THE ORDINAL INPUT FOR CARDINAL OUTPUT APPROACH OF NON-COMPENSATORY COMPOSITE INDICATORS: THE PROMETHEE SCORING METHOD

SALVATORE GRECO[a,b], ALESSIO ISHIZAKA[d], MENELAOS TASIOU[b,c], and GIANPIERO TORRISI[a,c]

[a]*Department of Economics and Business, University of Catania, Corso Italia 55, 95129 Catania, Italy*
[b]*University of Portsmouth, Portsmouth Business School, Centre of Operations Research and Logistics, PO1 3DE, Portsmouth, UK*
[c]*University of Portsmouth, Portsmouth Business School, PO1 3AH, Portsmouth, UK*
[d]*Neoma Business School, 76130 Mont-Saint-Aignan, France*

ABSTRACT

Despite serious threats as to their soundness, the adoption of composite indicators is constantly growing alongside their popularity, especially when it comes to their adoption in policy-making exercises. This study presents a robust non-compensatory approach to construct composite indicators that is mainly based, at least with respect to the basic ideas, on the classic Borda scoring procedure. The non-compensatory indicators we are proposing can be seen as aggregation of ordinal non-compensatory preferences between considered units supplying a numerical cardinal comprehensive evaluation. For this reason, we define our methodology, the ordinal input for cardinal output non-compensatory approach for composite indicators. To take into account hesitation, imprecision and ill-determination in defining preference relations with respect to the elementary indices, we adopt the PROMETHEE methods, whose net flow score can be seen as an extension to the fuzzy preferences of the Borda score. Moreover, we systematically deal with robustness of the results with respect to weighting and parameters such as indifference and preference thresholds, allowing to define preference relations of elementary indices. In this regard, we couple PROMETHEE methods with the recently proposed $\sigma - \mu$ approach, which permits to explore the whole domain of feasible preference parameters mentioned above, giving a synthetic representation of the distribution of the values assumed by the composite indicators in terms of mean, μ, and standard deviation, σ. μ and σ are also used to define a comprehensive overall composite indicator. Finally, we enrich the results of this analysis with a set of graphical visualizations based on principal component analysis applied to the PROMETHEE methods with the GAIA technique, providing better understanding of the outcomes of our approach. To illustrate its assets, we provide a case study of inclusive development evaluation, based on the data of the homonymous report produced by the World Economic Forum.

Keywords: Multiple-Criteria Analysis · Composite Indicators · Non-compensatory Aggregation · PROMETHEE methods · Inclusive Development Index

---

✉ Menelaos Tasiou
    menelaos.tasiou@port.ac.uk

Salvatore Greco
salgreco@unict.it

Alessio Ishizaka
alessio.ishizaka@neoma-bs.fr

Gianpiero Torrisi
gianpiero.torrisi@port.ac.uk

1   Introduction

The adoption of composite indicators in policy analysis and public communication is constantly growing in popularity (OECD, 2008). Their use by global institutions (e.g. the OECD, World Bank, EU, etc.) and the interest shown by the media and policy-makers around the globe gave rise to their adoption in several domains of academic research, as this is witnessed by the exponential increase of studies in the literature (Greco et al., 2019a). As their name suggests, these measures provide a value that encompasses in itself the information of a set of underlying sub-indicators. Understandably, these synthetic and opaque measures could sweep methodological issues in their underlying framework under the carpet that can nonetheless largely distort the outcome. This could result in sending "*misleading, non-robust policy messages*" if they are poorly constructed (Saisana et al., 2005, p.308), while there is considerable room in their framework for "*manipulation*" (Grupp and Schubert, 2010, p.69). This is detrimental for an analysis based on such measures, and it comes naturally given the plethora of steps needed to be meticulously followed in their construction (see 'checklist' in the construction handbook provided by the OECD, 2008, p.20). Nonetheless, two steps in this checklist are arguably of utmost importance when it comes to the development process of a composite index, and these are namely the *weighting* (and, more in general, selection of the parameters required by the composite index) and *aggregation* of the sub-indicators.

These two steps are intrinsically related under some aggregation settings, and choices as to their methodological aspects may radically alter the results. The reason is that composite indicators are ultimately sole values, produced under a type of aggregation, the form of which deems which Decision Making Unit (DMU) evaluated could be 'under', or 'over'-represented (always subject to the hypotheses of the type of aggregation chosen). When it comes to choosing the type of aggregation, the difference between *compensatory* and *non-compensatory*, (Fishburn 1974; 1975; 1976, Plott 1975, Bouyssou and Vansnick 1986, Bouyssou 1986) simply boils down to whether one permits compensation among attributes, i.e. a unit can 'offset' a loss in a sub-indicator with a gain in another. Despite the so many proposals of non-compensatory composite indices (see e.g. Munda and Nardo 2009, Mazziotta and Pareto 2016, Attardi et al., 2018) in the literature, we believe that this point is rather delicate and deserves an accurate discussion.

Munda (2012, p.338) considers an example of a hypothetical sustainability index, in which a classic composite indicator setting (i.e. that of a weighted additive model) could allow trade-offs among economic growth and environmental destruction; or, a more 'extreme' case within the latter dimension, he adds: 'clean air' could compensate for a loss in 'potable water'. Understandably, and as the author acknowledges, these situations are not desirable, and this takes a developer of an index to another route: considering a non-compensatory aggregation model. Despite the prior urge of several advocates in the literature (see Munda, 2007; 2009; 2012; Billaut et al., 2010; Paruolo et al., 2013), the domain of composite indicators remained resilient, holding onto the typical weighted average (see Bandura, 2011, for an inventory of over 400 documented composite indicators evaluating a single or a group of countries jointly or individually on a socio-economic, political or environmental aspect). Still, recent proposals employing composite indicators to assess urban planning (Attardi et al., 2018) and low-carbon performance (Zhang and Zhou, 2018) offer a viable alternative to this classic setting. Both studies use ELECTRE methods (Roy, 1990; Figueira et al., 2013; 2016) as a non-compensatory aggregation method in their evaluation; though unless someone is interested in outranking relationships (as it is indeed the case in the Multiple Criteria Decision Aiding (MCDA) environment), they do not provide a sole value acting as an estimation -i.e. a literal meaning of a 'composite index'.

In this study, we introduce a novel definition of non-compensatory composite indicators based on the classic axiomatic foundations of non-compensatory preferences. In this perspective, we propose the use of an MCDA method that can be interpreted as an extension of the classic Borda score (Borda, 1781); that is the PROMETHEE family of methods (Brans and Vincke, 1985; Brans and De Smet, 2016; see Marchant, 1998 for the identification of PROMETHEE net flow score in terms of Borda score) as an effective option for constructing non-compensatory

composite indicators. Let us point out that our approach is based on the aggregation of ordinal preferences on elementary indices to get basically cardinal numerical overall evaluations. The ordinal preferences in input permit us to define this approach as non-compensatory, while the cardinal nature of the overall evaluations in output -in agreement with the basic intuitive idea of composite indicators-, allows for comparison of the difference in the overall evaluations of considered units, going far beyond their merely ordinal final ranking. We believe that these are essential characteristics of a genuine non-compensatory composite indicator and, consequently, we define our proposal as the ordinal input for cardinal output approach.

We also advocate the adoption of the SMAA-PROMETHEE variant (Corrente et al., 2014) to take into account any sources of uncertainty arising during the development of a composite index, some conceptual issues regarding the representation of the population interested in the index (Greco et al., 2018), or simply to further enhance the transparency of these opaque measures in general. In addition, we present another SMAA variant of GAIA (Mareschal and Brans, 1988) delineating cardinal information, which is well in line with the meaning of composite indicators (Booysen, 2002). To illustrate the assets of the proposed method over its compensatory alternatives, we apply it to a case study evaluating the inclusive growth and development of 108 economies based on the homonymous index produced by the World Economic Forum (WEF) (Samans et al., 2017).

The remaining of this paper is structured as follows: Section 2 provides the necessary preliminaries for this study with a discussion of the nature of non-compensatory composite indicators. Section 3 contains the proposal for a non-compensatory setting for composite indicators based on SMAA-PROMETHEE as well as a modification of SMAA-GAIA for analytical visuals. Section 4 contains a case study on the World Economic Forum's 'Inclusive Development Index', and Section 5 contains a discussion and some concluding remarks about the future direction of research.

2   Non-compensatory composite indicators

Consider a set of units $A = \{a_1, \ldots, a_n\}$ to be evaluated on a set of elementary indicators $G = \{g_1, \ldots, g_m\}$, where $g_j : A \to X_j \subseteq \mathbb{R}$, $j \in J = \{1, \ldots, m\}$. Without loss of generality, one can assume that criteria $g_j \in G$ are increasing with respect to preferences. Each unit $a \in A$ is associated with a vector $\mathbf{g}(a)$ of performances with respect to the elementary indicators, that is, $\mathbf{g}(a) = [g_1(a), \ldots, g_m(a)] \in X$, with X denoting the set of all feasible vectors of evaluations, that is, $X = X_1 \times \ldots \times X_m$. For each $g_j \in G$, a valued preference function is a function $P_j : A \times A \to [0, 1]$ such that, for all $a, a' \in A$, $P_j(a, a') = f_j(g_j(a), g_j(a'))$ with $f_j : X_j \times X_j \to [0, 1]$ being a function non-decreasing in its first argument, non-increasing in its second argument and, such that if $f(x_j, x'_j) = 1$, then $f(x'_j, x_j) = 0$ for all $x_j, x'_j \in X_j$, so that, if $P_j(a, a') = 1$, then $P_j(a', a) = 0$, and such that $f_j(x_j, x_j) = 0$ for all $x_j \in X_j$; that is $P_j(a, a) = 0$ for all $a \in A$. For all $a, a' \in A$, $P_j(a, a')$ expresses the credibility of the preference of a over a' with respect to the elementary indicator g_j. If function f_j can take only values 0 or 1, then P_j is a crisp preference relation, otherwise it is a valued or fuzzy preference relation. An overall preference is a function $P : A \times A \to [0, 1]$, such that there exist $F : [0, 1]^{2m} \to [0, 1]$ for which $P(a, a') = F(P_1(a, a'), \ldots, P_m(a, a'), P_1(a', a), \ldots, P_m(a', a))$. It is reasonable to require following conditions for function F:

- F is non-decreasing in its first m arguments, that is, for all $a, a' \in A$ and for all $g_j \in G$, the increase in the preferences $P_j(a, a')$ cannot decrease the overall preference $P(a, a')$,

- F is non-increasing in its second m arguments, that is, for all $a, a' \in A$ and for all $g_j \in G$, the increase in the preferences $P_j(a', a)$ cannot increase the overall preference $P(a, a')$,

- $F(\underbrace{1, \ldots, 1}_{m}, \underbrace{0, \ldots, 0}_{m}) = 1$, so that, for all $a, a' \in A$, if $P_1(a, a') = 1, \ldots, P_m(a, a') = 1$, then $P(a, a') = 1$, that is, if there is full preference for a over a' with respect to all $g_j \in G$, then there is also full overall preference for a over a',

- $F(\underbrace{0, \ldots, 0}_{m}, \underbrace{a_1, \ldots, a_m}_{m}) = 0$, for all $[a_1, \ldots, a_m] \in [0, 1]^m$, so that, for all $a, a' \in A$, if $P_1(a, a') = 0, \ldots, P_m(a, a') = 0$, then

$P(a, a') = 0$, that is, if there is null preference for a over a' with respect to all $g_j \in G$, then there is also null overall preference for a over a',

- if $F(\underbrace{a_1, \ldots, a_m}_{m}, \underbrace{b_1, \ldots, b_m}_{m}) = 1$, then $F(\underbrace{b_1, \ldots, b_m}_{m}, \underbrace{a_1, \ldots, a_m}_{m}) = 0$, for all $[a_1, \ldots, a_m], [b_1, \ldots, b_m] \in [0, 1]^m$, that is, if there is full preference for a over a', there must be a null preference for a' over a.

For all $a, a' \in A$, $P(a, a')$ expresses the credibility of the comprehensive preference of a over a'.

According to the definition proposed independently by Fishburn (1974; 1975; 1976) and Plott (1975) and further extensively discussed in Bouyssou (1986) and Bouyssou and Vansnick (1986), an aggregation procedure is non-compensatory if in the overall final ranking $\succsim$ the comparison of the two alternatives a and a' depends only on the two sets of criteria $\mathscr{P}(a, a')$ for which a is preferred to a' and $\mathscr{P}(a', a)$ for which a' is preferred to a. This amounts to the following assumptions:

- $P_j(a, a') \in \{0, 1\}$ for all $g_j \in G$ and all $a, a' \in A$,

- $P(a, a') \in \{0, 1\}$ for all $a, a' \in A$,

- $a \succ a'$ if and only if $P(a, a') = 1$ (with $\succ$ being the asymmetric part of $\succsim$, that is, for all $a, a' \in A$, $a \succ a'$ if and ony if $a \succsim a'$ and not $a' \succsim a$).

Observe, however, that both Fishburn (1975) and Plott et al. (1975) proved that, under some mild assumptions, the only aggregation procedure providing a weak order -that is a strongly complete and transitive binary preference relation- on the set of alternatives is the lexicographic order. This seems a rather restrictive result that would definitely close the discussion on interesting non-compensatory scoring procedures; particularly if they should be used to define a composite indicator. Indeed, giving such a great importance to the most important criterion seems to be contradicting the general philosophy of composite indicators that, instead, aims to give a comprehensive synthesis of the evaluations the units of interest get on all the elementary indicators. In this perspective, with the aim of constructing composite indicators maintaining the initial idea of non-compensatory preferences, we propose a definition of non-compensatory composite indicator $U(a), a \in A$, as aggregation for all $a' \in A - \{a\}$ of the overall non-compensatory preferences $P(a, a')$ and $P(a', a)$. As definition of non-compensatory preference we assume only the essential point that $\Pi(a, a') = V(P_1(a, a'), \ldots, P_m(a, a'), P_1(a', a), \ldots, P_m(a', a))$ with V being non-decreasing in its first m arguments and non-increasing in its second m arguments, with $V : [0, 1]^{2m} \to \mathbb{R}$ such that $\Pi(a, a') = -\Pi(a', a)$, with $\Pi(a, a')$ measuring the overall preference of a over a' if $\Pi(a, a') > 0$, and $|\Pi(a, a')| = \Pi(a', a)$ measuring the overall preference of a' over a if $\Pi(a, a') < 0$. On this basis, a non-compensatory composite indicator is a function $U : A \to \mathbf{R}$ for which there is a function $H : [0, 1]^{n-1} \to \mathbf{R}$ such that, for all $a \in A$

$$U(a) = H\left(\Pi(a, a')_{a' \neq a}\right)$$

with H satisfying the following conditions:

- H is non-decreasing in its arguments, so that, for all a, the increase of $\Pi(a, a'), a' \neq, a$ cannot decrease the overall evaluation $U(a)$,

- for any permutation π on $\{1, \ldots, n\}$ and for all $[a_1, \ldots, a_{n-1}] \in \mathbb{R}^{n-1}$,

$$H\left(a_{\pi(1)}, \ldots, a_{\pi(n-1)}\right) = H(a_1, \ldots, a_{n-1})$$

so that for any permutation σ on A, putting $\Pi^\sigma(\sigma(a), \sigma(a')) = \Pi(a, a')$ for all $a, a' \in A$, we have

$$U^\sigma(\sigma(a)) = H\left(\Pi^\sigma(\sigma(a), \sigma(a'))_{a' \neq a}\right) = H\left(\Pi(a, a')_{a' \neq a}\right) = U(a).$$

The last condition expresses neutrality, according to which the overall evaluation given by the composite indicator U does not discriminate between units because of their labels.

Observe that, the above definition of non-compensatory composite indicator can be extended considering fuzzy preferences $P_j(a, a')$, as well as fuzzy overall preferences $P(a, a'), a, a' \in A$, in which case we can speak of generalized non-compensatory composite indicators.

To illustrate the idea of the non-compensatory composite indicator we are proposing, let us consider the Borda rule (Borda, 1781), according to which each alternative $a \in A$ is assigned the following evaluation (Nitzan and Rubinstein, 1981) called Borda score:

$$U_{Borda}(a) = \sum_{g_j \in G} \left| \left\{ a' \in A : g_j(a) > g_j(a') \right\} \right|, \tag{1.1}$$

which, in case there are no ex-aequo in the order established by $g_j, j = 1, \ldots, m$ -that is there is no $a, a' \in A$ for which $g_j(a) = g_j(a')$ for all $j = 1, \ldots, m$-, can be rewritten as in Black (1976)

$$U_{Borda}(a) = \frac{\sum_{a' \in A \setminus \{a\}} \Pi(a, a')}{2} + \frac{n(m-1)}{2}, \tag{1.2}$$

with

$$\Pi(a_i, a_{i'}) = \sum_{g_j \in G} P_j(a_i, a_{i'}) - \sum_{g_j \in G} P_j(a_{i'}, a_i)$$

and $P_j(a_i, a_{i'}) = 1$ if $g_j(a_i) > g_j(a_{i'})$, and $P_j(a_i, a_{i'}) = 0$ otherwise.

In fact (1.2) holds also in case there are ex-aequo in the order established by $g_j, j = 1, \ldots, m$, provided that U_{Borda} is opportunely extended (Black 1976). Therefore, since it gives the same ranking, in the following, when considering the Borda rule, we shall refer to the following formulation of the Borda score:

$$U^*_{Borda}(a) = \sum_{a' \in A \setminus \{a\}} \Pi(a, a'), \tag{1.3}$$

Observe that, according to the above definition, the Borda score $U^*_{Borda}(a)$ is a non-compensatory composite indicator and, in particular, we have

$$H\left(\Pi(a, a')_{a' \neq a} \right) = \sum_{a' \in A \setminus \{a\}} \Pi(a, a').$$

Some remarks are now in order:

- Borda score has been already used in the domain of composite indicators since the work of Dasgupta and Weale (1992), which clearly explains the reason to prefer such aggregation procedure as follows:

  "The nature of the data being what it is for a great many of the countries, it is unwise to rely on their cardinal magnitudes. We will therefore base our comparison on ordinal measures. This way, systematic biases in claims about achievement across countries will not affect the international comparison. But first, we need an ordinal aggregator. Of the many we may devise, the one most well known and most studied is the Borda Rule."

