Chapter 4 Aggregation of indicators
1. Aggregation, a necessary yet delicate step

In the previous chapter the indicators selected for the FEEM SI have been normalised introducing in the analysis some reference values for each indicator, in order to recalculate them on a common scale between 0 and 1. The normalised scores allow direct comparison even between extremely different aspects of sustainability, but can say very little on the overall appraisal of a country.

The contribution of this work to the climate-related and sustainability literature is the construction of a composite index, which aggregates together the different components into one single measure, easily comparable across countries and in time. Yet, aggregating indicators that are very different in nature into one single measure may be seen as an oversimplification of the issues included in the different components.

The aggregation step is therefore a very delicate part of the construction of an index, which needs to be undertaken with particular care. This chapter will account for the aggregation procedure as follows. Firstly, the literature on composite indices and aggregation methodology will be reviewed and briefly summarised, in order to present the methodological possibilities available to this study and highlight drawbacks and opportunities offered by composite indices. Secondly, the methodology applied to the FEEM SI will be presented both at theoretical and practical levels. Finally, a description of an introductory robustness analysis concludes.

2. Composite indices: methodological review

In order to aggregate indicators, some weights need to be given to each of them that express their importance with respect to the final composite index. Once weights are defined, different techniques are used to combine the weighted indicators into one single measure. A broadly used aggregation technique is the Equally Weighted Average (EWA), which, as its

---

1 This section will only provide an overview of the more commonly used methodology to aggregate sustainability indicators in literature. For a more thorough review of weighting and aggregation methodologies the reader should refer to the Handbook on composite indicators (Saisana et al., 2005).
name suggests, relies on a simple mathematical operation in which all indicators are given the same weight. Giving the same weight suggests perfect substitutability among indicators, which is not a very reasonable assumption given the nature of the indicators and could lead to double counting. Nevertheless, it also implies recognizing the lack of empirical scientific proofs on the relative importance of different sustainability indicators (Saisana et al. 2005) and is considered the most transparent way of producing aggregate indices (Yale Center for Environmental Law and Policy, 2005). In real world applications, EWA may be applied with particular care only to the case where no interactions exist among the criteria—an unlikely and quite rare situation. In fact, the perfect substitutability very often fails to be satisfied, which means that the compensative assumption (technically, the Preferential Independent axiom) is unsatisfied; therefore, EWA cannot be used².

In order to address the situations where EWA cannot be used, many methods have been proposed in the Multi Attribute Value Theory context (MAVT), such as the multiplicative approach, the compensation operator (Von Altrock, 1995) or the Ordered Weighted Averaging (OWA) operator (Yager, 1993)³. Nowadays, it is widely recognized that the non additive measures approach (NAM) satisfies many theoretical requirements, and, at the same time, it is sufficiently general to cover a lot of preference structures of the Decision Maker (DM). Many types of interactions can be modelled in this way, like pessimistic or optimistic behaviour of the DM, that indicate respectively that the satisfaction degree is high only if all the criteria are satisfied (corresponding to the logical conjunction operator AND), or if at least one of them is high (corresponding to the logical disjunction operator OR). Note that both cases cannot be implemented by EWA.

Many aggregation operators can be implemented to calculate the values of such measures starting from one extreme case, the AND, up to the other extreme case, the OR operator. Thus, we can obtain the MIN (corresponding to the AND), the MAX (corresponding to the OR), the k-statistic, and even the OWA and the EWA, which are then sub-cases of non additive measure-based aggregation operators⁴. The required parameters can be easily obtained by means of a simple questionnaire. The price to pay consists into an exponentially increasing

