Ecological Indicators 39 (2014) 189–202

Contents lists available at ScienceDirect

Ecological Indicators

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

Constructing the FEEM sustainability index: A Choquet integral application☆

Mehmet Pinar[a,*], Caterina Cruciani[b,c], Silvio Giove[b,c], Matteo Sostero[b,d]

[a] Edge Hill University, Ormskirk, United Kingdom
[b] Fondazione Eni Enrico Mattei, Venice, Italy
[c] Ca' Foscari University of Venice, Venice, Italy
[d] Sant'Anna School of Advanced Studies, Pisa, Italy

ARTICLE INFO

Article history:
Received 23 January 2013
Received in revised form 3 December 2013
Accepted 12 December 2013

JEL classification:
C43
C44
Q01
Q56

Keywords:
Sustainability indices
Composite indicators
Choquet integral
Multi-attribute value theory

ABSTRACT

This paper presents the development of the FEEM sustainability index (FEEM SI), a composite index including 19 different indicators grouped in the three classical pillars of sustainability – economic, social and environmental. We present the relevance of multi-attribute aggregation methodologies when dealing with such complex concepts and apply an aggregation methodology used for this case study: the Choquet integral operator. First, we normalize each sustainability indicator with the use of a benchmarking procedure with a smooth target of sustainability. We then develop an aggregation tree of sustainability criteria and a questionnaire to measure the values that experts attribute to individual sustainability criteria and their interaction. This survey suggests that a majority of experts consider sustainability criteria as complementary to each other. After combining the preferences of different experts to establish a consensus, we construct the FEEM SI using the Choquet integral aggregation procedure. The results for sustainability levels show that countries that are ranked at higher (lower) positions are those that have better (worse) outcomes in at least in two final pillars, respectively. Finally, we conduct a robustness analysis by repeating the aggregation procedure with different convex combinations of experts' preferences. The results indicate that, while sustainability levels of countries do vary with the expert preferences, countries' respective rankings remain mainly the same, irrespective of the combination of experts' preferences.

1. Introduction

Sustainability is a somewhat elusive concept: although its main message is widely understood, it is quite unusual to encounter two identical descriptions of it, when it comes to spelling out its different components. The most widely used definition of sustainable development is given in the Bruntland report, which defines it as a "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Bruntland, 1987). Achieving a sustainable development has been one of the major concerns of modern societies, which have long been interested in understanding and governing the multi-faceted issue of development (see, e.g. Fleurbaey, 2009; Fleurbaey and Blanchet, 2013 among many others). Thus, a comprehensive assessment of sustainability has become crucial to measure progress, identify areas to be addressed and evaluate policy outcomes. The need to find ways to measure sustainability translated into a multitude of approaches and sustainability indicators that have been aggregated in different ways to arrive at composite indices. For a methodological review on the definition of sustainability, see Bossel (1999), OECD JRC (2008), and Singh et al. (2009); and for list of sustainability indicators refer to the EU core set of indicators (EEA, 2005), and the UN Commission on Sustainable Development (UN CSD, 2005).

This paper focuses on the methodological issues in the construction of a composite sustainability index, an area that has been gaining interest in empirical literature due to its high policy potential (Saltelli, 2007). In fact, a composite index allows for a quick assessment of sustainability performance across different countries and moments in time. Moreover, sustainability indices convey a straightforward message to stakeholders and policy makers, and are able to highlight best practices and weaknesses of sustainability strategies (Ness et al., 2007).

☆ This paper belongs to the research activities of the Climate Change and Sustainable Development Programme of the Fondazione ENI Enrico Mattei. The authors are grateful to Carlo Carraro, Fabio Eboli, Francesco Bosello, Ramiro Parrado, Elisa Lanzi, Lorenza Campagnolo and all participants at seminars in Venice and Milan in which a preliminary version was presented. The usual disclaimer applies.

* Corresponding author at: Business School, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, United Kingdom. Tel.: +44 01695657629; fax: +44 01695584675.
E-mail address: mehmet.pinar@edgehill.ac.uk (M. Pinar).

1470-160X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ecolind.2013.12.012

Such a sustainability index needs to be constructed very carefully, using procedures as transparent as possible, in order to gain trust in a policy environment (Saisana et al., 2005). Moreover, sustainability is characterized with many different aspects that are somewhat interlinked with each other (e.g. economic growth and greenhouse gas emissions). Simple aggregation techniques disallow this interaction possibility, which may discard too much information regarding potential interactions among indicators (Munda and Nardo, 2009). For example, a linear aggregation would evaluate indicators that are substitutes or complements of each other in the same manner without considering interactions among them. Therefore, construction of a sustainability index should rely on non-linear aggregation methodology that takes into account the potential interactions between sustainability indicators.

There exist many examples of composite sustainability indices in the literature (see, e.g. Singh et al., 2009 for a detailed list of composite indices). However, none of those indices captures inter-relations (i.e. synergies or redundancies) when indicators are aggregated (see, e.g. Panayotou, 1993; Grossman and Krueger, 1993; Selden and Song, 1994; Arrow et al., 2004; Millennium Ecosystem Assessment, 2005; World Bank, 2010; Agliardi, 2011; Arrow et al., 2012 for discussion on the relationship between economic growth and environmental quality). In other words, a linear composite index employed by the literature (such as the Environmental Sustainability Index (ESI, 2005)), implies constant substitutability among indicators; this feature is not desirable in the context of sustainability indicators, for which the relation tends to be non-compensative (see, e.g. Munda, 2005; Munda and Nardo, 2009).

Since there is no clear agreement in the literature, this paper aims to establish the relation between different sustainability indicators – in terms of complementarity and substitutability – as well as estimating their relative importance in the definition of sustainability.

To address these issues – for the construction of the FEEM sustainability index (FEEM SI hereafter) – we identify 19 sustainability indicators for three branches of sustainability: economic, social and environmental. We design a hierarchical *aggregation tree* that combines the indicators into aggregate indices in distinct nodes. We then develop a custom questionnaire to elicit the preferences of experts (expressed in terms of *measures*) at each node of the tree, to assess both sustainability criteria in isolation and their interactions. As the evaluation of sustainability indicators might differ across experts, we derive a set of *consensus measures* from a combination of experts' preferences at each node of the aggregation tree. We then normalize all indicators composing the FEEM SI according to a *benchmarking* methodology, allowing us to identify the best and worst practices for every indicator, but also to give an appraisal of the relative distance to the relevant sustainable target. As there are many potential normalization techniques (see e.g., OECD JRC, 2008), we also discuss why *benchmarking* is the most appropriate method when one aims to measure sustainability.

The first part of the paper reviews the theoretical features of the Choquet integral (Choquet, 1953) as an aggregation methodology in the sustainability context, pointing out useful properties and features. Then, we provide the detailed stages of the aggregation framework for the FEEM SI. After undertaking all stages of the aggregation, we provide the sustainability scores for countries and macro-regions and conduct robustness analysis. Finally, we discuss the possible extensions of the current study.

2. Multi-attribute aggregation

Sustainability evaluation is a multi-attribute problem: it is characterized by many different components (called *criteria*) that can interact with each other (Munda, 1997, 2005; Ulengin et al., 2001). The literature suggests several approaches to deal with multi-attribute problems, each characterized by specific mathematical properties, which have very different implications. In this section, we briefly review possible aggregation options; we provide elements to explain why sustainability cannot be fully addressed by some of them and argue why the Choquet integral is better suited for the aggregation of sustainability indicators.

Vincke (1999) classifies the multi-attribute aggregation approaches into three categories: *multi-attribute value theory* (MAVT), *outranking* approaches, and *interactive* approaches. In general, the MAVT methods – the most widely used in multi-attribute problems – use an aggregation algorithm to compute a score for each alternative under consideration (Klement et al., 2000). A MAVT method is characterized by two subsequent phases. In the first, all the criteria are normalized in a common scale, usually [0,1], in order to allow direct comparisons.[1] In the second phase, the normalized values are aggregated using an aggregation operator: a monotonic function, where, ceteris paribus, "more" is preferred to "less".

A broadly used MAVT-based aggregation technique is the weighted average, which relies on the arithmetic weighted average of the (normalized) indicator values (e.g., Human Development Index, HDI, until its 2010 release, and ESI, 2005). The most common case is the one where the weights are the same for all the criteria. Despite the fact that this method is simple and intuitive, the linearity of the aggregation function implies constant substitutability among the criteria – which is not an appropriate assumption, in the context of sustainability – and could lead to double counting (Nardo et al., 2005). Weighted average should be taken with particular care since it assumes no interactions among the indicators – also an unlikely characteristic in the context of sustainability (Munda, 2005). Consider for instance two criteria, which are used to construct a composite index by weighted average. By assigning equal weights and considering linear aggregation, two countries may be assigned the same outcomes even though one achieves, say, a high level in one criterion and a poor level in another one, whereas the other country achieves moderate levels in both criteria. If one were to consider possible interaction among these criteria, these two countries should not be assigned same outcomes. For example, if there is a positive interaction between these two criteria, then one should rank the latter country before the former. In other words, having a balanced achievement of criteria should be considered better than that of a good achievement in one criterion and a bad one in another (see Dale et al., 2013 on how complementary indicators are treated within sustainability context). On the other hand, if criteria are perfect substitutes, then one should rank the former before the latter (see McGillivray, 1991; McGillivray and White, 1993 for redundancy problem of linear composite indices when indicators are highly correlated). Nevertheless, the general tendency is to use the equally weighted indices because of a lack of data on the relative importance of different criteria; it is also considered the most transparent way of producing aggregate indices (ESI, 2005).