- One can imagine to generalize the concept of non-compensatory composite indicators taking into account imprecision and inaccurate determination in the decision model, so that it is reasonable to define fuzzy preference relations $P_j : A \times A \to [0, 1], g_j \in G$ on considered criteria. Thus, for $(a, a') \in A \times A, P_j(a, a')$ gives the credibility that a is preferred over a' on criterion g_j. In this context, we can extend the concept of non-compensatory aggregation procedure admitting that the overall preference of a over a' depends on the values $P_j(a, a')$ and $P_j(a', a)$ for all $g_j \in G$. In this perspective the Borda score can be reformulated as follows:

$$\tilde{U}^*_{Borda}(a) = \sum_{a' \in A \setminus \{a\}} \left[\sum_{g_i \in G} P_i(a, a') - \sum_{g_i \in G} P_i(a', a) \right]. \tag{1.6}$$

- In the domain of social choice, where the Borda procedure has been mainly studied, anonymity is a basic assumption, so that all the "criteria" -that is, all the voters- have the same importance. Of course, this is not the case in multiple criteria decision-making situations, such as the case of definition of composite indicators. In this context, to give a specific weight to each criterion seems definitely appropriate, so that, supposing one gives the weights $w_j \geq 0$, $j = 1, \ldots, m$, $w_1 + \ldots w_m = 1$, to the criteria $g_1, \ldots, g_m$, we can redefine the Borda score U^*_{Borda} as follows:

$$U^*_{Borda}(a) = \sum_{a' \in A \setminus \{a\}} \left[\sum_{g_j \in P(a,a')} w_j - \sum_{g_j \in P(a',a)} w_j \right], \tag{1.7}$$

and, also taking into account valued preferences,

$$\tilde{U}^*_{Borda}(a) = \sum_{a' \in A \setminus \{a\}} \left[\sum_{g_j \in G} w_j P_j(a, a') - \sum_{g_j \in G} w_j P_j(a', a) \right], \tag{1.8}$$

Since formulation 1.7 is a particular case of formulation 1.8 when $P_j(a, a')$ can only take values 0 or 1 for all $g_j \in G$ and for all $(a, a') \in A \times A$, in the following we shall refer only to 1.8.

- Considering some psychological aspects of decision making such as regret (Bell, 1982; Loomes and Sugden, 1982), the specific formulation of the Borda score suggests to split the value $\tilde{U}_{Borda}(a)$ in the two components

$$\tilde{U}^+_{Borda}(a) = \sum_{a' \in A \setminus \{a\}} \sum_{g_j \in G} w_j P_j(a, a') \quad \text{and} \quad \tilde{U}^-_{Borda}(a) = - \sum_{a' \in A \setminus \{a\}} \sum_{g_j \in G} w_j P_j(a', a), \tag{1.9}$$

and interpret $\tilde{U}^+_{Borda}(a)$ and $\tilde{U}^-_{Borda}(a)$ as levels of rejoice and regret derived from preferring alternative a to other alternatives (see, e.g., Özerol and Karasakal, 2008).

- In fact, $\tilde{U}_{Borda}(a)$, $\tilde{U}^+_{Borda}(a)$ and $\tilde{U}^-_{Borda}(a)$ are the net flow score, the outflow and the inflow of the PROMETHEE methods (Brans and Vincke, 1985; Brans and De Smet, 2016), being a very well-known and appreciated family of methods for Multiple Criteria Decision Aiding (MCDA; see Ishizaka and Nemery, 2013; Greco et al., 2016). In fact, the identification of the net flow score of PROMETHEE methods with the Borda score is proposed for the first time in Marchant (1998), where a discussion on the cardinal nature of the Borda score is proposed. In this perspective, the Borda score can be seen as a function returning a real valued evaluation of considered alternatives on an interval scale, so that, if for alternatives a, b, c, d $\in A$ one has

$$\tilde{U}_{Borda}(a) - \tilde{U}_{Borda}(b) = \zeta \left(\tilde{U}_{Borda}(c) - \tilde{U}_{Borda}(d) \right),$$

hence, it is meaningful (in the sense of measurement theory, see Roberts, 1985) to say that the preference of a over b is ζ times, $\zeta \in \mathbb{R}^+$, greater than the preference of c over d. This cardinal property of the net flow score of PROMETHEE methods seems quite important for composite indicators that aim to give a numerical evaluation -and not only an ordinal ranking- to the alternatives under analysis (for a further discussion on the cardinal properties of net flow score of PROMETHEE see also Marchant, 2000). In this perspective, the approach we are proposing seems quite appealing, because it conjugates the basic ordinality of the inputs (that can be mitigated with fuzzy preferences to take into account imprecision) with basic cardinality of the output, which seems quite relevant because of the evaluations on a numerical scale expected from composite indicators. For this reason, we shall refer to our definition of non-compensatory composite indicators with the expression "ordinal input

for cardinal output approach", which, in our opinion, expresses well the basic idea and the main advantages of the proposed methodology.

On the basis of the above remarks, we propose to use PROMETHEE methods to construct non-compensatory composite indicators as detailed in the following section.

3 Basic concepts of PROMETHEE

3.1 The PROMETHEE methods

Let us briefly describe the PROMETHEE methods I & II (Brans and Vincke, 1985; Brans et al., 1986) that consist the base of our proposal and, as such, preliminaries for the upcoming sections. Consider a set of alternatives $A = \{a_1, \ldots, a_n\}$ to be evaluated according to criteria $G = \{g_1, \ldots, g_m\}$, where $g_j : A \to \mathbb{R}$, $j \in J = \{1, \ldots, m\}$. For each criterion $g_j \in G$, PROMETHEE methods use a function $P_j(a_i, a_{i'})$, $i \neq i'$ that represents the degree of preference of a_i over $a_{i'}$ on criterion g_j being a non-decreasing function of $d_j(a_i, a_{i'}) = g_j(a_i) - g_j(a_{i'})$. There are six different functions that could be chosen for each criterion by the decision-maker (hereafter, 'DM') (see Brans and De Smet, 2016, for a recent review of the PROMETHEE methods), but for the sake of simplicity we will only use the commonly-used piecewise linear function defined as follows:

$$P_j(a_i, a_{i'}) = \begin{pmatrix} 0 & \text{if} & d_j(a_i, a_{i'}) \leq q_j \\ \frac{d_j(a_i, a_{i'}) - q_j}{p_j - q_j} & \text{if} & q_j < d_j(a_i, a_{i'}) < p_j \\ 1 & \text{if} & d_j(a_i, a_{i'}) \geq p_j \end{pmatrix}, \tag{3.1.1}$$

where q_j and p_j are the indifference and preference thresholds accordingly, as these are set by the DM for each criterion $g_j \in G$. Given that each criterion g_j is assigned a weight w_j (reflecting its importance instead of a trade-off in this exercise), with $w_j \geq 0$ and $\sum_{j=1}^{m} w_j = 1$; for each pair of alternatives $(a_i, a_{i'}) \in A \times A$, PROMETHEE methods compute how much a_i is preferred over $a_{i'}$ taking into account all criteria $g \in G$ as follows:

$$\pi(a_i, a_{i'}) = \sum_{j=1}^{m} w_j P_j(a_i, a_{i'}),$$

with values of $\pi(a_i, a_{i'})$ ranging between 0 and 1. Moreover, higher values denote higher preference of a_i over $a_{i'}$ and vice versa. To compare an alternative, say a_i, with all other alternatives $a_{i'}$, $i \neq i'$, PROMETHEE methods compute the positive and negative flows as follows:

$$\phi^-(a_i) = \frac{1}{n-1} \sum_{a_{i'} \in A \setminus \{a_i\}} \pi(a_{i'}, a_i) \quad \text{and} \quad \phi^+(a_i) = \frac{1}{n-1} \sum_{a_{i'} \in A \setminus \{a_i\}} \pi(a_i, a_{i'}),$$

where $\phi^-(a_i)$ (negative flow) shows how much all the other alternatives, $a_{i'} \in A \setminus \{a_i\}$, are preferred over a_i on average, and $\phi^+(a_i)$ (positive flow) shows how much a_i is preferred over the others instead. Understandably, the smaller an alternative's, say a_i, $\phi^-(a_i)$ and the larger its $\phi^+(a_i)$, the better is its performance over all other alternatives $a_{i'} \in A \setminus \{a_i\}$ and vice versa. Understandably, PROMETHEE I gives us two bipolar scores that show the dominating and dominated status of each alternative. Ordinal inferences can be made on the basis of these two scores through the PROMETHEE I partial ranking (P^I, I^I, R^I). For instance, suppose that we would like to infer some ordinal information about two alternatives, say a_i and $a_{i'}$ on the basis of the PROMETHEE I partial ranking. That could be accomplished as follows:

$$\begin{cases} a_i P^I a_{i'} & \text{iff} & \begin{cases} \phi^+(a_i) > \phi^+(a_{i'}) \text{ and } \phi^-(a_i) < \phi^-(a_{i'}), \text{ or} \\ \phi^+(a_i) = \phi^+(a_{i'}) \text{ and } \phi^-(a_i) < \phi^-(a_{i'}), \text{ or} \\ \phi^+(a_i) > \phi^+(a_{i'}) \text{ and } \phi^-(a_i) = \phi^-(a_{i'}) \end{cases} \\ a_i I^I a_{i'} & \text{iff} & \phi^+(a_i) = \phi^+(a_{i'}) \text{ and } \phi^-(a_i) = \phi^-(a_{i'}) \\ a_i R^I a_{i'} & \text{iff} & \begin{cases} \phi^+(a_i) > \phi^+(a_{i'}) \text{ and } \phi^-(a_i) > \phi^-(a_{i'}) \\ \phi^+(a_i) < \phi^+(a_{i'}) \text{ and } \phi^-(a_i) < \phi^-(a_{i'}) \end{cases} \end{cases} , \tag{3.1.2}$$

where P^I, I^I and R^I denote *preference*, *indifference* and *incomparability* respectively. When incomparabilities among alternatives (see $aR^I b$ above) exist, the use of PROMETHEE II alleviates this issue by providing a unipolar scoring. More detailed, PROMETHEE II method computes the net-flow of bipolar (PROMETHEE I) scores for each alternative a_i as follows:

$$\phi(a_i) = \phi^+(a_i) - \phi^-(a_i), \tag{3.1.3}$$

which permits the ranking of alternatives in a complete pre-order based on the preference and indifference (P^I, I^I) among them as follows:

$$\begin{cases} a_i P^{II} a_{i'} & \text{iff} & \phi(a_i) > \phi(a_{i'}) \\ a_i I^{II} a_{i'} & \text{iff} & \phi(a_i) = \phi(a_{i'}) \end{cases} . \tag{3.1.4}$$

PROMETHEE II score (net-flow) is defined in the range $[-1, 1]$, and essentially shows on average how much an alternative a_i is preferred over all other the others $a_{i'}$, taking into account how much it is dominated at the same time. This offers a score that can be used as ordinal information (i.e. to provide a ranking) showing a complete pre-order of each alternative. However, as it will be shown in Section 5, one could use the scores instead as a means to provide cardinal information. Obviously, the higher the score the better an alternative is performing and thus preferred over the rest.

3.2 The SMAA-PROMETHEE method

Developed by Corrente et al. (2014), the SMAA-PROMETHEE method is a fusion of the classic PROMETHEE and the SMAA (see Lahdelma and Salminen, 1998; 2001) methods, designed to deal with uncertainty and imprecisions in real world decision-making problems. SMAA considers a probability distribution f_w over the space of all possible weight vectors, and two probability distributions f_q and f_p over the space of potential dominance $d \subset \mathbb{R}$ in the elementary set of indicators (i.e. d comprised of: $d_j(a_i, a_{i'}) = g_j(a_i) - g_j(a_{i'})$, $i \neq i'$, $j \in J$). Of course, imprecisions in the criteria could be modelled accordingly considering a probability distribution f_χ over the space $\chi \subset \mathbb{R}^{m \times n}$ of the alternatives' evaluations $g_j(a_i)$, with $j \in J$ and $a_i \in A$. However, in this paper we are solely engrossed with the former three sources of uncertainty, and as such, we leave this case outside the scope of this analysis.

The above-mentioned sources of uncertainty could be handled in two distinct ways. First, in the *lack of information* regarding the preferences of the DM, all three sources could be declared as uncertain, and thereby randomly estimated (uniformly, in the lack of information from the DM to suggest otherwise) in a Monte Carlo simulation environment. This would imply the creation of the following three $m \times s$ matrices to be used as inputs[1], where m is the number of criteria and $mc = 1, \dots, s$ is the number of Monte Carlo simulations[2]:

---

[1]Understandably, if one considers more sources of uncertainty (e.g. functions, imprecisions in the data etc.), the number of matrices grows accordingly.

[2]While there is no standard practice to choosing the number of simulations, i.e. parameter s, Tervonen and Lahdelma (2007) suggest a value of 10,000 simulations to be adequate for robust results.

- an $m \times s$ matrix W containing the weight vectors

$$W = \left\{ \mathbf{w} = [w_1, \ldots, w_m] : w_j \geqslant 0, j = 1, \ldots, m, \sum_{j=1}^{m} w_j = 1 \right\}, \tag{3.2.1}$$

- an $m \times s$ matrix P containing the vectors of the preference thresholds

$$P = \left\{ \mathbf{p} = [p_1, \ldots, p_m] : \min_{a_i, a_{i'}} |d_j(a_i, a_{i'})| \leq p_j, j = 1, \ldots, m \right\}, \tag{3.2.2}$$

- an $m \times s$ vector Q containing the vectors of the indifference thresholds

$$Q = \left\{ \mathbf{q} = [q_1, \ldots, q_m] : q_j \leq p_j, j = 1, \ldots, m \right\}. \tag{3.2.3}$$

The second case regards a DM that is able to provide *some* information about the sources of uncertainty. This information could then be used to adjust the above-mentioned inputs, and could regard anything from the distribution to be chosen, to restrictions in the space of possible outcomes. For instance, the DM could provide information regarding the weighting preferences among the criteria at hand. This could happen e.g. in the form of linear inequalities (e.g. $w_1 > w_2 > \ldots > w_m$), or assurance regions (e.g. $a \leq w_j \leq b$, where $a, b \in [0, 1]$, $a < b$) etc.[3] Of course, this would adjust the space of weights accordingly. For instance, in the case of the DM providing information in the form of linear inequalities, as in the example above, the space of weights would be transformed and thus matrix W would be adjusted as follows:

$$W = \left\{ \mathbf{w} = [w_1, \ldots, w_m] : w_{g_1} > \ldots > w_{g_m}, w_j \geq 0, j = 1, \ldots, m, \sum_{j=1}^{m} w_j = 1 \right\}. \tag{3.2.4}$$

Turning to the output of the SMAA-PROMETHEE method, because $\phi^+(a_i)$ and $\phi^-(a_i)$ (in the case of PROMETHEE I) or $\phi(a_i)$ (in the case of PROMETHEE II), $a_i \in A$, provide a ranking for each $\mathbf{w}$ in W, $\mathbf{q}$ in Q and $\mathbf{p}$ in P, SMAA gives the ranking of each alternative a_i for every $mc = 1, \ldots, s$. This permits computing the *rank acceptability index*, the *central weight vector* and the *pairwise winning index*. We give a brief description of their use below, though for a detailed analysis and their computation process, we refer the reader to the studies of Lahdelma and Salminen (1998; 2001) for the SMAA, and Corrente et al. (2014) for the SMAA-PROMETHEE method in particular.

- **Rank acceptability index**

  The *rank acceptability index* (RAI) essentially shows the shares of parameters (in this case q_j, p_j and w_j, $j \in J$) that give an alternative, a_i, the rth place. Suppose that we annotate RAI with b_i^r; then b_i^1 shows the shares of parameters giving the alternative a_i the 1st place. The RAIs of all alternatives are typically presented in an $n \times n$ table, where each row is an alternative and each column is the probability of it attaining a given rank, i.e. $r = 1, \ldots, n$, in the s simulations.

- **Central weight vector**

  The *central weight vector* (CWV) illustrates the weight preferences of a typical DM ($\mathbf{w}$) that makes an alternative, a_i, the most preferred. The CWVs of all alternatives are typically disclosed in an $n \times m$ table, where rows point to the alternative $a_i, i = 1, \ldots, n$ and columns illustrate the weight of criterion $j = 1, \ldots, m$.

- **Pairwise winning index**

---

[3]Obviously, the other two sources of uncertainty, namely the indifference and preference thresholds, could be treated similarly.

The *pairwise winning index* (PWI) is used to compare an alternative a_i to another one, say $a_{i'}$, showing the probability the former is preferred to the latter. It is typically disclosed in an $n \times n$ table, where each row shows the probability that this alternative beats its counterpart in a given column.

To better understand the above three SMAA outputs, we give three visuals (Figs. 1, 2, 3) reflecting the outputs of SMAA in the case of the G-10 countries' evaluation in the WEF's Inclusive Development Index (IDI) that will be formally discussed in Section 4 where we introduce the case study. For reasons of simplicity, the only source of uncertainty remains the criteria weights, while the preference function is the piecewise linear, described in eq.(3.1.1), with indifference thresholds set to 0, and preference ones set to $\max|d_j(a_i, a_{i'})|$ for each criterion g_j.

Figure 1. Central Weight Vector (CWV) for Switzerland and Sweden.

