ORIGINAL RESEARCH

CrossMark

A Hybrid Enhanced Scatter Search—Composite I-Distance Indicator (eSS-CIDI) Optimization Approach for Determining Weights Within Composite Indicators

Milica Maricic[1] · Jose A. Egea[2] · Veljko Jeremic[1]

Abstract
Considering the impact composite indicators can have on public opinion and policy development, the need for their frameworks to be methodologically sound and statistically verified is growing daily. One of the issues in the process of composite indicator construction which has generated much debate is how to choose the weighting scheme. To address this slippery step, we propose an optimization approach based on the enhanced Scatter Search (eSS) metaheuristic. In this paper, the eSS algorithm is applied to obtain a weighting scheme which will increase the stability of the composite indicator. The objective function is based on the relative contributions of indicators, while the problem constraints rely on the bootstrap Composite I-distance Indicator (CIDI) approach which is also proposed herein. The eSS-CIDI approach combines the exploration capability of eSS and the data-driven constraints devised from the bootstrap CIDI. This novel weighting approach was tested on two acknowledged composite indicators: the *Academic Ranking of World Universities* (ARWU) and the *Networked Readiness Index* (NRI). Results indicate that the composite indicators created using the eSS-CIDI weighting approach are more stable than the official ones.

✉ Milica Maricic
milica.maricic@fon.bg.ac.rs

Jose A. Egea
josea.egea@upct.es

Veljko Jeremic
veljko.jeremic@fon.bg.ac.rs

[1] Faculty of Organizational Sciences, University of Belgrade, Jove Ilica 154, Belgrade 11000, Serbia

[2] Department of Applied Mathematics and Statistics, Technical University of Cartagena, Av/Dr. Fleming, ETSII, Cartagena 30202, Spain

1 Introduction

The Organisation for Economic Co-operation and Development (2004) delineates composite indicators as metrics formed "*when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured*". At the very outset, this definition provokes a number of questions: Which individual indicators to choose? How to compile the chosen indicators? Whether or not to apply weights to the indicators? How to devise an underlying model which will best describe the observed phenomenon? It can be observed that the creation of a composite metric might be a frustrating process since there are many questions with different possible subjective answers.

Nevertheless, composite indicators or composite indexes have become popular and widely used metrics for the assessment of complex and sophisticated phenomena which cannot be captured with a single variable. According to Saisana and Tarantola (2002), there are several reasons for the increased popularity of composite indicators. Namely, they provide a single number which summarizes a multi-dimensional phenomenon, making it easier to understand and communicate. Furthermore, a single number provides a 'bigger picture', offers the possibility of comparison and ranking between entities, and can be used to assess progress of entities over time.

On the other hand, there are drawbacks of the idea to compile individual indicators. Namely, the results of composite indicators should be interpreted with caution and policymakers and other stakeholders who interpret them should have prior knowledge of the measured multi-dimensional phenomenon (Zornic et al. 2015). Next, they may be misused to gauge the desired policy or may even lead to inappropriate policies if certain dimensions or indicators are not included in the original framework (Nardo et al. 2005). Also, as there is no simple answer to the methodological choices, various experts can easily initiate debate (e.g. Cherchye et al. 2007; Decancq and Lugo 2013; Jeremic et al. 2011; Ray 2007).

Having in mind the pros and cons of composite indicators, their possible policy implications should be more closely observed. According to Birkmann (2007), the purposes of assessment using composite indicators are to assist policy makers in defining strategies, developing policies, and identifying priorities, and to promote the exchange of information between the government and the wider public. Also, composite indicators can be used to initiate public and political discourse on various policies, for example on national university systems (Saisana et al. 2011). It can be concluded that various stakeholders establish opinions and make decisions based on these ranking results. Therefore, composite indicators can have significant policy implications and should be created and interpreted with caution. Grupp and Mogee (2004) warn that results of composite indicators should not be blindly followed as good policymaking in one country may be poor in another one. As there is absence of transparency and clear rule for the methodological choices, the results of composite indicators should have a complementary function in policymaking (Grupp and Schubert 2010; Saltelli 2007; Sébastien et al. 2014).

