - \overline{x}_i * \epsilon_i$ be the lower bound of measures of indicator i. Also, let $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$ and $\check{x} = (\check{x}_1, \ldots, \check{x}_n)$. With different relative errors for each indicator, the largest relative error among the set of relative errors determines the bound. Let $\epsilon_{max} = \max\{\epsilon_1,..., \epsilon_n\}$, let $\hat{x}_i^{max} = \overline{x}_i + \overline{x}_i * \epsilon_{max}$, let $\check{x}_i^{max} = \overline{x}_i - \overline{x}_i * \epsilon_{max}$, and let $\hat{\boldsymbol{x}}_{\boldsymbol{max}} = (\hat{x}_1^{max}, \ldots, \hat{x}_n^{max})$ and $\check{\boldsymbol{x}}_{\boldsymbol{max}} = (\check{x}_1^{max}, \ldots, \check{x}_n^{max})$. It follows that $\lVert \hat{\boldsymbol{x}} - \check{\boldsymbol{x}} \rVert_1 \leq \lVert \hat{\boldsymbol{x}}_{\boldsymbol{max}} - \check{\boldsymbol{x}}_{\boldsymbol{max}} \rVert_1$ and one can obtain a bound in the aggregate value, similar to the bound above, given by:

$$\begin{aligned} \lvert A(\hat{\mathbf{x}}) - A(\check{\boldsymbol{x}}) \rvert \leq \frac{1}{n} \lVert \hat{\boldsymbol{x}} - \check{\boldsymbol{x}} \rVert_1 &\leq \frac{1}{n} \lVert \hat{\boldsymbol{x}}_{\boldsymbol{max}} - \check{\boldsymbol{x}}_{\boldsymbol{max}} \rVert_1 \\ &= \frac{2\lvert \epsilon_{max} \rvert}{n} \lVert \overline{\boldsymbol{x}} \rVert_1. \end{aligned} \tag{3}$$

In this case if indicator measures are normalized using distance-to-target to fall in the interval between 0 and 1, then $\lVert \overline{\boldsymbol{x}} \rVert_1 \leq n$, leading to $\lvert A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}}) \rvert \leq 2\lvert \epsilon_{max} \rvert$.

Inequalities (2) and (3) show, for a set of indicator variables and some known relative error in their measurement, there is a precise bound for the arithmetic mean maximum range of aggregate output. Although this formulation is not relevant for unbounded error terms such as when ϵ is assumed to follow some probability distribution, the result is useful for sensitivity analysis as well as the quantification of uncertainty in aggregate output given the uncertainty in indicator input values, both of which are key components in sustainability assessment.

2.4. Internality, Conjunctivity, and Disjunctivity Properties

The degree to which, and if, compensation should occur between indicator variables to be aggregated is often contentious (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008). These disagreements are based on the fact that compensation between indicator variables implies that the quantities represented by those variables are in some sense substitutable. The use of the term *compensatory* or *compensation* in this paper describes the ability, in the aggregate output, of 'high' input component values to offset 'low' input component values, and vice-versa. Hacking and Guthrie (2008) point out that the Living Planet Index (McRae et al., 2012) utilizes a mathematically appropriate aggregation function but is flawed in that it assumes the substitutability of different species. The properties of internality, conjunctivity, and disjunctivity (defined in Table 3) are all related to how functions deal with extreme elements of input vectors and if the aggregate output falls strictly within the range of component input values or not. Thus, this set of properties is directly related to the compensatory behavior of the aggregation function.

Given the relationship between the Min(x) and Max(x) functions and the order statistic functions (see Table 1), the notion of conjunctivity and disjunctivity can be generalized to *k-conjunctive* (or *k-disjunctive*) functions. For example, a k-conjunctive function remains unchanged when any of the ordered components, $x_{(k+1)}, \ldots, x_{(n)}$ are replaced with values greater than or equal to the *kth* smallest ordered element, $x_{(k)}$. These types of functions can be used to 'ignore' the upper $n - k$ elements of an input vector.

[1] Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be vectors in $\mathbb{R}^n$. The L_1 norm of the difference between x and y denoted $\lVert x - y \rVert_1$, is given by $\lVert x - y \rVert_1 = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$ (also see Table 2 for a brief explanation of a *norm*).

Table 3
Internality, conjunctivity, and disjunctivity properties.

Property	Definition	Interpretation/notes
Conjunctive	[a] $F:\mathbb{I}^n \to \overline{\mathbb{R}}, x \in \mathbb{I}^n$ F is *conjunctive* if $\inf \mathbb{I} \le F(x) \le \mathrm{Min}(x)$	The output of the function F must be bounded (above) by the Min(x) function. This condition means that, in a conjunctive function, no low input component can be compensated for by a high input component.
Disjunctive	$F:\mathbb{I}^n \to \overline{\mathbb{R}}, x \in \mathbb{I}^n$ F is *disjunctive* if $\mathrm{Max}(x) \le F(x) \le \sup \mathbb{I}$	Similar to conjunctivity, but the output of the function F must be bounded (below) by the Max(x) function. Meaning that no low input component values may compensate for a high input component value.
Internal	$F:\mathbb{I}^n \to \overline{\mathbb{R}}, x \in \mathbb{I}^n$ F is *internal* if $\mathrm{Min}(x) \le F(x) \le \mathrm{Max}(x)$	Internal aggregation functions allow for *compensatory effects* between input component values. Here compensatory effects are taken to mean those that allow, for example, high input components to offset low input components in the aggregate output. Averages or mean aggregation functions are internal functions.

[a] $\overline{\mathbb{R}} = [-\infty, \infty]$ represents the *extended real line* and $\mathbb{I} \subseteq \overline{\mathbb{R}}$.

An intuitive definition of an aggregation function includes the mathematical concept of internality[2], given the prevalence of means and averages in assessment aggregation. If non-compensatory aggregation is sought in sustainability assessment, then further investigation of the types of aggregation functions that are not internal is necessary (see Grabisch et al., 2009, Chapter 3 for a treatment of conjunctive and disjunctive aggregation functions).

2.5. Example: Weak Versus Strong Sustainability and Internality

The concepts of *weak* and *strong sustainability* were introduced to capture the idea that certain natural capital stocks are unique, essential, and their loss could have irreversible effects on human well-being (Pearce et al., 1994). An 'overall capital stock' approach that allows for compensation between these unique, non-substitutable natural capital stocks and capital stocks, not categorized as such, was deemed to fall within a *weak sustainability* framework. A framework that respected the non-substitutability of these natural stocks was deemed as adhering to *strong sustainability*.

The concepts presented in Pearce et al. (1994) have since been used to categorize sustainability assessments (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008). A weak sustainability assessment approach permits compensation of indicators across the three pillars into a representative aggregate value. A strong sustainability assessment, on the other hand, specifies that no aggregation of indicators across the three pillars should be allowed during assessment (Mori and Christodoulou, 2012). The link between compensation and aggregation is substantial, since most aggregation functions used in practice are compensatory. However, not all aggregation functions are compensatory, allowing high values of inputs to offset low values of others. With this in mind, the framing of weak versus strong sustainability may be seen not necessarily as a matter of aggregation but more of compensation within the aggregate output.