This figure shows the preferences of a typical DM as to the choices that will make Switzerland or Sweden the best-performing country (i.e. ranked 1st). The horizontal axis shows the typical weight (%) of each criterion g_j portrayed on the vertical axis. Indicators are coloured based on the higher dimension in which they belong (See Table 4 for more details).
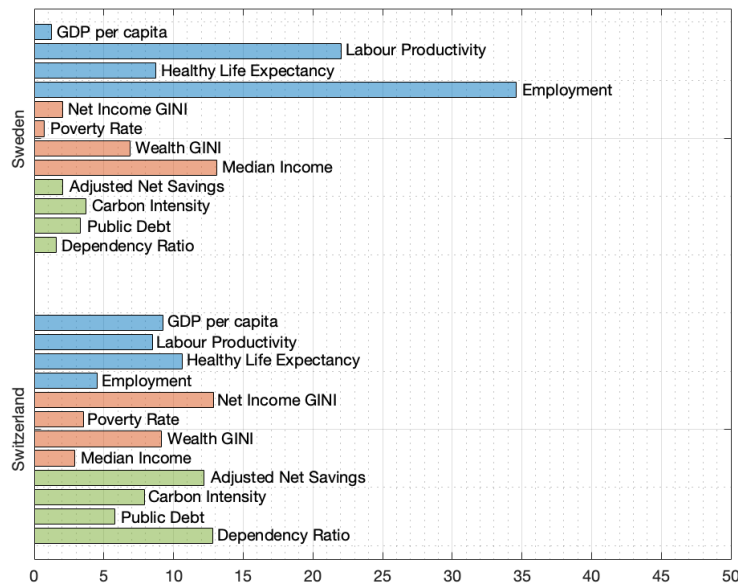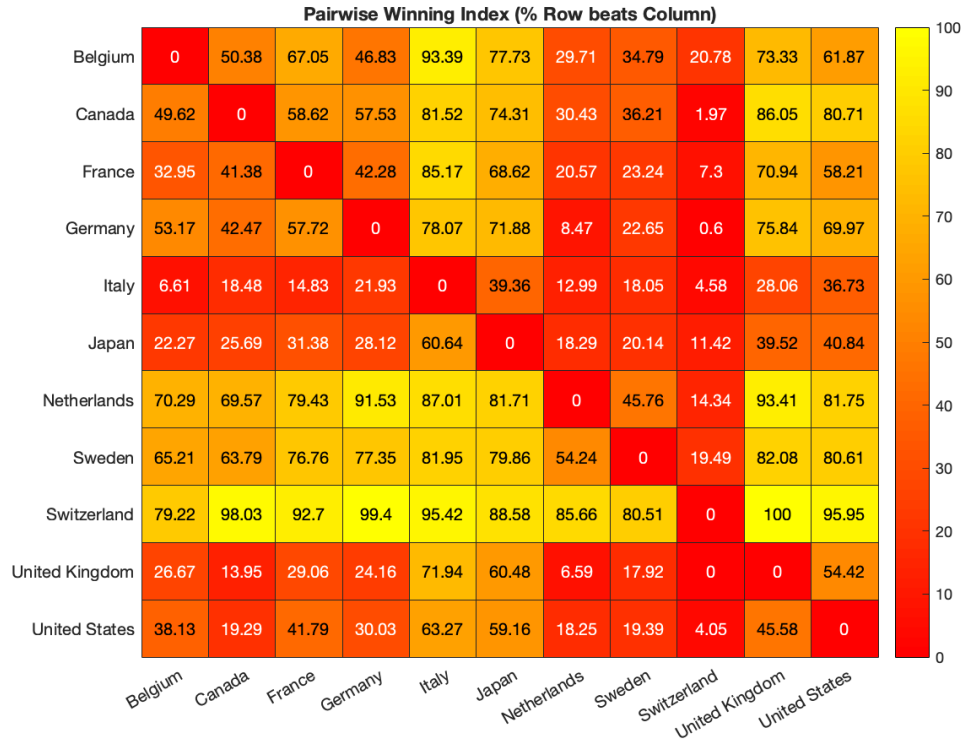

3.3 GAIA

GAIA, developed by Mareschal and Brans (1988), is a visual interactive module often implemented alongside PROMETHEE methods, and recently migrated to the AHP family of methods (Ishizaka et al., 2016). It provides DMs with a clear view of how each alternative performs in each of the considered criteria. Essentially, GAIA is an implementation of Principal Component Analysis (PCA) on the unicriterion net-flow matrix[4]. In particular, the two eigenvectors with the two largest values are selected and plotted on a 2-dimensional (most common) or a 3-dimensional (3D) plot, thus collapsing the m-dimensional space in a plot that is visually clearer to make inferences from. The 3D plot is usually preferred in cases that one may wish to explore the m-dimensional space in three coordinates (x, y, z) and get a better grip of the dynamics from the inclusion of the z-th dimension, or when the explained variance of the two eigenvectors alone is not enough by the standards of PCA to explain the original m-dimensional space; that is, the explained variance from the two principal components is less than 60%.

---

[4]As it is briefly introduced in Section 4.3, this is a $n \times m$ matrix showing the non-weighted net flows of each alternative with respect to the remaining $n-1$ alternatives (diagonal of this matrix equals 0), in each criterion g_j. Essentially, it represents how an alternative outranks ($u(a) > 0$; eq.4.3.1) or is outranked ($u(a) < 0$; eq.4.3.1) by the remaining $n-1$ alternatives in each criterion g_j.
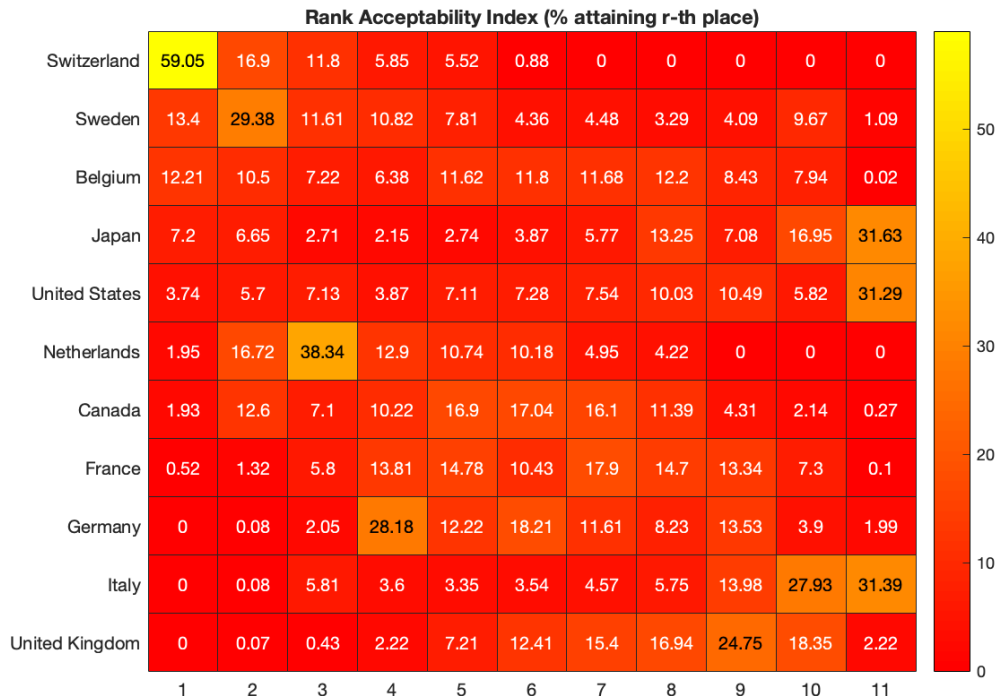
Figure 2. Pairwise Winning Index (PWI) for the G-10 countries.

This figure shows the probability that an alternative (row) beats the rest of the alternatives (in columns) (%).

Pairwise Winning Index (% Row beats Column)

	Belgium	Canada	France	Germany	Italy	Japan	Netherlands	Sweden	Switzerland	United Kingdom	United States
Belgium	0	50.38	67.05	46.83	93.39	77.73	29.71	34.79	20.78	73.33	61.87
Canada	49.62	0	58.62	57.53	81.52	74.31	30.43	36.21	1.97	86.05	80.71
France	32.95	41.38	0	42.28	85.17	68.62	20.57	23.24	7.3	70.94	58.21
Germany	53.17	42.47	57.72	0	78.07	71.88	8.47	22.65	0.6	75.84	69.97
Italy	6.61	18.48	14.83	21.93	0	39.36	12.99	18.05	4.58	28.06	36.73
Japan	22.27	25.69	31.38	28.12	60.64	0	18.29	20.14	11.42	39.52	40.84
Netherlands	70.29	69.57	79.43	91.53	87.01	81.71	0	45.76	14.34	93.41	81.75
Sweden	65.21	63.79	76.76	77.35	81.95	79.86	54.24	0	19.49	82.08	80.61
Switzerland	79.22	98.03	92.7	99.4	95.42	88.58	85.66	80.51	0	100	95.95
United Kingdom	26.67	13.95	29.06	24.16	71.94	60.48	6.59	17.92	0	0	54.42
United States	38.13	19.29	41.79	30.03	63.27	59.16	18.25	19.39	4.05	45.58	0

Figure 3. Rank Acceptability Index (RAI) for the G-10 countries.

This figure shows the probability (%) that an alternative (row) is positioned in the r-th place.

Rank Acceptability Index (% attaining r-th place)

	1	2	3	4	5	6	7	8	9	10	11
Switzerland	59.05	16.9	11.8	5.85	5.52	0.88	0	0	0	0	0
Sweden	13.4	29.38	11.61	10.82	7.81	4.36	4.48	3.29	4.09	9.67	1.09
Belgium	12.21	10.5	7.22	6.38	11.62	11.8	11.68	12.2	8.43	7.94	0.02
Japan	7.2	6.65	2.71	2.15	2.74	3.87	5.77	13.25	7.08	16.95	31.63
United States	3.74	5.7	7.13	3.87	7.11	7.28	7.54	10.03	10.49	5.82	31.29
Netherlands	1.95	16.72	38.34	12.9	10.74	10.18	4.95	4.22	0	0	0
Canada	1.93	12.6	7.1	10.22	16.9	17.04	16.1	11.39	4.31	2.14	0.27
France	0.52	1.32	5.8	13.81	14.78	10.43	17.9	14.7	13.34	7.3	0.1
Germany	0	0.08	2.05	28.18	12.22	18.21	11.61	8.23	13.53	3.9	1.99
Italy	0	0.08	5.81	3.6	3.35	3.54	4.57	5.75	13.98	27.93	31.39
United Kingdom	0	0.07	0.43	2.22	7.21	12.41	15.4	16.94	24.75	18.35	2.22

Figure 4. GAIA plane for the G-10 countries' evaluation on WEF's IDI.
This figure shows the 2D (left) and 3D (right) version of the GAIA plane. Triangles reflect the alternatives (G-10 countries). Dotted lines reflect the attributes (see Table 4 for more information). The 'decision stick' is a vector of equal weights (8.25% per criterion).
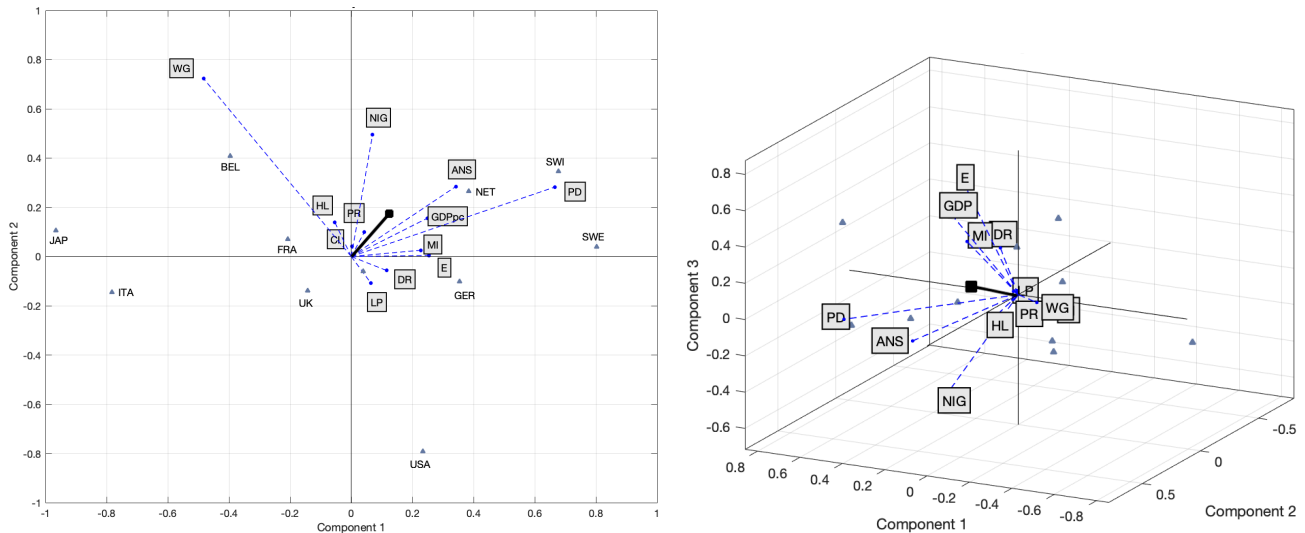
We give an example of the GAIA plane in the case of the G-10 evaluation on WEF's Inclusive Development Index (IDI) in Fig.4.

Dashed lines show the direction of each criterion g_j. If an alternative is close to or towards the same direction of a criterion, it means that it performs well on it. On the contrary, if it is plotted the opposite way (180 degrees), it means that its performance is poor on this criterion. Criteria extending in an orthogonal way between them seem to be simply unrelated to each other. The solid plotted line with the square marker reflects the 'decision stick', and is essentially the weight vector (hereby set to equal weights) of the criteria. To give an example, in Fig. 4, Switzerland (SWI) seems to perform well on criterion 'PD', and adequately well on criteria towards the same direction (i.e. 'GDP', 'MI', 'ANS')[5]. Last but not least, the variance explained with the 2D visualisation is 79.5% and just over 89.5% for the 3D version.

Given that uncertainties may arise in the decision-making process, thus making SMAA-PROMETHEE crucial in such respect, a GAIA variant dealing with uncertainty followed suit (see e.g. Hubinont, 2016; Arcidiacono et al., 2018, for extentions of GAIA to the SMAA variant of PROMETHEE and the bipolar PROMETHEE methods accordingly). Hubinont (2016) applies a bivariate kernel density on the stochastic net flows for each alternative, estimating the proportions of the projections around each noodle with the Parzen method. Arcidiacono et al. (2018) shows how a cloud of points could be plotted on the GAIA plane symbolizing the weight vectors taken into account in the SMAA evaluation. As the latter version is the one we will build upon later on in Section 4.3, we show an example of its output and a couple extensions of its reasoning in Fig. 5.

More detailed, there are two versions provided in Fig. 5 (left and right column sub-plots). The left one shows the unconstrained weight space (matrix W - eq.3.2.1), whereas the right shows how the same space is constrained as discussed in the same section (i.e. dimension 3 is more important than dimension 2, which in turn is more important than dimension 1 - eq.3.2.4). The bottom part of the figure (i.e. bottom sub-plots) show how ordinal information could enrich the SMAA-GAIA plane, illustrating for instance the weight space for which Belgium

---

[5]For more insights of the GAIA plane, we refer the interested reader to the paper of Mareschal and Brans (1988).

(blue) or Switzerland (green) is 1st.

4 PROMETHEE methods for scoring

While PROMETHEE methods are often implemented to provide ordinal information, e.g. in the form of a ranking of the considered alternatives and as discussed up to this point; they could equally be used to provide cardinal information that conveys more information about the magnitude of each alternative's performance. Composite indicators are based on this property, as they are cardinal in nature (Booysen, 2002), and as such, PROMETHEE methods could be another tool in their toolbox. In particular, PROMETHEE methods have recently been used in the field of composite indicators for robustness purposes, or to choose among alternatives of composite indicators constructed with other methods (see, e.g., De Mare et al., 2015; Antanasijevic et al., 2017; Rosic et al., 2017). Nonetheless, we would like to highlight in more detail first (subsection 4.1), how PROMETHEE methods could be used for scoring in this domain and second (subsection 4.2), and most importantly, to extend this to the case of the SMAA-PROMETHEE method that takes into account crucial issues in the construction of composite indicators. Before we begin our analysis, let us give some brief remarks/caveats that the DM should have in mind when designing composite indicators with the PROMETHEE methods.

First and foremost, we should note that the PROMETHEE methods will ensure that weights will act as 'importance coefficients' rather than trade-offs, contrary to other types of aggregation approaches (e.g. the simple additive model). This essentially eliminates the conceptual issue apparent in the development of composite indicators using additive utility aggregators, in which DMs are setting the weights as importance coefficients, while they end up being used as trade-offs between pairs of indicators. Moreover, the full compensation among criteria (apparent in the additive utility function) is now moderated according to our definition of non-compensatory aggregation given in Section 2. Nonetheless, such benefits come at a trade-off. In particular, the input required on behalf of the DM in the construction of the index is enlarged as opposed to other aggregation approaches. The reason being PROMETHEE methods require three additional choices besides the weights of the attributes; these are namely the choice of a preference function and the indifference and preference thresholds[6]. These shall be set individually for every attribute. Thus, the DM should be carefully choosing these three inputs in the creation of the index and justify them accordingly.

4.1 Developing composite indicators with the PROMETHEE I & II methods

4.1.1 Bipolar Scoring

PROMETHEE I provides a bipolar type of scoring. In particular, two outputs namely, the *positive outranking flow or outflow* ($\phi^-(a_i)$) and the *negative outranking flow* or *inflow* ($\phi^+(a_i)$) are obtained, showing two distinct scores for each alternative $a_i \in A$ for two different in principle, but essentially complementary concepts. For instance, the negative flow ($\phi^-(a_i)$) expresses in a $[0,1]$ scale how much an alternative is dominated by the remaining $n-1$ alternatives on average. A unity score in this output would indicate complete domination by all alternatives in all criteria, whereas a zero value would imply zero domination accordingly. This indicator would be in line with the theory of regret aversion or anticipated regret (see e.g. Loomes and Sugden, 1982; Bell, 1982; Fishburn, 2013), in the sense that the higher this output, the higher the regret of an individual choosing this alternative over a different option. On the other hand, the positive flow ($\phi^+(a_i)$) shows the degree of preference of an alternative over the remaining ones. Similarly to the inflow, outflow is expressed in a $[0,1]$ scale, with higher values exhibiting higher preference and vice versa.