On one hand, there is a clear demand for composite measures, while on the other hand the process of their creation is not straightforward and involves assumptions (Saltelli 2007). Policymakers and other stakeholders are in need of stable and methodologically sound composite indicators. This provoked a new research direction in the process of composite indicator creation: the robustness analysis of the composite indicator through uncertainty and sensitivity analysis of methodological assumptions (Saisana et al. 2005). The main purpose of the uncertainty and sensitivity analysis is to act as a 'quality assurance' tool

which should indicate how sensitive the ranking is to methodological choices (Greco et al. 2018a). Although both analyses provide important insights on the robustness of the created metric, this step is still not universally conducted in the process of composite indicator creation (Burgass et al. 2017; Freudenberg 2003; Nardo et al. 2005; Saisana et al. 2005).

The methodological assumption within the creation of composite indicators that attracted our attention is the weighting scheme and its impact on the overall stability of the metric. The question we aim to answer is how to assign weights to indicators within a composite indicator, in a manner which will improve the stability of the metric. Therefore, herein the focus will be how to minimize the effects of the weighting scheme in terms of rank sensitivity and rank reversal. This study attempts to provide an optimization approach that proposes a new weighting scheme, which is data-driven and maximizes the stability of the composite indicator considering specific weight constraints.

The purpose of this study is to devise a fresh weighting approach using non-participatory unsupervised methods with the goal to improve the robustness of the final metric. Such methods do not depend on expert opinion inputs and therefore reduce the possible bias (Singh et al. 2007). However, one should have in mind that the results of such methods should act as a starting point for experts and decision makers upon which they will form the final decision on the weighting scheme (Burgass et al. 2017; Zanakis et al. 2016). Namely, our approach proposes a weighting scheme which should inform the experts from which 'neighbourhood' or 'environment' they should aim to choose the final weights so as to be sure that their composite indicator will be stable. One of the benefits of the proposed approach is that it might accelerate the process of assigning weights which includes expert opinion.

The proposed approach is named enhanced Scatter Search—Composite I-distance Indicator (eSS-CIDI) approach. In order to obtain weight constraints, we applied the bootstrap Composite I-distance Indicator (CIDI), to form the objective function we used relative contributions (Sect. 3.1) and to solve the optimization problem we chose the enhanced Scatter Search (eSS) (Sect. 3.2).

The optimization problem that this study addresses is formulated as:

Minimize   (a) Sum of standard deviations of relative contributions of indicators which make a composite indicator through weighting scheme

Subject to   (b) Each weight should be within the interval based on the bootstrap CIDI results

        (c) Sum of weights must equal 1

        (d) Weights are greater or equal to 0

The question which arises is where does this study position itself in the literature. Is it a contribution to the field of weighting schemes or the field of robustness analysis? Having in mind the rising importance of both fields in the novel literature (e.g. Becker et al. 2017; Dobbie and Dail 2013; Foster et al. 2013; Greco et al. 2018a; Van Puyenbroeck and Rogge 2017) and the need to develop tools to reduce uncertainty and sensitivity of composite scores (Dialga and Thi Hang Giang 2017) we aimed to propose an approach which will encompass both issues at the same time. Namely, in the previous researches usually a novel weighting scheme is proposed, and in the next step, the robustness analysis is performed. Such an approach has been taken in several researches so far (e.g. Dobrota et al. 2015,

2016; Saisana and D'Hombres 2008). The eSS-CIDI approach, on the other hand, does the two at once and can thus speed up the process of composite indicator creation. It proposes a weighting scheme and automatically verifies that it is the most stable one if the robustness analysis is conducted.