Sustainability indices should allow for redundancies and positive interactions among different criteria, meaning that the *compensative* assumption (i.e., *the mutual preferential independence axiom* in Marichal, 2000) will not be satisfied, which implies that the weighted sum operator is no longer applicable (see also Marichal and Roubens, 2000). The literature presents several operators that allow for interaction among criteria (Yager, 1993; Marichal, 1998; Klement et al., 2000; Grabisch et al., 2009). For the FEEM SI we adopted one such operator, the Choquet integral, which allows a

[1] Where 0 represents "very bad" outcome, 1 represents "very good" outcome for any indicator which is increasing in a social 'good'.

preference-based construction of a sustainability index. This aggregation operator – which can be seen as a versatile extension of the weighted sum – is mathematically well-characterized (see Marichal, 2000) and allows for a straightforward specification of a preference structure with no loss of generality. Specifically, it allows us to obtain the preferences of experts based on their personal assessments about the sustainability of different hypothetical societies. In a recent application of the Choquet integral in a multi-attribute assessment of well-being, Meyer and Ponthière (2011) applied such a methodology and found the existence of complementarities and redundancies between dimensions of standard of living, as well as identifying a strong heterogeneity in individual preferences.

In this paper, we apply the Choquet integral-based multi-attribute theory to every node of an *aggregation tree* (see Section 3.1 for the construction of the aggregation tree) on the issue of sustainability. The same procedure is repeated for every country and every year, thus obtaining sustainability outcomes and country rankings, accounting for the individual preferences of experts over the definition of sustainability. In the next section, we introduce the formal definition of the Choquet integral, its components and its main properties, in order to show how it applies to the issue at hand.

2.1. Choquet integral as an aggregation operator

This section presents the formal definitions pertaining to the Choquet integral, which is a function of several criteria and is characterized by a set of parameters, called *monotonic measures*. This aggregation operator is applied to each of the aggregation node of the aggregation tree presented in Section 3.1, in a bottom-up sequence. At every *aggregation node* of the tree, the sub-nodes represent the criteria under consideration. For any given node, starting from those at the lowest layer (the *sustainability indicators*) we aggregate the criteria node-wise into a *composite index* for the respective node, which is then taken as a criterion of a hierarchically superior node. This bottom-up aggregation algorithm continues until the final node (the FEEM SI composite index) is evaluated. We thus use the term "*criteria*" to denote both the normalized indicator values and the values of the intermediate nodes of the tree, which themselves aggregate indices. The monotonic measures, for their part, are used to represent the preferences of the experts at any given aggregation node: they represent the valuation given by the expert to the fulfilment of every criterion in isolation, as well as valuations for every coalition of criteria. These monotonic measures are obtained by the procedure of expert elicitation presented in Section 3.3 and are formally defined as follows:

Definition 1. Let $N = \{1, 2, \ldots, n\}$ be the set of normalized criteria. A *monotonic measure* (also referred simply as "measure" in what follows) is a set function $m : S \subseteq N \to [0, 1]$, which satisfies:

(i) $m(\varnothing) = 0$, $m(N) = 1$,
(ii) $\forall S, T \subseteq N : S \subseteq T \Rightarrow m(S) \leq m(T)$.

The *measure* of any subset of criteria can be interpreted as the "importance" of that subset, when all criteria in the set are fully satisfied (i.e., when their *normalized values* are 1) and all the criteria that do not belong to the subset are not satisfied (their *normalized values* are 0). The first two constraints are the two *border conditions*. The second constraint represents the monotonicity condition, a typical requirement for any aggregation operator.

Given that the measure needs to be defined for every subset, if n is the number of the criteria, the specification of a set of monotonic measures required to map each of the 2^n possible subsets of criteria into a value [0,1], while the weighted average only requires allocation of n weights.[2] To obtain these measures, we conducted a questionnaire where the experts were asked to evaluate all possible combinations of criteria at any node of the aggregation tree (we provide the details of the expert elicitation and their preferences about the measures in Section 3.3). Nevertheless, the complexity of having too many subsets can be overcome by limiting the number of criteria in each node of the decision tree to a small number. Therefore, we designed the aggregation tree in such a way that every aggregation node features at most three criteria.

The issue of complementarity or substitutability among criteria is reflected in the measures as follows: for all disjoint subsets of $S, T \subseteq N$, (i.e. $S \cap T = \varnothing$) a measure is said to be *additive*, *sub-additive (redundant)* or *super-additive (synergic)* if $m(S \cup T) = m(S) + m(T)$, $m(S \cup T) < m(S) + m(T)$, and $m(S \cup T) > m(S) + m(T)$, respectively.

This specification allows for great generality in the expression of preferences in a multi-attribute problem. For instance, consider the case of two hypothetical criteria: longevity and standard of living. If an expert deemed that the achievement of good outcomes in both criteria at the same time (importance of the pair) was independent from achieving good outcomes in each criterion separately (importance of criterion longevity and standard of living), the evaluation of criteria would be considered to be additive and the importance of the pair would be obtained by adding the importance of each criterion. However, if an expert deemed that a society is better when good outcomes in standard of living and longevity achieved simultaneously, the importance of the pair would be higher than the sum of individual components. Similarly, if an expert considered longevity and standard of living as substitutes, the importance of the pair would be less than the sum of importance of each criterion alone (see Marichal, 2000 for detailed discussion).

Definition 2. To every set of monotonic measures $\{m(S)\}$, we can associate univocally its Möbius transform, defined as:

$$\alpha_m(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} m(T) \quad \text{for} \quad \forall S \subseteq N$$

The argument of the Choquet integral function – aside from the monotonic measures that parameterize it – is the value of the criteria. We now describe the structure of the Choquet integral with respect to this. Let $(x_1, x_2, \ldots, x_n)$ be the values of the criteria[3] and $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ be the ordered vector obtained from $(x_1, x_2, \ldots, x_n)$ by a suitable permutation of indices, so that $0 = x_{(0)} \leq x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ and $x_{(i)} \in (x_{(0)}, x_{(1)}, \ldots, x_{(n)})$ for $\forall i = 1, 2, \ldots, n$.

Definition 3. The Choquet integral of vector criteria $(x_1, x_2, \ldots, x_n)$, $x_i \in [0, 1]$ with respect to the measure of the vector of criteria $m : S \subseteq N \to [0, 1]$ is given by:

$$C_m(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} (x_{(i)} - x_{(i-1)}) \cdot m(A_{(i)})$$

where $A_{(i)} = \{i, i+1, \ldots, n\}$, is the set of indices successive to i and $A_{(n+1)} = \varnothing$.

[2] To be exact, there are $2^n - 2$ required parameters since the "border" conditions are already predetermined in which the empty set is null and the universal set is one.

[3] In the case of the lower-level aggregation nodes, the criteria are the *normalised* indicators. For any hierarchically superior node, the criteria are the values of the aggregate indices in the sub-nodes, which are already in the scale [0,1] by construction.

The Choquet integral can be alternatively computed using the Möbius transform ($\alpha_m(S)$) of the measure, as follows:

$$C_m(x_1, x_2, \ldots, x_n) = \sum_{T \subseteq N} \alpha_m(T) \cdot \min_{i \in T}\{x_i\},$$

where $\alpha_m(S)$ are the Möbius coefficients associated to the measure m (see Grabisch et al., 2009).

The flexibility of the Choquet aggregation methodology allows us to achieve not only a summarized measure of sustainability, but also to look in further details at the interaction of the different elements of the aggregation methodology, such as addressing the (relative) importance of criteria. In the next subsections, we provide some of the properties of the Choquet integral, which may be helpful in understanding the preferences of experts in the next sections when they evaluate all subsets of the criteria.

2.2. Relative importance of indicators (Shapley value)

The Shapley value (Shapley, 1953) characterizes the "relative importance" of each criterion under consideration, from the set of monotonic measure. For a given criterion i, the Shapley value is calculated by comparing the value of the measure of every set including i with every set that does not include it (thus obtaining "marginal gains" of i) and averaging the results (Grabisch, 1995, 1996).

For the ith criterion, the Shapley value is calculated as follows:

$$v(i) = \sum_{T \subseteq N \setminus i} \frac{(n-t-1)!\,t!}{n!} \left[m(T \cup i) - m(T)\right],$$

$$\text{where} \quad t = card(T) \quad \text{and} \quad \sum_{i=1}^{n} v(i) = 1.$$

It is possible to verify that the Shapley values vary between 0 and 1, and the higher value the higher the importance of that indicator.

2.3. Orness and andness indices

One can further characterize the measures provided by the experts by calculating the *andness* or *orness* indices. Consider a node consisting of the criteria $\{x_1, x_2, x_3\}$. The corresponding vector of measures is:

$$[m(\varnothing), m(x_1), m(x_2), m(x_3), m(x_1, x_2), m(x_1, x_3), m(x_2, x_3), m(x_1, x_2, x_3)].$$

When the vector of measures is (0, 1, 1, 1, 1, 1, 1, 1), the expert allocates measure 1 to all combinations of satisfied criteria except for the empty set. In this case, the Choquet aggregation operator corresponds to the maximum operator and the implicit expert behaviour is said to be *fully compensative* (i.e., the criteria are *perfect substitutes* to each other) and the Choquet integral of the criteria will be the maximum of the criteria. Conversely, when the vector of measures is (0, 0, 0, 0, 0, 0, 0, 1), the expert behaviour is said to be *fully non-compensative* and the Choquet integral corresponds to the minimum operator (i.e., the criteria are *perfect complements*).

The above conditions are extreme cases and one can characterize whether the expert follows a more "pessimistic" or "optimistic" behaviour in each respective case by computing two indices, the *orness* or *andness* indices, depending solely on the values of the monotonic measures provided by the expert. The *orness* index measures the extent to which the expert's preferences – represented by monotonic measures provided – allow criteria to compensate each other, while the second one, measures the extent to which the expert considers the criteria to be non-compensative. The *orness* index is computed from the Möbius transform of the measures as[4]:

$$ORNESS_m(i) = \frac{1}{n-1} \sum_{T \subseteq N} \frac{n-t}{t+1} \alpha(T)$$

If *orness* = 1, then the expert's measures are *fully compensative* (i.e., criteria are perfect substitutes of each other). Whereas, if *orness* = 0 (i.e. *andness* = 1 since *orness* + *andness* = 1), then the expert's preferences about criteria suggest that they are *perfect complements*, and if *orness* = 0.5 the expert has additive preferences on average.