---

² The compensative assumption is rarely tested in practical applications, but missing this check can induce a strong distortion in the decisional process.
³ Please refer to Klement (2000) for a more detailed theoretical framework on multi-attribute aggregation techniques.
⁴ The OWA operator is obtained when the measure of each coalition with the same cardinality is the same, while the EWA when its equals the sum of the measures of every sub-coalition which define a partition.
numerical complexity due to the number of parameters involved. In fact, if \( n \) is the number of the criteria, a NAM requires the specification of \( 2^n \) parameters, i.e. the number of all the subsets of the \( n \) criteria. This is different and more complex than the EWA approach, which only needs \( n \) parameters\(^5\). In fact, this method is based on the observation that the linearity implicitly assumed in EWA can be violated, that is, the “weight” of a coalition of sub-criteria can be greater or less than the sum of the weights of each of the sub-criterion belonging to the coalition itself. Thus, the main idea consists into assigning a weight to every possible subset of the criteria which refer to the considered node in the tree. Subsequently, a simple algorithm, the so-called Choquet Integral (De Waegenaere, 2001 and Murofushi, 1994), computes a weighted averaging of the values of all the subsets, extending EWA by taking the coalitions into account, instead of taking the single criteria only (singleton). Naturally, if for every coalition the weight (or the importance) of each coalition is formed by the sum of the weights of each sub-set of its criteria forming a partition, we obtain again the EWA. Conversely, if for a coalition its weight is inferior to such a sum, a redundant interaction exists among the included criteria, while if it is greater than the sum, a synergic interaction exists. Some slightly different versions of the basic Choquet integral aggregation exists, for example the multi-linear operator (Grabish, 1995 and 1996), a smoother algorithm than the Choquet integral\(^6\).

Because of the importance of considering the interactions between the indicators that compose the FEEM SI, it has been decided to use a non-additive aggregation measure. However, compared to those described above, in the case of the FEEM SI a further requirement has been set for the analysis, namely the respect of the monotonicity criterion. This principle implies that the importance of a coalition cannot be less than the minimum of the weight of each sub-coalition included in it. It is clear that this assumption is quite natural, for most applications and in particular for the present work.

To obtain the values of the measure, a suitable questionnaire was developed by Despic (2000), where all the (0/1) combinations of criteria was considered (for a given node in the tree), meaning that “0” corresponds to the WORST case, and “1” to the BEST one. The questionnaire is a list of some possible scenarios, i.e. all the combinations of BEST and WORST values. If \( n \) is the number of sub-criteria for the considered node, \( 2^n \) is the number of

\(^5\) To be exact, they are \( 2^n - 2 \) since the border condition, which is the measure relative of the case in which the empty set is null and the measure of the case in which the universal set is one, reduce the required number of parameters, see Annex II for further details.

\(^6\) Please refer to Annex II for a more detailed description of this algorithm.
possible questions. As soon as the questionnaire is fulfilled, the NAM is defined, and an algorithm computes the value of the multi-linear operator.

Let us observe that, to render the values of the NAM as objective as possible, more than one DM is interviewed and asked to provide his/her measure values. The analysis of the DM behaviour (optimism, pessimism, accord) will be further developed in section 4.

3. Weighting and aggregation for the FEEM SI

In order to perform the analysis presented in theory in the previous section, the indicators selected for the FEEM SI have been organised into a decision tree, in which partial aggregation take place at different levels.

**FIGURE 3.1 FEEM SI AGGREGATION TREE**

The decision tree should be read from bottom (leaves) to top (final node) and features a different colour coding for aggregations taking place at different times. The tree respects the
three-pillar structure introduced in the previous chapters, with the final node producing the aggregate index. Differently from other procedures, where a relative weight is defined for each indicator with respect to the others, the decision tree requires to attribute weights to the coalitions of indicators present at each node. In order to facilitate this evaluation the indicators present at each node will be interpreted “at the edge” - either at their best or worst level - creating a matrix of all the possible combinations between these two levels of the indicators. These best and worst levels are not defined in quantitative terms in order to avoid bias deriving from respondents disagreeing with the judgement given. It is up to the respondent to “imagine” what “best” and “worst” mean in each case. The indicator-coalition matrices are presented to in the form of a questionnaire\(^7\) to a panel of respondents (“experts”) that will determine the final weights for the aggregation of the FEEM SI.

I. Features of the questionnaire

The questionnaire is implemented in a software that elaborates the weights given at every row of the matrix and computes one single weight at each node for every Expert. The Experts are allowed to give scores between 0 and 100 for each row, except for the first and the last (where indicators are respectively all “worst” and all “best”) which are given 0 and 100 by default. Moreover, the weights given at each row of every matrix need to respect the monotonicity criterion introduced in the previous section. In the case of this questionnaire, respecting this criterion implies that, if a combination where only one indicator is “best” is given a certain weight x, all combinations including that indicator in the “best” case should be given a weight at least equal to x.