The paper is organized as follows: Sect. 2 presents a literature review on the issues in the process of composite indicator creation with a special overview on the weighting approaches based on optimization. The same Section also provides insights on the robustness analysis and the issue of measuring composite indicator stability. Section 3 features the theoretical background of the chosen methods which form the proposed approach, presented in detail in Sect. 4.1. In Sect. 5.3, the results of the application of the proposed approach on the *Academic Ranking of World Universities (ARWU) 2017* and the *Networked Readiness Index (NRI) 2016* are presented. In Sect. 6 we analyse the prospects of future research, and finally, we provide concluding remarks in Sect. 7.

2 Related Work

2.1 Issues in the process of composite indicator creation

Although composite indicators are appealing and informative, several steps in their creation raise significant concerns among academics and researchers (see e.g. Paruolo et al. 2013; Saisana and D'Hombres 2008). The commonly cited slippery steps are the choice of method of normalisation, weighting approach, and aggregation. Therefore, the uncertainty and sensitivity analysis are commonly conducted to explore the effects of different approaches to these steps (Cherchye et al. 2008a; Paruolo et al. 2013; Saisana et al. 2011, 2005). Herein, we provide a brief literature review on the most commonly chosen approaches for the above-mentioned steps.

2.1.1 Normalization

After choosing individual indicators and prior to their weighting and aggregation, the raw data usually measured in different scales (euros, percentages, meters, …) should be put on a standard scale. The step which deals with this issue is normalization. Again, the question arises which method of normalization to use? Freudenberg (2003) lists out several methods: standard deviation from the mean (z-scores), distance from the group leader, distance from the mean, distance from the best and worst performers (min–max), and categorical scale. The normalization method does not affect the ranking of entities per individual indicators, but how does it impact the overall composite indicator value? As different normalization methods produce different results (Freudenberg 2003) they might have significant effects on composite indicator scores (Cherchye et al. 2007; Ebert and Welsch 2004).

Several studies have been conducted so as to explore the impact of the normalization method on the overall composite indicator values. For example, Jovanovic et al. (2012) showed that depending on whether raw or normalized data is used different rankings could be obtained. Pollesch and Dale (2016) conducted an analysis using partial derivatives of the aggregation and normalization functions to inspect their impact. They showed that the overall composite indicator value significantly changes if the value of the indicator changes for a unit depending on the normalization method employed. Talukder et al. (2017) tested and compared various normalization and aggregation techniques for developing composite

indicators and also showed that that the normalization gravely impacts the overall results. The conclusion can be made that whichever normalization method is employed it should be chosen with care (Nardo et al. 2005) and in line with the weighting approach which is to be used (Mazziotta and Pareto 2012).

2.1.2 Weighting

According to the OECD *Handbook on constructing composite indicators*, weights should ideally reflect the contribution of each indicator to the overall composite indicator (Nardo et al. 2005). Therefore, prior to aggregation, the composite indicator creators are given the possibility to assign weight to indicators which will reflect their relative importance to the overall composite indicator (Booysen 2002). The question arises how to choose an appropriate weighting scheme as undeniably, the weighting scheme might have an impact on the entities ranked (Greco et al. 2018a).

Decancq and Lugo (2013) distinguish three essential classes of approach to assigning weights: data-driven, normative, and hybrid. Data-driven weights allow the collected data to determine the weights associated with each indicator, thus bypassing the value judgments of experts. Data-driven methods are usually based on a statistical approach, such as Principal Component Analysis (PCA) or an optimization method such as Data Envelopment Analysis (DEA), Linear programming, or Goal programming (Nardo et al. 2005). As weighting approaches based on optimization are of great importance for this paper, we elaborate more on them in Sect. 2.2.1. Normative approaches, in contrast, rely solely on the judgements of the surveyed experts or stakeholders about the importance of the indicators which make up the composite indicator. Some of the normative weighting approaches include equal weighting, Analytic Hierarchy Process (AHP), Conjoint Analysis (CA), public opinion, budget allocation and others (Singh et al. 2007). Hybrid approaches integrate data-driven and normative approaches. Namely, weights at one level of the composite indicator can be data-driven and expert-driven at the other (Maricic et al. 2015). Another type of hybrid weighting is when expert opinion acts as a constraint for data-driven optimization procedures (Joro and Viitala 2004; Reggi et al. 2014). Although the results of data-driven and hybrid weighting approaches are promising, new approaches are needed to develop a weighting scheme which will differentiate the indicators by importance, but which will also consider the stability of the composite indicator.