An example of how aggregation theory may be applied to give a mathematical definition to weak versus strong sustainability is useful. Assume that a 'low' value represents a nadir (or anti-ideal) state for a given indicator, and a 'high' value represents an ideal state. Within a strong sustainability assessment framework, for example, no low indicator should offset a high economic or social indicator level and no low social indicator should offset a high environmental or economic indicator level, and so forth. Such compensatory effects, from an aggregation theoretic perspective, fall under the set of properties related to internality, conjunctivity, and disjunctivity. With this in mind, the following definition is suggested:

Let $\mathbb{I}^i_{env}, \mathbb{I}^j_{soc}, \mathbb{I}^k_{econ}$ represent disjoint subsets of indicators containing environmental, social, and economic indicators, respectively. Let $\mathbb{I}^n_{sust} = \mathbb{I}^i_{env} \cup \mathbb{I}^j_{soc} \cup \mathbb{I}^k_{econ}$ represent the entire set of sustainability indicators and let $F_{(x)}$ be an aggregation function s.t. $F:\mathbb{I}^n_{sust} \to \mathbb{I}$, then:

Definition. An aggregation function F will satisfy *strong sustainability* for any $x \in \mathbb{I}^n_{sust}$ if F is *conjunctive* or *disjunctive*.

This definition states that only conjunctive or disjunctive aggregation functions will satisfy strong sustainability when acting on input vectors that include components from more than one of the three pillars of sustainability.

With respect to the concept of weak and strong sustainability, this example can easily be extended to use different indicator categorizations. For instance, as opposed to assuming that social, environmental, and economic capital stocks are non-substitutable across the three pillars as categories (and are completely substitutable within), one can utilize indicator categorizations that identify the non-substitutable indicator components specifically. This approach would enable moving beyond the necessity of categorizing all indicators as being environmental, social, or economic, while still following the conceptual guidelines for strong sustainability as presented in Pearce et al. (1994).

An interesting example of a function in the sustainability assessment literature that can be either internal or conjunctive is provided by Díaz-Balteiro and Romero (2004). The authors consider n systems being assessed by m indicators. Letting W_j be a weight, or relative importance factor, for indicator j, and $\overline{R}_{ij}$ represent the normalized value for system i and indicator j, they propose the following index function, IS_i:

$$IS_i = (1-\lambda)\left[\min_j\left(W_j\overline{R}_{ij}\right)\right] + \lambda\sum_{j=1}^{m} W_j\overline{R}_{ij}$$

where $\lambda \in [0, 1]$ represents a *compensation parameter*. Notice that the internality/conjunctivity of this function is controlled by the value of the parameter λ. This function is a convex combination of the internal function $\sum_{j=1}^{m} W_j\overline{R}_{ij}$, the weighted arithmetic mean, and the conjunctive $\min_j\left(W_j\overline{R}_{ij}\right)$ function. In this case, any $\lambda \in (0, 1]$ will make the function IS_i internal; when $\lambda = 0$ the function IS_i is conjunctive. Also, it can be shown that IS_i satisfies the formal conditions (see Section 2.1) of an aggregation function.

Using the IS_i function from Díaz-Balteiro and Romero (2004), a simplified numerical example to emphasize the concepts presented is provided next. Consider a single bioenergy production site ($i = 1$) that has been assessed for environmental, social, and economic sustainability ($j = 3$). Let x_{env} be the environmental sustainability score, let x_{soc} be the social sustainability score, and let x_{econ} be the economic sustainability score and assume each value for these scores is on the scale from 0 to 1 where 0 represents a non-ideal state, and 1 represents an ideal sustainability state. If the overall sustainability of this site is to be assessed using these scores and equal weighting of scores is assumed (so $W_j = \frac{1}{j} = \frac{1}{3}$), then IS_i may be rewritten as:

$$IS_i = (1-\lambda)\min\{x_{env}, x_{soc}, x_{econ}\} + \lambda\left(\frac{1}{3}(x_{env} + x_{econ} + x_{soc})\right).$$

Assume further that the system has the following assessment scores: $x_{env} = 0.8$, $x_{soc} = 0.33$, $x_{econ} = 0.98$. Assuming complete compensation and substitutability between the three pillars, then $\lambda = 0$, and

$$IS_i = 0 + 1\left(\frac{1}{3}(0.8 + 0.33 + 0.98)\right) = 0.703$$

[2] The term *internality* used in this paper is purely mathematical, describing the behavior of the aggregation function in mapping an input vector to an output value that falls within the range of input component values. It is not related to the economic concept of imposition of costs as being internal or external that utilizes the same terminology.

If the compensation parameter is set $\lambda = 0$, then no compensation between the three pillars is assumed, and, in the aggregate, only the minimum value will be used to represent the system, specifically,

$$IS_i = 1(\min\{0.8, 0.33, 0.98\}) + 0 = 0.33$$

As the compensation parameter λ is varied between 1 and 0, IS_i increases (linearly) between the minimum value of the three and their arithmetic mean. Through this example, one can see that the concept of compensation between input values in an aggregation function relates directly to the properties of internality, conjunctivity, and disjunctivity of the aggregation function employed.

2.6. Grouping Based Properties

A key component of sustainability assessment is indicator identification and categorization. Categorization of indicators takes place at multiple levels using a variety of attributes of the indicator variables. Once categorization has occurred, this can lead to groups and subgroups of indicators. An ideal assessment method allows investigators or policy makers to both compress information when simplicity is demanded and also control the aggregation to maintain acceptable levels of information retained versus lost. Such a method may utilize aggregated values of groups or categories of indicators to focus on particular dimensions of the assessment. Additionally, this flexible assessment approach may have aggregation taking place multiple times and at multiple levels of the data structure, which can lead to inconsistent results depending on the aggregation function chosen. Properties related to the behavior of an aggregation function with respect to grouping and aggregation at multiple levels is found in Table 4. The behavior of the aggregation function with respect to the ordering of inputs (or groups of inputs) is captured by symmetry related properties presented in Table 5. These properties and behaviors are discussed below.

2.6.1. Properties Related to Aggregation at Multiple Levels

Repeated aggregation, or aggregation at multiple levels, is common in sustainability assessment. It arises when aggregate values are used to calculate other aggregate values. One example is when indices are used within indices. The inclusion of any biodiversity index as an indicator in a further aggregate value is an exemplar. The *Environmental Sustainability Index* calculation uses the *National Biodiversity Index* as an indicator. More recently Dobbs et al. (2011) propose that Shannon diversity and evenness index be used as an indicator for urban forest ecosystem services and goods assessment (Esty et al., 2005; Dobbs et al., 2011). Another example can be found in Gómez-Limón and Sanchez-Fernandez (2010), where the *risk of abandonment of agricultural activity* is an index included in their composite indicator of agricultural sustainability. The mathematical properties of *associativity* and *decomposability* (Table 4) are related to the behavior of function output with respect to aggregation at multiple levels.

Table 4
Associativity and decomposability properties.