2.4. Interaction among criteria

In order to further characterize the expert's preferences, we can look at the way in which the monotonic measures suggest a degree of "interaction" among criteria, which may affect policymaking or the general understanding of the problem at hand. This can be measured with the *interaction index*, which is computed in a similar way as the Shapley index, considering the joint contribution of indicators – as opposed to considering the individual contributions. Let us consider two criteria, i and j: if $m(i, j) > m(i) + m(j)$, then the monotonic measure shows a complementary effect between i and j. Similarly, if $(i, j) < m(i) + m(j)$, then i and j are deemed to be substitutive. Finally, if $m(i, j) = m(i) + m(j)$, the criteria i and j do not interact, that is, they are deemed to be independent.

To measure all possible interactions of two criteria with the remaining ones, the average interaction between two criteria i and j is calculated with the *interaction index* (see Murofushi and Soneda, 1993) defined as follows:

$$I_m(ij) = \sum_{T \subseteq N \setminus ij} \frac{(n-t-2)!\,t!}{(n-1)!} \left[m(T \cup ij) - m(T \cup i) - m(T \cup j) + m(T)\right]$$

$$\text{where} \quad t = card(T)$$

The value of $I_m(ij)$ can be considered as a measure of the average marginal interaction between i and j. One of the important properties of the interaction index is that $I_m(ij) \in [-1, 1]$ for all i and j. The interaction index being 1 (−1) represents perfect complementarity (substitutability) between i and j (see Grabisch, 1997).

The wide range of interactions allowed by the measures implies that the Choquet integral is able to represent all sorts of interactions among criteria. Given the nature of the problem at hand, we believe that experts evaluating the issue of sustainability might be more inclined towards *andness* and more complementary-oriented behaviour among indicators (i.e., positive interaction indices), as complementary implicitly requires a balanced development across its different components (see Munda, 2005 for discussion on non-compensative nature of sustainability indicators). This can be evaluated after conducting the expert elicitation and collecting the relevant monotonic measures on all the nodes of the aggregation tree. Thus, the Choquet integral approach not only allows us to obtain final sustainability values but also gives some insight on how the experts consider sustainability, which in itself gives insight on policymaking.

[4] The *andness index* can be computed also using the measure values, but the computation is more complicated, and it is not reported.

3. Conceptualizing sustainability: the FEEM sustainability index

In this section we introduce the aggregation methodology developed for the FEEM SI, an aggregate sustainability index characterized by 19 indicators belonging to the three pillars of sustainability (i.e., economic, social and environmental). The construction of the FEEM SI is characterized by an indicator selection phase; followed by a normalization procedure, in order to bring all the selected indicators to a comparable numerical scale; and the aggregation phase. Prior to the aggregation of the normalized indicators, we conducted a questionnaire where the experts were asked to evaluate the chosen criteria and their interactions, from which we derived the relevant monotonic measures. Since all experts' evaluations differ from each other, we computed a set of "consensus" monotonic measures to provide the necessary parameters for every node of the aggregation tree. Finally, by using the normalized indicators and the consensus measures, we employ the Choquet integral to aggregate the sustainability indicators, in order to arrive to sustainability outcomes. Figure outlines the stages to obtain final sustainability outcomes where these sections will be discussed in the next sections (Fig. 1).

3.1. Indicator selection

The FEEM SI indicators have been chosen after a thorough review of the literature on sustainability assessment in order to address its relevant components and offer insights to the main policy questions regarding sustainability assessment and management. Table 1 reports an overview of the main sources used to draw information for the indicator selection stage and the list of selected indicators (please refer to Carraro et al., 2012 for further details and see Table 1.3 of the FEEM, 2011 for the detailed explanation and literature reference of each indicator).

The indicators are constructed within a recursive-dynamic general equilibrium model ICES-SI (see Eboli et al., 2010; Carraro et al., 2012) which produces future projections of all indicators in the time frame 2011–2020 that can be used in comparative static policy analysis and provides useful policy implications (Bohringer and Loschel, 2006). Since the FEEM SI deals with global sustainability, the subjects of this sustainability analysis are countries and macro regions.

Indicators are organized in a decision tree along the three main pillars of sustainability in nodes including either two or three indicators. On one side, this allows it to address sustainability not only at aggregate level, but also in a theme-based fashion, coherently with the UN Commission on Sustainable Development (UN CSD, 2005) and the European Union Sustainable Development Strategy 2006 (EU SDS, 2006), which offer a theme-based indicator set for assessing the sustainability levels. Furthermore, Environmental Performance Index (EPI) includes 25 performance indicators tracked across ten policy categories covering both environmental

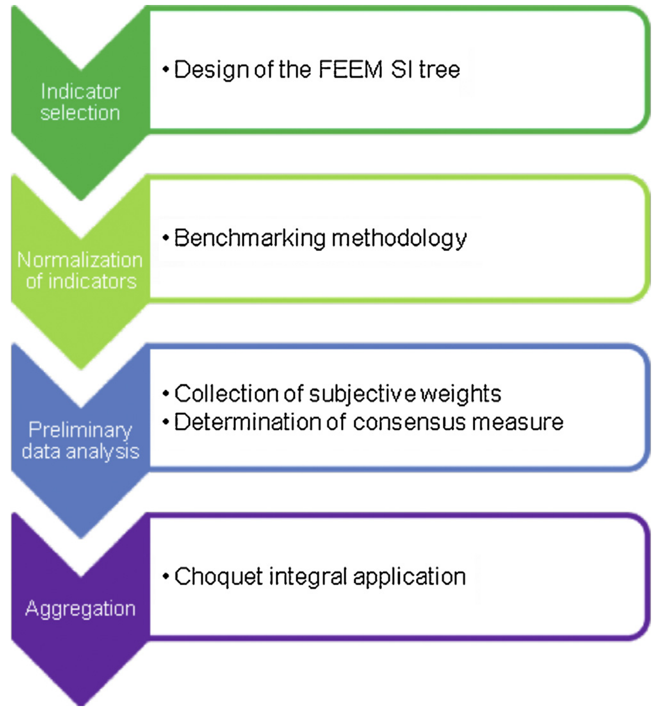

Fig. 1. Stages to obtain sustainability outcomes of the FEEM SI.

Table 1
Indicators and their description.

Dimension	Node	Indicators		Description
	Growth drivers	1. R&D		R&D Expenditure/GDP
		2. Investment		Net investment/capital stock
Economic dimension	GDP per capita	3. GDP per capita		GDP/population
	Exposure	4. Relative trade balance		Net exports/(exports + imports)
		5. Public debt		Government debt/GDP
	Population density	6. Population density		Total population/total area of the country (in kilometre square)
	Wellbeing	7. Education		Expenditure on education/GDP
Social dimension		8. Health		Expenditure on health/GDP
	Vulnerability	9. Food relevance		Total food expenditure/total expenditure
		10. Private health		Private health expenditure/GDP
		Energy security	11. Energy imported	Energy imported/total energy consumption
			12. Energy Access	Population that has access to electricity/total population
	Air pollution	13. GHG per capita		Total GHG emissions/total population
		14. CO_2 intensity		Total CO_2 emissions/total primary energy consumption
Environmental dimension	**Energy**	15. Energy intensity		Total primary energy supply/GDP
		16. Renewables		Renewable energy consumption/total primary energy consumption
	Natural endowment	17. Water		Total water use/total renewable water resources available
		Biodiversity	18. Animals	Endangered Species/total species
			19. Plants	Endangered Species/total species

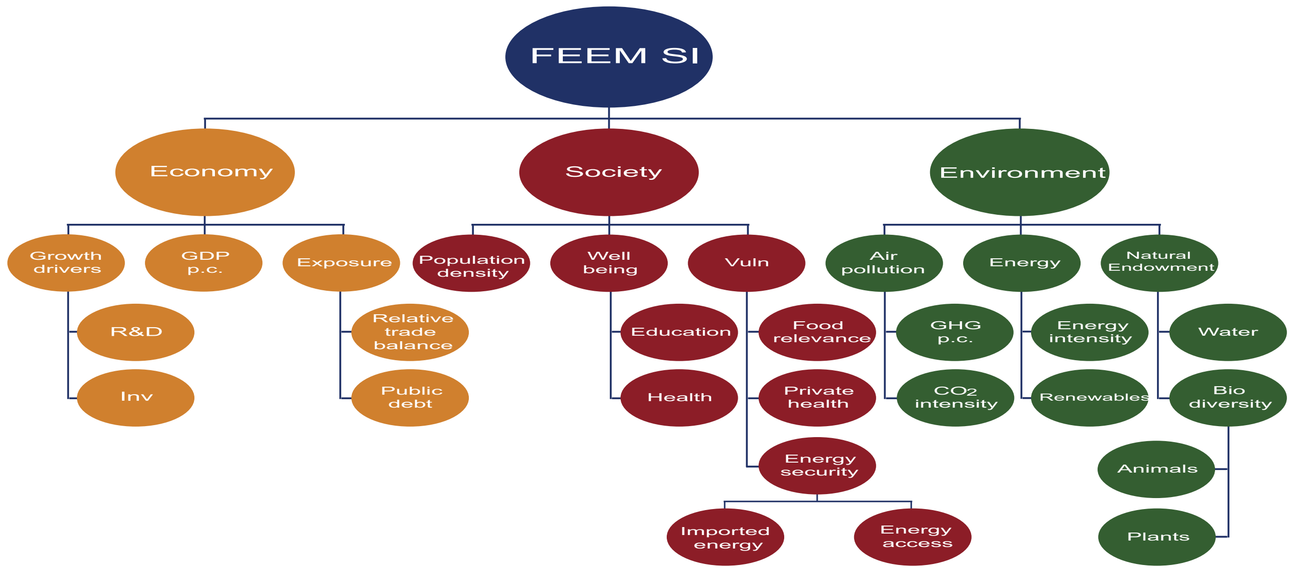

Fig. 2. FEEM SI aggregation tree.

public health and ecosystem vitality assessing the closeness of a country to environmental policy goals (EPI, 2010). More importantly, the organization of the aggregation tree in nodes of two or three indicators allows it to exploit the potential of subjective weights, elicited from a set of experts, while avoiding the exponentially increasing complexity arising from the implementation of Choquet integral aggregation method with too many indicators at each aggregation node. The indicators selected for the FEEM SI introduced earlier have been organized into a decision tree, in which partial aggregation take place at different levels, leading to the hierarchical decomposition of Fig. 2.