Neither weighting approach is flawless (Greco et al. 2018a). Namely, data-driven methods tend to exclude expert opinion completely and introduce a high level of conceptual rigidity into the metric generated (Booysen 2002). Moreover, they are sometimes based on an over-complex multivariate analysis which final users do not understand (Cox et al. 1992). Also, the measured statistical relationships between indicators might not always indicate an actual relation between them (Saisana and Tarantola 2002). On the other hand, expert-driven weighting is often characterized by strong inter-individual disagreement (Rogge 2012) which can lead to conflicting opinions that may, in turn, compromise the validity of the suggested weights (Giannetti et al. 2009). When it comes to equal weighting, it is criticized for not being adequately justified (Greco et al. 2018a) and because it may imply unequal weighting to indicators. Initially, the equal weighting scheme considers all dimensions in the composite indicator equally important. If dimensions consist of a different number of indicators, the indicators in the dimension with the largest number of indicators will be given less weight in the calculation of the overall metric (Dobbie and Dail 2013). However, if a composite indicator has a hierarchy, meaning that each dimension

consists of the same number of indicators, all indicators will be equally weighted. Another issue of equal weighting is "double counting", which occurs when two collinear indicators are included in the composite indicator. In that case, a unique dimension of the observed multi-dimensional concept will be assigned double weight (Nardo et al. 2005).

2.1.3 Aggregation

After weighting of individual indicators, aggregation comes as a final step before obtaining the results of the composite indicator. As in case with normalization and weighting, there are multiple approaches to choose from. The first issue is how to divide aggregation methods into categories. Booysen (2002) states they can be additive or functional in nature. The OECD observes them as linear, geometric, and multi-criteria (Nardo et al. 2005), while Munda (2005a) divides them on compensatory and non-compensatory. Nevertheless, in their review on aggregation methods, Greco et al. (2018a) also list "mixed strategies" which are both weighting and aggregation and could not perfectly fit in neither category. Examples of such approaches are Mazziotta–Pareto Index (MPI) (Mazziotta and Pareto 2007), Penalty for a Bottleneck methodology (Ács et al. 2014), and Mean–Min Function (Casadio Tarabusi and Guarini 2013). Herein, we will briefly present and compare the linear and geometric methods, as they are still the most commonly used aggregation methods in the composite indicator literature (Greco et al. 2018a).

The linear approach allows trade-off, while the geometric does not. Namely, if the linear approach is taken, low results in one indicator can easily be compensated with higher result in the other (Munda and Nardo 2009). Therefore, one should be cautioned when assigning weights led by the idea of giving importance to indicators, as the perceived importance will be transformed to trade-off (Paruolo et al. 2013). A mean of skipping the possibility of constant trade-off is to use the geometric mean, as it offers inferior compensability (Van Puyenbroeck and Rogge 2017). Also, if the geometric mean is used, entities with lower results would be motivated more to increase their results, as a small increase in the indicator value will have more effect on the final results than for the entities with a higher rank (Dobbie and Dail 2013; Greco et al. 2018a).

2.2 Weighting Approaches Based on Optimization

The approach based on optimization caught our attention from different data-driven and non-participatory weighting approaches. The idea of assigning weights through optimization is to maximize or minimize a specific objective function in which the decision variables are the indicator weights. Our literature review shows that this can be achieved through the use of Data Envelopment Analysis (DEA) and various metaheuristics.

