Measuring the sustainability of a natural system by using multi-criteria distance function methods: Some critical issues

L. Diaz-Balteiro , P. Belavenutti , M. Ezquerro , J. Gonzalez-Pachon , S. Ribeiro Nobre , C. Romero

A B S T R A C T

There is an important body of literature using multi-criteria distance function methods for the aggregation of a battery of sustainability indicators in order to obtain a composite index. This index is considered to be a proxy of the sustainability goodness of a natural system. Although this approach has been profusely used in the literature, it is not exempt from difficulties and potential pitfalls. Thus, in this paper, a significant number of critical issues have been identified showing different procedures capable of avoiding, or at least of mitigating, the inherent potential pitfalls associated with each one. The recommendations made in the paper could increase the theoretical soundness of the multi-criteria distance function methods when this type of approach is applied in the sustainability field, thus increasing the accuracy and realism of the sustainability measurements obtained.

1. Introduction

The origin of the concept of sustainability has been traced by Pretzch (2014) referring to key research undertaken by Central European foresters more than 300 years ago (von Carlowitz, 1713). The basic idea of these pioneers was to request from a forest a long-term stable supply of a flow of goods and services necessary for the welfare of human beings. Later on, this concept of sustainability was transferred from forestry to other natural systems. Until very recently the commodities and services demanded from natural systems was always understood as being within a context of mono-functionality. That is, the basic purpose of natural systems is to provide inputs that the production system transforms into goods and services, whose consumption satisfies primary human needs.

The above approach for measuring the sustainability of a natural system was not questioned until the second part of the last century, when it was recognized that the environment had physical limits, making it necessary to move to a new context of multi-functionality. In this direction the seminal work by Gregory (1955) should be cited. Although this work is focused on forest systems, its basic ideas are straightforwardly applied to the multi-functionality of any natural systems. In fact, modern societies require not only the long-term durable provision throughout the time of raw materials from natural systems to be transformed into outputs, but also a supply of an important number of environmental goods and services like biodiversity conservation, carbon sequestration, soil erosion, etc.

The above mentioned shift towards the multi-functionality of natural systems implies key changes in the concept and measurement of sustainability. In other words, nowadays any specific natural system should supply a lasting broad spectrum of goods and services of a very different nature. Some of them are valued by the markets, but others, although no less essential, have no market value.

Within this new scenario it seems sensible to use a battery of indicators as a proxy of the different functions provided by a natural system. Thus, any sustainability measurement is established by the aggregation of those indicators into a single composite index, the value achieved by this index being a proxy of the respective sustainability goodness.

Following the above direction, a promising line of research is to treat the indicators as criteria. In this way, it is possible to resort to the work done in the Multiple Criteria Decision Making (MCDM) field. In fact, the purpose of MCDM is to aggregate in one way or another a set of conflicting criteria, so that it is only needed to change criteria for indicators and to proceed to their aggregation. Given the correspondence between the aggregation of criteria and that of sustainability indicators, in the last decades a substantial body of literature applying MCDM techniques to the measurement of sustainability has appeared. In a recent review (Diaz-Balteiro et al., 2017a) more than 270 papers included in the ISI Web of Science database dealing with sustainability from a multi-criteria perspective have been found, with a considerable increase in papers published in recent years.

Among the different MCDM approaches used for addressing sustainability issues, the multi-criteria methods based on the minimization of distance functions (MDF) have achieved paramount importance (Lozano-Oyola et al., 2012). However, despite the interest and success of this orientation, many of the applications reported in the literature are not exempt from difficulties, which imply poor modeling practices, leading to possibly erroneous results. The main aim of this paper is to detect several types of vital issues regarding the application of different MDF techniques for measuring natural system sustainability. Besides this, several methods to overcome these potential insufficiencies are proposed. The practical effectiveness of this type of approach could therefore be increased. It should be noted that the idea of this manuscript was not to review MDF methods applied in sustainability issues, but to make some reflections on the use of these methods, in order to gain some insight and make recommendations for future work.

The paper is organized as follows: in Section 2 the basic aspects of MDF methods are presented; after that, a set of major issues is given, proposing methods for overcoming possible misconceptions and pitfalls. The paper ends up raising some basic conclusions.