The decision tree should be read from bottom (leaves) to top (final node) and is characterized by three successive decomposition levels. The tree respects the three main pillar structure that is common in most sustainability studies (see e.g., UN CSD, 2005; Global Reporting Initiative (GRI), 2010; Krajnc and Glavic, 2005), with the final node producing the aggregate index.

3.2. Normalization procedure

Since each indicator not only is measured in different metrics but also different levels for each indicator represents different sustainability outcomes, we first normalize each indicator, using a *policy-oriented benchmarking technique* developed for all the indicators of the FEEM SI prior to the aggregation of all indicators into a single composite index.

According to the OECD's *Handbook on constructing composite indicators* (OECD JRC, 2008), "normalization is required prior to any data aggregation as the indicators in a data set often have different measurement units". The normalization approach taken for all indicators in the FEEM SI is the *benchmarking*. Such method starts by individuating a best practice or target level to be used as a benchmark and is very appropriate especially in the case of those indicators for which an agreed target (at EU or global level, for instance) exists. The *benchmarking procedure* normally assigns only two values, 1 and 0, according to whether a given indicator meets a chosen reference level or not. This method allows comparison through time and across countries, whilst supplying a policy-based normalization, which is particularly suitable for the construction of the FEEM SI.

Since the purpose of creating a sustainability index is not only to identify best and worst practices, but also to give an appraisal of the relative distance to the sustainable target, all of the sustainability indicators (see Table 1 for the list of all indicators) are normalized according to a *benchmark function*. In this normalization procedure, each sustainability indicator passes through its respective five reference levels that correspond to a normalized value comprised between 1 and 0 as presented in Table 2.

Table 2
Normalization of sustainability indicators.

Normalized value	Sustainability level
0	Extremely unsustainable
0.25	Still not sustainable but not as severely as in the previous case
0.50	Discrete level of sustainability, but still far from target
0.75	Satisfactory level of sustainability, yet not on target
1	Fully sustainable

The normalized values for each indicator correspond to specific levels of the indicator in the original measurement unit (see Section 3.3 of FEEM, 2011 for benchmarks of each sustainability indicator). Such levels are defined according to reliable and authoritative literature and international legislation sources to increase the acceptability of the methodology. Whenever possible, the objectives outlined in the EU Sustainable Development Strategy or in the Europe 2020 (a follow-up of the Lisbon Strategy) have been used to define one or more level of the benchmarking function. In all other cases, broader EU policy objectives and international standards from established institutions such as the OECD, World Bank, UN, and International Monetary Fund have been taken as primary source of information.[5]

The *benchmarking technique*, used to normalize sustainability indicators, appears to be a better choice than commonly used alternatives for the construction of the FEEM SI (see OECD JRC, 2008 for detailed normalization techniques). Unlike many alternatives, it is based on exogenous sustainability benchmarks: therefore, a change in the normalized values always corresponds to an improvement in sustainability (unlike the *percentage differences* technique). Moreover, it does not relate the value of sustainability of a given country to its position in the indicator ranking (unlike

[5] A more thorough discussion on the specific benchmarking functions defined for each indicator can be found in Carraro et al. (2012).

Table 3
Construction of indicator-coalition matrix.

Economic	Social	Environmental	Measure
Worst	Worst	Worst	0
Best	Worst	Worst	20
Worst	Best	Worst	50
Worst	Worst	Best	30
Best	Best	Worst	$X \geq 50$
Best	Worst	Best	$X \geq 30$
Worst	Best	Best	$X \geq 50$
Best	Best	Best	100

the *ranking* or *categorical scales* techniques), or its position in the sample distribution (unlike the *standardization* technique). Thus, the benchmarking technique used for all FEEM SI indicators allows for meaningful, robust, cross-sectional and time comparison of countries for every indicator.

3.3. Expert elicitation

In order to obtain the measures that are necessary for the aggregation, we prepared a paper-based questionnaire, which includes a decision matrix for each one of the thirteen decomposition nodes of the aggregation tree. The questionnaire represents a list of the possible scenarios with two defined qualitative levels of the criteria – i.e. all the combinations of "best" and "worst" values – drawing from Despic and Simonovic (2000). If n is the number of criteria in the node under consideration of the aggregation tree, the decision matrix will then have 2^n rows, thus requiring the same number of evaluations by the expert.

The experts had to provide numerical valuations (the *monotonic measures* presented in Section 2.1, multiplied by 100) for each row of the decision matrix. This was done for all 13 aggregation nodes, by choosing a value between 0 and 100 for each row, except for the first and the last, where indicators are respectively all "worst" and all "best" and are given 0 and 100 by default. Moreover, the measures given at each row of every matrix need to respect the monotonicity criterion. This implies that, if a combination where only one indicator is at its "best" is given a certain measure x, all combinations including that indicator in the "best" case should be given a weight at least equal to x, as Table 3 shows.

Measures for the aggregation methodology have been collected in a pilot study, which was implemented using the QUALTRICS software.[6] Participants that are from different background, expertise and profession were contacted and asked to fill the questionnaire between the beginning of May 2011 and the end of July 2011. Besides responding to the questionnaire, respondents were asked to provide their names, type of institution they work and the country of residence. At the initial stage of the questionnaire, we introduced control questionnaires that were designed to explore whether the respondent understood the monotonicity rule as explained above.

Overall, 20 experts participated in the questionnaire and fulfilled the monotonicity axiom as required. An overview of the expert pool used to collect the necessary data for the aggregation methodology is provided in Fig. 3. Experts have a different geographical location (Europe, USA and Asia) and different backgrounds: 40% of the experts are affiliated to academia, another 40% of the experts are affiliated to international organizations and the remaining 20% of the experts are part of a think tank organization.

[6] Qualtrics is a private research software which enables one to build web-base surveys which is easy to distribute and allow worldwide participation, through providing secure online access to surveys.

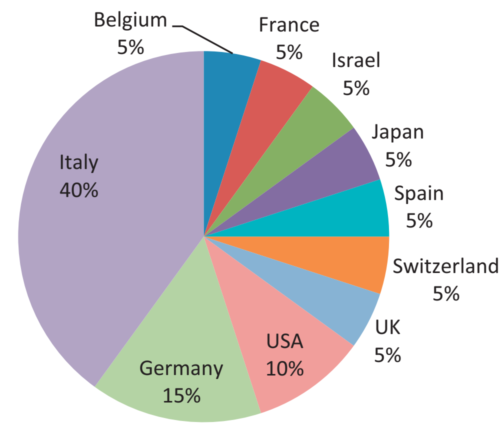

Fig. 3. Overview of the expert pool.

After conducting a control questionnaire stage, experts were provided the FEEM SI tree and were aware how the indicators are allocated within the tree. Therefore, the experts were made aware about the allocation of indicators, therefore were aware of the hierarchical ordering of the indicators and might give higher measures to a given node to increase the importance assigned to indicators that are located at lower leaves of the tree.

3.4. Consensus measures

In order to obtain the single set of monotonic measures needed for the aggregation of indicators, the information collected from all experts had to be further processed to derive a set of "consensus" measures for all criteria and their coalitions at every node of the aggregation tree. Since the mathematical properties of monotonic measures are preserved under linear aggregation operators, there may be several procedures to combine the preferences (i.e., expert valuations, expressed in measures) of different experts, the simplest being an arithmetic average of their elicited measures. Given the level of complexity of the issue of sustainability and the difference in background of the experts involved, however, one may argue that an aggregated valuation should take account of the level of "agreement" among experts. This makes the reference measures resistant to isolated and drastically dissenting expert valuations. This mitigates the bias that may potentially result from the selection of a sample of experts, while including all of their preferences. The choice of using consensus measures that do not treat all expert opinions equally, in the case of idiosyncratic valuations, is in line with Budnitz et al. (1997). Furthermore, in a recent study, Etminani et al. (2013) suggest that the idea of using equally weighted averages of preferences to arrive to a consensus, as preferences might include outliers and that of giving equal weights to each expert is a risky choice. Therefore, rather than determining the weight that is assigned to outlier expert subjectively, we derive a mathematical consensus metric which implicitly assigns lower weight to outlier experts.

For the FEEM SI, we determined a set of *consensus monotonic measures* among respondents, using the pairwise metric distance among each of the measures provided by the experts. This rewards valuations in agreement with one another (i.e., having lower distance measure) and penalizes the ones that differ sizably from every other (i.e., having a higher pairwise distance measure). By doing so, we obtain consensus measures for each sustainability indicator and their coalitions at every node of FEEM SI. In Section 4, we test the robustness of this aggregation procedure by considering different combinations of expert valuations.