2.2.1 Data Envelopment Analysis (DEA) and DEA-Like Approaches

The basic idea which underpins DEA, devised by Charnes et al. (1978), is to calculate the relative efficiency of decision-making units (DMUs) based on their inputs and outputs. DEA can be used in the process of composite indicator development to devise an optimal set of weights. Namely, a series of multiplicative DEA models that can be transformed into equivalent linear programs are solved (Zhou et al. 2010). The key obstacle in the application of DEA and DEA-like models in composite indicator creation is that without constraints, all the observed entities can achieve a maximum or close to maximum

score (Hatefi and Torabi 2010). So far, several solutions to this problem have emerged. For example, Despotis (2005) proposed a two-phase method: first, the application of a standard DEA-like model and then the application of parametric goal programming to discriminate entities which receive a performance score of 1. Zhou et al. (2010) suggested the creation of a composite indicator which will be the weighted sum of the "best" and "worst" case scenario. Both scenarios are created using a multiplicative DEA model, whereas the "best" case scenario aims to assign weights that maximise the index value, while the "worst" aims for the opposite. Their approach is valuable as it avoids the inclusion of constraints but takes the extremes into account. On the other hand, Ramón et al. (2012) suggested a DEA approach which is based on the idea of minimizing the deviations of the common set of weights from DEA profiles of weights provided by the CCR model for efficient DMUs. Another, more recent approach is to use the results of the multivariate Composite I-distance Indicator (CIDI) approach as constraints for the DEA model (Radojicic et al. 2015, 2018).

Led by the idea of DEA, Melyn and Moesen (1991) proposed the Benefit-of-Doubt (BoD) model. Namely, in essence, the BoD model is a DEA CCR model with unitary inputs. Therefore, BoD is sometimes defined as output-oriented DEA (Rogge 2012). The BoD model aims at maximizing the overall composite indicator value of each entity without prior information on the weighting scheme. Clearly, there are conceptual similarities between DEA and BoD: first, between their goals and second, in the lack of available information on weights (Cherchye et al. 2007). Recently, the entity-specific BoD weighting technique has become an established method in the composite indicator literature (Amado et al. 2016; Mariano et al. 2015; Van Puyenbroeck and Rogge 2017). Because the BoD model is a linear programming problem (Rogge 2012), the question of additional model constraints arises. There is a number of approaches to model restriction. Cherchye et al. (2007) presented four different models of constraints to the BoD: absolute restrictions, ordinal sub-indicator share restrictions, relative restrictions, and proportional sub-indicator share restrictions. Blancas et al. (2013) followed the idea of weight restrictions and suggested a common-weight model, inspired by the DEA, but included an objective in the determination of weights: to minimize the number of ties. Giambona and Vassallo (2013) proposed to restrict the weights using the proportional share of index dimensions. Maricic et al. (2016) presented the possibility of integrating the BoD and the CIDI approach. In addition to the original BoD, new types of BoD model have emerged. Directional-BoD, based on a directional distance function model (Zanella et al. 2015), the Meta-Goal Programming BoD with two sets of goals and two meta-goals (Sayed et al. 2015), and the Goal Programming BoD (GP-BoD) (Sayed et al. 2018) which aims to obtain consistent and stable rankings through BoD weights.

2.2.2 Application of Metaheuristics

Metaheuristics can be defined as solution methods that conduct an interaction between the local improvement procedures and complex strategies to create a method able to move from local optima and perform a robust search of the solution space (Glover and Kochenberger 2003). Blum and Roli (2003) outlined fundamental properties of metaheuristics. They state that metaheuristics are strategies that "guide" or "lead" the search process with the goal to find (near-) optimal solutions, that the complexity of their algorithms vary, that they are not problem specific and can easily be modified to solve a particular problem, and that some of them can incorporate mechanisms to avoid getting trapped in the local optima.