---

[6]We should note that these are only used in five out of the six preference functions. For instance, the 'Usual' preference function does not require any kind of threshold, though it is mainly used for qualitative attributes. Additionally, in the 'Gaussian' preference function an intermediate value between q and p (namely, 's') has to be set as well to shape the curve of the Gaussian function. For a more detailed analysis, we refer the interested reader to Brans and De Smet (2016).

Figure 5. SMAA-GAIA plane for the G-10 countries' evaluation on WEF's IDI.

This figure shows the unconstrained (left) SMAA-PROMETHEE evaluation, in which all set of plausible weight vectors are sampled randomly and unconditionally, the constrained (right) evaluation, in which dimension 3 is weighted higher than dimension 2 and in turn dimension 1. Finally, the sub-figures at the bottom shows the weight vectors for which Belgium is ranked 1st, in comparison to the weight vectors for which Switzerland is ranked 1st.
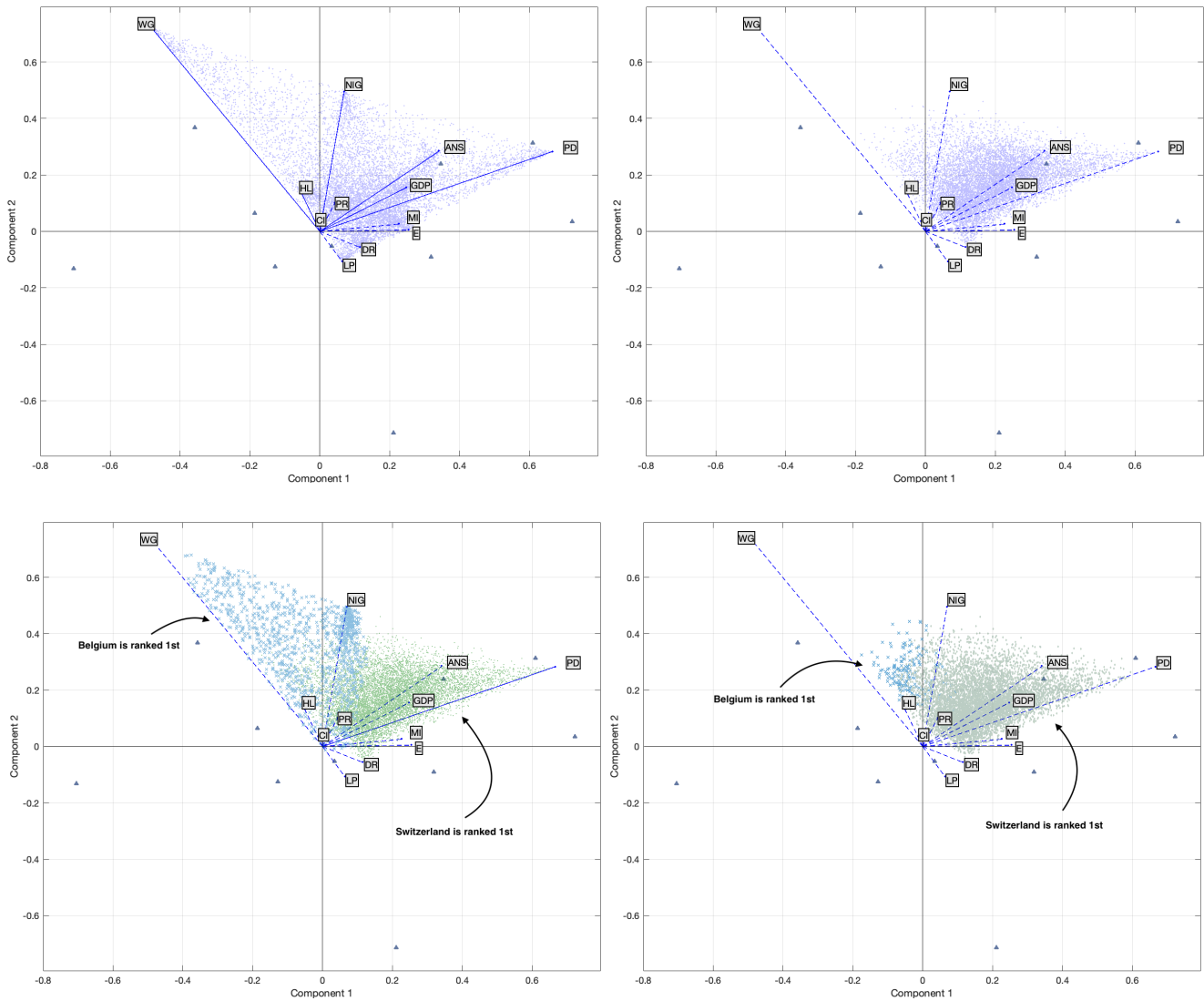
Understandably, it is not necessary that both types of flows will give the same results. If someone is solely interested in insights from the one or the other, then one could observe either. Nonetheless, if the desire is to make inferences based on these, e.g. to get an insight on the preference of an alternative over another, the intersection of the two flows should be considered to provide a 'unipolar' scoring. This is abridged in the following section.

4.1.2 Unipolar Scoring

Following on from the output of PROMETHEE I, the PROMETHEE II method provides the unipolar scoring (eq.3.1.3). It essentially consists a global score that provides a balance between the positive and the negative flows, in a sort of a net (unipolar) scoring that encapsulates both types of information discussed above; namely, the "regret" factor of choosing an alternative (i.e. $\phi^-(a_i)$) and the benefit of doing so without considering the regret factor (i.e. $\phi^+(a_i)$). The unipolar (PROMETHEE II) score bears the following two properties:

$$\begin{cases} -1 \le \phi(a_i) \le 1, \forall\ a_i \in A \\ \quad \sum_{a_i \in A} \phi(a_i) = 0 \end{cases}. \tag{4.1.2.2}$$

There is a trade-off inherent in using PROMETHEE II. That is one gains incomparability to cease, but at the cost of loss of information. For instance, considering the PROMETHEE II score of two alternatives, say a_i and $a_{i'}$: is the former preferred to the latter due to its superior performance or its lower regret? By looking at the two flows, one may infer such information (always in case of comparable alternatives (see eq.3.1.2)). As Brans and De Smet (2016, p.174) argue: "In real-world applications, we recommend to both the analysts and the decision-makers to consider both PROMETHEE I and PROMETHEE II." In fact, it is reasonable to use both types of information to get some inferences out of how the global score was constructed. Such an example can be given by looking at the PROMETHEE I & II results[7] for the G-10 countries in Table 1.

Table 1: PROMETHEE I & II scores for the G-10.

Country	ϕ	ϕ^+	ϕ^-
Switzerland	0.090	0.100	0.009
Netherlands	0.039	0.058	0.019
Sweden	0.034	0.067	0.033
Belgium	0.014	0.051	0.037
Canada	0.008	0.040	0.032
Germany	0.003	0.039	0.036
France	-0.012	0.031	0.043
United Kingdom	-0.028	0.023	0.051
United States	-0.040	0.033	0.073
Japan	-0.045	0.035	0.080
Italy	-0.064	0.020	0.084

Seemingly, United Kingdom performs better than the United States in terms of the unipolar score (that is ϕ), though we can see that this comes from its lower regret factor ($\phi^-(UK) < \phi^-(US)$) rather than its superior performance in the attributes ($\phi^+(US) > \phi^+(UK)$). Of course, in the case of PROMETHEE I, we wouldn't be able

---

[7]For reasons of simplicity, we have used equal weights across all dimensions, piecewise linear function with zero indifference thresholds (q) for all criteria, and $p_{g_j} = \max|(d_{g_j}(a_i, a_{i'}))|$. For an outline of the criteria (formally to be discussed in Section 5) see Table 4.

to make inferences about a preference relationship, as this is an example of an incomparability situation (i.e. UK R^I US).

4.2 Developing composite indicators with the SMAA-PROMETHEE methods

The issue with using the classic PROMETHEE I and II methods to construct composite indicators is that of using a precise set of parameters (i.e. **w**, **p**, **q**: eq. 3.1.1). First, it is very difficult for a DM to come up with a very precise such set of parameters for every criterion. Second, even if the DM does indeed come up with a set of parameters, this is supposed to be representative of the whole population interested in the composite indicator being provided. In brief[8], the considered set of parameters, even if it is fully justifiable by the DM setting it, remains subjective to its full extent. Of course, we should note here that MCDA is in itself inherently subjective. We are not to argue against subjectivity, rather the contrary; in exercises where it is needed, we want to make it transparent by increasing that subjectivity to involve all potential parties interested in this evaluation. As the example mentioned in Greco et al. (2019b), in an exercise involving the evaluation of a country's performance in a socio-economic aspect, the set of potential decision-makers could involve policy-makers, analysts and practitioners, or even citizens to whom the evaluation is targeted at and concern. That said, we do support, in this section and onward, that a multiplicity of viewpoints should be considered when it comes to such evaluation practices.

Generally speaking, the utilization of the SMAA variant of PROMETHEE (Corrente et al., 2014) permits the inclusion of a plethora of weight vectors, indifference and preference thresholds. In particular, as many as the number of simulations. Understandably, at the same time it creates as many outcomes and, of course, as many rankings accordingly for each unit evaluated. This is both an advantage and a drawback of this method for the creation of composite indicators. On the one hand, this increases the transparency of the evaluation process, showing the larger picture, along with which parameters give each alternative a specific place (probabilistic outcomes, see Section 4.2). This is of utmost importance in the development of composite indicators, and in fact, it is a special case of uncertainty analysis (Saisana et al., 2005) that should be accompanying the results of every composite index (OECD, 2008; Doumpos et al., 2016; Greco et al., 2019a). Indeed, the use of SMAA in this case seems alluring as it encapsulates a type of uncertainty analysis, but -perhaps most importantly- it permits dealing with the issue of the representative agent (see Greco et al., 2018, p.587) inherent in the development process of a composite index. On the other hand, this creates an issue as to the consolidation of these results into a single index that encompasses all this information. Towards the solution of this issue, Greco et al. (2019b) propose another variant in the family of SMAA called "$\sigma - \mu-$SMAA". We abridge its preliminaries in Subsection 4.2.1, though for a detailed analysis we refer the reader to the original study. Subsections 4.2.2 and 4.2.3 build upon the preceded preliminaries, adjusting the $\sigma - \mu$ approach to the PROMETHEE methods I & II respectively.

4.2.1 The Sigma-Mu approach: Preliminaries and intuition

Starting from a theoretical point of view, unlike other variants in the SMAA family, the $\sigma - \mu$ variant does not focus on probabilistic outcomes or shares of inputs leading to these outcomes accordingly. Rather, it takes into account the distribution of composite indicator values collected within SMAA for each alternative, considering its arithmetic average, μ, and its standard deviation, σ, to use those for the subsequent part of the analysis. Essentially, these two parameters illustrate the typical evaluation of an alternative -taking into account all potential decision-makers' preferences- (using μ), and the inverse robustness of that measure (using σ), larger values of which denote greater instability as to the degree of dominance of an alternative in question. To better understand the intuition behind these two parameters, Greco et al. (see 2019b, Section 5) give an example on how they could be conceptualised from a neo-Benthamite perspective in a case of a socio-economic cross-country evaluation. In particular, given that the end evaluation in their case study concerns the well-being of countries in which citizens live, one may consider each simulation $mc = 1,\ldots,s$ as an alternative set of preferences that is expressed from

---

[8]For a more detailed conversation about this issue, we refer the interested reader to the studies of Greco et al. (2018, 2019b).

different citizens. As such, *s* subjective evaluations occur from *s* different preferences, with their average score per country (μ) illustrating its typical well-being, and the standard deviation (σ) denoting a measure of well-being inequality for that country. The higher the latter is, the higher that country's inequality as regards its citizens' well-being, as there is huge dispersion to how much its citizens are satisfied.

Turning to the computation aspects, these two parameters of interest (σ, μ) can be adjusted to the SMAA-PROMETHEE outputs as follows. Assuming here a piecewise linear preference PROMETHEE function (although, without loss of generality, other preference functions can be used accordingly), and given the spaces of weights (*W*), preferences (*P*) and indifferences (*Q*) (see eqs.3.2.1 to 3.2.3); one may consider for each alternative $a_i \in A$ the PROMETHEE's *positive*, *negative* and *net* flows (i.e. $\phi^+(a_i), \phi^-(a_i), \phi(a_i)$) in this space and compute the respective arithmetic average, μ, to define a typical flow, as shown in equations 4.2.1.4a to 4.2.1.4c) below:

$$\mu_i^{\phi^+} = \int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \phi^+(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}, \tag{4.2.1.4a}$$

$$\mu_i^{\phi^-} = \int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \phi^-(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}, \tag{4.2.1.4b}$$

$$\mu_i^{\phi} = \int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \phi(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}, \tag{4.2.1.4c}$$

and the standard deviation, σ, to measure the overall dispersion as in equations 4.2.1.5a to 4.2.1.5c below:

$$\sigma_i^{\phi^+} = \sqrt{\int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \left[\phi(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) - \mu_i^{\phi^+} \right]^2 \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}}, \tag{4.2.1.5a}$$

$$\sigma_i^{\phi^-} = \sqrt{\int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \left[\phi(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) - \mu_i^{\phi^-} \right]^2 \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}}, \tag{4.2.1.5b}$$

$$\sigma_i^{\phi} = \sqrt{\int_{\mathbf{p} \in P} f(\mathbf{p}) \int_{\mathbf{q} \in Q} f(\mathbf{q}) \int_{\mathbf{w} \in W} f(\mathbf{w}) \left[\phi(a_i, \mathbf{p}, \mathbf{q}, \mathbf{w}) - \mu_i^{\phi} \right]^2 \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{w}}. \tag{4.2.1.5c}$$

Of course, in real-world situations, these integrals can be approximated via the use of a Monte-Carlo simulation. Assuming complete lack of information from the decision-maker(s), three $m \times s$ matrices **RW**, **RP** and **RQ** can be defined through unconditional random sampling, showing the attribute weights, preference and indifference thresholds in the $mc = 1, \dots, s$ simulations (according to eqs.3.2.1 to 3.2.3), with *s* being a relatively large number, as follows:

$$\underset{m \times s}{\mathbf{RW}} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1s} \\ w_{21} & w_{22} & \cdots & w_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{ms} \end{pmatrix}, \underset{m \times s}{\mathbf{RP}} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{ms} \end{pmatrix}, \underset{m \times s}{\mathbf{RQ}} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1s} \\ q_{21} & q_{22} & \cdots & q_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{ms} \end{pmatrix}.$$

Understandably, any information about the distribution or potential constraints among the attributes can shape these matrices respectively. Following their computation, they will consist the inputs to the creation of the following three $n \times s$ matrices that collect the results of PROMETHEE I & II methods accordingly:

$$\Phi^+_{n\times s} = \begin{pmatrix} \phi^+(\mathbf{a}_1,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^+(\mathbf{a}_1,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi^+(\mathbf{a}_1,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \phi^+(\mathbf{a}_2,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^+(\mathbf{a}_2,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \ldots & \phi^+(\mathbf{a}_2,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \vdots & \vdots & \ldots & \vdots \\ \phi^+(\mathbf{a}_n,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^+(\mathbf{a}_n,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi^+(\mathbf{a}_n,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \end{pmatrix}, \tag{4.2.1.6a}$$

$$\Phi^-_{n\times s} = \begin{pmatrix} \phi^-(\mathbf{a}_1,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^-(\mathbf{a}_1,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi^-(\mathbf{a}_1,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \phi^-(\mathbf{a}_2,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^-(\mathbf{a}_2,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \ldots & \phi^-(\mathbf{a}_2,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \vdots & \vdots & \ldots & \vdots \\ \phi^-(\mathbf{a}_n,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi^-(\mathbf{a}_n,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi^-(\mathbf{a}_n,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \end{pmatrix}, \tag{4.2.1.6b}$$

$$\Phi_{n\times s} = \begin{pmatrix} \phi(\mathbf{a}_1,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi(\mathbf{a}_1,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi(\mathbf{a}_1,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \phi(\mathbf{a}_2,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi(\mathbf{a}_2,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \ldots & \phi(\mathbf{a}_2,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \\ \vdots & \vdots & \ldots & \vdots \\ \phi(\mathbf{a}_n,\mathbf{w}_1,\mathbf{p}_1,\mathbf{q}_1) & \phi(\mathbf{a}_n,\mathbf{w}_2,\mathbf{p}_2,\mathbf{q}_2) & \cdots & \phi(\mathbf{a}_n,\mathbf{w}_s,\mathbf{p}_s,\mathbf{q}_s) \end{pmatrix}. \tag{4.2.1.6c}$$

Essentially, these matrices collect a representative sample of all potential values (for all three types of flows) based on all potential preferences (from weights to preference and indifference thresholds where applicable) for every alternative $a_i \in A$. Then, one may simply approximate the integrals in equations (4.2.1.4) and (4.2.1.5) for each alternative a_i, $i \in I$ computing the arithmetic mean ($\tilde{\mu}_i$) and standard deviation ($\tilde{\sigma}_i$) of each row of the matrices (4.2.1.6a) to (4.2.1.6c) accordingly ($\tilde{\mu}_i^\phi \approx \mu_i^\phi$ and $\tilde{\sigma}_i^\phi \approx \sigma_i^\phi$). For instance, for the case of ϕ, $\tilde{\mu}_i^\phi$ and $\tilde{\sigma}_i^\phi$ would equal:

$$\tilde{\mu}_i^\phi = \frac{1}{s}\sum_{mc=1}^{s}\phi(a_i,\mathbf{w}_{mc},\mathbf{p}_{mc},\mathbf{q}_{mc}), \forall i = 1,\ldots,n, \tag{4.2.1.6d}$$

$$\tilde{\sigma}_i^\phi = \sqrt{\frac{1}{s}\sum_{mc=1}^{s}\left(\phi(a_i,\mathbf{w}_{mc},\mathbf{p}_{mc},\mathbf{q}_{mc})-\tilde{\mu}_i^\phi\right)^2}, \forall i = 1,\ldots,n. \tag{4.2.1.6e}$$

The next step builds on these two parameters to arrive at an overall score by considering a definition of dominance. In particular, plotting each alternative on a 2-dimensional plane (called the $\sigma - \mu$ plane) with coordinates (σ_i, μ_i), on the basis of the concept of Pareto-Koopmans efficiency and the objective to maximize μ and minimize σ; through a set of linear programming (LP) formulations, Greco et al. (2019b) provide two types of estimators, each denoting a different concept of efficiency. These are the *local* and *global* efficiency scores, the intuition of which is explained on abstract grounds below, whilst we adjust them to the PROMETHEE I & II methods shortly afterwards.