2. Multi-criteria distance function models: some basic aspects

In this section, the core of the multi-criteria distance function (MDF) approach referring to a sustainability setting will be briefly presented. Thus, we have $i = 1,2, \ldots n$ systems to be evaluated according to $j = 1,2, \ldots m$ sustainability indicators. R_{ij} measures the value reached by the ith system when it is evaluated according to the jth indicator. W_j represents a preferential weight, that is, the relative importance attached by the Decision Maker (DM) to the jth indicator with respect to the others. Also, p is the topologic metric defining the L_p family of distance functions. K_j is a normalizing constant and, finally, $\widehat{R}_j$ represents a desirable level of achievement for the jth indicator. X_i is a binary variable, taking the value 1 when the ith system is the most sustainable one, otherwise $X_i = 0$, being the value achieved by the objective function the measurement of the sustainability of the ith system. Thus, by solving model (1)–(4) n times, incorporating in each iteration the additional constraint $X_k = 1$, when the kth system is the most sustainable among the remaining ones, then the ranking of the n systems in terms of sustainability is obtained. The values achieved by the respective objective functions provide the sustainability measurement for each system. See Diaz-Balteiro and Romero (2004a) for a detailed explanation of the whole optimization process. Once all the variables and parameters have been defined, the analytical expression for the MDF model is:

$$Min\ L_p = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} W_j^p \left| \frac{\left(\widehat{R}_j - R_{ij}\right) X_i}{K_j} \right|^p \right]^{1/p} \quad (1)$$

Underlying the general MDF model given by (1) are an important number of specific models according to the value of the metric p and according to the meaning given to the desirable level of achievement $\widehat{R}_j$. Thus, if $\widehat{R}_j$ is an ideal value, model (1) derives towards a compromise programming approach, if $\widehat{R}_j$ is a "satisficing" target model (2) derives towards a goal programming approach, etc. Regarding the value of the metric, for $p = 1$ model (1) turns into the following one:

$$Min\ L_1 = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} W_j^p \left| \frac{\left(\widehat{R}_j - R_{ij}\right) X_i}{K_j} \right| \right] \quad (2)$$

Model (2) implies optimizing the average providing the best aggregate performance. However, this solution can give poor results for some of the indicators. This bias can make this type of solution unacceptable from a sustainability perspective. For metric $p = \infty$ model (1) turns into the following one:

$$Min\ L_\infty = D$$

Subject to:

$$\sum_{j=1}^{m} W_j \left| \frac{\left(\widehat{R}_j - R_{ij}\right) X_i}{K_j} \right| - D \le 0 \quad i \in \{1, 2, \ldots n\} \quad (3)$$

D being the maximum deviation. In this way, the "most balanced" solution is obtained. This solution is an appealing one due to its balancing character but it can produce poor "average" results. It is tempting to use metric p as a "balancing factor" between "average" and "balance". However, this type of orientation leads to having to solve complex non-linear and non-convex mathematical programming problems. That is why the above conflict within the MDF approach is usually treated by trading-off L_1 and L_∞, with the help of the following model:

$$Min\ L_\lambda = (1-\lambda)D + \lambda \left[\sum_{i=1}^{n} \sum_{j=1}^{m} W_j \left| \frac{\left(\widehat{R}_j - R_{ij}\right) X_i}{K_j} \right| \right]$$

Subject to: (4)

$$\sum_{j=1}^{m} W_j \left| \frac{\left(\widehat{R}_j - R_{ij}\right) X_i}{K_j} \right| - D \le 0 \quad i \in \{1, 2, \ldots n\}$$

In model (4) λ plays the role of a control parameter. Thus, when $\lambda = 1$ model (4) turns into model (2) and the "average" is optimized and when $\lambda = 0$, then model (4) turns into model (3) and the "balance" is optimized. For control parameter values belonging to the open interval (0,1) compromise solutions between L_1 and L_∞ can be obtained if they exist. One example of these compromise solutions within a sustainability context can be seen in Gómez-Limón and Sanchez-Fernandez (2010). Hence, control parameter λ can be interpreted as being a marginal rate of transformation between "average" and "balance". Some clarifications on the meaning and implementation of MDF model in the sustainability field are made in Section 9.

3. Indicators and criteria

In talking about criteria and indicators, it is necessary to point out that there is no unanimous consensus in scientific literature on when to use either word or the other. Thus, many works have not differentiated between criteria and indicators (in Diaz-Balteiro et al. (2017a), 25.8% of the articles analyzed did not make that differentiation), and on many occasions they were taken to be synonyms. There is also a frequent hierarchy between the two terms, so that, if it is aimed to analyze the sustainability of a natural system, a set of indicators is defined for each of the criteria.

The same as happens in the different ways of denominating that aggregate index (index or composite indicator, synthetic index, aggregate index, etc.), other authors use other denominations to name the criteria and indicators. For example, the economic, social and environmental dimensions of sustainability are often called "pillars", whereas, in other works, those dimensions are only treated as criteria. In another sense, Wang (2015) talks about sub-indicators, and Kuzdas et al. (2016) about principles. Besides, in complex problems, more levels than that of a criterion or indicator are defined. It is fitting to insist that these circumstances are produced in academic works since, in institutional spheres (international organizations, government agencies, etc.), it is indeed permitted to analyze sustainability by means of a closed set of criteria and indicators.