The possibility of emerging from the local optima and finding a better solution is what has made metaheuristics remarkably effective and appealing (Gendreau and Potvin 2010). Due to many benefits that metaheuristics offer in solving complex problems, this research field is in constant expansion since 1980s (Sörensen et al. 2018).

Various efforts at obtaining weights using metaheuristics have been summarized in Table 1. Kim and Han (2000) proposed the application of genetic algorithm (GA) approach to feature discretization and determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Socha and Blum (2007), on the other hand, used Ant Colony Optimization (ACO) to train feed-forward neural networks for pattern classification. Again, to solve the issue of connection weights in ANNs Karaboga et al. (2007) applied Artificial Bee Colony (ABC) algorithm while Mirjalili et al. (2012) used Particle Swarm Optimization (PSO) and the Gravitational Search algorithm (GSA).

Chowdhury et al. (2008) suggested using the Predator-Prey optimization (PPO) to solve the multi-objective optimization problem. The idea in multi-objective optimization is to assign different weights to each objective so as to minimize the objective function. Again, to solve the problem of weights within a multi-objective function Taghdisian et al. (2015) used GA, Macedo et al. (2017) employed Evolutionary Algorithms (EA), while Dubey et al. (2016) used Ant Lion optimization (ALO).

Jain et al. (2015) aimed to resolve the weights flexibility problem of DEA. They implemented a GA approach to find a set of weights which maximize the DMUs' efficiency and which are at a minimum distance from all the decision makers' preferences.

Two especially interesting papers are those by Grupp and Schubert (2010) and Becker et al. (2017). Both aim to optimize an aspect of the composite indicator through indicator weights using different approaches. The first seeks to minimize the difference between the best and the worst achievable rank of each ranked entity when weights change arbitrarily. They formulate the problem as a non-linear, mixed-integer problem which they solve using GA. Becker et al. (2017) based their approach on a normalized correlation ratio. The objective function aims to minimize the sum of square difference between the normalized correlation ratio based on the weights suggested by experts and the normalized correlation ratio based on the newly proposed weights. To solve the problem they utilized the Nelder-Mead simplex search (Lagarias et al. 1998).

The presented overview of methods employed to optimize a goal function through the weighting scheme shows that the DEA and DEA-like approaches have been used more often in the process of creating composite indicators while metaheuristics have been employed with a lot of success to optimize weights in ANNs and in multi-objective functions. A research direction which started to develop recently is the application of metaheuristics to optimize the weighting scheme and the structure of composite indicators and DEA models. A valuable insight provided additionally by Table 1 is that the metaheuristics most commonly used to solve the issue of assigning weights are population-based metaheuristics. Research shows that the ANNs which used the weights devised from population-based metaheuristics outperformed the gradient descent algorithm (Gupta and Sexton 1999) and random search and Levenberg–Marquardt algorithms (Socha and Blum 2007). Such a result can indicate that metaheuristics can be used with success in the process of establishing weighting schemes.

Table 1 Literature overview of the employed metaheuristics to optimize the weighting scheme sorted by publication year

	Authors (year)	Metaheuristics	Applications
1.	Kim and Han (2000)	GA	Connection weights for artificial neural networks (ANNs)
2.	Socha and Blum (2007)	ACO	Connection weights for artificial neural networks (ANNs)
3.	Karaboga et al. (2007)	ABC	Connection weights for artificial neural networks (ANNs)
4.	Chowdhury et al. (2008)	PPO	Weights of specific objective functions within a multi-objective function
5.	Grupp and Schubert (2010)	GA	Weighting scheme of the composite indicator
6.	Mirjalili et al. (2012)	PSO, GSA	Connection weights for artificial neural networks (ANNs)
7.	Jain et al. (2015)	GA	Weight restrictions for DEA
8.	Taghdisian et al. (2015)	GA	Weights of specific objective functions within a multi-objective function
9.	Dubey et al. (2016)	ALO	Weights of specific objective functions within a multi-objective function
10.	Becker et al. (2017)	NMSS	Weighting scheme of the composite indicator
11.	Macedo et al. (2017)	EA	Weights of specific objective functions within a multi-objective function