The consensus measures are obtained as follows: let m_{ki} is the measure provided by kth expert for the ith coalition of criteria at a given node of the aggregation tree. One can calculate the total absolute distance of kth expert's measures to all other experts' measures, D_k, as follows:

$$D_k = \sum_{i=1}^{i=2^n} \sum_{l=1}^{l=e} \left| m_{ki} - m_{li} \right|$$

where $l \neq k$, and $l = 1, 2, \ldots, e$ are the experts and n is the number of criteria at a given node of the aggregation tree. After calculating the absolute distance for each expert, the sum of absolute distances of all experts is defined as:

$$\bar{D} = \sum_{k=1}^{k=e} D_k$$

These two distances are then combined to compute the weight attached to the expert's valuation for the purposes of aggregating the expert's measures, as follows:

$$W_k = \left(\frac{D_k}{\bar{D}} \right)^{-1}$$

These are then normalized, so they are bounded in [0, 1]:

$$w_k = \frac{W_k}{\sum_{k=e}^{k=1} W_k}, \quad \text{where} \quad \sum_{k=1}^{k=e} w_k = 1$$

The set of *consensus monotonic measures*, m_i^c, for all possible criteria at a given node are calculated by the weighted average of experts' measures as:

$$m_i^c = \sum_{k=1}^{k=e} w_k m_{ki} \quad \text{for} \quad \forall i, \quad \text{where} \quad i = 1, 2, \ldots, 2^n.$$

After obtaining consensus measures for each coalition of indicators at each node of the FEEM SI tree, the Choquet integral is used to aggregate all indicators to an overall index where the aggregation takes place at different stages starting from bottom nodes and ending at the final node.

3.5. Characteristic of consensus measures

In this section, we first offer the consensus measures for each sustainability indicator and for their interactions. The right and left panels of Table 4 offer the consensus measures for nodes that have 3 and 2 indicators respectively. As described in the previous sections, when a node has 3 criteria, there are 8 possible combinations of criteria; the first column of the table provides all possible combinations of criteria at a given node. For example, the FEEM SI node has economic, social and environmental pillars as criteria and every combination of them is given. Similar sets of combinations are given for the nodes with 2-indicators at the right panel of the table.

The characteristics of consensus measures provide some insight on the experts' preferences and on their propensity to consider criteria as complements or substitutes as well as their relative importance. We will highlight the important characteristics of consensus measures in this section and give insights about how the majority of the experts consider the sustainability. Panels A and B of Table 5 represent the *andness* degree and the interaction indices among indicators at each node for the consensus expert measures. Panel A presents the *andness* degrees and interaction indices for the three-indicator nodes and panel B offers the same information for the two-indicator nodes.

The consensus expert measures suggest a tendency of being more *andness*-oriented, showing positive interaction behaviour towards sustainability indicators in all nodes with the exception of the final node of FEEM SI. This suggests that the "consensus expert" prizes balanced values in sustainability indicators more than unbalanced ones.

Let us say that we have two indicators: indicator A and B, and two countries, country X and Y, which have normalized criteria $x_1 = (0.5, 0.5)$ and $x_2 = (0.6, 0.4)$. Let us consider three possible measure allocations: $m_1 = (0, 0.3, 0.3, 1)$, $m_2 = (0, 0.5, 0.5, 1)$, and $m_3 = (0, 0.7, 0.7, 1)$. The measures represent the valuations given by the expert when two the criteria are at their "worst" levels; when first criterion is at its "best" and second is at its "worst"; when the first criterion is at its "worst" and second is at its "best"; and when both criteria are at their "best" levels. The *andness* measures for m_1, m_2, and m_3 are 0.7, 0.5, and 0.3, respectively. Using the criteria $\{x_1, x_2\}$ as arguments of the Choquet integral and the alternating the measures m_1, m_2 and m_3 as its sets of parameters, country X's aggregated index would be (0.5, 0.5, 0.5), for each of the corresponding sets of measures, whereas country Y's would be (0.46, 0.5, 0.54), respectively.

As one can see when *andness* measure is the highest (0.7 for the first measure allocation), this type of choice of measures gives higher outcomes for countries which have balanced criteria values (i.e., 0.5 and 0.46 for country X and Y respectively). Whereas, if the measures are additive (i.e. *andness* score is 0.5 and the aggregation pins down to weighted average), there exists no interaction among indicators and both countries' valuation is the same. Finally, when the *andness* score is less than 0.5 (i.e. the third measure allocation), the indicators are treated as compensative. In this case, a country having a higher value in one dimension would benefit a higher outcome (i.e. country Y would have a higher outcome than country X). Similarly, interaction indices could be considered in the same way. A positive interaction index between two criteria would suggest that the aggregation would favour countries with balanced values in both criteria. Negative interaction index would suggest that it is not necessarily important how balanced the criteria values are to achieve a higher outcome. The effect would obviously change depending on how strong the *andness* and interaction index is between these indicators.

The final node of FEEM SI has an *andness* degree of 0.493, which represents a slightly *compensative* attitude towards the final node of the FEEM SI tree. Moreover, interaction index between economic and social pillars (environmental and social pillars) is −0.024 (−0.019), which suggests that the consensus expert evaluates those interactions as slightly *competitive* (or substitutes). In other words, in general terms, majority of the experts considers that economic deterioration could be substituted with a better outcome in social pillar (and vice versa). On the other hand, the interaction index between the economic and environmental pillar is 0.020, and the consensus expert evaluates those pillars as slightly complementary indicators. In this case, experts consider a country more sustainable when balanced values are obtained in both economic and environment pillars. For the remaining nodes, the reference expert features an *andness* index that is greater than 0.5 (i.e. more *non-compensative* attitude towards the nodes) and a positive interaction index value between two indicators at a given node (i.e. two indicator being more *complementary*). In other words, to have higher sustainability outcomes, countries need to have values that are more balanced. Overall, these results are in line with the current literature, which assumes a non-compensative set of aggregation operators for sustainability indices (Munda, 2005, 2012; Munda and Nardo, 2009).

Moreover, given considering the consensus measures in the context of the aggregation tree, it is possible to determine the *implied relative importance* of each indicator and node. This can be achieved by computing the *Shapley values* (presented in Section 2.1) of the criteria at any given node, as computed in Table 6.

These values reflect the *local* relative importance of criteria at every level of the aggregation tree. For instance, it is possible to see

Table 4
Consensus measures from expert elicitation for all nodes.

3-Indicator nodes Indicators and their combinations		2-Indicator nodes Indicators and their combinations	
FEEM SI node	**Consensus measures**	**Growth drivers node**	**Consensus measures**
{∅}	0	{∅}	0
Economic	0.358	R&D	0.493
Social	0.361	Capital accumulation	0.449
Environmental	0.376	R&D, capital accumulation	1
Economic, social	0.623	**Exposure node**	
Economic, environmental	0.681	{∅}	0
Social, environmental	0.646	Relative trade	0.46
Economic, social, environmental	1	Public debt	0.353
Economic pillar		Relative trade, public debt	1
{∅}	0	**Well-being node**	
Growth drivers	0.369	{∅}	0
GDP pc	0.339	Education	0.493
Exposure	0.260	Health	0.478
Growth drivers, GDP pc	0.673	Education, health	1
Growth drivers, exposure	0.574	**Energy security node**	
GDP pc, exposure	0.558	{∅}	0
GDP pc, growth drivers, exposure	1	Imported energy	0.500
Social pillar		Energy access	0.500
{∅}	0	Imported energy, energy access	1
Pop. density	0.258	**Air pollution node**	
Well being	0.430	{∅}	0
Vulnerability	0.333	GHG p.c.	0.428
Pop. density, well being	0.607	CO_2 intensity	0.389
Pop. density, vulnerability	0.535	GHG p.c., CO_2 intensity	1
Well being, vulnerability	0.686	**Energy use node**	
Pop. density, well being, vulnerability	1	{∅}	0
Environmental pillar		Energy intensity	0.432
{∅}	0	Renewables	0.515
Air pollution	0.369	Energy intensity, renewables	1
Energy	0.347	**Endowments node**	
Endowments	0.328	{∅}	0
Air pollution, energy	0.598	Biodiversity	0.426
Air pollution, endowments	0.595	Water	0.516
Energy, endowments	0.573	Biodiversity, water	1
Air pollution, energy, endowments	1	**Biodiversity node**	
Vulnerability node		{∅}	0
{∅}	0	Animal	0.431
Food	0.397	Plant	0.398
Private health	0.276	Animal, plant	1
Energy security	0.340		
Food, private health	0.616		
Food, energy security	0.662		
Private health, energy security	0.541		
Food, private health, energy security	1		

that at the highest node (FEEM SI), greater relative importance is given to environmental sustainability (0.352) than to economic sustainability (0.332) or social sustainability (0.316). It is also possible to combine these results in a linear fashion in order to approximate how much each indicator contributes towards the determination of the final FEEM SI values. By multiplying the Shapley values of every hierarchically superior criterion, from the bottom of the aggregation tree to the top, we are able to determine the overall importance of each indicator. For instance, the contribution of "health" is calculated by multiplying the Shapley values of "health", "well-being" and "social" pillars, since the "health" indicator is under the node of "well-being" which is a node of "social" pillar. Overall, Shapley values indicate how important each sustainability indicator is when one considers its marginal contribution to the overall index.

4. Results and robustness analysis

Given the normalized sustainability indicators and the consensus measures assigned to all possible combinations of sustainability criteria, one can aggregate the sustainability indicators to an overall index by employing the Choquet integral aggregation operator. In this section, we will present only some of the FEEM SI results in order to describe the impacts of the aggregation methodology and provide examples of the importance of such methods in evaluating policy choices; for a more complete overview of the FEEM SI results please refer to the material available online.[7]

4.1. FEEM SI results using the Choquet integral as an aggregation operator

The hierarchical structure used to construct the FEEM SI allows us to obtain the sustainability ranking for each year of analysis, including future projections of the sustainability levels, enlarging the scope of the analysis to policy implications. Since this paper focuses on the role of the aggregation methodology in dealing with sustainability, reported results refer only to the baseline scenario.[8]

For each year of the analysis, the aggregation tree and the consensus measures are used to determine the FEEM SI, which summarizes the overall sustainability of any country. Given the normalized sustainability indicators and consensus measures, we now can aggregate the sustainability indicators to obtain an overall sustainability outcome for each country and macro-regions. Table 7

[7] http://www.feemsi.org/

[8] For further details and policy implications please refer to Carraro et al. (2012).

Table 5
(Panel A) Interaction indices and *andness* degree at 3-indicator nodes. (Panel B) Interaction indices and *andness* degree at 2-indicator nodes.