Property	Definition	Interpretation/notes
Associativity	$F : \cup_{n\in\mathbb{N}}\mathbb{I}^n \rightarrow \mathbb{I}$ F is *Associative* if $F(x) = x$ for all $x \in \mathbb{I}$ and if $F(x, x') = F(F(x), F(x'))$ for all $x, x' \in \cup_{n\in\mathbb{N}}\mathbb{I}^n$	Associativity preserves the output of an aggregation of n indicators under the situation where first a subset of k components ($k < n$) are aggregated, then that output value is aggregated with the rest of the components.
Decomposability	[a]$F : \cup_{n\in\mathbb{N}}\mathbb{I}^n \rightarrow \mathbb{I}$ F is *Decomposable* if $F(x) = x$ for all $x \in \mathbb{I}$ and if [b]$F(x, x') = F(k \cdot F(x), k' \cdot F(x'))$ for all k, k' non-negative integers, and $x \in \mathbb{I}^k, x' \in \mathbb{I}^{k'}$	Decomposability is a similar property to associativity but requires knowledge of how many input values, k and k', the aggregate values to be aggregated again contain.

[a] The notation, $\cup_{n\in\mathbb{N}}\mathbb{I}^n$, is used to represent that the function F can map input vectors of varying number of arguments, n for any $n \in \mathbb{N}$, to the output interval $\mathbb{I}$.

[b] The notation $F(k \cdot F(x), k' \cdot F(x')) = F\Big(\underbrace{F(x), \ldots, F(x)}_{k \text{ of these}}, \underbrace{F(x'), \ldots, F(x')}_{k' \text{of these}}\Big)$.

The definitions given in Table 4 use notation that may be unfamiliar. To further decode the notation and what the definitions mean, consider the following example. Let $\{x_1, \ldots, x_{30}\}$ be a set of 30 sustainability indicator variables to be aggregated. Where the first 5 indicator values, $\{x_1, \ldots, x_5\}$ are all related to air quality, the next 10 are related to water quality, $\{x_6, \ldots, x_{15}\}$ and the final 15 indicators, $\{x_{16}, \ldots, x_{30}\}$ are related to soil quality. Let F be the aggregation function, and let $x = (x_1, \ldots, x_5)$, $x' = (x_6, \ldots, x_{15})$, $x'' = (x_{16}, \ldots, x_{30})$, so $x \in \mathbb{I}^5, x' \in \mathbb{I}^{10}$, and $x'' \in \mathbb{I}^{15}$. The aggregate value of the air, water, and soil quality indicators individually is given by $F(x)$, $F(x')$, and $F(x'')$, respectively. $F(x, x', x'')$ gives the total aggregate value for all indicators, where F is now taking as input, these three vectors of different dimensions, where the dimensions are related to number of indicators in the different groupings of the total set of indicators. The value, $F(F(x), F(x'), F(x''))$, may also be used to represent the total aggregate value of all indicators. In this case, F is only taking 3 input values, namely, the three values found to represent each group individually. The notation $F : \cup_{n\in\mathbb{N}}\mathbb{I}^n \rightarrow \bar{\mathbb{I}}$ indicates that our aggregation function F needs to have the flexibility to take as input arguments of varying numbers of indicators. And to reiterate, if it is found that $F(x, x', x'') = F(F(x), F(x'), F(x''))$, then the aggregation function F is associative. If it is the case that $F(x, x', x'') = F(5 \cdot F(x), 10 \cdot F(x'), 15 \cdot F(x''))$, then the aggregation function F is decomposable (see Table 4).

To give specific examples of aggregation functions with associativity or decomposability as properties, $\prod_i x_i$ and $\sum_i x_i$ are associative, while aggregation functions such as the arithmetic and geometric mean, $\frac{1}{n}\sum_{i=1}^{n} x_i$ and $(\prod_{i=1}^{n} x_i)^{1\backslash n}$ are decomposable but not associative. The following example shows how inconsistency in aggregate value output can arise if associativity and decomposability of the aggregation function are ignored.

2.6.2. Example: Aggregation of Subsets of Indicators and Associativity

The larger the scope and the more data included in the assessment can lead to aggregation taking place multiple times to produce assessment results. In addition to the examples of indices within indices, often, there is statistical analysis of replicates of data measurements for a given indicator to produce a single representative measurement for that indicator. Specifically, mean or median values of repeated indicator measurements are often used to calculate a composite indicator index rather than the raw data directly. The following example highlights how using a common aggregation function, the arithmetic mean, may cause inconsistencies under two different approaches to find a representative value for a data set.

Consider the following example using a simplified assessment scenario of two indicators. Let x_1 and x_2 represent the indicators and assume each indicator has multiple measurement observations. Indicator x_1 has been measured 5 times to give the following observations, (0.4,0.6,0.5,0.5,0.5). Indicator x_2 has been measured 4 times, with observations of its measure as (0.1,0.1,0.3,0.3). If one is interested in an aggregate value for either indicator x_1 or x_2 individually, then the arithmetic mean of the measures of x_1 is 0.5, and for x_2 the arithmetic mean is of its measures is 0.2.

If a composite (or aggregate) value is sought for the two indicators x_1 and x_2 together, two approaches are:

Approach A. Aggregating the mean values of the measurements for x_1 and x_2, 0.5 and 0.2, respectively. Using the arithmetic mean results in an overall representative value of 0.35.

Table 5
Symmetry properties.

Property	Definition	Interpretation/notes
Symmetry	$F : \mathbb{I}^n \to \overline{\mathbb{R}}$ F is *symmetric* if [a] $F(x) = F([x]_\sigma)$ for all $x \in \mathbb{I}^n$ and $\sigma \in \Sigma_{[n]}$	Symmetry preserves the output of an aggregation of n indicators under any permutation of the input components.
Bisymmetry	$F : \mathbb{I}^n \to \overline{\mathbb{R}}$ F is *bisymmetric* if $F(F(x_{11}, \ldots x_{1n}),..., F(x_{n1},..., x_{nn})) = F(F(x_{11}, \ldots x_{n1}),..., F(x_{1n},..., x_{nn}))$ for all $n \times n$ square matrices $\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix} \in \mathbb{I}^{n \times n}$	Within the matrix framework, bisymmetry ensures that the output of the function being applied to the function values determined by the row entries is equivalent to the output value determined by the function being applied to the function values determined by the column entries.
Strong bisymmetry	$F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$ F is *strongly bisymmetric* if $F(x) = x$ for all $x \in \mathbb{I}$ and if for any $n, p \in \mathbb{N}$ we have $F(F(x_{11}, \ldots x_{1n}), \ldots, F(x_{p1}, \ldots, x_{pn})) = F(F(x_{11}, \ldots x_{p1}), \ldots, F(x_{1n}, \ldots, x_{pn}))$ for all $p \times n$ matrices $\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pn} \end{pmatrix} \in \mathbb{I}^{p \times n}$	Strong Bisymmetry extends the property of bisymmetry to rectangular $n \times p$ matrices for $n \neq p$.

[a] The notation, $[x]_\sigma$ represents a permutation, σ, of the components of the input variable, x. $\Sigma_{[n]}$ is used to denote the set of permutations of the n components.

Approach B. Aggregating all the measurement data to find the arithmetic mean of the entire list of measures for both indicators, (0.5,0.4,0.6,0.5,0.5,0.1,0.1,0.3,0.3). Following this approach gives an overall representative value of $0.36\overline{6}$.