The *local* scores (δ_{ih}) are essentially collected in vectors of efficiency measures (one vector for each unit). In particular, by decomposing the $\sigma - \mu$ plane into a family of Pareto-Koopmans frontiers (**PKF** $= \{PKF_1, PKF_2, \ldots, PKF_k\}$) it is straightforward to measure the efficiency of each unit $i \in I$ with respect to each PKF_h, $h = 1, \ldots, k$ in the plane. This is based on the concept of 'context-dependent Data Envelopment Analysis' originally developed by Seiford and Zhu (2003). Generally speaking, the local $\sigma - \mu$ efficiency scores are found by solving the following LP formulation:

$$\delta_{ih} = \underset{\alpha,\beta}{\text{Max}}\, \delta$$

$$\text{s.t.} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.2.1.7)$$
$$\begin{cases} \alpha\mu_i - \beta\sigma_i \geq \alpha\mu_{i'} - \beta\sigma_{i'} + \delta, \forall\, i' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h \\ \alpha, \beta \geq 0 \\ \alpha + \beta = 1 \end{cases}$$

A positive value (i.e. $\delta_{ih} > 0$) denotes efficiency of a unit i with respect to PKF_h, whilst negative values (i.e. $\delta_{ih} < 0$) denote inefficiency respectively, the magnitude of which is $|\delta_{ih}|$. Of course, the larger the positive (negative) value of δ_{ih} is the greater its (in)efficiency. Solutions to the LP formulation in eq.(4.2.1.7) consist the local efficiency scores δ_{ih}. These help identify how each unit is benchmarked against each frontier that is assumed to be a different context. One may think of each frontier as a different level of competition around each unit, with closer frontiers being the nearest level of competition and vice versa. Of course, on their own, local efficiencies do not give us an aggregate picture of the overall performance of a unit. To this end, *global* efficiencies (sm_i) take into account the spatial information in the plane by aggregating the local efficiencies. These are computed for each unit i as follows:

$$sm_i = \sum_{h=1}^{k} \delta_{ih}. \qquad\qquad\qquad\qquad (4.2.1.8)$$

Essentially, sm_i is defined in the $(-\infty, +\infty)$ space and illustrates the overall spatial dominance in the $\sigma - \mu$ plane. A value of $sm_i = 0$ shows that unit i is equally dominated as it dominates the remaining units i'. Of course, the larger the value of sm_i the greater its overall dominance and vice versa. Greco et al. (2019b) suggest normalisation of sm_i scores in the $[0, 1]$ space to resemble more the usual scale encountered in the literature of composite indicators.

Having provided the general definitions of dominance and the intuition behind each step in the $\sigma - \mu$ approach, we now tailor them accordingly to the PROMETHEE I & II methods in the following subsections.

4.2.2 Sigma-Mu applied to PROMETHEE I

In this subsection, we detail two ways the PROMETHEE I outputs can be used in the $\sigma - \mu$ approach. For both cases that we will forthwith discuss, we assume that matrices **RW**, **RP** and **RQ**, as well as matrices Φ^+ and Φ^- that were discussed in Section 4.2.1 are already computed.

In the first case, $\sigma - \mu$ can be individually applied to the two flows computed with SMAA-PROMETHEE I and collected in Φ^+ and Φ^-. In particular, for each alternative $a_i \in A$, two pairs of coordinates can be obtained, $(\sigma_i^{\phi^+}, \mu_i^{\phi^+})$ and $(\sigma_i^{\phi^-}, \mu_i^{\phi^-})$ accordingly, which summarize the distributions of the evaluations of each alternative in the two matrices Φ^+ and Φ^- respectively. The case of the positive flow is straightforward in the sense that it is in complete agreement with the LP formulation in eq. (4.2.1.7). That is, $\mu_i^{\phi^+}$ should be maximised as it denotes the overall score of dominance for alternative a_i (with respect to the remaining $n-1$ alternatives) taking into account all potential preferences declared in the SMAA evaluation, i.e. **w**, **p** and **q**. On the contrary, $\sigma_i^{\phi^+}$ shall be minimized, as it denotes an inverse measure of robustness such that the larger it is, the more disperse the alternative's evaluations ($\phi^+(\mathbf{a}_i, \mathbf{w}_{mc}, \mathbf{p}_{mc}, \mathbf{q}_{mc}), mc = 1, \ldots, s$). The underlying reason is that it relies on a particular set of preferences to exhibit a great performance, with slight deviations from this set radically altering this alternative's score. That said, formulation (4.2.1.7) can be simply adjusted to the current mathematical notation written below, with everything else (concept-wise) remaining the same.

$$\delta_{ik}^+ = \underset{\alpha^+,\beta^+}{\text{Max}} \, \delta^+$$

$$\text{s.t.} \quad \begin{cases} \alpha^+ \mu_i^{\phi^+} - \beta^+ \sigma_i^{\phi^+} \geq \alpha^+ \mu_{i'}^{\phi^+} - \beta^+ \sigma_{i'}^{\phi^+} + \delta^+, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h \\ \alpha^+, \beta^+ \geq 0 \\ \alpha^+ + \beta^+ = 1 \end{cases} \qquad (4.2.2.1)$$

Solutions to (3.2.2.1) provide the local efficiencies for Φ^+ (i.e. δ_{ik}^+) for every $PKF_h, h = 1,\ldots,k$. Of course, these can be then aggregated to compute the SMAA-PROMETHEE I global positive flow efficiencies, sm_i^+, as in eq.4.2.1.8, i.e.:

$$sm_i^+ = \sum_{h=1}^k \delta_{ik}^+. \qquad (4.2.2.2)$$

These global positive flow efficiencies provide a more holistic score that encapsulates the SMAA-PROMETHEE I positive flow scores, as well as the spatial information of the $\sigma - \mu$ plane into a single value. As they are defined in the $(-\infty, +\infty)$ space, one may re-scale them to vary in the $[0,1]$ range (e.g. through 'min-max' normalization) to better resemble the classic PROMETHEE I scale of outflows ϕ^+.

Turning to the negative flow of PROMETHEE I, as discussed in Section 4.1.1, it is in line with the theory of regret aversion (or anticipated regret). For instance, a score of $\phi^-(a) = 1$ means that an alternative is dominated by all remaining ones and in all criteria, so this would certainly be a regretful decision over other, better alternatives. In particular, defined in the $[0,1]$ space, one may think of $\phi^-(a)$ as a number, a high value of which means the regret factor (by not choosing a different alternative with a lower $\phi^-(a)$ value) is increasing. As $\sigma - \mu$ analysis provides efficiency scores, its intuition in the case of the negative flow is that of a 'regret' measure. Thus, the LP formulation as described in eq. (4.2.1.7) -adjusted for the notation of the inflow- is the following:

$$\delta_{ik}^- = \underset{\alpha^-,\beta^-}{\text{Max}} \, \delta^-$$

$$\text{s.t.} \quad \begin{cases} \alpha^- \mu_i^{\phi^-} - \beta^- \sigma_i^{\phi^-} \geq \alpha^- \mu_{i'}^{\phi^-} - \beta^- \sigma_{i'}^{\phi^-} + \delta^-, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h \\ \alpha^-, \beta^- \geq 0 \\ \alpha^- + \beta^- = 1 \end{cases} \qquad (4.2.2.3)$$

the solutions to which provide the PKF_h and the individual local efficiencies for every unit. We should note here that higher local "efficiencies" mean higher regret and vice versa. That said, the global efficiencies are computed accordingly as follows:

$$sm_i^- = \sum_{h=1}^k \delta_{ik}^-. \qquad (4.2.2.4)$$

Observe that, in this interpretation, we are considering the standard deviation $\sigma^{\phi^-}(a_i)$ as a measure of dispersion of the negative flow score $\phi^-(a)$ that it is preferable to be increased, which is in agreement with experimental evidence of prospect theory (Kahneman and Tversky, 1979, 1984; Tversky and Kahneman, 1981) for

Figure 6. The $\sigma - \mu$ plane for each flow.

This figure shows how the G-10 countries are evaluated in the $\sigma - \mu$ plane based on their *positive* (left) and *negative* (right) flows. Note: Efficiency in the negative flows means more regret. Both axes are standardized using the Z-scores.
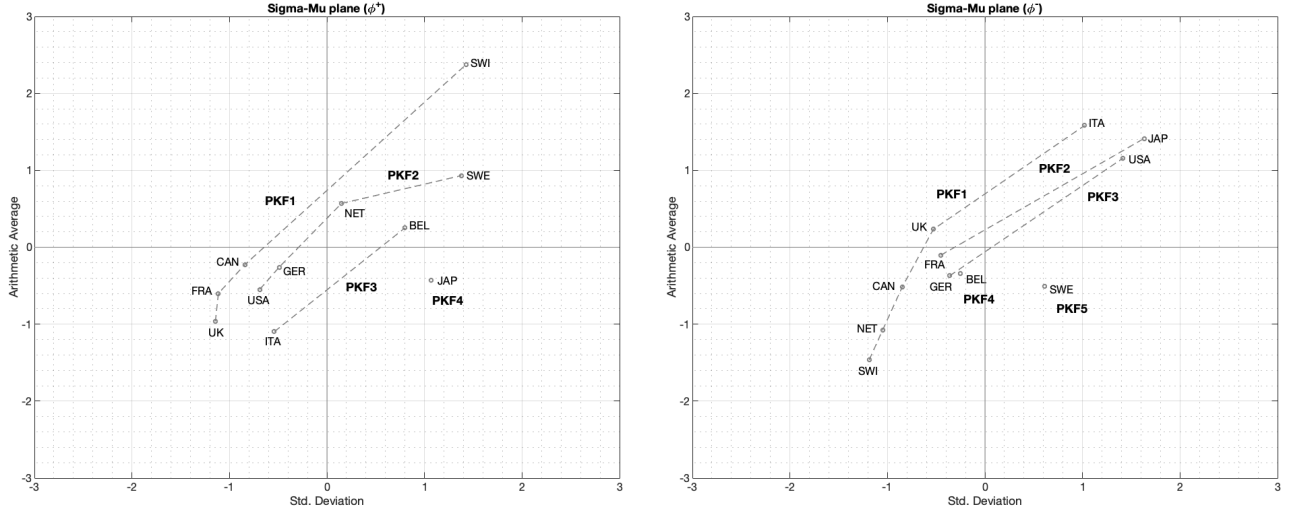
which people are risk averse in case of gains and risk-seeking in case of losses. Indeed, $\phi^+(a_i)$ can be considered as a gain so that the greater $\sigma_i^{\phi^+}$, the smaller the global score sm_i^+ which is undesirable. Whereas, $\phi^-(a_i)$ can be considered as a loss so that, the greater $\sigma_i^{\phi^-}$, the smaller the global score sm_i^-, which is desirable. Observe, however, that while there is definitely a natural tendency to be risk averse for the gains and, consequently, in our context, to reduce $\sigma_i^{\phi^+}$; this is not the case for the risk seeking in case of losses, because also to reduce the variability (in our context $\sigma_i^{\phi^-}$) could be reasonable in line with limitation of greater losses, as it is the case in finance when measures of risk are minimized (see, for example Jorion, 2000; Artzner et al., 1999). In this case, the constraints comparing unit i with all other units i' should be reformulated as follows:

$$\alpha^- \mu_i^{\phi^-} + \beta^- \sigma_i^{\phi^-} \geq \alpha^- \mu_{i'}^{\phi^-} + \beta^- \sigma_{i'}^{\phi^-} + \delta^-, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h.$$

We shall adopt the latter perspective in this section when we shall define of an overall efficiency index taking into account both positive flows ϕ_i^+ and negative flows ϕ_i^-.

In Fig. 6 we show a side-by-side evaluation of the G-10 countries as to their *positive* ($\phi^+(a)$) and *negative* ($\phi^-(a)$) flows, the (normalised using 'min-max') global scores of which are given in Table 2. Essentially, these two scores (sm_i^+ and sm_i^-) are the more holistic equivalent of the $\phi^+(a_i)$ and $\phi^-(a_i)$ outputs in the PROMETHEE I method, in the sense that they encapsulate the whole space of preferences, as this is proxied by the defined criteria weights and respective preference and indifference thresholds accounted for within SMAA.

According to the output in Fig. 6, there exist four PKF in the left plane (i.e. Sigma-Mu ϕ^+), and five PKF in the right one (i.e. ϕ^-). Global efficiencies as to each flow are provided in Table 2. As previously discussed, the $\sigma - \mu$ positive flow global score (i.e. sm^+) is a mere measure of performance evaluation that takes into account three key objectives: (i) the overall performance of a DMU (i.e. μ), (ii) how balanced its performance is to satisfy all potential viewpoints taken into account in the evaluation phase (i.e. σ), as well as (iii) how (in)efficient it is with respect to its inner (closer) and outer (further) competition (i.e. δ_{ih}, not reported in Table 2 to conserve space) as proxied by the PKFs. In that regard, according to the sm^+ metric, Switzerland is seemingly the best performing

Table 2: Global efficiencies for the G-10 countries.

Positive Flow (ϕ^+)			Negative Flow (ϕ^-)		
Country	sm^+	Rank	Country	sm^+	Rank
Switzerland	1.000	1	Italy	1.000	1
France	0.403	2	United States	0.888	2
United Kingdom	0.399	3	Japan	0.855	3
Sweden	0.392	4	Switzerland	0.777	4
Canada	0.377	5	Netherlands	0.700	5
Netherlands	0.343	6	Canada	0.614	6
Italy	0.331	7	United Kingdom	0.597	7
United States	0.324	8	France	0.469	8
Germany	0.306	9	Belgium	0.409	9
Belgium	0.249	10	Germany	0.376	10
Japan	0.000	11	Sweden	0.000	11

G-10 Country, followed by France, UK and Sweden. Japan is placed last, and is preceded by Belgium and Germany. What is noteworthy, global scores (sm^+) show that top-performer (i.e. Switzerland) aside, most countries are very close performance-wise. For instance, the difference between France and UK (ranked 2nd and 3rd accordingly) is just a mere 1% ($sm^+_{\text{France}} = 0.403$, $sm^+_{\text{UK}} = 0.399$), with more or less similar score differences for the remaining countries.

Turning to the negative flow evaluations, we are now seeing the opposite picture of performance, i.e. that of regret. As discussed in this section, the $\sigma - \mu$ global score of the negative flow (sm^-) differs from the above-discussed output in two ways. First, its first component (μ) shows an evaluation of regret instead of performance, which is essentially the other side of the coin in that an alternative is dominated by all remaining ones. Yet, its second component (σ) shows the variability in how this regret changes taking into account all potential viewpoints in the evaluation. In particular, how slight deviations in the preferences of a DM may vastly increase or decrease the regret factor, proving this alternative as highly sensitive and imbalanced. Here, the scores (sm^-, Table 2) show a higher variability compared to the positive flow previously discussed. Seemingly, the most regretful alternative from the G-10 countries seem to be Italy, followed by the US and Japan. On the other side of this ranking lies Sweden, which is seemingly the least regretful country.

Although both above outputs are greatly informative on their own to obtain a better insight about the sheer performance or regret of each alternative compared to the remaining ones; one thing worth noting is the following. These two rankings or magnitudes presented in Table 2 closely follow the partial rankings denoted in (3.1.2). That is, in order for an alternative to be preferred to another one, it has to dominate in at least one flow, and weakly dominate in the other one. However, this means that a lot of inconsistencies could arise, making real-world scenarios (involving a large number of alternatives) difficult to process. For this reason, a unipolar scoring taking into account the two flows' distributions simultaneously could be computed. This can be accomplished by combining equations (4.2.2.1) and (4.2.2.3), forming a different LP formulation that could take into account both previous formulations as follows:

$$\delta_{ih}^{PI} = \underset{\alpha^-,\alpha^+,\beta^-,\beta^+}{\text{Max}} \delta^{PI}$$

$$\text{s.t.} \tag{4.2.2.5}$$

$$\begin{cases} \alpha^+ \mu_i^{\phi^+} - \alpha^- \mu_i^{\phi^-} - \beta^+ \sigma_i^{\phi^+} - \beta^- \sigma_i^{\phi^-} \geq \alpha^+ \mu_{i'}^{\phi^+} - \alpha^- \mu_{i'}^{\phi^-} - \beta^+ \sigma_{i'}^{\phi^+} - \beta^- \sigma_{i'}^{\phi^-} + \delta^{PI}, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h \\ \alpha^-,\alpha^+,\beta^+,\beta^- \geq 0 \\ \alpha^- + \beta^- = 1 \\ \alpha^+ + \beta^+ = 1 \end{cases},$$

where μ^{ϕ^+} is the overall dominance score that is supposed to be maximised, whilst its dispersion, σ^{ϕ^+}, should be minimised as larger values denote instability due to the change of preferences. Likewise, μ^{ϕ^-} shall be minimised as it denotes the overall regret in the whole space of preferences and so does its dispersion, i.e. σ^{ϕ^-}. The reason is that, if one wants to minimize the regret factor of an alternative, both its average regret and its dispersion need to be minimized to achieve a more balanced and non-regretful performance.