GA genetic algorithm, *ACO* ant colony optimization, *ABC* Artificial Bee Colony algorithm, *PSO* particle swarm optimization, *GSA* gravitational search algorithm, *PPO* predator pray optimization, *ALO* ant lion optimization, *EA* evolutionary algorithm, *NMSS* Nelder-Mead Simplex Search

2.3 Robustness Analysis of Composite Indicators

Robustness and stability are considered to be important issues in the process of composite indicator development (Dobrota et al. 2015). Nardo et al. (2005) state that there are multiple approaches to be considered when constructing a composite indicator. Therefore, analyses should be undertaken to assess the impact of the different methodological approaches employed on the values and ranks of entities. According to Greco et al. (2018a), there are three ways to conduct the robustness analysis using: uncertainty and sensitivity analysis, stochastic multicriteria acceptability analysis, and other approaches which are mostly based on linear programming and optimization. Herein, we provide a literature review on the first two approaches, as the third group will be more closely observed in the following subsection as it is based on the minimization of rank-reversals.

The most commonly used robustness analyses to evaluate the stability of a particular composite indicator are uncertainty and sensitivity analysis (Dijkstra et al. 2011; Paruolo et al. 2013; Saisana and D'Hombres 2008; Saltelli et al. 2000). Usually, uncertainty analysis is performed first to quantify the impact of alternative models on the rankings. Each model is, in fact, a different composite indicator, in which the normalization method, weighting approach, the aggregation method or another factor have been randomly chosen from the predefined methods for each of the methodological assumptions (Freudenberg 2003). Alongside uncertainty analysis lies sensitivity analysis. The goal here is to present, qualitatively and quantitatively, the variability of the scores and ranks which occur due to different methodological assumptions (Saltelli et al. 2007). The sensitivity analysis can be used to test composite indicators for robustness as it shows how much an individual source of uncertainty influences the output variance (Saisana et al. 2005). There are multiple examples in which the uncertainty and sensitivity analysis have been conducted. For example, Dijkstra et al. (2011) observed the uncertainty of the threshold for the definition of the development stage of countries and the weighting scheme of the Regional Competitiveness Index (RCI). Saisana and Saltelli (2014) observed the uncertainty of the aggregation method, missing data imputation method, and the weighting scheme of the Rule of Law Index (RoL), while Cherchye et al. (2008a) had 23 uncertain input factors in the analysis of the Technology Achievement Index (TAI). Importantly, both uncertainty and sensitivity analysis can be used to assess a single methodological assumption, for example solely for the normalization method. Namely, sometimes, it is of interest to explore the robustness of the results when solely the weights are considered. Dobrota et al. (2015) inspected the stability of the information and communication technology (ICT) Development Index (IDI), Dobrota and Dobrota (2016) observed the stability of ARWU and Alternative ARWU ranking, while Dobrota and Jeremic (2017) analysed the robustness of the Quacquarelli Symonds (QS) and University Ranking by Academic Performance (URAP) rankings. All three papers measured the stability through relative contributions of indicators (Sect. 3.1) and indicated that the original weighting scheme could be altered so as to improve the stability of the composite indicator.