Although in its beginnings sustainability had been defined on the basis of environmental, economic and social criteria, since then its casuistry has become more extensive and works in which other types of criteria are employed are easily found; for example, those including an institutional dimension (Valentin and Spangenberg, 2000). The inclusion of these criteria ad hoc is often explained by the difficulty in finding data for some criteria and/or indicators, or because it is attempted to specify much better the scope of each element intervening in the evaluation of the sustainability. However, for some authors and under certain circumstances, the environmental criteria have a pre-emptive importance with respect to the social and the economic ones (e.g., Bathrellos et al., 2017). Finally, it should be noted that the application of MDF methods requires the classification of the indicators into two analytical categories, "less is better" or "more is better", according to different publications (e.g., Voces et al., 2012).

There is another issue of interest in the use of MCDM models for evaluating the sustainability of a natural system. To be specific, we refer to the number of criteria and indicators employed in the analysis. Some data available show us that, in each work, the average of criteria used is over 4, whereas that of indicators is close to 19. In fact, it is not infrequent to read works in which the number of indicators used exceeds 40 as takes place in Diaz-Balteiro et al. (2017a). This practice often signifies employing in the analysis indicators that may be correlated with others. That is why, before application of the MDF methods, it is recommended to make a statistical analysis beforehand in order to measure the correlation between pairs of indicators and, in the case of exceeding a certain threshold, the number of indicators could be reduced (OECD, 2008; Singh et al., 2009). Some reviews carried out have shown that few works follow this trend (Diaz-Balteiro et al., 2017a), while other authors suggest that, if there is a correlation between the indicators, some MDCM methods cannot be applied (Salvado et al., 2015).

4. Strong and weak sustainability

The distinction between a strong and weak sustainability is one of the precisions most abundantly cited in the literature in relation to the concept of the sustainability of a natural system (Neumayer, 2013). Although other authors describe MDCM methods with which they report being able to analyze strong sustainability (Shmelev, 2012), mainly through compensatory aggregation methods, this is not common to all the MDF methodologies. Indeed, most MDF methodologies start from the hypothesis that there is a certain degree of compensation between criteria and indicators (C/I), so that, in principle, on using these techniques we should always be talking about a weak sustainability. However, exceptions to this rule have been found in works like Blancas et al. (2010), using goal programming, and in Ruiz et al. (2011) where, instead of an aggregate index, two synthetic ones are proposed: one with a weak sustainability orientation and the other a strong one.

Additionally, it should be noted that the degree of compensation between C/I can always be specified in MDF methods (Ruiz et al., 2017). For instance, in defining a lexicographic goal programming model, it can always be guaranteed that certain indicators reach values corresponding to the levels aspired. Namely, with this technique, the decision-making center ensures that there can be no finite trade-off between the values of some indicators/criteria as is shown in Romero (1991). Also, through interaction with the stakeholders, the exchange between the set of indicators can be limited, in such a way that exchange is not always feasible. Thus, a certain natural system can be obliged to be sustainable if it reaches some minimum reference values for each indicator (Diaz-Balteiro et al., 2016a). In synthesis, addressing sustainability by assuming that, in principle, it is not a strong case sustainability does not imply that the possible exchanges between the indicators are always perfect. This idea is in agreement with the one proposed in Neumayer (2013) defining hybrid indicators for strong sustainability, which combines physical indicators (in theory the only indicators permitted when strong sustainability is analyzed) with monetary valuations.

5. Static and dynamic indicators

Within this context, we usually think in terms of a static-type measurement (measuring over time). However, sustainability in itself is a dynamic concept (Moldan et al., 2012), permitting changes in criteria and indicators throughout the time (Popp et al., 2001). For that reason, it is often of great interest to articulate a procedure allowing an adequate aggregation of the values expected from the indicators over a period of time. This focus would permit, for example, the carrying out of simulation exercises in order to verify how the sustainability of a natural system would vary in diverse scenarios. This is of great importance in natural systems, in which handling alternatives promoting a more sustainable management have to be evaluated on a long-term basis, such as in the case of forest systems shown in Hickey and Innes (2008).

MDF methodologies can be modified to integrate this new perspective and, in principle, two distinct problems can be defined. The first one would be to dispose of a historical series of a set of indicators and to calculate how the sustainability varies throughout the time. Here the value of the aggregate indicator could be valid in order to see how the sustainability varies over the time provided that certain hypotheses are assumed. For instance, it has to be presumed that the preferences of the decision-making centers for the different indicators are not modified (see sections 8 and 9).

On the contrary, for the second type of problem, the criteria and indicator values would have to be simulated in a planning horizon against diverse hypotheses. This implies that a greater care should be taken when selecting the indicators and adequately justifying how the values of each indicator are calculated in the future. On these lines, in Diaz-Balteiro et al. (2017b), for a forest management exercise, the values of six indicators for the next hundred years were calculated. By means of a model proceeding from eq. (4), three management alternatives were defined and the aim was to select the best one in the face of six different climate change scenarios. That is to say, that through the aggregation of indicators, the best management alternative can be chosen for each scenario proposed initially.