This LP formulation does indeed take into account both flows and allows some flexibility on the trade-offs between each flow's μ and σ parameters. The global scores (sm_i^{PI}; arising from the summation of δ_{ih}^{PI}) are a 'loose' global evaluation, in the sense that they permit some flexibility on how each flow is taken into account, as do note that in the absence of further constraints, α or β of a particular flow could be zero. If one does not wish to permit such a possibility, and in accordance to the ability to give a complete pre-order inherent in PROMETHEE II; we provide a stricter, though more straightforward formulation in the following section, where we show how the $\sigma - \mu -$SMAA approach can be applied to PROMETHEE II directly, which takes both flows implicitly into account. Let us note that we have hereby defined the global scores sm_i^{PI} in the perspective of risk aversion both for gains, $\phi^+(a_i)$, and for losses, $\phi^-(a_i)$. Indeed, the greater $\sigma_i^{\phi^+}$ and $\sigma_i^{\phi^-}$, the smaller the global score sm_i^{PI} which is not desirable. Of course, the same index could be computed (obtaining different results) in the perspective of the prospect theory, with risk aversion in case of gains and risk-seeking in case of losses, so that, the greater $\sigma_i^{\phi^+}$ and the smaller $\sigma_i^{\phi^-}$, the smaller the global score sm_i^{PI}. In this case, the constraints comparing unit i with all other units i' should be reformulated as follows:

$$\alpha^+ \mu_i^{\phi^+} - \alpha^- \mu_i^{\phi^-} - \beta^+ \sigma_i^{\phi^+} + \beta^- \sigma_i^{\phi^-} \geq \alpha^+ \mu_{i'}^{\phi^+} - \alpha^- \mu_{i'}^{\phi^-} - \beta^+ \sigma_{i'}^{\phi^+} + \beta^- \sigma_{i'}^{\phi^-} + \delta^{PI}, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h.$$

4.2.3 Sigma-Mu applied to PROMETHEE II

Assuming all necessary steps to apply SMAA-PROMETHEE II to a dataset are accomplished -that is: matrices **RW**, **RP** and **RQ** are constructed to compute matrices Φ^+, Φ^- and eventually Φ-, it is straightforward to compute the two parameters of interest, μ^ϕ and σ^ϕ for every alternative $a_i \in A$ as in equations (4.2.1.6d) and (4.2.1.6e) accordingly. These two parameters are based on the net flows, hence the regret factor is already taken into account in the intrinsic values implicitly -$\phi(a_i, \mathbf{w}_{mc}, \mathbf{p}_{mc}, \mathbf{q}_{mc}), \forall i = 1, \ldots, n, mc = 1, \ldots, s$- and thus, in the parameters proxying their distribution, i.e. $\mu_i^\phi, \sigma_i^\phi$. This is the fundamental difference with LP formulation (4.2.2.5), which includes all components and thus the possibility to a more 'loose' trade-off among the two flows. That said, the LP formulation for the $\sigma - \mu - $SMAA-PROMETHEE is the following:

$$\delta_{ih}^{PII} = \underset{\alpha, \beta}{\text{Max}} \, \delta^{PII}$$

$$\text{s.t.} \qquad\qquad\qquad\qquad \cdot \qquad\qquad\qquad (4.2.3.1)$$
$$\begin{cases} \alpha\mu_i^\phi - \beta\sigma_i^\phi \geq \alpha\mu_{i'}^\phi - \beta\sigma_{i'}^\phi + \delta^{PII}, \forall i' \in A \setminus \bigcup_{h=1}^{k-1} PKF_h \\ \alpha, \beta \geq 0 \\ \alpha + \beta = 1 \end{cases}$$

with the global net flow efficiencies (sm_i^{PII}) arising naturally as the sum of the local efficiencies (i.e. δ_{ih}^{PII}) for every alternative $i \in I$.

Considering again the G-10 data presented up to this point as an illustrative example, five PKF were found solving (4.2.3.1), which are illustrated in Fig. 7. Global efficiencies (sm^{PII}) are given in Table 3. A comparison with the more 'loose' formulation (sm^{PI}, eq. 4.2.2.5) is provided, with both rankings being fairly similar. Furthermore, for reasons of comparability, we provide two outputs of the SMAA-PROMETHEE II method (Corrente et al., 2014); the expected rank (see 'holistic acceptability index' in Lahdelma and Salminen, 2001, p.449) and the rank with the highest probability (i.e. highest 'rank acceptability index' - or modal rank). As far as the ranking of G-10 countries is concerned, SMAA-PROMETHEE II provides a probabilistic one (as also visualised in Fig. 3), yet at this case it is fairly inconclusive. The reason is that the highest ranking acceptability index is that of Switzerland attaining the first rank with a 59.05% probability, with all the remaining alternatives achieving probabilistic ranking with a certainty between 12.21% and 38.34% (Table 3). This is admittedly a very low probability to be acceptable evidence of a country being ranked at that place.

However, this also highlights how the $\sigma - \mu$ efficiency analysis method complements the SMAA variant of PROMETHEE by permitting encapsulation of the vagueness associated with a ranking (for ordinal outcomes) or the magnitude (for cardinal outcomes) of an alternative into a single value. Of course, that is not to say that uncertainty analysis should be neglected, rather the contrary. One might obtain the single estimators, but can always go back to the SMAA-PROMETHEE II outputs to obtain more interesting insights of how this single value was obtained, as well as comparative benchmarks (such as the pairwise winning index or ranking acceptability indices) that give the DM a more evident overview of what lies underneath these values, as well as comparative insights between the alternatives being evaluated; particularly when it comes to their visual exploration through the SMAA-GAIA method that is introduced in the following section.

4.3   A cardinal version of SMAA-GAIA

In Section 3.3 we briefly discussed the concepts of GAIA and its SMAA variant as given in Arcidiacono et al. (2018). In this section, we complement the latter study in two ways. First, we introduce an alternative visualisation of GAIA for SMAA-PROMETHEE that displays cardinal information for a unit of interest on the plane. Second, we embed the two inputs (μ and σ) and the global output (sm) of our above proposed approach in the plane, in order to provide analytical insights of their relation to the rest of the criteria that formed them. Before we begin introducing these concepts, let us briefly give a few necessary preliminaries.

Consider the following $n \times m$ unicriterion flow matrix U:

Figure 7. The $\sigma - \mu$ plane for net flow.

This figure shows how the G-10 Countries are evaluated in the $\sigma - \mu$ plane based on their *net* flows. Both axes are standardized using the Z-scores.
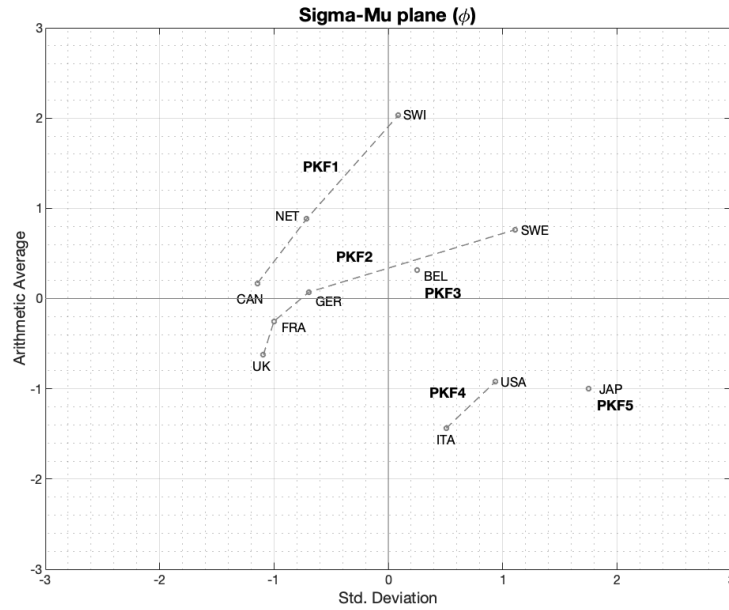

Table 3: Global efficiencies for the G-10 Countries

Country	$\sigma - \mu$ PII		$\sigma - \mu$ PI		SMAA-PROMETHEE II		
	sm^{PII}	Rank	sm^{PI}	Rank	Expected Rank	Rank with Highest Probability	Probability
Belgium	0.489	5	0.421	8	5	7	12.21%
Canada	0.512	4	0.516	4	5	6	17.04%
France	0.462	7	0.456	5	6	7	17.90%
Germany	0.441	8	0.442	7	6	4	28.18%
Italy	0.181	9	0.159	10	9	11	31.39%
Japan	0.000	11	0.000	11	8	11	31.63%
Netherlands	0.601	2	0.615	2	4	3	38.34%
Sweden	0.518	3	0.524	3	4	2	29.38%
Switzerland	1.000	1	1.000	1	2	1	59.05%
United Kingdom	0.467	6	0.455	6	8	9	24.75%
United States	0.128	10	0.172	9	8	11	31.29%

Note: $\sigma - \mu$ PII refers to the global efficiencies of LP formulation (4.2.3.1), whilst $\sigma - \mu$ PI refers to the global efficiencies from LP formulation (4.2.2.5). Expected rank is the rank taking into account all probabilistic outcomes (i.e. for each $i = 1,\ldots,n$: Expected Rank $= \sum_{r=1}^{n} p(i = r) \times r$), rounded to no decimals. Rank with the highest probability is, as its name suggests, the rank for which RAI is the max for each Country, i.e. the modal rank a country achieves in the SMAA. For more details on the outputs of this method, see the original paper by Corrente et al. (2014).

$$\mathbf{U}_{n \times m} = \begin{pmatrix} u_{\mathrm{a}_1,g_1} & u_{\mathrm{a}_1,g_{12}} & \cdots & u_{\mathrm{a}_1,g_m} \\ u_{\mathrm{a}_2,g_1} & u_{\mathrm{a}_2,g_{12}} & \cdots & u_{\mathrm{a}_2,g_m} \\ \vdots & \vdots & \cdots & \vdots \\ u_{\mathrm{a}_n,g_1} & u_{\mathrm{a}_n,g_{12}} & \cdots & u_{\mathrm{a}_n,g_m} \end{pmatrix}, \tag{4.3.1}$$

where: $u_{\mathrm{a}_i,g_j} = \frac{1}{n-1} \sum_{i=1}^{n} \left[P_j(\mathrm{a}_i, \mathrm{a}_{i'}) - P_j(\mathrm{a}_{i'}, \mathrm{a}_i) \right]$, $i' \neq i$. GAIA is essentially an application of PCA on $\mathbf{U}$, reducing the m-dimensional space to just a two or three-dimensional plane that is visually clear to the keen eye. In particular, consider that we want to construct a two-dimensional GAIA plane, with λ_1, λ_2 the two largest eigenvalues and $\mathbf{e_1}$, $\mathbf{e_2}$ the corresponding eigenvectors, all obtained from applying PCA to (4.3.1). Considering that the explained variance (i.e. $\delta = \frac{\lambda_1 + \lambda_2}{\sum_{r=1}^{m} \lambda_r}$) is at least 60% (Brans and Mareschal, 1995), the GAIA visual consists of a two-dimensional plane on which:

- Each criterion g_j is plotted with coordinates $(\mathbf{e}_{1(j)}, \mathbf{e}_{2(j)})$, with a line linking it to the origin of the plane, i.e. (0,0).

- Each alternative is plotted using its principal component scores as coordinates.

- The 'decision stick' is plotted using $(\mathbf{w}^\top \mathbf{e_1}, \mathbf{w}^\top \mathbf{e_2})$ as coordinate, with $\mathbf{w}$ the weight vector chosen. Again, a line connects these coordinates to the origin of the plane, i.e. (0,0).

Arcidiacono et al. (2018) propose a SMAA variant of GAIA where instead of one 'decision stick' -as in regular PROMETHEE methods-, we have one for each weight vector, all of which can be plotted on the plane (see upper two plots of Fig. 5). This shows how the preferences taken into account in the SMAA are dispersed along the criteria. Moreover, in line with the ranking acceptability indices, the authors propose highlighting those weight vectors for which an alternative is ranked at a given place (e.g. 1st, 2nd and so on place) (see lower two plots of Fig. 5) which showcases ordinal information on the plane. Building upon their contribution, we give cardinal meaning to the SMAA-GAIA plane by highlighting each weight vector with a particular color corresponding to a rich gradient that is linked to an alternative's net flow (ϕ). We forthwith explain how this is attainable and give a brief example with the G-10 countries' evaluations discussed thus far.

Consider a SMAA-GAIA representation of an alternative say a_i. The GAIA plane is defined exactly as mentioned in the list above. Now, each weight vector $\mathbf{w} \in W$ is plotted with coordinates $(\mathbf{w}^\top \mathbf{e_1}, \mathbf{w}^\top \mathbf{e_2})$ and a color in the RGB gamut of preference that depends on the alternative's net flow score for that particular vector. For instance, consider that alternative a_i takes net flow scores between 0.5 and 1 (in a normalized $[0,1]$ scale for simplicity). One could visualize this in a gradient of one's choice, e.g. black color equals 0.5, white equals 1, and every value in-between takes a linear combination of these RGB codes' values[9].

To give an example of our proposal, Fig. 8 delineates the SMAA-GAIA plane for the G-10 evaluation using SMAA-PROMETHEE, depicting the space of weight vectors included in our analysis and, based on each vector, the net flow scores that Switzerland achieves in this evaluation. According to the plot, Switzerland takes net flow values (normalized in the $[0,1]$ range) between just under 0.5 and up to 1, with the latter value being the norm. Particularly, as it is clear from the plot, it is consistently achieving a unity (top) score (yellow areas) in a vast part of the included weight vectors. Unless the preferences lean significantly more towards the 'WG', 'NIG' or 'LP' criteria (blue-cyan areas), it achieves a top, or a near top performance compared to the rest of the G-10 countries. Of course, in the software, one could use this figure in a more interactive way, e.g. by zooming in and exploring the relationships accordingly. For instance, the right subplot of Fig. 8 shows a 40% zoomed frame of the original figure.

---

[9] For instance, consider a value of 0.5 would be linked to a pure black color with an RGB code of [0,0,0], whereas a value of 1 would correspond to a pure white color, with a code of [255,255,255]. A value of 0.75 would be linearly interpolated to the RGB code of [128,128,128] which is the grey color standing right in the middle of this grayscale chart. MATLAB automatically applies a color gradient of preference (see e.g. 'colormap' function) easily implemented through its 'scatter' function.

Figure 8. The SMAA-GAIA plane.

This figure shows how the G-10 countries are evaluated in the GAIA plane. The SMAA-PROMETHEE weight vectors are plotted, colored according to the net flows Switzerland achieves based on these preferences. The right subplot is a 40% zoomed-in version of the left one.
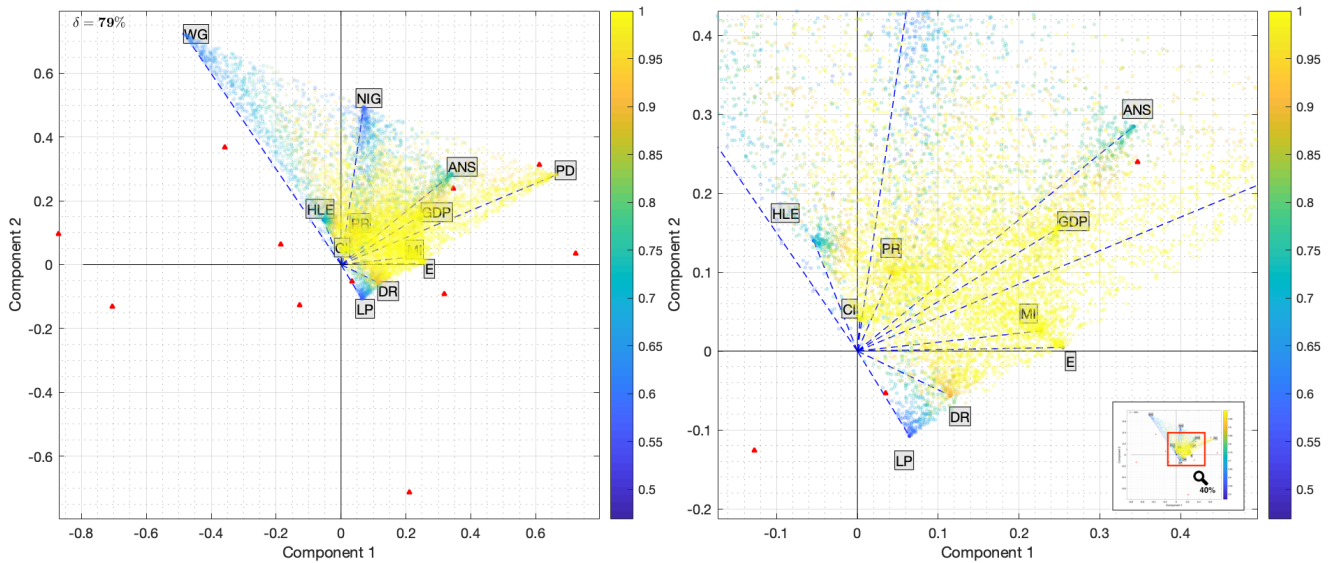
Looking at it, its clear that no matter which linear combination of weights among 'CI', 'PR', 'CR', 'MI', 'GDP' or 'PD' criteria this country is weighted more in, it still achieves a top score (pure yellow highlighted area).

While not shown due to space constraints, other variants of Fig. 8 could provide further insights. For instance, one could be interested in visualising e.g. which preference combinations would yield a score of between 0.80 and 1 for Switzerland. This would require plotting fewer weight vectors, whilst highlighting more important areas for the DM. On a similar note, the DM could be interested in those preferences that put Switzerland's score in at least the top 10th percentile, or the other way around; that is, which preferences make Switzerland performing poorly, putting its performance in the bottom 10%. Last, but not least, should the DM like to benchmark how changes in preferences affect the scores between a unit of interest (say one country) and another (say a close-performing peer of that country) this could be feasible as well (for an example see Fig. 13). What is more, it could be combined with the classic outputs of the SMAA-PROMETHEE, such as the central weight vectors, permitting the DM to see the typical preference for that space of interest.