The following example will clarify this process.

---

\(^7\) For complete version of the questionnaire please refer to Annex 2.
The application of this decision process in the context of sustainability deserves some remarks. First of all, the term Expert needs some explanations. In the line of principle, it seems that different real Experts, assign their judgment to each coalition, meaning that a Group of Experts cooperate together, blindly or not, to a Multi Person Session, as in Delphi procedure. Nevertheless, this process can be criticized and requires particular care. In fact, a deep explanation is required to assure the transparency and the comprehension of the spirit of the methodology. In particular, for the higher level of the hierarchy, the parameters are to be tuned on the basis of each Experts’ own preference structure, and not to the expertise in strict sense. That is, the decision has to be intended more political (subjective) than technical (objective). In this sense, the term Expert can be better intended as Stakeholder. Even if, the border between the two roles, each of them related more or less to technical or political aspect, is not so strict and clear. This is a first and important aspect that is sometimes missing in the specific literature on Decision Making processes, as well as in the Decision Support System literature which applies similar - even if simpler - hierarchical tools.

In the WA approach, comparing two criteria together, a greater weight for the first criterion usually means that, in the Stakeholder opinion, this criterion is more important than

| RESTAURANTS |
|-------------|-------------|-------------|---|
| Price       | Quality     | Kindness    | Weights |
| Worst       | Worst       | Worst       | 0    |
| Best        | Worst       | Worst       | 20   |
| Worst       | Best        | Worst       | 50   |
| Worst       | Worst       | Best        | 30   |
| Best        | Best        | Worst       | $\geq 50$ |
| Best        | Worst       | Best        | $\geq 30$ |
| Worst       | Best        | Best        | $\geq 50$ |
| Best        | Best        | Best        | 100  |

The coalitions where more than one indicator is “best” must be at least valued at the highest between the scores given to the coalition where only one of the indicators was “best”, as the numerical example shows.
the second one. For the highest level in the hierarchy in the construction of the FEEM SI, this decision is more subjective and depends on each Stakeholder’s preference\(^8\).

It is important to underline that the sustainability approach requires a non-compensative aggregation between the criteria. For instance, a sustainable policy requires an economical development not to be obtained at the price of a critical environmental pollution. This is a strong motivation in support of the present methodological proposal. A low level of compensation cannot be achieved by **EWA**, but neither by **WA** with different values of the weights. Instead, a non additive measures approach encounters the sustainability philosophy if the measures are selected in such a way that compensation cannot be completely fulfilled. The level of compensation implicit in the weights assigned by the experts can be quantified by means of two complementary indices, the **ANDNESS** and **ORNESS** indices measure respectively these behaviours. The first corresponds to an optimistic Expert, whilst the second to a pessimistic one. The sum of these indices is always 1, with each of them being given a score between 0 and 1. An ANDNESS degree close to 1 indicates that the Expert/Stakeholder tends to be non-compensative, meaning that he/she cares for all sub-criteria present at a specific node to be high (“best” in the case of the questionnaire used for the FEEM SI). On the contrary, an ORNESS degree close to 1 indicates that the DM is satisfied even if only one sub-criterion is “best”. Being the andness and orness measures complementary to 1, studying only one of them is sufficient. In the case of sustainability, it is more interesting to study whether the Expert/Stakeholder tend to be compensative and consider excellence in a sub-indicator as a convenient compensation for low levels in other sub-components, as this behaviour does not seem very much in line with the idea of sustainability, where total compensation among different components is not considered feasible.

Thus, for addressing sustainability, we expect that the measures assigned by each Expert/Stakeholder should be of **ANDNESS** type. This, together with the difficulty in finding an objective Expert/Stakeholder able to assess a multidisciplinary subject such as sustainability, leads us to underline two important implications. Firstly, it is clearly impossible to propose a questionnaire to a sufficient number of Stakeholders, so as to include most of the current opinions about sustainability. Given the critical issue of the questionnaire