6. The necessity of a normalization system

The sustainability indicators used in a particular problem have to be normalized for the following two reasons:

a) The indicators are generally measured in different units, so they are not commensurable. In short, to apply an aggregation rule to a battery of sustainable indicators without a previous normalization is like adding "cubic meters of timber with Euros of net present value". If we were to operate in this way, the final figures obtained would have no logical meaning.
b) Given the broad nature of indicators used in the sustainability field, the values achieved by them are, in general, very different. In this situation, the solution provided by any aggregation rule could be biased as more implicit importance is given to the indicators with higher numerical values than those with lower ones; that is, the former indicators receive something like an artificial extra-weight with respect the latter ones, which does not reflect the relative importance attached by the DM to the different indicators considered.

Despite the above two reasons, in the literature covering the measuring of the sustainability of several natural systems there are a significant number of papers ignoring the necessity of implementing a proper normalization of the indicators. Thus, Ibáñez-Forés et al. (2014) state that, in around 70% of the papers analyzed in their work, no normalization system is implemented. However, in Diaz-Balteiro et al. (2017a), this figure decreases to around 30% of the reviewed papers.

Many possible normalization systems can be applied, and choosing one of them involves different implications (Pollesch and Dale, 2016). After all, from a mathematical perspective, normalizing implies dividing a vector by a norm; i.e, parameters K_j of model (1). In our context, vector ith represents the values achieved by the ith system for the m indicators considered. On the other hand, a norm is a scalar attached to a vector with the respective mapping holding some intuitive axioms. Given these considerations, one possible normalization system that is mathematically valid and profusely used in the sustainability field is the following (Diaz-Balteiro and Romero, 2004b):

$$\overline{R}_{ij} = 1 - \frac{R_j^* - R_{ij}}{R_j^* - R_{*j}} = \frac{R_{ij} - R_{*j}}{R_j^* - R_{*j}} \quad \forall i, j \tag{5}$$

where all the variables were defined in section 2, R_j^* and R_{*j} being the best (ideal), and the worst (anti-ideal) values achieved by the jth indicator. In this way, $\overline{R}_{ij}$ represents the normalized value attached to the ith system when it is evaluated according to the jth indicator. With this normalization procedure, the indicator values have no dimension and all of them are bounded between 0 and 1. Hence, the two problems pointed out above are overcome. In fact, there are no commensurability problems, and the possible effect of different numerical values also disappears.

There are other potential normalization systems like fixing an Euclidian norm, summation normalization, the zero-one procedure, etc. (see Tamiz et al., 1998, pp. 572–573 for technical details). Which of them is the best normalization system is still an open question in the sustainability field. The right choice basically depends on the characteristics of the specific problem analyzed. But, within a sustainability context, the indicators used must be normalized in one way or another if we want that the final solutions reflect the reality analyzed with a minimum of accuracy.

7. The weighting question

Except in the case of using tools like DEA to obtain the weights (Zhou et al., 2007), in all multi-criteria MDFs the parameters W_i aim to capture the relative preference attached to the DM to each one of the m indicators considered in a particular problem. Basically, the following three strategies have been used for dealing with this preferential issue:

a) To attach the same preferential weight to all the indicators. Even though this orientation is objective and clear it hides crucial aspects of the reality analyzed. In fact, it is fairly obvious that, in general, the relative importance of the different indicators within a sustainability context might be significantly different. However, this is the strategy most widely used for aggregating sustainability indicators with an MDF approach (e.g., Diaz-Balteiro and Romero, 2004; Lopez-Ridaura et al., 2009; Giménez et al., 2013). This might be due to the influence of the famous Brundtland report, which implicitly confires the same importance to the aspects characterizing the sustainability of a system as proposed in Valentin and Spangenberg (2000).
b) To implement a sensitivity analysis with the values of the preferential weights. In this way, the level of robustness of the sustainable measures obtained with respect to changes in the values of the preferential weights can be checked. This approach is an improvement on the latter one, and is adopted in many works like that of Ruiz et al. (2011); Doukas et al. (2010); Streimikiene et al. (2012). In a recent paper, Becker et al. (2017) deal with a modification of this approach in order to measure the importance of each indicator in a composite index. However, the sensitivity analysis does not tackle the actual underlying problem: how to incorporate a measurement of the real preferences of the DM into the sustainability analysis.
c) To estimate the value of the preferential weight through an interactive process with the DM. This orientation is the most ambitious and also the most realistic one. However, given the usual high number of indicators and/or sometimes the use of a significant number of stakeholders, it would seem advisable to resort to relatively straightforward interactive methods. That is, methods requiring a large amount of information from the DM do not work in practice. In this sense, the classic interactive methods proposed by the multi-criteria school, although theoretically sound, do not appear to be suitable in a sustainability context (e.g., Zionts and Wallenius, 1976, 1983). More pragmatic methods like those based on a "pairwise" comparison format, seem more advisable. In fact, with this type of interaction a global preferential assessment of the m indicators considered is not required from the DM, but a local assessment, which is much easier to provide. Some papers following this orientation are, among others, Gómez-Limón and Riesgo (2009); Diaz-Balteiro et al. (2011, 2016b).