A second interesting use of GAIA could involve visualising the relationships between the elementary criteria and the inputs (σ, μ) and global output (sm) of our proposed method. In particular, one could be interested in how the very basic 'raw material' forming the subsequent part of the analysis that we presented relate to it. Put simply, we're looking to delineate the relationships between these two sets. This does not involve any modification of GAIA at all. In particular, we can do this by horizontally concatenating a matrix containing the unicriteria net flows of the three measures (μ, σ, sm) with the matrix **U** (4.3.1). Then, as we are not interested in projecting any cloud of weight vectors, the procedure described in the beginning of this section runs with the newly formed $n \times (m + 3)$ matrix being projected (through PCA) in a two-dimensional plane. In the case of the G-10 countries' evaluation, this would produce Fig. 9.

There are a few key observations to be made from this figure. First, μ and sm seem to be driven towards the same direction, whilst σ is located exactly opposite to sm. This is of course expected as μ is supposed to

Figure 9. Sigma-Mu and the GAIA plane.

This figure shows how the sigma-mu analysis inputs (σ, μ) and global output (sm) can be embedded in the GAIA plane, providing the DM with further insights on the relationships between elementary criteria that formed those, as well as between them.
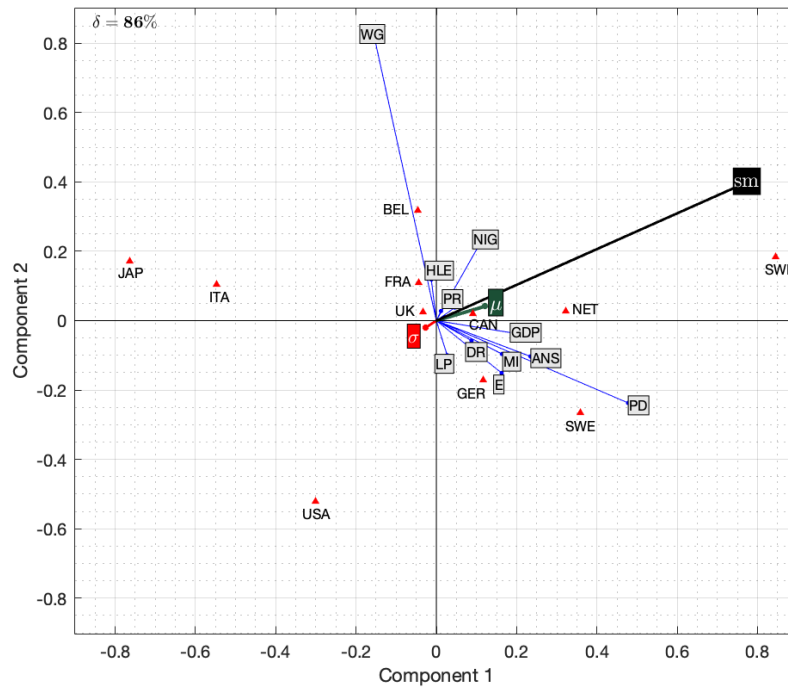
maximise an alternative's score, and σ to penalize it. Second, with the exception of 'LP', 'HLE' and 'WG', which are orthogonal to σ (thereby not relating to it), two criteria are completely in the opposite direction to σ (i.e. 'PR' and 'NIG'). This implies that the former two criteria are reducing σ in the majority of alternatives, while there's a mixed case for the remaining ones not mentioned above. Of course, similar notes can be inferred from the relationship between μ (or even sm) and the remaining criteria, though we avoid it to conserve space.

Last, but not least, embedding the three key measures outlined in this study into the GAIA plane lets us directly observe the performance (though in terms of dominance as to the remaining alternatives) of the evaluated alternatives as to both the inputs (σ, μ) of sigma-mu and the global output (sm). For instance, the further a G-10 country is located towards the same direction of σ, the bigger its dispersion is compared to the other G-10 countries. An example is that of Japan (JAP), Italy (ITA) and USA (USA) having a noticeably higher dispersion to other alternatives (e.g. the UK, CAN, FRA etc.), whilst at the same time, the Netherlands (NET) and Switzerland (SWI) seem to dominate other G-10 countries both in terms of μ and of global scores (sm).

Essentially, one could think of the above outputs (i.e. Figs. 8 and 9) as a visual aid tool in the hands of the DM in the following way. Consider that the DM is interested in evaluating the set of G10 countries, with her main interest lying in the case of Switzerland (SWI). Fig. 9 straightforwardly gives the DM the information that SWI is dominating the remaining alternatives by a great deal in terms of global output (sm). SWI has the highest μ as well, whilst should the DM wants to see how this score is achieved (e.g. is it consistent with the majority of preferences, or is it due to outliers - i.e. a good performance due to preferences concentrated in some criteria for which a unit is performing very good), one may look at Fig. 8. Of course, the same could happen for a different country, and with different variations of this plot. To give an example, Fig. 8 could be highlighting the net flow scores of a different country, or the dimensions on which SWI performs poorly (e.g. maybe a cloud of points corresponding to the bottom 10% of a country's performance), so that the policy-maker could focus on improving

those dimensions that have the greater effect lifting an alternative's performance *ceteris paribus.*

5   Case Study: The Inclusive Development Index

The need against a solely economic-oriented measure of growth, such as the GDP, is well-advocated in the literature (e.g. see, among other influential studies, Stiglitz et al., 2009; Costanza et al., 2009; Kubiszewski et al., 2013). Their advocates do not protest the use of GDP to measure economic growth, rather its association with the measurenent of a nation's welfare; something that is noted even during its very conception by Simon Kuznets (1934) dated in the far 1934. Several attempts have been made by global organizations and institutions to measure welfare individually (e.g. UNDP's 'World Happiness' report), or jointly with GDP in a more socio-economic inclusive growth index (e.g. OECD's 'Better Life Index' (BLI), WEF's 'Inclusive Development Index' (IDI)). The former is carried out in the form of surveys, while the latter two are presented as composite indicators that rank the OECD and 108 economies respectively, on the basis of 12 and 3 dimensions accordingly. Being composites of an additive type and no decisive judgement on a differential weighting, means both the BLI and IDI indicators bear the issues discussed in the introduction of this study. Interestingly, while both start from equal weighting to form their baseline results, they leave the choice of a different weight vector to the end user through their interactive platforms on their official websites. The BLI has been extensively discussed before (for a comprehensive review of the literature and a methodological proposal see Greco et al., 2017). Thus, in this study we are engrossed with WEF's IDI that we briefly describe in the following (for an extensive description, see the full report from Samans et al., 2017).

The Inclusive Development Index is hierarchical in that it consists of three dimensions, each of which contains four sub-indicators (see Table 4 for an outline). According to the report's authors (Samans et al., 2017, p.9) this set of indicators, namely 'National Key Performance Indicators' provides "*[...] a more complete picture of national economic performance than that provided by GDP alone, particularly if the ultimate objective of development is understood to be sustained, broad-based advancement of living standards rather than increased production of goods and services, per se*". They claim this index is overall useful for governments and stakeholders to determine the effect of changes in policy and conditions within a typical political cycle. Taken into consideration with the report's policy framework and metrics consisting of seven pillars - and offering a relative demonstration of institutional strength enabling environment conditions in fifteen of the most relevant policy domains for inclusive growth (see Samans et al., 2017, Fig.1)-; one could monitor both the output (that is the inclusive growth index hereby studied) and the input (that is the environment laying the foundations to inclusive growth as witnessed by the seven pillars) of each of the 108 economies analysed in the report. Of course, in this study we are interested in the analysis of the output measure, that is the inclusive development index (hereafter referred to as 'IDI').

The report's 'scoreboards' are based on equal weighting, dimension and sub-indicators-wise, which means that each dimension is given 33.3% of weight and each sub-indicator 8.25% weight accordingly. We originally construct the IDI using the PROMETHEE II method with equal weights to be consistent with the report as to the preferences on criteria importance[10]. This index will act as a comparative metric against which we will compare the results of our proposed approach. We annotate the obtained index as IDI_P.

Using the $\sigma - \mu$ SMAA-PROMETHEE methods discussed in section 4, we construct another version of IDI, taking into account the whole space of weight vectors this time. This allows the developer to extend the analysis above and beyond the issue of the representative agent inherent in the classic analysis of composite indicators (see Greco et al., 2018, p.587 for a discussion), whilst it also encapsulates a basic form of uncertainty and sensitivity analysis (see Saisana et al., 2005) that is frequently found to be missing from the development of composite indicators, despite its importance (Burgass et al., 2017). Using 10,000 randomly (uniformly) simulated weight

---

[10]For reasons of simplicity, we use the linear function, and for each criterion we set a zero indifference threshold (q) and the max of the differences among alternatives as the preference thresholds (p_{g_j}) accordingly.

Table 4: Inclusive Development Index (IDI)

Dimensions	Sub-Indicators
Growth and Development	GDP per capita (GDP)
	Labor Productivity (LP)
	Employment (E)
	Healthy Life Expectancy (HL)
Inclusion	Net Income GINI (NIG)
	Poverty Rate (PR)
	Wealth GINI (WG)
	Median Income (MI)
Intergenerational Equity and Sustainability	Adjusted Net Savings (ANS)
	Carbon Intensity (CI)
	Public Debt (PD)
	Dependency Ratio (DR)

For an extensive description of the sub-indicators and their sources, we refer the reader to the original report (Samans et al., 2017), or the official website of the WEF at: https://goo.gl/2wrF7K.

vectors as potential preferences in the SMAA-PROMETHEE approach, we apply the $\sigma - \mu$ PROMETHEE I (eq.4.2.2.5, Section 4.2.2) and PROMETHEE II (eq. 4.2.3.1, Section 4.2.3) approaches to this set of data. We do remind that these two approaches are similar, in the sense that they take into account incomparability in the evaluation, though the former is more flexible than the latter, as it gives the benefit of the doubt to the unit being evaluated as to the balance between performance and regret. Instead, the latter takes these aspects implicitly into account for all units with the same rate (hence, no flexibility -e.g. through weights α^+, β^+ and α^-, β^- - in that regard).

Carrying out the above analysis, we find that the 108 countries are scattered in 27 PKF, visualised in Fig. 10. The global scores obtained through the $\sigma - \mu$ PROMETHEE II (sm^{PII}) are delineated in a world heat-map in Fig. 11. Very similar results were obtained with the $\sigma - \mu$ PROMETHEE I approach (Spearman's correlation: 99.7%, Kendal's Tau: 98.09%) thus we do not differentiate between the two approaches, but we only report and discuss the former one. Given the fact that analysing and reporting table results for 108 countries would need a fair amount of space, we only focus on those countries that made the top 15 list (Table 5), providing the full set of results in an online supplementary appendix.

According to the results in Table 5, the rankings of the two variants of $\sigma - \mu$ applied to SMAA-PROMETHEE (i.e. IDI$_{SMPI}$ and IDI$_{SMPII}$) are identical (which is reasonably expected given their very high correlation). The top country according to its socio-economic inclusive development is Norway, something that is confirmed through all models, as well as probabilistic outcomes (i.e. SMAA-PROMETHEE II output - unreported here for brevity). In fact, the countries making it to the Top-8 list are consistently ranked at that place even with equal weights (i.e. IDI$_P$), while there's a small reshuffle experienced in the remaining seven positions. Out of the top fifteen countries, ten central and northern European countries made it to this list, another four countries from the southern hemisphere (Asia, Australia & Oceania) and one from North America (Canada). Whilst unreported in this list, United States was ranked 35th according to both variants of our proposed approach, and 30th according to the WEF's preferences (i.e. equal weights).

Understandably, providing a series of rankings as we do here raises the question of which one an interested party needs to take into account. Admittedly, there is no such thing as a 'correct' or 'false' ranking, but rather a

Figure 10. The Sigma-Mu plane.

This figure shows how the 108 countries in our sample are evaluated in the sigma-mu plane. 27 PKF are found. Axes are normalized according to their Z-scores.
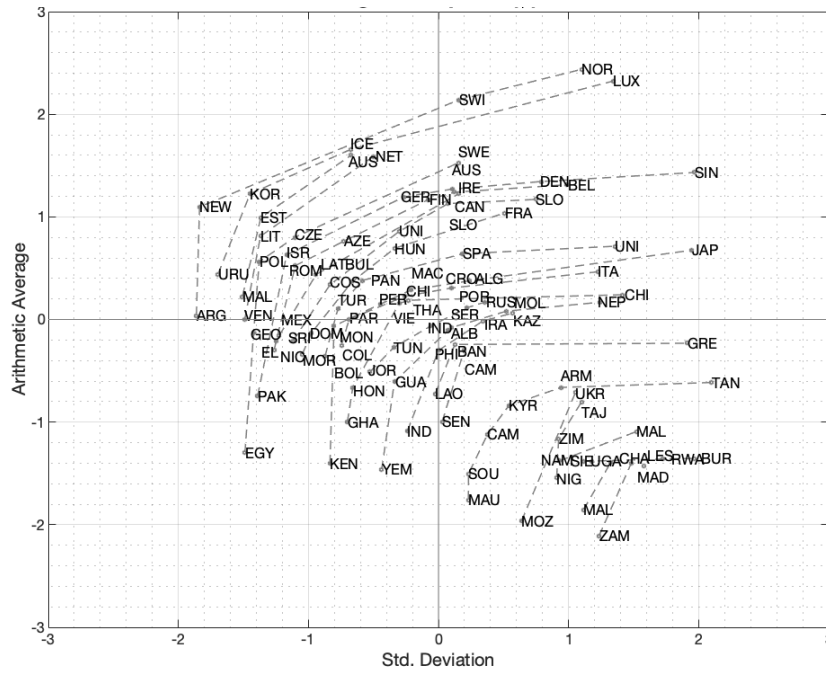
Figure 11. Inclusive Development - Global scores.

This figure shows how the 108 countries in our sample are evaluated according to their inclusive development. Both size and colour delineated in the heatmap shows the global score a country achieves according to the 12 criteria (see Table 4) the WEF provides as indicators to inclusive development.
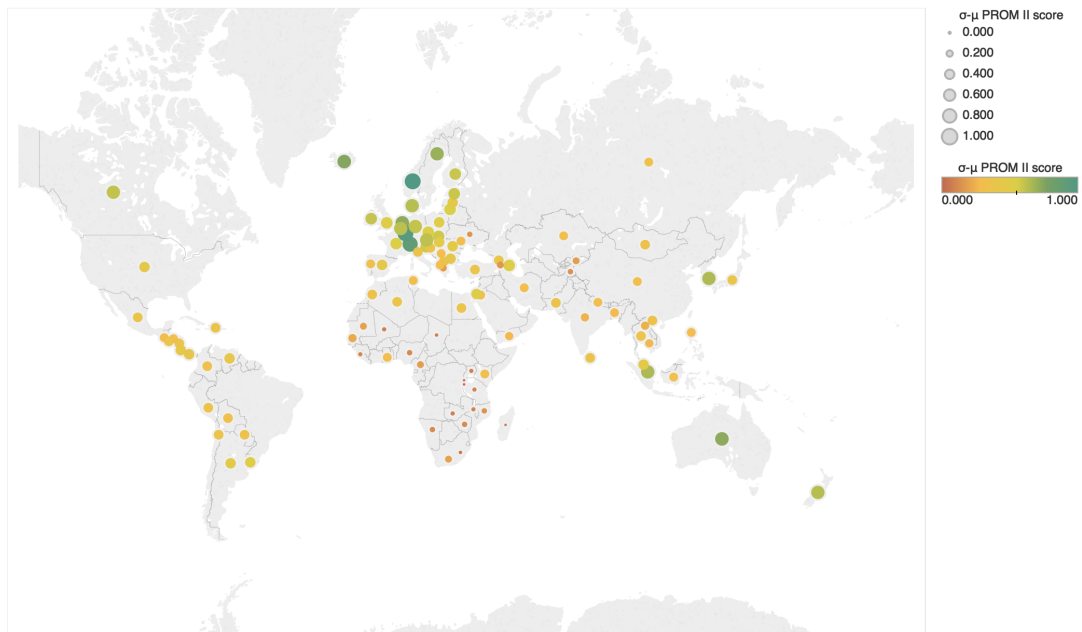
Table 5: Inclusive Development Index (IDI) for the Top 15

Country	IDI$_{SMPI}$		IDI$_{SMPII}$		IDI$_P$	
	sm^+	Ranking	sm	Ranking	ϕ	Ranking
Norway	1.00	1	1.00	1	1.00	1
Luxembourg	0.97	2	0.97	2	0.97	2
Switzerland	0.91	3	0.90	3	0.93	3
Iceland	0.75	4	0.75	4	0.83	4
Australia	0.74	5	0.73	5	0.81	5
Netherlands	0.73	6	0.72	6	0.81	6
Sweden	0.70	7	0.70	7	0.80	7
Singapore	0.66	8	0.66	8	0.78	8
Korea, Rep.	0.66	9	0.65	9	0.73	13
New Zealand	0.64	10	0.63	10	0.70	15
Denmark	0.63	11	0.63	11	0.76	9
Belgium	0.62	12	0.62	12	0.75	10
Austria	0.62	13	0.62	13	0.74	11
Canada	0.61	14	0.61	14	0.73	12
Germany	0.61	15	0.60	15	0.72	14

This table shows the estimators and the rankings of the Top-15 countries, achieved with equal weights (IDI$_P$), and taking into account the whole space of weight vectors using the $\sigma - \mu$ SMAA-PROMETHEE I (IDI$_{SMPI}$), and II (i.e. IDI$_{SMPI}$).

different underlying assumption inherent in it. We do believe that the one we provide here under the 'IDI$_{SMPII}$' label is more holistic in the sense that it implicitly takes a few important things into account: a multiplicity of viewpoints in the evaluation exercise, and spatial information about the competition surrounding each alternative in the $\sigma - \mu$ plane. Moreover, compared to a set of estimators obtained through a single weight vector, both our estimators and the rankings based on these are 'corrected' for uncertainty, as imbalanced units are being penalized more in their final global scores.