Panel A

Node	Interaction indices			*andness* degree
FEEM SI	Economic	Social	Environmental	0.493
Economic	NA	−0.024	0.020	
Social		NA	−0.019	
Environmental			NA	
Economic	Growth drivers	GDP pc	Exposure	0.538
Growth drivers	NA	0.047	0.026	
GDP pc		NA	0.041	
Exposure			NA	
Social	Pop. Density	Well being	Vulnerability	0.525
Pop. density	NA	0.016	0.041	
Well being		NA	0.020	
Vulnerability			NA	
Environmental	Air pollution	Energy	Endowments	0.532
Air pollution	NA	0.021	0.037	
Energy		NA	0.037	
Endowments			NA	
Vulnerability	Food	Private Health	Energy Security	0.528
Food	NA	0.040	0.022	
Private health		NA	0.022	
Energy security			NA	

Panel B

Node	Indicators	Interaction index	*andness* degree
Growth drivers	R&D, capital accumulation	0.058	0.5290
Exposure	Relative trade, public debt	0.187	0.5935
Well being	Education, health	0.029	0.5145
Energy security	Imp. energy, energy access	0.000	0.5000
Air pollution	GHG p.c., CO_2 intensity	0.183	0.5915
Energy use	Energy intensity, renewables	0.053	0.5265
Endowments	Biodiversity, water	0.058	0.5290
Biodiversity	Animal, plant	0.171	0.5855

presents the overall, economic, social and environmental sustainability levels and rankings of countries in each respective index in the year 2011.[9] The first two columns are the FEEM SI and its ranking for the countries and macro-regions and the remaining columns represent the economic, social and environmental sustainability levels and their respective rankings.

One interesting aspect of this ranking is that countries that are ranked at the higher (lower) positions are the ones that have better (worse) outcomes in at least in two final pillars, respectively. For example, Norway and Sweden not only have outstanding sustainability levels in the social pillar, but also have also quite good performances both in the economic and environmental pillars. Among the lower-ranking countries, India has a poor performance in the social pillar due to high levels of population density, private health spending and lower levels of public spending in education and health sectors, whereas a moderate achievements in economic and environmental pillars. China has a moderate economic performance, but features low social and environmental sustainability outcomes. Both Rest of Asia and Indonesia have a low performance in the economic and social pillars and moderate environmental performances. On the other hand, some countries achieve good results in some pillar(s), while their remaining pillar(s) lag behind from many countries. For example, USA and Australia have better sustainability levels in economic and social pillars, but very poor levels of environmental performance. Moreover, Korea only achieves a better economic sustainability level, but has very poor performances in social and environmental aspect.

4.2. Robustness analysis

In a complex aggregation such as the one used for the FEEM SI, the preferences of the consensus expert are a key component of the procedure. Thus, in the construction of a composite index it is important to check how robust the ranking is to a change in the determination of consensus measures (Saltelli et al., 2004; Saisana et al., 2005).

There exists many ways to aggregate the measures provided by the experts: a straightforward way is to consider every experts' preferences as a point in a measure space. Then, one can perform a robustness analysis by building a linear convex combination of the measures that are provided by each expert and running a significant number of simulations. We implement the robustness analysis by generating 1000 sets of measures for each node that are necessary to aggregate all the indicators into the final FEEM SI. Each of these sets constitutes, for any practical purposes, a set of internally consistent valuations of sustainability, observationally equivalent to what is provided by experts. These sets of measure have thus been called "artificial experts" (AEs). In this particular application, each AE represents a univocal instance of consensus among "real" experts, whose measure allocations in each node have been aggregated by giving random weights to each expert's valuations, in a similar way to how the reference measures have been constructed. The measures contained in the artificial experts have been used to aggregate the indicators into the FEEM SI with the Choquet integral. The process results in a distribution of final FEEM SI for each country considered, which can then be ranked

[9] Current analysis considers individual countries (e.g., Norway) and macro-regions (e.g., Rest of Latin America). For detailed country and macro-region classification, see Table A.1.

Table 6
Relative importance of each indicator at a given node.

Node	Indicator (or node)	Shapley value
FEEM SI	Economic	0.332
	Social	0.316
	Environmental	0.352
Economic	Growth drivers	0.378
	GDP per capita	0.355
	Exposure	0.267
Social	Population density	0.254
	Well being	0.415
	Vulnerability	0.331
Environment	Air pollution	0.351
	Energy	0.330
	Natural endowment	0.319
Growth drivers	R&D	0.522
	Investment	0.478
Exposure	Relative trade balance	0.554
	National debt	0.446
Well being	Education	0.508
	Health	0.492
Vulnerability	Food relevance	0.395
	Energy security	0.275
	Private health	0.330
Energy security	Energy imported	0.500
	Energy access	0.500
Air pollution	GHG per capita	0.520
	CO_2 intensity	0.480
Energy	Energy intensity	0.458
	Renewables	0.542
Natural endowment	Biodiversity	0.455
	Water	0.545
Biodiversity	Animals	0.516
	Plants	0.484

according to the relative dominance measure, ρ (derivation of the measure described in Appendix A). The results of this simulation, on the 2011 FEEM SI data, are provided in Fig. 4.

The scatterplot displays the simulated FEEM SI values according to every artificial expert. The distribution of these is summarized by box-plots for every country. It can be seen that, within the "consensus" between experts – reflected in any AE, which results in a point of the distribution – some countries or groups of countries clearly "dominate" others in the ranking. One should be careful, however, whenever drawing inferences from this analysis, since the distributions of simulated FEEM SI values are not independent from one another. This means that analysing the ranking results

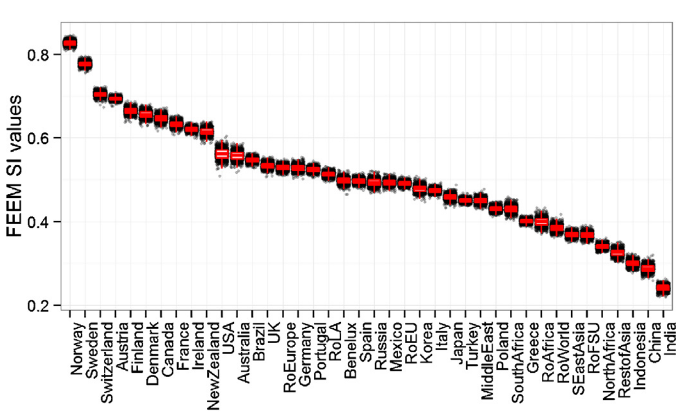

Fig. 4. Distribution of FEEM SI values according to 1000 different consensus measures among experts.

merely by comparing the features of the distributions would not fully take into account the nature of the data and could possibly lead to misleading interpretations.

In particular, whenever any two given countries have simulated distributions that partially overlap each other, this does not imply that there is an underlying ambiguity over how these two countries are ranked according to the measures provided by AEs. This result stems from the fact that measures provided by single AE (constructed from "real" expert measure allocations) contributes to determine the FEEM SI values for every country. It is therefore necessary to analyze the simulation results using a measurement that takes into account the relationship between countries across simulations. The matrix, Δ and the indices ρ^+, ρ^- and ρ are designed for this purpose (see Appendix for the derivation of these indices). These represent the average cardinal dominance; the degree to which a given country i dominates every other country; the degree to which a given country i is dominated by every other country; and the relative dominance level of the country in question, respectively. These results, obtained for those indices for each country, for the 2011 FEEM SI data, are also reported in Appendix in Table A.2.

Considering the dominance analysis together with the plot of simulated distributions unveils some interesting results. For instance, it is clear that the leading countries, Norway and Sweden, are quite set apart from the rest of the group – and from each other. In fact, Norway is never dominated by any other country across all simulations, a quite remarkable result given the variability introduced by the simulations. Norway and Sweden are both followed by a group of eight countries (Switzerland, Austria, Finland, Denmark, Canada, France, Ireland and New Zealand) that constitute a fraction of relatively high-scoring countries. These feature a consolidated ranking among themselves, as measured by the dominance index across simulations, which is stable by construction. These countries are followed by two somehow discontinuous clusters of countries (from USA to South Africa and from Greece to China) featuring a less dramatic discontinuity among clusters. In last position, India never dominates any other country across simulations.

By nature of the dominance analysis, these results tend to produce a robust ranking and illustrate the extent to which a change in "consensus" between experts can result in variability in the score of countries, thereby adding a valuable complement to the consensus measures.

5. Discussion

This paper aimed at proposing an application of the multi-attribute value theory to the sustainability literature, extending the current work in this field to address the intrinsic complexity underlying in the sustainability concept, in order to develop an aggregated sustainability index, the FEEM SI. The aggregation approach was inspired by two considerations: firstly, the non-compensative nature of the sustainability concept, fraught with inter-linkages and synergies across its different components. Secondly, the clear policy relevance of any sustainability analysis requires the subjective judgements of policy makers and relevant stakeholders in order to define a feasible plan for the implementation of a new definition of world progress. This requires managing the subjective character of the decision support tool. Combining the nonlinear Choquet operator – a novelty in the field of sustainability analysis – with robustness analysis, a well-known approach for simulation, the scoring system for sustainability assessment has been improved with respect to competing approaches. Despite the unavoidable partial uncertainty of any scoring system, the method proposed fulfils two requirements that are necessary for a rational sustainability analysis: the monotonicity and the non-compensability assumptions. Robust options are enhanced by

Table 7
Sustainability pillars: Rankings in FEEM SI, Economic, Social and Environmental Pillars in 2011.