8. The role of the stakeholders: searching for social weights

The current complexity associated with the determination of the individual preferential weights has increased nowadays, due to the manner in which different social groups or stakeholders perceive the relative importance of the different sustainable indicators considered. In fact, in the field studied, there are stakeholders like farmers, consumers, environmentalists, public decision-makers, etc. with different perceptions of the relative importance of the indicators considered (e.g., Bausch et al., 2014). In short, sustainable measurements are researched nowadays within a participatory decision making context.

The above considerations imply the need to aggregate in one way or another individual preferential weights into collective or social weights, which, in the best possible way, reflect the preferences of the different stakeholders as a whole; i.e., parameters W_i

should reflect collective preferences instead of individual judgments. In short, it is necessary to enter into the field of group decision making or social choice. Within this field, the well-known Arrow impossibility theorem shows how there is no aggregation rule for the individual preferences holding a set of sensible conditions or axioms when individual preferences are stated in an ordinal way. However, within a sustainable context, preferences are normally stated with a certain degree of cardinality. Thus, the information of interest is not to know whether one indicator is preferred or indifferent with respect to another one, but to state the intensity of the respective preference. It is more complicated to provide cardinal preferential information, but, paradoxically, the latter is more easily aggregated. Thus, Keeney and Kirkwood (1975) and Keeney (1976) demonstrated that when the individual preferences are stated through cardinal utility functions, there are collective choice rules, like the multi-linear utility function, fulfilling the Arrow conditions.

On the other hand, it is important to be aware that, within a sustainable context, the number of indicators considered is normally very high and the number of stakeholders is not negligible. Therefore, it does not seem advisable to resort to approaches based on multi-attribute utility functions for not being implementable in our context. Within a sustainable context it is necessary to select the most pragmatic approaches capable of determining collective preferences.

Within this pragmatic focus, seemingly, a promising tactic would be to request from the stakeholders their preferences with respect to the different indicators following a “pairwise” comparison format. In this way, the aggregation procedure would consist of the aggregation of a number of “pairwise” comparison matrices equal to the number of stakeholders considered. This task is affordable from the modeling point of view as well as from a computational perspective. See, in this direction, the works by González-Pachón and Romero (2007); Diaz-Balteiro et al. (2009); Nordström et al. (2009), among others.

In any case, how to implement an aggregation rule for a set of individual preferences in our field is still a wide open issue, requiring more research (Rönnqvist et al., 2015). However, this being a controversial topic, the consideration of a sustainable assessment within a participatory decision-making context is nowadays of paramount importance.

9. The selection of an aggregation rule: some preferential considerations

The MDF outlined in Section 2 implies a significant number of aggregation rules depending on the value chosen for metric p or for control parameter λ. However, in the sustainability literature, the selection of the aggregation rule is generally made mechanistically, without taking into account the different structures of DM's preferences underlying the aggregation rule chosen. Thus, Burgman et al. (2014) state that it is usual to choose an aggregation system without any theoretical or empirical justification. For instance, within the field of sustainability, Salvado et al. (2015) chose the additive weighting method as an aggregation rule, under no theoretical or empirical justification conditions.

The above issue is quite crucial if we want to obtain composite indices and rankings compatible with the actual preferences of the DM. This argument is developed in what follows. Thus, as mentioned in Section 2, it is well-known in the multi-criteria literature that metric p or control parameter λ play the role of a balancing factor between the optimization of the average achievement ($p = 1$ or $\lambda = 1$) among indicators, or the minimization of the value achieved by the indicator most displaced with respect to the solution obtained. This latter solution is the most balanced option ($p = \infty$ or $\lambda = 0$). On the other hand, for values of metric p belonging to the open interval (0, ∞) or to the control parameter λ belonging to the open interval (0,1) compromise composite indices, if they exist between the average and the balance orientations, can be obtained. See Yu (1973, 1985), André and Romero (2008) for technical details. However, it is important to note that, for computational reasons, the determination of the compromise solutions is usually made with the help of control parameter λ instead of metric p. In fact with this strategy it is only necessary to solve linear programming models of a moderate size.

Following this type of argument, it is important to realize that, in a sustainability context, the optimization of the averages may provide problematic solutions. This idea can be illustrated with the help of a simple numerical example. Thus, let us assume that I_{EC} and I_{EN} represent two criteria measuring the economic and environmental achievement of a set of natural systems, respectively. Both criteria are measured according to a scale $[0, 10]$ in the sense “more is better”, with the following scoring for three feasible systems:

$$10 + 0, \qquad 0 + 10, \qquad 4.5 + 4.5$$

It is obvious that a MDF for metric $p = 1$ or $\lambda = 1$ will choose the first or the second one (alternate optima) as being the most sustainable system. However, it is very likely that neither system will be acceptable from a sustainable perspective, the first one for environmental reasons and the second one for economic ones. However, the third one, due to its balanced character seems to be implementable in a real sustainability scenario.