The inconsistency in results between approaches grows as the disparity in the number of observed measures for each indicator increases. For example, if indicator x_1 is measured five additional times at a value of 0.5, to give ten total observations, the arithmetic mean of the observations of x_1 is still 0.5. Approach A yields the same overall result of 0.35 for indicators x_1 and x_2. However, Approach B, which reports the arithmetic of all measurements for both indicators, yields an overall result of 0.41.

The discrepancy arising in the above example is because the arithmetic mean is not *associative*. The arithmetic mean is, however, *decomposable* and therefore can be used consistently in the two different approaches given in the example as long as the number of observed measures used to arrive at the mean values of 0.5 and 0.2 are known (see Table 4). The decomposability property applied to this circumstance gives that since five measures contribute to the mean value of 0.5 and four measures contribute to the mean value of 0.2, the aggregate value of[3] $(5 \cdot 0.5, 4 \cdot 0.2) = (0.5, 0.5, 0.5, 0.5, 0.5, 0.2, 0.2, 0.2, 0.2)$ will be same as the aggregate value of (0.5,0.4,0.6,0.5,0.5,0.1,0.1,0.3,0.3) using the arithmetic mean.

The assessment scenario just presented is abridged for clarity. However, variation in the number of measurements of different indicators is a common occurrence, given the diversity of indicators included in sustainability assessments and the differing measurement techniques that accompany each indicator. This example highlights that if one wishes to carry out further aggregations using aggregate values, then the grouping properties of aggregation functions becomes important in order to arrive at consistent values.

2.6.3. Symmetry Properties

The order in which the indicators appear in the input vector may or may not influence the aggregate value. Practically speaking, in the case when equal weights are assumed among indicators, the ordering of indicator appearance in the input vector should have no impact on the output value. The dependence of aggregate output value on input value ordering is captured by symmetry-related properties.

Symmetry is also extended to two-dimensional arrays, or matrices, of indicator values. In this setting, the property of *bisymmetry* is used to describe aggregate-value behavior under both varied grouping and varied ordering to arrive at aggregate values. Specifically, one can consider a set of indicator values organized in a $n \times n$ or $n \times p$ dimensional matrix and ask if the total aggregate value is consistent with the aggregate value found by first aggregating the rows and then the columns as well as the value found by first aggregating the columns and then the rows. This type of question might arise when considering a set of indicators to be aggregated for multiple sites to arrive at a total aggregate value for all sites.

To understand what symmetry of an aggregation function guarantees, consider the following example for a three dimensional input vector. Let $x = (x_1, x_2, x_3) \in \mathbb{I}^3$ and $A(x)$ be our aggregation function, if $A(x)$ is symmetric then $A(x_1, x_2, x_3) = A(x_1, x_3, x_2) = A(x_2, x_1, x_3) = A(x_2, x_3, x_1) = A(x_3, x_1, x_2) = A(x_3, x_2, x_1)$.

2.6.4. Example: Strong Bisymmetry and Multiple Site, Multiple Indicator Aggregation

The final example provided in this section concerns the aggregation of multiple indicators across multiple sites. For this example, consider sustainability of a bioenergy production site as measured by n indicators and p different production sites that are being assessed using the same set of indicators. Let $x_{ij} \in \mathbb{I}$ be the measure of indicator i at some site j. For an aggregation function, $F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$, let $F_{*j} = F(x_{1j}, x_{2j},..., x_{nj})$ be the aggregate value of all n indicators at site j, and let $F_{i*} = F(x_{i1}, x_{i2},..., x_{ip})$ be the aggregate value of a single indicator, i, at all p sites.

If a researcher is interested in finding one representative value for the sustainability of the all the p production sites within the region, they can follow different approaches:

Approach A. First aggregate a single indicator across all sites, and then aggregate that value across indicators.

Approach B. First aggregate all indicators for a particular site, and then aggregate those values for all sites.

The pertinent question is whether both approaches yield the same overall output from the aggregation. Or, symbolically:

$$F\Big(F(x_{11}, \ldots, x_{n1}), \ldots, F\Big(x_{1p}, \ldots, x_{np}\Big)\Big) \stackrel{?}{=} F\Big(F\Big(x_{11}, \ldots, x_{1p}\Big), \ldots, F\Big(x_{n1}, \ldots, x_{np}\Big)\Big).$$

If the function used to carry out the aggregation is *strongly bisymmetric*, then consistency is guaranteed. Otherwise, these two approaches may result in different overall assessments of the all the sites, using all the indicators, depending on the order in which aggregation takes places.

[3] The notation $(5{\cdot}0.5, 4{\cdot}0.2)$ is defined in Table 5.

3. Invariant and Meaningful Aggregation Functions by Level of Measurability

Given the variety of sustainability assessment approaches, developing procedures for the construction of aggregation methodologies adds uniformity and rigor to the field of research. However, in order for aggregation procedures to be derived, commonalities among the diverse approaches must be identified.

Ebert and Welsch (2004) provide an excellent example. Their work to define a *meaningful environmental index* is derived from concepts of *invariance* and *meaningfulness* and has since been used as a standard in the comprehensive evaluation of sustainability indices in Böhringer and Jochem (2007) and Singh et al. (2009). More recently, the work of Roberts (2014) discusses meaningfulness related to landscape ecology and biodiversity measures. Ebert and Welsch (2004) note that the topics of meaningfulness and invariance are well-known in social choice theory, and Roberts has developed a body of work discussing meaningfulness in relation to aggregation in multiple scientific contexts.

Ebert and Welsch (2004) has its basis in the theory of measurement. The *level of measurability* or *scale of measurability* is a classification of a given indicator variable based on the way in which the indicator can be quantified[4]. The classic examples of *levels of measurability* for variables are nominal, ordinal, interval, and ratio (Stevens, 1946). This fundamental classification of indicator variables and the properties associated with each level of measurability give rise to the concepts of invariance and meaningfulness for a given scale of measurability. After a brief introduction to the theory underlying the work of Ebert and Welsch (2004), the rest of this section connects their work to the larger context of aggregation functions, ending with examples in which invariance and meaningfulness can be applied to the development of sustainability assessment methodologies.

3.1. Admissible Transformations

Transforming variables between different units of measure is something that scientists often do. Whether it be from U.S. customary units to metric units for mass or length or from Celsius to Fahrenheit, these different units are seen as equivalent. However, when aggregation occurs among these variables, inconsistencies can arise if the scales on which they are measured are not taken into account.

3.1.1. Example: Inconsistent Aggregate Output Under Measurement Unit Transformations

In the creation of a sustainability assessment tool, allowing for transformation of data measurements between equivalent units, such as inches to centimeters or parts per million to parts per billion, is certainly desirable. In this example, consider a researcher who is comparing two different bioenergy production sites, site A and site B. The researcher wishes to track the total impact of nitrogen (N) and phosphorus (P) concentration in streams adjacent to the production site and then focus mitigation efforts on the site with the larger loading of nitrogen and phosphorus. Nitrogen concentrations are taken in units of milligram per liter (mg/L) and phosphorus in centigram per liter (cg/L). Site A has a nitrogen concentration of 0.970 mg/L and phosphorus concentration of 0.0051 cg/L. Site B has a nitrogen concentration of 0.950 mg/L and phosphorus concentration of 0.0082 cg/L.