FEEM SI rank	FEEM SI	Country	Economic	Econ. rank	Social	Social rank	Environ.	Envi. rank
1	0.823	Norway	0.752	3	0.985	1	0.718	1
2	0.774	Sweden	0.728	5	0.922	2	0.664	2
3	0.700	Switzerland	0.766	1	0.668	12	0.661	3
4	0.691	Austria	0.700	7	0.755	9	0.623	5
5	0.661	Finland	0.686	8	0.799	6	0.512	10
6	0.653	Denmark	0.663	10	0.837	4	0.469	15
7	0.641	Canada	0.566	19	0.845	3	0.499	12
8	0.630	France	0.584	15	0.789	8	0.509	11
9	0.620	Ireland	0.666	9	0.683	11	0.528	8
10	0.609	New Zealand	0.591	13	0.829	5	0.411	24
11	0.554	USA	0.725	6	0.790	7	0.210	39
12	0.553	Australia	0.737	4	0.734	10	0.251	36
13	0.546	Brazil	0.446	26	0.603	17	0.597	6
14	0.531	UK	0.577	17	0.582	19	0.451	16
15	0.529	RoEurope	0.433	28	0.519	24	0.625	4
16	0.525	Germany	0.617	11	0.618	15	0.372	30
17	0.522	Portugal	0.458	23	0.646	14	0.449	17
18	0.512	RoLA	0.392	31	0.570	20	0.585	7
19	0.497	Spain	0.575	18	0.597	18	0.347	31
20	0.495	Benelux	0.611	12	0.480	29	0.396	26
21	0.493	Russia	0.586	14	0.511	25	0.393	27
22	0.493	RoEU	0.491	21	0.499	26	0.487	13
23	0.492	Mexico	0.435	27	0.656	13	0.374	29
24	0.477	Korea	0.761	2	0.330	34	0.312	33
25	0.472	Italy	0.404	30	0.559	21	0.446	19
26	0.456	Japan	0.581	16	0.351	33	0.420	22
27	0.453	Turkey	0.417	29	0.491	27	0.448	18
28	0.450	Middle East	0.558	20	0.543	22	0.283	35
29	0.430	Poland	0.463	22	0.538	23	0.304	34
30	0.426	South Africa	0.454	25	0.612	16	0.230	38
31	0.399	Greece	0.354	34	0.439	30	0.402	25
32	0.398	RoAfrica	0.279	40	0.378	32	0.523	9
33	0.385	RoWorld	0.306	37	0.405	31	0.445	20
34	0.368	SEastAsia	0.390	32	0.261	36	0.440	21
35	0.367	RoFSU	0.386	33	0.482	28	0.244	37
36	0.342	North Africa	0.350	35	0.285	35	0.385	28
37	0.325	RoAsia	0.285	39	0.185	38	0.477	14
38	0.299	Indonesia	0.331	36	0.127	39	0.419	23
39	0.287	China	0.455	24	0.260	37	0.147	40
40	0.240	India	0.301	38	0.077	40	0.328	32

Benelux: Belgium, Netherlands, and Luxembourg; **RoAfrica:** Rest of Africa; RoAsia: Rest of Asia; **RoEU:** Rest of European Union; **RoEurope:** Rest of Europe; **RoFSU:** Rest of Former Soviet Union; **RoLA:** Rest of Latin America; **RoWorld:** Rest of World; **SEastAsia:** Southeast Asia.

numerical simulation, as soon as some pillars are defined as basic measures with respect to such requirements. It is quite important that these properties be fully understood and accepted.

The proposed approach also has interesting policy-making potential. In fact, using the method proposed in this paper, a complete sustainability ranking of the regions of the world has been proposed for the year of 2011. However, we also provide future sustainability projections by exploiting the features of the ICES computable general equilibrium model where the detailed sustainability outcomes could be referred at the project website that is provided in the paper. Thus, comparative statistical analysis both across countries and through time has been made possible with the proposed methodology: a novelty in the field of sustainability assessment that may have important policy-making applications.

The analysis has been completed by three further investigations: through the computation of the Shapley index, it has been possible to address the relative importance of different indicators, which could also be used in the future to refine the current sustainability tree. Secondly, *andness* and interaction indices highlight that the consensus expert evaluates the majority of the sustainability indicators as being more *complementary* and therefore, for a country to have a higher sustainability level, it needs to perform well in all indicators rather than simply having a satisfactory performance in only one of those. This finding could be considered within the policy-agenda, as the majority of the experts value more a balanced achievement than that of achieving extremely well in some dimensions and badly in other dimensions. In other words, the set of experts value societies that achieved a balanced level of sustainability when compared with unbalanced ones. This pilot evaluation of sustainability indicators might be extended to more participants such as the general public or stakeholders, who have power to implement policies, to analyze whether the opinion of public (or stakeholders) is similar to that of the ones found in this paper.

A robustness analysis has been conducted to examine the variation in the sustainability outcomes and respective country rakings when different expert preferences are used to aggregate the sustainability indicators. This robustness analysis led to some variation in the levels of sustainability for a majority of the countries but rankings remain mainly the same. For example, Norway was the most sustainable country in the world in 2011 irrespective of the preferences of experts. In other words, even though experts have different preferences over sustainability attributes, Norway had the most sustainable outcomes, and therefore could be considered a role model country by others if they were to achieve better sustainability outcomes.

Finally, despite the importance of extending the current pool of decision makers involved in the analysis, the method proposed is already able to capture important information about sustainability, economizing on computational time without sacrificing too much information – another important feature for policy-making applications.

Appendix A.

A.1. Dominance analysis

The analysis of the simulation results should take into account the fact that the distributions of simulated FEEM SI values are not independent from one another since the data provided by single AE contribute to determine the FEEM SI values for every country. In order to describe more accurately the simulation results, the following metrics have been implemented to compare any two countries i and j included in the ranking:

$$\Delta(i,j) = \frac{1}{N}\sum_{k=1}^{k} F[R_k(i) - R_k(j)]$$

where N is the number of countries included in the ranking, $R(i)$ and $R(j)$ are the FEEM SI values for the ith and jth country respectively. k is the number of simulations and $F(x)$ takes the form:

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 0 \\ x & \text{if} \quad x > 0 \end{cases}$$

When constructed in this way, $\Delta(i,j)$ represents the "average cardinal dominance" of country i on country j. That is, the measurement expresses by how much, on average, the ith country dominates the jth across simulations. The overall dominance measure of country i on every other country is given by:

$$\rho^{+}(i) = \frac{1}{N-1}\sum_{j=1}^{N} \Delta(i,j)$$

Whereas the degree to which country i is dominated by every other country is given by:

$$\rho^{-}(i) = \frac{1}{N-1}\sum_{j=1}^{N} \Delta(j,i)$$

We can thus construct the following measure:

$$\rho(i) = \frac{\rho^{+}(i)}{\rho^{+}(i) + \rho^{-}(i)}$$

which indicates the extent of relative dominance of the ith country. This will be equal to 1 if the country in question dominates any other across all simulations and measures 0 if country i is being dominated by all other countries. Being within the [0,1] range, its interpretation is quite straightforward.

Table A.1
List of countries and macro-regions.

No.	Macro-regions	Countries
1	Australia	Australia
2	New Zealand	New Zealand
3	Japan	Japan
4	Korea	Korea
5	China	China, Hong Kong, Taiwan
6	India	Indonesia
7	Indonesia	India
8	SEastAsia	Malaysia, Philippines, Singapore, Thailand, Vietnam
9	RoAsia	Afghanistan, Bangladesh, Bhutan, Brunei Darassalam, Cambodia, Democratic Republic of Korea, Lao People's Democratic Republic, Macau, Maldives, Mongolia, Myanmar, Nepal, Pakistan, Sri Lanka, Timor Leste
10	USA	USA
11	Canada	Canada
12	Mexico	Mexico
13	Brazil	Brazil
14	RoLA	Argentina, Bolivia, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, Falkland Islands (Malvinas), French Guiana, Guyana, Suriname, Costa Rica, Guatemala, Nicaragua, Panama, Belize, El Salvador, Honduras, Saint Vincent and the Grenadines, Trinidad and Tobago, Turks and Caicos, Anguilla, Antigua & Barbuda, Aruba, Bahamas, Barbados, Cayman Islands, Cuba, Dominica, Dominican Republic, Grenada, Guadeloupe, Haiti, Jamaica, Martinique, Montserrat, Netherlands Antilles, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Virgin Islands (British), Virgin Islands (U.S.)
15	Austria	Austria
16	Benelux	Belgium, Luxembourg, Netherlands
17	Denmark	Denmark
18	Finland	Finland
19	France	France
20	Germany	Germany
21	Greece	Greece
22	Ireland	Ireland
23	Italy	Italy
24	Poland	Poland
25	Portugal	Portugal
26	Spain	Spain
27	Sweden	Sweden
28	UK	UK
29	RoEU	Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Malta, Slovakia, Slovenia, Bulgaria, Romania
30	Switzerland	Switzerland
31	Norway	Norway
32	RoEurope	Albania, Andorra, Bosnia and Herzegovina, Croatia, Faroe Islands, Gibraltar, Iceland, Liechtenstein, Macedonia, the former Yugoslav Republic of, Monaco, San Marino, Serbia and Montenegro
33	Russia	Russia
34	RoFSU	Belarus, Ukraine, Moldova, Republic of, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, Armenia, Azerbaijan, Georgia
35	Turkey	Turkey
36	MiddleEast	Bahrain, Islamic Republic of Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Occupied Palestinian Territory, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, United Arab Emirates, Yemen
37	North Africa	Algeria, Egypt, Libyan Arab Jamahiriya, Morocco, Tunisia
38	RoAfrica	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Democratic Republic of the Congo, Cote d'Ivoire, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Mozambique, Niger, Nigeria, Reunion, Rwanda, Saint Helena, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
39	South Africa	South Africa
40	RoWorld	American Samoa, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia, Federated States of, Nauru, New Caledonia, Norfolk Island, Northern Mariana Islands, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu, Island of Wallis and Futuna, Bermuda, Greenland, Saint Pierre and Miquelon

Table A.2
Ranking of countries according to average dominance index, ρ, for 2011 FEEM SI.