The so-called paradox of the averages is crucial in many contexts and very especially within an environmental sustainability context. It is important to note that this type of paradox implicit in some MDF models also underlies several additive approaches applied in the sustainability field like the AHP or MAUT ones, when the assumption of the independence of preferences is accepted. For those reasons, the selection of a rule for the proper aggregation of a set of indicators must not be made mechanistically if we wish the composite indices obtained to be accepted in sustainability scenarios by real decision-makers.

10. Conclusions

MDFs have been profusely used for measuring the sustainability of natural systems. The respective aggregation of indicators into a composite index is undertaken by minimizing topological distances between the values achieved by the different indicators and a reference vector of desirable values. The value achieved by the composite index measures the sustainable goodness of a natural system. This type of quantitative information might be useful, among other things, for designing sensible environmental policies.

MDF is a well-established method in the management science/operational research discipline, enjoying good theoretical properties and with its application only requiring a light computational burden. However, this paper clearly shows a significant number of oddities and misconceptions when this approach is applied within a sustainability context. This paper not only detects these critical issues but also proposes ideas and methods for circumventing them. Some of them are easily addressed, such as: the normalization of the values achieved by the indicators, the preferential meaning underlying any aggregation rule, the incorporation of preferential weights, etc. On the other hand, others like the consideration of the role of the stakeholders in the analysis is still an open problem requiring more research. Anyway, the consideration by the analysts of the recommendations made in this paper will increase the accuracy and realism of the sustainability exercises undertaken with this type of approach.

Acknowledgments

The work of Carlos Romero and Luis Diaz-Balteiro was funded by the Ministry of Economy and Competitiveness of Spain under project AGL2015-68657-R. Pedro Belavenutti and Silvana Ribeiro Nobre have received funding from the SuFoRun Project ("Models and decision Support tools for integrated Forest policy development under global change and associated Risk and Uncertainty"), funded by a Marie Curie International Research Staff Exchange Scheme Fellowship within the H2020 European Union Framework Program (H2020-MSCA-RISE-2015). Comments raised by the reviewers are highly appreciated. Finally, thanks are given to Diana Badder for editing the English.