The researcher calculates the arithmetic mean of nitrogen and phosphorus indicators and finds that site A has a value of $(0.970 + 0.0051)/2 = 0.488$ and that site B has a value of $(0.950 + 0.0082)/2 = 0.479$. The researcher concludes that *site A has a higher aggregate value of nitrogen and phosphorus loading than site B* and thus focuses mitigation efforts on site A.

Consider now that the researcher decides to record both phosphorus and nitrogen measures using the same units, milligrams per liter. After a quick change of units, phosphorus is now recorded as 0.0510 mg/L at site A and 0.0820 mg/L at site B. The arithmetic means are taken again. Site A has an aggregate value of $(0.970 + 0.0510) / 2 = 0.511$, while site B has an aggregate value of $(0.950 + 0.0820) / 2 = 0.516$. The researcher now concludes that *site B has a higher aggregate value of nitrogen and phosphorus loading than site A*. The researcher is left with a contradiction.

Both mg/L and cg/L are ratio scale measurable units. The inconsistent result shown in this example is due to changing between different ratio scale measurable units while using the arithmetic mean to aggregate. In order to ensure consistency under unit transformations that are often taken for granted, measurability scale invariant transformations need to be defined and aggregation functions identified that respect those transformations. In this example, had an aggregation function been chosen that is *meaningful on independent ratio scales* (the geometric mean for example), this inconsistent site ranking would not have occurred.

3.2. Admissible Transformation Formulations by Scale of Measurability

Ordinal, interval, and ratio scale measurable data all appear in sustainability assessments. Ordinal scale data represent a ranking or an order, but differences between numbers do not have meaning. Many surveys utilize ordinal scale measurable data. In recent research, Kopmann and Rehdanz (2013) include *Life Satisfaction* measured on a scale from 1 to 10 (where 1 = very dissatisfied and 10 = very satisfied) in their human well-being approach to assess value of natural land areas. Interval scale data are similar to an ordinal scale, except that differences between data points are meaningful. Interval scale measurable data also have arbitrary zero values that do not indicate the absence of the measured variable. Temperature as measured in Celsius is a classic example of interval scale measurement, since the difference between 20 and 21 degrees is the same as between 6 and 7, but 0 degrees does not represent the absence of temperature. Variables measured on a ratio scale are similar to interval scale measurable variables, except that the zero value is unique and non-arbitrary. Many indicators used for sustainability assessment are ratio scale measurable variables (see example in Section 3.3), such as bulk density measurements in soils and measurements of CO_2 emissions from a power plant over a given period of time. Besides nominal, ordinal, interval, and ratio, additional scales of measurability have been defined. Scales are also not fixed, as some data may be transformed between scales, temperature in Celsius (interval scale) to temperature in Kelvin (ratio scale) is an example.

Each scale of measurability has its own set of transformations that maintain the information contained in the data. Since the ordering or ranking is the information stored by ordinal scale data, functions that transform data on the ordinal-scale need to maintain that order. An *admissible transformation on the ordinal scale* is of the form

$$x \mapsto \varphi(x), \text{ where } \varphi \text{ is any strictly increasing function.}^{5} \quad (4)$$

Interval scale measurable data have the same restriction as ordinal scale data, but they must also maintain the distance between measurements. An *admissible transformation on the interval scale* is of the form,

$$x \mapsto rx + s, \text{ where } r > 0 \text{ and } s \in \mathbb{R}. \quad (5)$$

[4] One should be clear that *scale of measurability* of an indicator variable has no connection to the spatial or temporal extent to which a variable belongs.

[5] A function F is strictly increasing if $a < b \Rightarrow F(a) < F(b)$.

For ratio scale measurable data, order, distance, and the unique zero point must be maintained through transformations of the data. An *admissible transformation on the ratio scale* is of the form

$$x \mapsto rx, \text{ where } r > 0. \tag{6}$$

As an example to motivate consideration of admissible transformations, recall the *Life Satisfaction* ordinal response scale from 1 to 10, where 1 = very dissatisfied and 10 = very satisfied, from Kopmann and Rehdanz (2013). For simplicity, assume that there are 3 respondents to this survey question and their responses are {1,9,3}. Here respondent 1 is the least satisfied, respondent 2 is the most satisfied, and respondent 3 is somewhere in between. Formulation (4) states that in order to maintain the information in the response data, any later transformation must be a strictly increasing function. Choosing $x \in \mathbb{I}^n$ as a simple example of a strictly increasing function, under the use of φ, the responses are now transformed from $\{1,9,3\} \xrightarrow{\varphi} \{2,18,6\}$. Respondent 1 is still the least satisfied, respondent 2 the most satisfied, and the transformed data set has the same rank order in responses. If a function that is not strictly increasing was chosen, such as $\hat{\varphi}(x) = (x-5)^2$ then $\{1,9,3\} \xrightarrow{\hat{\varphi}} \{16,16,4\}$ the result is that respondents 1 and 2 are equally satisfied, and respondent 3 is the least satisfied; the use of a non-strictly increasing function as a transformation has fundamentally changed the information contained within the data set.

3.2.1. Meaningfulness of Functions and Indices by Scale Type

Functions defined as *meaningful* obey the principle that an admissible transformation of the input variable(s) should lead to an admissible transformation of the output variable. This is known as Luce's principle (Grabisch et al., 2009; Luce, 1959). Consider a function F, a set of input variables, $x = (x_1, \ldots, x_n)$, a set of admissible transformations for input variables, $\varphi = (\varphi_1, \ldots, \varphi_n)$, and an admissible transformation for the output variable with respect to the set of transformations for the input variables, Ψ_φ.

A function F that satisfies the following equation:

$$F(\varphi_1(x_1), \ldots, \varphi_n(x_n)) = \Psi_\varphi(F(x_1, \ldots, x_n)) \tag{7}$$

is defined to be *meaningful* (Grabisch et al., 2009).

Formulation (7) is the most general condition for a meaningful function given some set of admissible transformations. Each input variable, x_i, can be measured on a different scale of measurability, and the output variable, $F(x)$, can be measured on a scale different from all the x_i. Each x_i can have its own class of admissible transformations, φ_i. The output variable, $F(x)$, can also have its own admissible transformation, Ψ_φ, different from all the x_i. Although abstract, Eq. (7) is used to derive the definitions that are given in the next section. Additionally, the general case for which all input variables and all output variables can be on any scale of measurability has simplifications that are utilized in practice. From here out, focus is placed on interval and ratio scales, and although similar results exist for ordinal scale meaningful functions, those results are omitted[6].