Turning to the utilization of the cardinal version of GAIA we provided in Section 4, Fig. 12 (left sub-plot) delineates how consistently Norway (ranked 1st through all specifications in Table 5) obtains a top score (normalised net flow figures in the $[0, 1]$ space) in the vast majority of the space W. It is apparent that almost no matter which linear combination between 'PR', 'HLE', 'NIG', 'LP', 'MI', 'GDP' and 'CI' (a staggering seven out of twelve criteria) is the choice of preferences, it achieves a top (unity) score. Its dominance over the 'PD' criterion is outstanding as well, whilst crucial improvement could be made with respect to 'E' and 'WG', criteria in which it is considerably dominated in by the remaining countries (where Norway's net flow score could even reach a low score of around the 40% mark).

An interesting insight that can be made from the cardinal information presented in the GAIA plane is that delineated in Fig. 12 (right sub-plot). Consider that a policy-maker in Luxembourg (consistently ranked below Norway in Table 5) would like to see areas of improvement having Norway as a benchmark. Of course, one could argue that this could be made by looking directly at the elementary indicators. The difference is that these do not provide any information about dominance, whereas a plotted preference (i.e. a weight vector) can show the evaluation of a country of interest (e.g. by highlighting this preference in a given colour) taking into account the underlying dominance (i.e. through the uni-criterion net flows). This can be done by highlighting a weight vector with a colour according to the difference in the net flow scores of the two alternatives achieved with that

Figure 12. The SMAA-GAIA plane: the case of Norway.

The left figure shows the space of weight vectors highlighted according to the evaluation of Norway's net flows. The right figure shows the plotted space of weights, constrained to only those for which Norway is at least 10% overall better than Luxembourg.
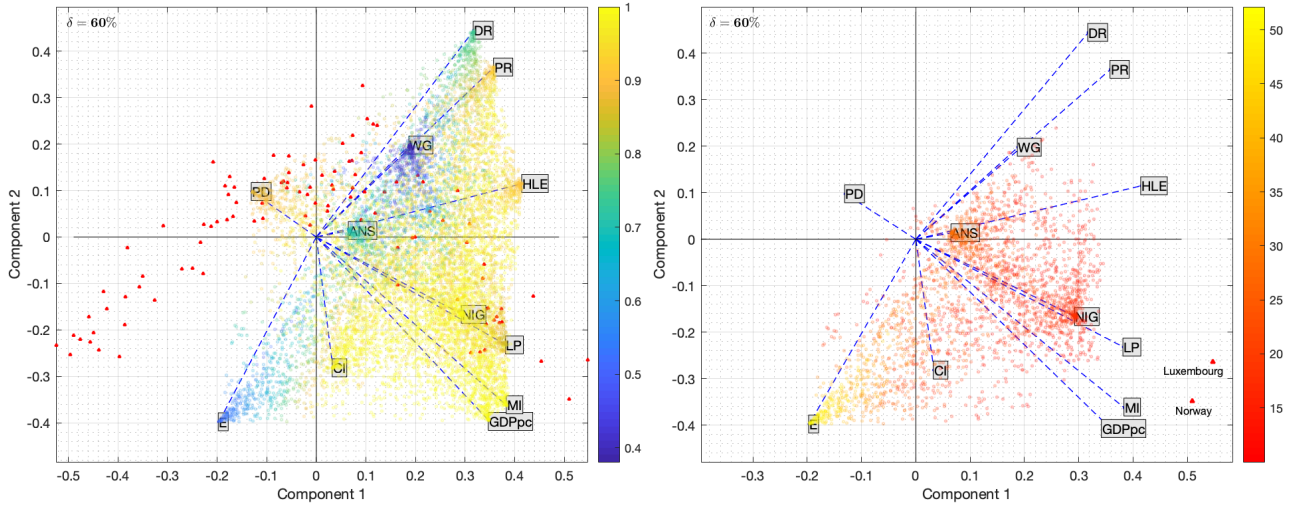
weight vector (i.e. $(\frac{\phi(\text{Nor})}{\phi(\text{Lux})} - 1) \times 100$). This would show the overall performance difference Norway attains against Luxembourg (%) according to that preference. As we are not interested in all the differences but only in areas of improvements for Luxembourg, we only show those vectors of preferences (i.e. clouds of points in the right sub-plot of Fig. 12) for which Norway's score is better than Luxembourg and for at least a 10% difference. As it seems from that figure, Norway is between at least 10% and 25% better than Luxembourg in criteria plotted towards the bottom half of the figure, with an extreme case scenario of the former being superior than the latter by 50% when the weight of preferences is solely focused around employment (i.e. criterion 'E'). Of course, the threshold of 10% could be removed/adjusted according to what the DM considers a difference big enough to take respective action to reduce the gap.

6 Conclusion

Composite indicators are still far from a perfect metric. The reason is that by involving a series of several steps -the most important being weighting and aggregation- they are fairly prone to error judgements, mistakes, uncertainty or even manipulation. Whilst it is a generally acceptable notion that no perfect aggregation will ever exist (Arrow and Raynaud, 1986), these composite and often opaque scores are, at the moment, the best and most popular metrics we may provide to summarise the multidimensionality of a phenomenon being evaluated.

An important issue in the construction of composite indicators is their compensatory nature for which some serious deficiency on one or more elementary indicators is counterbalanced by the performances of other elementary indicators, which can be questionable in several domains. In this perspective, we proposed a novel definition of non-compensatory composite indicator as aggregation of non-compensatory preferences of the considered units. In this context, we have seen that Borda count and its extensions, i.e. the PROMETHEE methods, constitute a valuable basis for constructing non-compensatory composite indicators. In particular, the approach we are proposing is characterized by:

- the basic ordinal nature of preferences on elementary indices (possibly mitigated by means of fuzzy preferences to take into account inaccurate determination, uncertainty and imprecision of original data);

- the basic cardinal nature of composite indicators, which is required to give the compared units an evaluation on a numerical scale and not simply an ordinal ranking.

We defined our methodology, the ordinal input for cardinal output approach to non-compensatory composite indicators, and we believe that it presents quite interesting properties and has a promising potential.

On this basis, we proposed a comprehensive methodology based on well-known operations research methodologies to construct non-compensatory composite indicators that offer the following advantages:

- Based on the SMAA methods, we enhance the transparency in an evaluation process. This is crucial as it shows how prone an alternative could be to changes in the parameters used to evaluate it. Moreover, SMAA permits going above and beyond the issue of the representative agent inherent in an evaluation exercise that concerns a population which is often unknown and thus almost impossible to guess the preferences of.

- Based on the PROMETHEE methods, which are based on a generalization of the classic Borda score, we construct our basic non-compensatory indices based on our approach of ordinal input for cardinal output. Moreover, in the step of aggregation, we disentangle and take into account both the performance and the regret factors of an alternative being evaluated. In this context of non-compensatory aggregation, normalization of elementary indicators is not needed, and weights act now as 'importance coefficients' instead of 'trade-offs' between pairs of indicators.

- Based on the 'Sigma-Mu efficiency analysis' approach, we are able to consolidate the breadth of information provided with the SMAA methods that, whilst greatly informative on its own, was not consolidating the output into a single value that acts as a performance metric. This approach takes into account the distribution of evaluations for each unit, essentially proxying for the whole population interested in the evaluation process. Furthermore, it takes into account the spatial information on the 'Sigma-Mu' plane, which adjusts the classic efficiency measurement to that of taking into account the distances from every single level of competition (as proxied by the many Pareto-Koopmans frontiers in the plane).

- Last, but certainly not least, based on the GAIA visual aid, we provide another SMAA-variant of this important tool in the hands of a decision-maker that is able to showcase cardinal information from the SMAA evaluation. In particular, it shows how an alternative's evaluation can change as a function of the preferences taken into account in that evaluation. Moreover, as showcased in an illustrative example in this paper, it could display areas of improvement for an alternative of interest compared to its closer competitive.

Closing this study, we would like to mention an important area of improvement in the construction of composite indicators. That is interactions among criteria. In particular, in this process we assume no externalities and interdependencies among criteria. In real world situations though, it is very probable that criteria (particularly those within the same dimension) can be mutually strengthening (or conflicting) the final score (for the basic theory on which this approach can be construct see Angilella et al. (2015) and Angilella et al. (2016) for the compensatory approach and Arcidiacono et al. (2018) for the non-compensatory approach, while for some first applications in this direction see Angilella et al. (2018) and Corrente et al. (2019)). This is, to our belief, an important and fruitful area of improvement that needs to be treated with caution.

Acknowledgements

References

Angilella, S., Catalfo, P., Corrente, S., Giarlotta, A., Greco, S., and Rizzo, M. (2018). Robust sustainable development assessment with composite indices aggregating interacting dimensions: The hierarchical-SMAA-Choquet integral approach. *Knowledge-Based Systems*, 158:136–153.

Angilella, S., Corrente, S., and Greco, S. (2015). Stochastic multiobjective acceptability analysis for the choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, 240(1):172–182.

Angilella, S., Corrente, S., Greco, S., and Słowiński, R. (2016). Robust Ordinal Regression and Stochastic Multiobjective Acceptability Analysis in multiple criteria hierarchy process for the Choquet integral preference model. *Omega*, 63:154–169.

Antanasijević, D., Pocajt, V., Ristić, M., and Perić-Grujić, A. (2017). A differential multi-criteria analysis for the assessment of sustainability performance of european countries: Beyond country ranking. *Journal of Cleaner Production*, 165:213–220.

Arcidiacono, S. G., Corrente, S., and Greco, S. (2018). GAIA-SMAA-PROMETHEE for a hierarchy of interacting criteria. *European Journal of Operational Research*, 270(2):606–624.

Arrow, K. J. and Raynaud, H. (1986). *Social choice and multicriterion decision-making*. MIT Press, Cambridge.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Attardi, R., Cerreta, M., Sannicandro, V., and Torre, C. M. (2018). Non-compensatory composite indicators for the evaluation of urban planning policy: The Land-Use Policy Efficiency Index (LUPEI). *European Journal of Operational Research*, 264(2):491–507.

Bandura, R. (2011). *Composite Indicators and Rankings: Inventory 2011*. Technical report, New York: Office of Development Studies, United Nations Development Programme (UNDP).

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research*, 30(5):961–981.

Billaut, J. C., Bouyssou, D., and Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84(1):237–263.

Black, D. (1976). Partial justification of the Borda count. *Public Choice*, 28(1):1–15.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, 59(2):115–151.

Borda, J. C. (1781). Histoire de I'Academie Royale des Sciences. 1781.

Bouyssou, D. (1986). Some remarks on the notion of compensation in MCDM. *European Journal of Operational Research*, 26(1):150–160.

Bouyssou, D. and Vansnick, J.-C. (1986). Noncompensatory and generalized noncompensatory preference structures. *Theory and Decision*, 21(3):251–266.

Brans, J.-P. and De Smet, Y. (2016). PROMETHEE Methods. In Greco, S., Ehrgott, M., and Figueira, J., editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 187–219. Springer New York, New York, NY.

Brans, J.-P. and Mareschal, B. (1995). The PROMETHEE VI procedure: how to differentiate hard from soft multicriteria problems. *Journal of Decision Systems*, 4(3):213–223.

Brans, J. P. and Vincke, P. (1985). Note—A Preference Ranking Organisation Method. *Management Science*, 31(6):647–656.

Brans, J.-P., Vincke, P., and Mareschal, B. (1986). How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2):228–238.

Burgass, M. J., Halpern, B. S., Nicholson, E., and Milner-Gulland, E. J. (2017). Navigating uncertainty in environmental composite indicators. *Ecological Indicators*, 75:268–278.

Corrente, S., Figueira, J. R., and Greco, S. (2014). The SMAA-PROMETHEE method. *European Journal of Operational Research*, 239(2):514–522.

Corrente, S., Greco, S., and Słowiński, R. (2019). Robust Ranking of Universities Evaluated by Hierarchical and Interacting Criteria. In Huber, S., Geiger, M., and de Almeida, A., editors, *Multiple Criteria Decision Making and Aiding. International Series in Operations Research & Management Science*, chapter 5, pages 145–192. Springer, Cham.

Costanza, R., Hart, M., Posner, S., and Talberth, J. (2009). *Beyond GDP: The Need for New Measures of Progress*. Pardee Center for the Study of the Longer-Range Future, Boston.

Dasgupta, P. and Weale, M. (1992). On measuring the quality of life. *World Development*, 20(1):119–131.

De Mare, G., Granata, M. F., and Nesticò, A. (2015). Weak and strong compensation for the prioritization of public investments: multidimensional analysis for pools. *Sustainability*, 7(12):16022–16038.

Doumpos, M., Gaganis, C., and Pasiouras, F. (2016). Bank Diversification and Overall Financial Strength: International Evidence. *Financial Markets, Institutions & Instruments*, 25(3):169–213.

Figueira, J. R., Greco, S., Roy, B., and Slowinski, R. (2013). An Overview of ELECTRE Methods and their Recent Extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2):61–85.

Figueira, J. R., Mousseau, V., and Roy, B. (2016). ELECTRE methods. In Greco, S., Ehrgott, M., and Figueira, J., editors, *Multiple criteria decision analysis: State of the art surveys*, pages 155–185. Springer.

Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, 20(11):1442–1471.

Fishburn, P. C. (1975). Axioms for lexicographic preferences. *The Review of Economic Studies*, 42(3):415–419.

Fishburn, P. C. (1976). Noncompensatory preferences. *Synthese*, 33(1):393–403.

Fishburn, P. C. (2013). *The foundations of expected utility*, volume 31. Springer Science & Business Media.

Greco, S., Ehrgott, M., and Figueira, J. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys.* International Series in Operations Research & Management Science. 2nd edition, New York: Springer.

Greco, S., Ishizaka, A., Matarazzo, B., and Torrisi, G. (2018). Stochastic multi-attribute acceptability analysis (SMAA): an application to the ranking of Italian regions, *Regional Studies*, 52(4), pp.585-600.

Greco, S., Ishizaka, A., Resce, G., and Torrisi, G. (2017). Is the grass always greener on the other side of the fence? composite index of well-being taking into account the local relative appreciations in better life index, *MPRA paper 82718*.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2019a). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1):61–94.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2019b). Sigma-mu efficiency analysis: A methodology for evaluating units through composite indicators. *European Journal of Operational Research*, 278(3):942–960.

Grupp, H. and Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, 39(1):67–78.

Hubinont, J.-P. (2016). SMAA-GAIA: a complementary tool of the SMAA-PROMETHEE method. *International Journal of Multicriteria Decision Making*, 6(3):237–246.

Ishizaka, A. and Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software.* Chichester, United Kingdom: John Wiley & Sons.

Ishizaka, A., Siraj, S., and Nemery, P. (2016). Which energy mix for the UK (United Kingdom)? An evolutive descriptive mapping with the integrated GAIA (graphical analysis for interactive aid)–AHP (analytic hierarchy process) visualization tool. *Energy*, 95:602–611.

Jorion, P. (2000). Value at risk: The New Benchmark for Managing Financial Risk. McGraw–Hill, New York.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39:341–350.

Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., and Aylmer, C. (2013). Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics*, 93:57–68.

Kuznets, S. (1934). National income, 1929-1932. In *National Income, 1929-1932*, pages 1–12. NBER.

Lahdelma, R., Hokkanen, J., and Salminen, P. (1998). SMAA - Stochastic Multiobjective Acceptability Analysis. *European Journal of Operational Research*, 106(1):137–143.

Lahdelma, R. and Salminen, P. (2001). SMAA-2 : Stochastic Multicriteria Acceptability Analysis for Group Decision Making. *Operations Research*, 49(3):444–454.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824.

Marchant, T. (1998). Cardinality and the Borda score. *European Journal of Operational Research*, 108(2):464–472.

Marchant, T. (2000). Does the Borda rule provide more than a ranking? *Social Choice and Welfare*, 17(3):381–391.

Mareschal, B. and Brans, J.-P. (1988). Geometrical representations for MCDA. *European Journal of Operational Research*, 34(1):69 – 77.

Mazziotta, M. and Pareto, A. (2016). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. *Social Indicators Research*, 127(3):983–1003.

Munda, G. (2007). *Social multi-criteria evaluation.* Springer-Verlag, Heidelberg, New York.

Munda, G. (2012). Choosing Aggregation Rules for Composite Indicators. *Social Indicators Research*, 109(3):337–354.

Munda, G. and Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics*, 41(12):1513–1523.

Nitzan, S. and Rubinstein, A. (1981). A further characterization of Borda ranking method. *Public Choice*, 36(1):153–158.

OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide.* OECD Publishing, Paris.

Özerol, G. and Karasakal, E. (2008). A parallel between regret theory and outranking methods for multicriteria decision making under imprecise information. *Theory and Decision*, 65(1):45–70.

Paruolo, P., Saisana, M., and Saltelli, A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3):609–634.

Plott, C. R., Little, J. T., and Parks, R. P. (1975). Individual choice when objects have" ordinal" properties. *The Review of Economic Studies*, 42(3):403–413.

Roberts, F. S. (1985). *Measurement Theory.* Cambridge University Press.

Rosić, M., Pešić, D., Kukić, D., Antić, B., and Božović, M. (2017). Method for selection of optimal road safety composite index with examples from DEA and TOPSIS method. *Accident Analysis & Prevention*, 98:277–286.

Roy, B. (1990). The outranking approach and the foundations of electre methods. In *Readings in multiple criteria decision aid*, pages 155–183. Springer.

Saisana, M., Saltelli, A., and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 168(2):307–323.

Samans, R., Blanke, J., Corrigan, G., and Drzeniek, M. (2017). The inclusive growth and development report 2017. In *Geneva: World Economic Forum.*

Seiford, L. M. and Zhu, J. (2003). Context-dependent Data Envelopment Analysis: Measuring attractiveness and progress. *Omega*, 31(5):397–408.

Stiglitz, J., Sen, A. K., and Fitoussi, J.-P. (2009). *The measurement of economic performance and social progress revisited: Reflections and Overview.* Commission on the Measurement of Economic Performance and Social Progress, Paris.

Tervonen, T. and Lahdelma, R. (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(2):500–513.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.

Zhang, L. and Zhou, P. (2018). A non-compensatory composite indicator approach to assessing low-carbon performance, *European Journal of Operational Research*, , 270(1), pp.352-361.