References

André, F.J., Romero, C., 2008. Computing compromise solutions: on the connections between compromise programming and composite programming. Appl. Math. Comput. 195, 1–10. https://doi.org/10.1016/j.amc.2007.04.064.
Bathrellos, G.D., Skilodimou, H.D., Chousianitis, K., Youssef, A.M., Pradhan, B., 2017. Suitability estimation for urban development using multi-hazard assessment map. Sci. Total Environ. 575, 119–134. https://doi.org/10.1016/j.scitotenv.2016.10.025.
Bausch, J.C., Bojórquez-Tapia, L., Eakin, H., 2014. Agro-environmental sustainability assessment using multicriteria decision analysis and system analysis. Sustain. Sci. 9, 303–319. https://doi.org/10.1007/s11625-014-0243-y.
Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I., 2017. Weights and importance in composite indicators: closing the gap. Ecol. Ind. 80, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056.
Blancas, F.J., Caballero, R., González, M., Lozano-Oyola, M., Pérez, F., 2010. Goal programming synthetic indicators: an application for sustainable tourism in Andalusian coastal counties. Ecol. Econ. 69, 2158–2172. https://doi.org/10.1016/j.ecolecon.2010.06.016.
Burgman, M.A., Regan, H.M., Maguire, L.A., Colyvan, M., Justus, J., Martin, T.G., Rothley, K., 2014. Voting systems for environmental decisions. Conserv. Biol. 28, 322–332. https://doi.org/10.1111/cobi.12209.
Diaz-Balteiro, L., Alfranca, O., Bertomeu, M., Ezquerro, M., Giménez, J.C., González-Pachón, J., Romero, C., 2016a. Using quantitative techniques to evaluate and explain the sustainability of forest plantations. Can. J. For. Res. 46, 1157–1166. https://doi.org/10.1139/cjfr-2015-0508.
Diaz-Balteiro, L., Alfranca, O., González-Pachón, J., Romero, C., 2016b. Ranking of industrial forest plantations in terms of sustainability: a multicriteria approach. J. Environ. Manag. 180, 123–132. https://doi.org/10.1016/j.jenvman.2016.05.022.
Diaz-Balteiro, L., Alonso, R., Martínez-Jáuregui, M., Pardos, M., 2017b. Selecting the best forest management alternative by aggregating ecosystem services indicators over time: a case study in central Spain. Ecol. Indic. 72, 322–329. https://doi.org/10.1016/j.ecolind.2016.06.025.
Diaz-Balteiro, L., González-Pachón, J., Romero, C., 2009. Forest management with multiple criteria and multiple stakeholders: an application to two public forests in Spain. Scand. J. For. Res. 24, 87–93. https://doi.org/10.1080/02827580802687440.
Diaz-Balteiro, L., González-Pachón, J., Romero, C., 2017a. Measuring systems sustainability with multi-criteria methods: a critical review. Eur. J. Oper. Res. 258, 607–616. https://doi.org/10.1016/j.ejor.2016.08.075.
Diaz-Balteiro, L., Romero, C., 2004a. Sustainability of forest management plans: a discrete goal programming approach. J. Environ. Manag. 71, 351–359. https://doi.org/10.1016/j.jenvman.2004.04.001.
Diaz-Balteiro, L., Romero, C., 2004b. In search of a natural systems sustainability index. Ecol. Econ. 49, 401–405. https://doi.org/10.1016/j.ecolecon.2004.02.005.
Diaz-Balteiro, L., Voces, R., Romero, C., 2011. Making sustainability rankings using compromise programming. An application to European paper industry. Silva Fenn. 45, 761–773. https://doi.org/10.14214/sf.103.
Doukas, H., Karakosta, C., Psarras, J., 2010. Computing with words to assess the sustainability of renewable energy options. Expert Syst. Appl. 37, 5491–5497. https://doi.org/10.1016/j.eswa.2010.02.061.
Giménez, J.C., Bertomeu, M., Diaz-Balteiro, L., Romero, C., 2013. Optimal harvest scheduling in Eucalyptus plantations under a sustainability perspective. For. Ecol. Manag. 291, 367–376. https://doi.org/10.1016/j.foreco.2012.11.045.
Gómez-Limón, J., Riesgo, L., 2009. Alternative approaches to the construction of a composite indicator of agricultural sustainability: an application to irrigated agriculture in the Duero basin in Spain. J. Environ. Manag. 90, 3345–3362. https://doi.org/10.1016/j.jenvman.2009.05.023.
Gómez-Limón, J.A., Sanchez-Fernandez, G., 2010. Empirical evaluation of agricultural sustainability using composite indicators. Ecol. Econ. 69, 1062–1075. https://doi.org/10.1016/j.ecolecon.2009.11.027.
González-Pachón, J., Romero, C., 2007. Inferring consensus weights from pairwise comparison matrices without suitable properties. Ann. Oper. Res. 154, 123–132. https://doi.org/10.1007/s10479-007-0182-4.
Gregory, G.R., 1955. An economic approach to multiple use. For. Sci. 1, 6–13.
Hickey, G.M., Innes, J.L., 2008. Indicators for demonstrating sustainable forest management in British Columbia, Canada: an international review. Ecol. Indic. 8, 131–140. https://doi.org/10.1016/j.ecolind.2006.11.005.
Ibáñez-Forés, V., Bovea, M.D., Pérez-Belis, V., 2014. A holistic review of applied methodologies for assessing and selecting the optimal technological alternative from a sustainability perspective. J. Clean. Prod. 70, 259–281. https://doi.org/10.1016/j.jclepro.2014.01.082.
Keeney, R.L., Kirkwood, C.W., 1975. Group decision making using cardinal social welfare functions. Manag. Sci. 22, 430–437. https://doi.org/10.1287/mnsc.22.4.430.
Keeney, R.L., 1976. A group preference axiomatization with cardinal utility. Manag. Sci. 23, 140–145. https://doi.org/10.1287/mnsc.23.2.140.