Simplifications exist to Eq. (7) that arise when input variables and the output variable are measured on the same scales, or, in the most restrictive case, when variables are on the same scale and use the same measurement units. Beginning with the assumption that all x_i and $F(x)$ are on the same scale of measurability (all ratio scale measurable for example), Grabisch et al. (2009) focus on three simplifications. When all input variables and output variables are measured on the same scale, but none share identical units of measurement, this is termed *meaningful on independent scales*. For example, the scenario of aggregating 3 indicators measured in parts per million (ppm), g/cm^3, and temperature in Kelvin, respectively, to arrive at an output variable that is measured on a ratio scale different from all these three would be captured by the term *meaningful on independent ratio scales*. A further simplification comes when input variables and the output variable are measured using the same scale and all input variables share the exact same unit of measurement. The term *meaningful on a single scale* is used to describe this case. If one sought to aggregate input variables all measured in Fahrenheit to an output variable measured in Celsius, a meaningful function in this case would be termed *meaningful on a single interval scale*. The final, and least general case, is when all the input and the output variables are measured on the same scale and in the same units. The term *invariant* is once again utilized to describe functions that are meaningful under this circumstance. If one wished to meaningfully aggregate input variables all measured in hectares to an output variable also measured in hectares, then, since hectares are ratio scale measurable units, this would be necessitate the use of a *ratio scale invariant* function. Although the final circumstance is the least general derived using Eq. (7), it frequently arises. When multiple samples are taken for a single indicator and aggregated to find a representative value for that indicator, this circumstance applies. The mathematical formulations that accompany the simplifications discussed in this paragraph for both ratio and interval scale measurable variables are presented next.

Admissible transformations for ratio scale measurable variables are classified in Eq. (6) above. Using this form of an admissible transformation, and the general form of meaningful functions presented in (7), the following definitions arise when all input and output variables are measured on ratio scales:

A function $F : \mathbb{I}^n \to \mathbb{R}$ is

Ratio scale invariant if for any $r > 0$,

$$F(rx) = rF(x) \tag{8}$$

for all $x \in \mathbb{I}^n$ such that $rx \in \mathbb{I}^n$.

Meaningful on a single ratio scale if for any $r > 0$, there exists $R(r) > 0$ such that,

$$F(rx) = R(r)F(x) \tag{9}$$

for all $x \in \mathbb{I}^n$ such that $rx \in \mathbb{I}^n$.

Meaningful on independent ratio scales if for any $r \in (0, \infty)^n$, there exists $R(r)$ such that,

$$F(rx) = R(r)F(x) \tag{10}$$

for all $x \in \mathbb{I}^n$ such that $rx \in \mathbb{I}^n$.

Admissible transformations for interval scale measurable variables are classified in Eq. (5) above. Using this form for admissible transformations, and the general forms of meaningful functions from Eq. (7), the following definitions arise for functions applied to interval scale measurable input and output variables:

A function $F : \mathbb{I}^n \to \mathbb{R}$ is

Interval scale invariant if for any $r > 0$ and $s \in \mathbb{R}$,

$$F(rx + s1) = rF(x) + s \tag{11}$$

for all $x \in \mathbb{I}^n$ such that $rx + s1 \in \mathbb{I}^n$.

Meaningful on a single interval scale if for any $r > 0$ and any $s \in \mathbb{R}$, there exists $R(r, s) > 0$ and $S(r, s) \in \mathbb{R}$ such that

$$F(rx + s1) = R(r,s)F(x) + S(r,s) \tag{12}$$

for all $x \in \mathbb{I}^n$ such that $rx + s1 \in \mathbb{I}^n$.

[6] For more about aggregation on ordinal scales see Grabisch et al. (2009), Chapter 8.

Meaningful on independent interval scales if for any $r \in (0, \infty)^n$ and any $s \in \mathbb{R}^n$, there exists $R(r, s)$ and $S(r, s) \in \mathbb{R}$ such that

$$(rx + s) = R(r, s)F(x) + S(r, s) \tag{13}$$

for all $x \in \mathbb{I}^n$, such that $rx + s \in \mathbb{I}^n$.

Although the definitions given above progress from least to most general in terms of the types of input and output variables that are considered, it is not the case that a function being meaningful on independent scales implies that the function is meaningful on a single scale or that a function being meaningful on a single scale implies it is also invariant on the scale. The relationship that holds between the definitions is that, if a function is scale invariant, it will also be meaningful on a single scale, and, if a function is meaningful on independent scales, it will also be meaningful on a single scale (Grabisch et al., 2009). With the definitions and conditions formalized for meaningful and invariant functions on given scales, the next section provides example functions that satisfy the six meaningful definitions given above.

3.2.2. Meaningful Aggregation Functions

General results for functions that satisfy the defined meaningfulness and invariance equations have been studied for nearly three decades. Results are available for all six of the different invariance and meaningfulness scenarios presented above, as well as others (Grabisch et al., 2009; Aczél and Roberts, 1989; Aczél et al., 1986). Using slightly different terminology, Ebert and Welsch (2004) present some of these results in the context of defining their meaningful environmental index as well.

Ebert and Welsch (2004) provided a derivation and examples of ratio noncomparable, ratio full comparable, interval noncomparable, and interval full comparable orderings that satisfy various forms of continuity and monotonicity. In their work, the term 'noncomparable' is similar to the term *independent* in classifying scales of input and output variables. The meaning of the term 'full comparable' is similar to the use of the word *single* above. The results of their paper have since been simplified and presented in the form of a compact table, which is given in Table 6 (Böhringer and Jochem, 2007).

If focusing on *aggregation functions*, then one only needs to consider functional forms that satisfy the different meaningful and invariance properties above and that are nondecreasing and fulfill the boundary conditions provided in Eq. (1). In doing so, nearly identical results arise as to those presented in Ebert and Welsch (2004). Table 7 is adapted from Grabisch et al. (2009) and presents aggregation functions that satisfy the different meaningfulness properties discussed thus far.

Grabisch et al. (2009) also provide deeper results for meaningfulness with respect to ratio and interval scale measurable variables than just examples of functions that satisfy different meaningfulness properties. Their results give complete descriptions for the types of aggregation functions that satisfy meaningfulness on independent scales for ratio (Eq. (10)) and interval (Eq. (13)) scale measurable variables.

3.2.2.1. Meaningful Aggregation Functions on Independent Ratio Scales

Proposition 7.8 (Grabisch et al., 2009). [7]Let $\mathbb{I} = |0, b|$ with $b \in (0, \infty]$. A function $F : \mathbb{I}^n \mapsto \mathbb{I}$ is a *meaningful aggregation function on independent ratio scales* if and only if

$$F(x) = a\prod_{i=1}^{n} x_i^{a_i} \tag{14}$$

where $a_1, \ldots, a_n \in [0, \infty)$, $\sum_{i=1}^{n} a_i > 0$ and $a > 0$ if $b = \infty$, while $a = b\prod_{i=1}^{n} b^{-a_i}$ if $b < \infty$.

[7] The notation $|0, b|$ represents any real interval with endpoints 0 and b.

Table 6
Aggregation rules for variables by Ebert and Welsch *via*?

	Noncomparable	Full comparable
Interval scale	Dictatorial ordering	Arithmetic mean
Ratio scale	Geometric mean	Any homothetic function

3.2.2.2. Meaningful Aggregation Functions on Independent Interval Scales

Proposition 7.34 (Grabisch et al., 2009). An aggregation function $F : \mathbb{I}^n \mapsto \mathbb{I}$ is *meaningful on independent interval scales* if and only if

$$F(x) = cx_i + d \tag{15}$$

for some i, where $c > 0$ and $d \in \mathbb{R}$ satisfy $ca + d = a, cb + d = b$ where $a = \inf \mathbb{I}, b = \sup \mathbb{I}$.