---

\(^8\) At lower levels of the hierarchy the weights to be given are theme-specific, thus it is easier for an Expert to give its opinion on a definite subject of his competence. However, at the top level, the aggregation requires interactions between the different themes of sustainability, thus in a more interdisciplinary approach. In this sense the opinion of the Experts may be biased by their own background and personal opinion. Nevertheless, a Stakeholder, who is more involved in political decision-making, will have a relevant opinion in the subject.
comprehension, this could be proposed to a restricted number of Stakeholders, which should be carefully selected, but this seems up to now a very datable approach. Secondly, even when forcing the measures to be ANDNESS oriented, infinite number of possibilities exist. Thus in our opinion, a correct approach in the line of sustainability context, is to choose some limited different scenarios characterized by rational values of weights for each coalition (monotonicity and ANDNESS have to be respected), and to move from one scenario to another one, suitably tuning some numerical parameters which weight more or less one scenario with respect to another one. In so doing, for each year every country can be ranked according to the FEEM SI with different scores depending on the combined scenario measures so obtained, in a way similar to a Monte Carlo approach. This also allows to test the robustness of the ranking.

In the current proposal, three scenarios were constructed assigning different values of the non additive measures. The measures were defined after some consultations among internal Researchers which cooperated to the project and were aware about the complexity of the proposed decision system. A first scenario was constructed, and afterward two more were obtained after suitable modifications of the first one. Figure 4.1 and Table 4.1 illustrate ANDNESS indices relative to these questionnaires.

**Figure 4.1 ANDNESS DEGREE OF RESPONDENTS AT FINAL NODES**

---

9 The next section will better explain the robustness analysis procedure.
### Table 4.1 Orness Degree Statistics

<table>
<thead>
<tr>
<th>NODE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEM SI</td>
<td>0.502</td>
<td>0.137</td>
</tr>
<tr>
<td>Economic</td>
<td>0.573</td>
<td>0.124</td>
</tr>
<tr>
<td>Social</td>
<td>0.492</td>
<td>0.123</td>
</tr>
<tr>
<td>Environmental</td>
<td>0.445</td>
<td>0.190</td>
</tr>
</tbody>
</table>

#### 4. Robustness analysis

As already introduced above, being the score of each country for each year depending on the values of the measures, it is of interest not only to observe how it changes with the different scenarios, but also how much it changes as soon as the measure values move from one scenario to another one. That is, many possible scenarios are computed, as combination of the three basic ones. To this purpose, a computational tool generates many sets of numbers between zero and one. Each of these numbers is multiplied to one of the measures of the initial questionnaires, and all the contribution are summed. In so doing, a new measure is created, which is more or less close to one of the basic scenarios, or to the average of all of them. To distinguish from the basic scenarios, we call these new artificially computed measures as simulated scenarios. Technically speaking, every new measure is a convex linear combination of the basic scenarios. For each of the simulated scenarios, a score is computed for each country and year. Given the unavoidable source of uncertainty in the judgements of each coalitions, this simulation approach seems to include the appealing sufficient generality and at the same time maintaining the non compensability requirement. In fact, it is possible to verify that a convex linear combination of non additive measures is a non-additive measure itself (monotonicity is preserved). Thus, the basic scenario can be selected just respecting the monotonicity requirement and the ANDNESS typology. Not much further effort is required. In fact, the robustness analysis provides by itself to answer questions like “what if the measures are changed a little bit?”. A user friendly graphical interface will provide to show the scores together with the range of variability, permitting the user to assess the sustainability levels as well as their relative robustness. This is a crucial information which can be successfully used for subsequent analysis. A more technical description of the robustness analysis is presented in the Annex II.
5. Future developments and conclusions

The aggregation approach was inspired by two considerations: the non compensative nature of the sustainability concept, together with the subjective character of every decision support tool. We feel that combining the non additive measure algorithm, a novelty in the field of sustainability analysis, with sensitivity analysis, a well known approach for simulation, we improved the scoring system with respect to other similar ones. In fact, we admit that a partial uncertainty is unavoidable for every scoring system, but two requirements are necessary for a rational sustainability analysis: the monotonicity and the non compensability. Robust options are enhanced by 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.

As future development, we propose to perform a deeper analysis about the relative importance, or trade-off, of each criterion (for a single node in the three). This can be done, by a simple computation, i.e. the Shapley index, which computes the average marginal gain obtained for all the coalitions not including the criterion, compared with the same coalition but with the criterion included in it. Even this, can be a useful information for analyzing and managing purposes.
REFERENCES