Country	ρ^{+}	ρ^{-}	ρ
Norway	8.272517	0	1
Sweden	7.017498	0.032180	0.995435
Switzerland	5.241969	0.125629	0.976595

Table A.2 (*Continued*)

Country	ρ^+	ρ^-	ρ
Austria	4.993397	0.145784	0.971633
Finland	4.322133	0.220431	0.951474
Denmark	4.124075	0.249059	0.943048
Canada	3.932928	0.282739	0.932931
France	3.632407	0.346158	0.912994
Ireland	3.381697	0.409414	0.892007
New Zealand	3.264846	0.443318	0.880448
USA	2.244132	0.783920	0.741114
Australia	2.171497	0.811651	0.727921
Brazil	1.979238	0.893875	0.688883
UK	1.738591	1.010101	0.632516
RoEurope	1.671711	1.049046	0.614429
Germany	1.658449	1.056061	0.610957
Portugal	1.592388	1.097545	0.591980
RoLA	1.427004	1.218559	0.539395
Benelux	1.221855	1.392268	0.467405
Spain	1.206080	1.406183	0.461699
Mexico	1.163242	1.451047	0.444955
Russia	1.163253	1.453589	0.444526
RoEU	1.130448	1.487507	0.431806
Korea	0.994921	1.667540	0.373685
Italy	0.950801	1.733224	0.354245
Japan	0.801836	1.979846	0.288256
Turkey	0.725650	2.124143	0.254632
MiddleEast	0.722531	2.130927	0.253212
Poland	0.572879	2.479087	0.187708
South Africa	0.568334	2.491081	0.185766
Greece	0.383163	3.049703	0.111616
RoAfrica	0.372474	3.087105	0.107665
RoWorld	0.294316	3.390436	0.079874
SEastAsia	0.226390	3.716297	0.057420
RoFSU	0.221337	3.745096	0.055803
North Africa	0.132058	4.360702	0.029393
Rest of Asia	0.094577	4.698028	0.019734
Indonesia	0.046588	5.290119	0.008730
China	0.029825	5.608623	0.005290
India	0	6.771014	0

References

Agliardi, E., 2011. Sustainability in uncertain economies. Environ. Resour. Econ. 48, 71–82.

Arrow, K.J., Dasgupta, P.S., Goulder, L.H., Daily, G., Ehrlich, P.R., Heal, G.M., Levin, S., Mäler, K.-G., Schneider, S., Starett, D.A., Walker, B., 2004. Are we consuming too much? J. Econ. Perspect. 18, 147–172.

Arrow, K.J., Dasgupta, P.S., Goulder, L.H., Mumford, K.J., Oleson, K., 2012. Sustainability and the measurement of wealth. Environ. Dev. Econ. 17, 317–353.

Bohringer, C., Loschel, A., 2006. Computable general equilibrium models for sustainability impact assessment: status quo and prospects. Ecol. Econ. 60, 49–64.

Bossel, H., 1999. Indicators for Sustainable Development: Theory, Method, Applications. International Institute for Sustainable Development, Winnipeg, Manitoba.

Bruntland, G. (Ed.), 1987. Our Common Future: The World Commission on Environment and Development. Oxford University Press, Oxford.

Budnitz, R., Apostolakis, G., Boore, D., Cluff, L., Coppersmith, K., Cornell, C., Morris, P., 1997. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and the Use of Experts. Report No.: NUREG/CR-6372. US Nuclear Regulatory Commission, Washington.

Carraro, C., Campagnolo, L., Eboli, F., Lanzi, E., Parrado, R., Portale, E., 2012. Quantifying Sustainability: A New Approach and World Rankings. Working Papers 94.2012. Fondazione Eni Enrico Mattei.

Choquet, G., 1953. Theory of capacities. Annales de l'Institut Fourier 5, 131–295.

Dale, V.H., Efroymson, R.A., Kline, K.L., Langholtz, M.H., Leiby, P.N., Oladosu, G.A., Davis, M.R., Downing, M.E., Hilliard, M.R., 2013. Indicators for assessing socio-economic sustainability of bioenergy systems: a shortlist of practical measures. Ecol. Indic. 26, 87–102.

Despic, O., Simonovic, S.P., 2000. Aggregation operators for decision making in water resources. Fuzzy Sets Syst. 115, 11–33.

Eboli, F., Parrado, R., Roson, R., 2010. Climate change feedback on economic growth: explorations with a dynamic general equilibrium model. Environ. Dev. Econ. 15, 515–533.

EPI, 2010. Environmental Performance Index: Summary for Policymakers. Yale and Columbia Universities.

ESI, 2005. Environmental Sustainability Index. Yale Center for Environmental Law & Policy, New Haven.

Etminani, K., Naghibzadeh, M., Peña, J.M., 2013. DemocraticOP: a democratic way of aggregating Bayesian network parameters. Int. J. Approx. Reason. 54, 602–614.

EEA, 2005. EEA Core Set of Indicators-Guide. Report No. 1/2005. European Environment Agency, Copenhagen.

EU SDS, 2006. Review of the European Union Sustainable Development Strategy–Renewed Strategy. Council of the European Union, Brussels.

FEEM, 2011. FEEM Sustainability Index Methodological Report 2011. Fondazione Eni Enrico Mattei, Available from: http://www.feemsi.org/documents/FEEM%20SI%202011%20Methodological%20Report%202011.pdf

Fleurbaey, M., 2009. Beyond the GDP: the quest for a measure of social welfare. J. Econ. Lit. 47, 1029–1075.

Fleurbaey, M., Blanchet, D., 2013. Beyond GDP: Measuring Welfare and Assessing Sustainability. Oxford University Press, New York.

Grabisch, M., 1995. Fuzzy integral in multicriteria decision making. Fuzzy Sets Syst. 69, 279–298.

Grabisch, M., 1996. The application of fuzzy integrals in multicriteria decision making. Eur. J. Oper. Res. 89, 445–456.

Grabisch, M., 1997. k-order additive discrete fuzzy measures and their representation. Fuzzy Sets Syst. 92, 167–189.

Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E., 2009. Aggregation Functions. Encyclopedia of Mathematics and its Applications, No 127. Cambridge University Press, Cambridge.

GRI, 2010. Global Reporting Initiative Sustainability Report 2009/2010. Global Reporting Initiative, Amsterdam.

Grossman, G., Krueger, A., 1993. Environmental impacts of a North American free trade agreement. In: Garber, P. (Ed.), The U.S.–Mexico Free Trade Agreement. MIT Press, Cambridge, pp. 13–56.

Klement, E.P., Mesiar, R., Pap, E., 2000. Triangular Norms. Kluwer Academic Publishers, Netherlands.

Krajnc, D., Glavic, P., 2005. A model for integrated assessment of sustainable development. Resour. Conserv. Recycl. 43, 189–208.

Marichal, J.L., (Ph.D. thesis) 1998. Aggregation Operators for Multicriteria Decision Aid. Institute of Mathematics, University of Liège.

Marichal, J.L., 2000. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. IEEE Trans. Fuzzy Syst. 8, 800–807.

Marichal, J.L., Roubens, M., 2000. Determination of weights of interacting criteria from a reference set. Eur. J. Oper. Res. 124, 641–650.

McGillivray, M., 1991. The human development index: yet another redundant composite development indicator? World Dev. 19, 1461–1468.

McGillivray, M., White, H., 1993. Measuring development? The UNDP's human development index. J. Int. Dev. 5, 183–192.

Meyer, P., Ponthière, G., 2011. Eliciting preferences on multiattribute societies with a Choquet integral. Comput. Econ. 37, 133–168.

Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-Being: Synthesis. Island Press, Washington.

Munda, G., 1997. Environmental economics, ecological economics, and the concept of sustainable development. Environ. Values 6, 213–233.

Munda, G., 2005. Measuring sustainability: a multi-criterion framework. Environ. Dev. Sustain. 7, 117–134.

Munda, G., Nardo, M., 2009. Non-compensatory/non-linear composite indicators for ranking countries: a defensible setting. Appl. Econ. 41, 1513–1523.

Munda, G., 2012. Choosing aggregation rules for composite indicators. Soc. Indic. Res. 109, 337–354.

Murofushi, T., Soneda, S., 1993. Techniques for reading fuzzy measures (iii): interaction index. In: 9th Fuzzy System Symposium, Saporo, Japan, pp. 693–696.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., 2005. Tools for Composite Indicators Building. European Commission, Ispra.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., Olsson, L., 2007. Categorising tools for sustainability assessment. Ecol. Econ. 60, 498–508.

OECD JRC, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD, Paris.

Panayotou, T., 1993. Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development. Working Paper WP238. Technology and Employment Programme, International Labor Office, Geneva.

Saisana, M., Tarantola, S., Saltelli, A., 2005. Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. J. Roy. Stat. Soc. A Sta. 168, 307–323.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley & Sons Publishers, New York.

Saltelli, A., 2007. Composite indicators between analysis and advocacy. Soc. Indic. Res. 81, 65–77.

Selden, T.M., Song, D., 1994. Environmental quality and development: is there a Kuznets curve for air pollution emissions? J. Environ. Econ. Manag. 27, 147–162.

Shapley, L.S., 1953. A value for *n*-person games. In: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games (Annals of Mathematics Studies 28). Princeton University Press, Princeton, pp. 307–317.

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

Ulengin, F., Topuc, Y.I., Sahin, S.O., 2001. An integrated decision aid system for Bosphorus water-crossing problem. Eur. J. Oper. Res. 134, 179–192.

UN CSD, 2005. Indicators of Sustainable Development – CSD Theme Indicator Framework. UN Commission on Sustainable Development, New York.

Vincke, P., 1999. Outranking approach. In: Gal, T., Stewart, T.J., Hanne, T. (Eds.), Multicriteria Decision Making. Advances in MCDM Models, Algorithms, Theory and Applications. Kluwer, Boston, pp. 305–333.

World Bank, 2010. The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium. World Bank, Washington.

Yager, R.R., 1993. Families of OWA operators. Fuzzy Sets Syst. 59, 125–148.