Kuzdas, C., Warner, B.P., Wiek, A., Vignola, R., Yglesias, M., Childers, D.L., 2016. Sustainability assessment of water governance alternatives: the case of Guanacaste Costa Rica. Sustain. Sci. 11, 231–247. https://doi.org/10.1007/s11625-015-0324-6.
Lopez-Ridaura, S., Van Keulen, H., van Ittersum, M.K., Leffelaar, P.A., 2009. Multiscale sustainability evaluation of natural resource management systems: quantifying indicators for different scales of analysis and their trade-offs using linear programming. Int. J. Sust. Dev. World 12, 81–97. https://doi.org/10.1080/13504500509469621.
Lozano-Oyola, M., Blancas, F.J., González, M., Caballero, R., 2012. Sustainable tourism indicators as planning tools in cultural destinations. Ecol. Indic. 18, 659–675. https://doi.org/10.1016/j.ecolind.2012.01.014.
Moldan, B., Janousková, S., Hák, T., 2012. How to understand and measure environmental sustainability: indicators and targets. Ecol. Indic. 17, 4–13. https://doi.org/10.1016/j.ecolind.2011.04.033.
Neumayer, E., 2013. Weak versus strong Sustainability, fourth ed. Edward Elgar, Cheltenham, UK, pp. 22–29.
Nordström, E.M., Romero, C., Eriksson, L.O., Öhman, K., 2009. Aggregation of preferences in participatory forest planning with multiple criteria: an application to the urban forest in Lycksele, Sweden. Can. J. For. Res. 39, 1979–1992. https://doi.org/10.1139/X09-107.
OECD, 2008. Organization for Economic Co-operation and Development. Handbook on Constructing Composite Indicators. Methodology and User Guide. OECD, Paris.
Pollesch, N.L., Dale, V.H., 2016. Normalization in sustainability assessment: methods and implications. Ecol. Econ. 130, 195–208. https://doi.org/10.1016/j.ecolecon.2016.06.018.
Popp, J., Hoag, D., Hyatt, E., 2001. Sustainability indices with multiple objectives. Ecol. Indic. 1, 37–47. https://doi.org/10.1016/S1470-160X(01)00006-1.
Pretzsch, J., 2014. Paradigms of tropical forestry in rural development. In: Pretzsch, J., Darr, D., Uibrig, H., Auch, E. (Eds.), Forests and Rural Development. Springer, Heidelberg, pp. 7–49.
Romero, C., 1991. Handbook of Critical Issues in Goal Programming. Pergamon Press, Oxford.
Rönnqvist, M., D'Amours, S., Weintraub, A., Jofre, A., Gunn, E., Haight, R.G., Martell, D., Murray, A.T., Romero, C., 2015. Operations research challenges in forestry: 33 open problems. Ann. Oper. Res. 232, 11–40. https://doi.org/10.1007/s10479-015-1907-4.
Ruiz, F., Cabello, J.M., Luque, M., 2011. An application of reference point techniques to the calculation of synthetic sustainability indicators. J. Oper. Res. Soc. 62, 189–197. https://doi.org/10.1057/jors.2009.187.
Ruiz, F., Cabello, J.M., Pérez-Gladish, B., 2017. Building ease-of-doing-business synthetic indicators using a double reference point approach. Technol. Forecast. Soc. https://doi.org/10.1016/j.techfore.2017.06.005.
Salvado, M.F., Azevedo, S.G., Matias, J.C.O., Ferreira, L.M., 2015. Proposal of a sustainability index for the automotive Industry. Sustainability 7, 2113–2144. https://doi.org/10.3390/su7022113.
Shmelev, S.E., 2012. Ecological Economics. Sustainability in Practice. Springer, Dordrecht, pp. 98–114.
Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An overview of sustainability assessment methodologies. Ecol. Indic. 9, 189–212. https://doi.org/10.1016/j.ecolind.2008.05.011.
Streimikiene, D., Balezentis, T., Krisciukaitienė, I., Balezentis, 2012. A Prioritizing sustainable electricity production technologies: MCDM approach. Renew. Sust. Energ. Rev. 16, 3302–3311. https://doi.org/10.1016/j.rser.2012.02.067.
Tamiz, M., Jones, D.F., Romero, C., 1998. Goal programming for decision making: an overview of the current state-of-the-art. Eur. J. Oper. Res. 111, 569–581. https://doi.org/10.1016/S0377-2217(97)00317-2.
Valentin, A., Spangenberg, J.H., 2000. A guide to community sustainability indicators. Environ. Impact Assess. Rev. 20, 381–392. https://doi.org/10.16/S0195-9255(00)00049-4.
Voces, R., Diaz-Balteiro, L., Romero, C., 2012. Characterization and explanation of the sustainability of the European wood manufacturing industries: a quantitative approach. Expert Syst. Appl. 39, 6618–6627. https://doi.org/10.1016/j.eswa.2011.12.040.
von Carlowitz, C., 1713. Sylviculture oeconomica oder haußwithliche nachricht und naturgemäßige. Anweisung zur Wilden Baum-Zucht: Johann Friedrich Braun 2. Leipzig, Germany.
Wang, H., 2015. A generalized MCDA-DEA (multi-criterion decision analysis-data envelopment analysis) approach to construct slacks-based composite indicator. Energy 80, 114–122. https://doi.org/10.1016/j.energy.2014.11.051.
Yu, P.L., 1973. A class of solutions for group decision problems. Manag. Sci. 19, 936–946. https://doi.org/10.1287/mnsc.19.8.936.
Yu, P.L., 1985. Multiple Criteria Decision Making: Concepts, Techniques and

Extensions. Plenum Press, New York.
Zhou, P., Ang, B.W., Poh, K.L., 2007. A mathematical programming approach to constructing composite indicators. Ecol. Econ. 62, 291–297. https://doi.org/10.1016/j.ecolecon.2006.12.020.
Zionts, S., Wallenius, J., 1976. An interactive programming method for solving the multiple criteria problem. Manag. Sci. 22, 652–663. https://doi.org/10.1287/mnsc.22.6.652.
Zionts, S., Wallenius, J., 1983. An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions. Manag. Sci. 29, 519–529. https://doi.org/10.1287/mnsc.29.5.519.