The major results of Ebert and Welsch (2004) are echoed in Table 7 and Propositions 7.34 and 7.8 from Grabisch et al. (2009). Specifically, Grabisch et al. (2009) give that, for Eq. (15), these solutions for meaningful aggregation functions on independent interval scales contained within bounded intervals are projections onto a single coordinate, $P_k(x)$ (all indicator variables will be contained in bounded intervals). Further, $P_k(x)$ may be seen as a *dictatorial ordering*, which is the terminology used by Ebert and Welsch (2004), because one chooses a single input element x_k to represent the rest of the input values by. The geometric mean and weighted geometric means are the only example aggregation functions that satisfy all three of the ratio scale meaningfulness properties; these clearly fit under the categorization of Eq. (14). $P_k(x)$ is the only example of an aggregation function that satisfies all six of the invariance and meaningfulness properties for interval and ratio scale measurable variables.

3.3. Application of Meaningfulness Properties to Sustainability Assessment

In order to apply successfully the invariance properties to develop an aggregation strategy, a proper classification of what measurability scenario exists for the chosen indicators is crucial. This necessitates categorization of the indicator variables included in the sustainability assessment into their scales of measurability. It further demands an understanding of how measurability scales change through any normalization or rescaling process.

Within sustainability assessment, one might encounter any of the meaningfulness or invariance scenarios presented above. Ratio scale measurable variables are common among indicator variables, since many scientific quantities fall on this scale naturally. As an example, Table 8 contains the list of recommended environmental sustainability indicator variables for assessing bioenergy sustainability as provided by McBride et al. (2011), and 18 out of the 19 indicators are ratio scale measurable. The indicators are from diverse categories such as air, soil,

Table 7
Meaningfulness of common aggregation functions *adapted from*?

Aggregation function	R.S.I.	S.R.S.	I.R.S.	I.S.I.	S.I.S.	I.I.S.
Arithmetic mean	✓	✓		✓	✓	
Geometric mean	✓	✓	✓			
[a]$P_k(x) := x_k$	✓	✓	✓	✓	✓	✓
[b]$OS_k(x) := x_{(k)}$	✓	✓		✓	✓	
Weighted arithmetic mean	✓	✓		✓	✓	
Weighted geometric mean	✓	✓	✓			
Ordered weighted average	✓	✓		✓	✓	
$\sum_{i=1}^{n} x_i$	✓	✓			✓	
$\prod_{i=1}^{n} x_i$	✓	✓				

Ratio scale invariant (R.S.I.), meaningful on a single ratio scale (S.R.S.) meaningful on independent ratio scales (I.R.S.), interval scale invariant (I.S.I), meaningful on a single interval scale (S.I.S.), and meaningful on an independent interval scales (I.I.S).

[a] $P_k(x)$ is the projection onto the kth element, x_k of the input vector x.

[b] $OS_k(x)$ is the projection on the kth ordered element, $x_{(k)}$ of the input vector x (All other function definitions may be found in Table 1).

Table 8
Recommended environmental indicators for bioenergy sustainability with measurability scales *adapted from*?

Category	Indicator	Units	Measurability scale
Soil quality	1. Total organic carbon (TOC)	Mg/ha	Ratio scale
	2. Total nitrogen (N)	Mg/ha	Ratio scale
	3. Extractable phosphorus (P)	Mg/ha	Ratio scale
	4. Bulk Density	g/cm^3	Ratio scale
Water quality and quantity	5. Nitrate concentration in streams (and export)	Concentration: mg/L; export: kg/ha/year	Ratio scale; ratio scale
	6. Total phosphorus (P) concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	7. Suspended sediment concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	8. Herbicide concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	9. Peak storm flow	L/s	Ratio scale
	10. Minimum base flow	L/s	Ratio scale
	11. Consumptive water use (incorporates base flow)	Feedstock production: m^3/ha/day; bioenergy: m^3/day	Ratio scale; ratio scale
Greenhouse gases	12. CO_2 equivalent emissions (CO_2 and N_2O)	kg C_{eq}/GJ	Ratio scale
Biodiversity	12. Presence of taxa of special concern	Presence	[a]
	14. Habitat area of taxa of special concern	ha	Ratio scale
Air quality	15. Tropospheric ozone	ppb	Ratio scale
	16. Carbon monoxide	ppm	Ratio scale
	17. Total particulate matter less than 2.5 µm diameter ($PM_{2.5}$)	$\mu g/m^3$	Ratio scale
	18. Total particulate matter less than 10 µm diameter (PM_{10})	$\mu g/m^3$	Ratio scale
Productivity	19. Aboveground net primary productivity (ANPP)/yield	g C/m^2/year	Ratio scale

[a] Due to variation of habitat for species of special concern in different contexts, this indicator does not have specified units of measurement.

and water quality, water quantity, greenhouse gas emissions as well as productivity and biodiversity.

Within some categories, such as soil and water quality, many of the variables are measured on the same scale in the same units, for example indicators 1–3 are all measured in Mg/ha, and indicators 5–8 are all measured in mg/L and kg/ha/year. This consistency allows one to utilize ratio scale invariant functions for aggregation within those particular groups. However, it may be simplest to choose a function that is meaningful on independent ratio scales, for then any aggregation on 18 of 19 indicator variables may be carried out using the same function. In the McBride et al. (2011) example, ratio scales dominate. Also, normalization procedures can produce measures on relative scales that are also ratio scale measurable.

Many sustainability assessment researchers argue that normalization by using a distance-to-target method is an appropriate way to deal with variables that reside on different scales of measurability (Mayer, 2008; Moldan et al., 2012). When it comes to normalized or relative values, Grabisch et al. (2009) point to the work of Roberts (1994) as identifying functions of the form given in Eq. (14) as appropriate aggregation functions on these scales. However, with respect to normalization, Ebert and Welsch (2004) contend that normalization introduces further ambiguities to the system when there are arbitrary normalization rules. They go on further to say that if one simply used unnormalized data measures that are on independent ratio scales, a meaningful environmental index would result by using a geometric mean as the aggregation function. Whether or how to normalize indicators is indeed something that varies by project, and, in either case, as long as one understands the scale of measurability on which the raw or normalized indicators fall, meaningful aggregation functions can be identified and used. Because of the importance of normalization as it relates to aggregation as well as to the assessment as a whole, the authors of this paper are investigating implications of aggregation theoretic properties as they relate to common normalizations procedures found in sustainability assessments.

Flexibility for assessing sustainability in different contexts is enhanced by defining site specific baselines and targets for distance-to-target normalization. If the same baselines and targets are set for all sites, then the normalized variables can be aggregated in a meaningful fashion using a ratio-scale invariant function. However, in a scenario where normalization using distance-to-target is used and each site has its own baselines and targets, the normalized variables no longer fall on identical scales, and so ratio scale invariant functions are no longer appropriate. This result occurs because, when different baselines and targets are used, a per unit change in the indicator at one site corresponds to a different change in the normalized variable than a per unit change in the same indicator at a different site. This observation highlights that even slightly different normalization by site has the ability to change the scales for identical indicators, and an aggregation function that is meaningful on independent ratios scales is needed for the normalized values from different sites.

4. Conclusion

Sustainability assessments are often complex, utilizing high-dimensional data sets and multifaceted analyses of the diverse indicator data. Aggregation is a key component in many sustainability assessments and a step that has large impact on the outcome of assessment results. To build upon the existing guidance for the construction of sustainability assessments, this paper introduces mathematical concepts that can be used to bolster the rigor and consistency of the aggregation component within the assessment. The concepts presented draw mostly from the mathematical study of aggregation functions, for which Grabisch et al. (2009) provide an excellent resource. Beyond the presentation and justification of relevant mathematical properties of aggregation functions, examples provide context and motivation for further investigation into this branch of mathematics that can be used in sustainability assessment. The paper concludes with a discussion of the work of Ebert and Welsch (2004) and meaningful indices and aggregation functions. It is shown that meaningful aggregation can take place using a variety of aggregation functions, depending on the scale in which the indicator variables are measured. Whether indicator data are normalized or not, meaningful aggregation functions can be defined and utilized for the synthesis and compression of high-dimensional assessment data.

As new sustainability assessments are constructed and existing assessment utilized, this paper should provide deeper understanding of how inconsistencies can arise in sustainability assessment in relation to the aggregation function(s) utilized. The properties of aggregation functions presented are by no means exhaustive and were chosen due to their particular relevance and to raise awareness of opportunities to introduce mathematical rigor within sustainability assessment.

Acknowledgments

This research was supported by the U.S. Department of Energy (DOE) under the Bioenergy Technologies Office. Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725. Comments by Lou Gross and Radko Mesiar on an earlier draft are also appreciated. Discussions with Keith Kline, Esther Parish, Jean-Luc Marichal, Bruce Peckham, and Suzanne Lenhart have been helpful.

References

Aczél, J., Roberts, F.S., 1989. On the possible merging functions. Math. Soc. Sci. 17 (3), 205–243.

Aczél, J., Roberts, F.S., Rosenbaum, Z., 1986. On scientific laws without dimensional constants. J. Math. Anal. Appl. 119 (1–2), 389–416.

Böhringer, C., Jochem, P.E., 2007. Measuring the immeasurable — a survey of sustainability indices. Ecol. Econ. 63 (1), 1–8.

Brundtland, G.H., World Commission on Environment and Development, 1987. Our common future vol. 383. Oxford University Press, Oxford.

Díaz-Balteiro, L., Romero, C., 2004. In search of a natural systems sustainability index. Ecol. Econ. 49 (3), 401–405.

Dobbs, C., Escobedo, F.J., Zipperer, W.C., 2011. A framework for developing urban forest ecosystem services and goods indicators. Landsc. Urban Plan. 99 (3), 196–206.

Ebert, U., Welsch, H., 2004. Meaningful environmental indices: a social choice approach. J. Environ. Econ. Manag. 47 (2), 270–283.

Efroymson, R.A., Dale, V.H., Kline, K.L., McBride, A.C., Bielicki, J.M., Smith, R.L., Parish, E.S., Schweizer, P.E., Shaw, D.M., 2013. Environmental indicators of biofuel sustainability: what about context? Environ. Manag. 51 (2), 291–306.

Esty, D.C., Levy, M., Srebotnjak, T., De Sherbinin, A., 2005. Environmental sustainability index: benchmarking national environmental stewardship. Yale Center for Environmental Law & Policy, New Haven, pp. 47–60.

Gómez-Limón, J.A., Sanchez-Fernandez, G., 2010. Empirical evaluation of agricultural sustainability using composite indicators. Ecol. Econ. 69 (5), 1062–1075.

Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E., 2009. Aggregation functions. 1 edition. Cambridge University Press, New York.

Hacking, T., Guthrie, P., 2008. A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment. Environ. Impact Assess. Rev. 28 (2–3), 73–89.

Kopmann, A., Rehdanz, K., 2013. A human well-being approach for assessing the value of natural land areas. Ecol. Econ. 93, 20–33.

Lele, S., Srinivasan, V., 2013. Disaggregated economic impact analysis incorporating ecological and social trade-offs and techno-institutional context: a case from the western ghats of India. Ecol. Econ. 91, 98–112.

Lenzen, M., 2007. Aggregation (in-) variance of shared responsibility: a case study of Australia. Ecol. Econ. 64 (1), 19–24.

Luce, R.D., 1959. On the possible psychophysical laws. Psychol. Rev. 66 (2), 81.

Marin, G., Mazzanti, M., Montini, A., 2012. Linking namea and input output for 'consumption vs. production perspective' analyses: evidence on emission efficiency and aggregation biases using the Italian and Spanish environmental accounts. Ecol. Econ. 74, 71–84.

Mayer, A.L., 2008. Strengths and weaknesses of common sustainability indices for multidimensional systems. Environ. Int. 34 (2), 277–291.

McBride, A.C., Dale, V.H., Baskaran, L.M., Downing, M.E., Eaton, L.M., Efroymson, R.A., Garten, C.T., Kline, K.L., Jager, H.I., Mulholland, P.J., Parish, E.S., Schweizer, P.E., Storey, J.M., 2011. Indicators to support environmental sustainability of bioenergy systems. Ecol. Indic. 11 (5), 1277–1289.

McRae, L., Collen, B., Deinet, S., Hill, P., Loh, J., Baillie, J., Price, V., 2012. The living planet index. In: Almond, R. (Ed.), The Living Planet Report 2012. WWF International, Gland, Switzerland.

Moldan, B., Janoušková, S., Hák, T., 2012. How to understand and measure environmental sustainability: indicators and targets. Ecol. Indic. 17, 4–13.

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.

Pearce, D.W., Atkinson, G.D., Dubourg, W.R., 1994. The economics of sustainable development. Annu. Rev. Energy Environ. 19 (1), 457–474.

Roberts, F.S., 1994. Operations research and the public sector. Handbooks in Operations Research and Management Science vol. 6. Elsevier.

Roberts, F.S., 2014. Meaningful and meaningless statements in landscape ecology and environmental sustainability. In: Pardalos, P., Goldengorin, B., Aleskerov, F. (Eds.), Clusters, Orders, Trees: Methods and Applications. Springer, pp. 297–312.

Singh, R.K., Murty, H., Gupta, S., Dikshit, A., 2009. An overview of sustainability assessment methodologies. Ecol. Indic. 9 (2), 189–212.

Stevens, S., 1946. On the theory of scales of measurement. Science (80-.) 103 (2684), 677–680.

Su, B., Ang, B., 2010. Input–output analysis of CO_2 emissions embodied in trade: the effects of spatial aggregation. Ecol. Econ. 70 (1), 10–18.

Tait, P., Baskaran, R., Cullen, R., Bicknell, K., 2012. Nonmarket valuation of water quality: addressing spatially heterogeneous preferences using gis and a random parameter logit model. Ecol. Econ. 75, 15–21.

Winands, S., Holm-Müller, K., Weikard, H.-P., 2013. The biodiversity conservation game with heterogeneous countries. Ecol. Econ. 89, 14–23.