Another approach is to apply the stochastic multicriteria acceptability analysis (SMAA) initially proposed by Lahdelma et al. (1998). The SMAA was first used in the field of multiple criteria decision making, but recently it has been used in the field of composite indicators. Greco et al. (2018b) observed the robustness through weights assigned to indicators. They ranked 20 Italian regions based on 65 indicators and used SMAA to analyse the whole space of possible weight vectors considering the spectrum of possible individual preferences. The result of their analysis is the probability of a region to be on a certain

rank. Corrente et al. (2018) proposed a novel technique, namely SMAA for strategic management analytics and assessment, or SMAA squared (SMAA-S) for ranking entrepreneurial ecosystems. Compared to the SMAA, the SMAA-S allows the definition of additional weight constraints. They ranked 23 European countries using 12 indicators and as a result, again obtained the probability of a country to be on a certain rank. Angilella et al. (2018) went a step further and proposed a hierarchical-SMAA-Choquet integral approach to create a composite index of sustainable development. Compared to the previous two versions of SMAA, this one is based on a more complex aggregation method, the Choquet integral which finds the balance between linear and geometric aggregation (Choquet 1953). They ranked 51 Italian municipalities using 10 indicators. We can conclude that the application of SMAA in the field of robustness analysis is a new promising field of study.

As it can be observed, the robustness analyses is commonly conducted to evaluate composite indicators as it allows the assessment to what extent, and for which entities, in particular, the modelling assumptions affect ranking results (Saisana and Saltelli 2014). Thus, the analysis provides valuable feedback to both composite indicator creators and policymakers. It can be argued that the stability of a composite indicator is highly important for policymakers as it assures them that the measure they are using is not susceptible to rank changes (Dobrota et al. 2015). Therefore, it can be debated that the statistical stability is one of the key attributes of any composite indicator (Renzi et al. 2017).

2.4 Measuring Composite Indicator Stability

The question which arises is how to define and measure the stability of a composite indicator? In this paper, we observed stability through the lens of rank reversals. Namely, the more rank reversals a weighting scheme yields compared to the official weighting scheme, the less stable it is. So far, the issue of rank reversal has been explored in the sphere of multicriteria decision making (MCDM). However, recent studies show that rank reversal in the sphere of composite indicators is slowly but surely becoming a topic of research interest. In the following paragraphs, we provide a literature review on the definition of stability through rank reversals in MCDM and composite indicators sphere and on the remedies to minimize it.

In the MCDM literature there are several definitions of rank reversal. One is that rank reversal is a ranking contradiction which occurs when the original set of alternatives alters by adding new alternatives or deleting one or more alternatives (Wang and Triantaphyllou 2008; Wang and Luo 2009). Macharis et al. (2004) define rank reversal as the phenomenon which occurs when a copy or a near copy of a choice is added to the set of options. The alternative term is equivalent to a country or entity in the composite indicator context (Sayed et al. 2018). According to Mousavi-Nasab and Sotoudeh-Anvari (2018) however the rank reversal is defined, it is an undesirable phenomenon that indicates unreliability. The same authors stress out that the research on rank reversal has great practical importance and should be continued.

The issue of rank reversal in MCDA has been a topic of thorough research as it is perceived as a drawback of almost all MCDA techniques. For example, Mousavi-Nasab and Sotoudeh-Anvari (2018) showed that COPRAS (Complex Proportional Assessment), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and VIKOR (VIseKriterijumska Optimizacija i Kompromisno Resenje, Serbian term) may produce rank reversals. Herein, we list several studies which tackled the issue of rank reversal occurring due to weighting scheme. In an interesting study by Ligmann-Zielinska and Jankowski (2008),

in which they presented a framework for sensitivity analysis in multiple criteria evaluation, they devoted special attention to weights as weights "have been most often criticized". They state that it should be possible to devise "critical weights" for which a relatively small change in the weight will cause minimal rank reversal. In their study on Analytical Hierarchy Process (AHP) Bojórquez-Tapia et al. (2005) chose, as the best solution, the structure which presented the lowest rank reversal sensitivity to weight uncertainty. Their analysis showed that the robustness of decisions based on the AHP depends on the relative weights. Munda (2005b) also pointed out the importance of weights by giving evidence that changing the weights in a multi-criterion framework for devising urban sustainability index significantly changes entity ranking.

In the composite indicator literature on rank reversal, there are several researches needed to be mentioned. The first one is the work of Cherchye et al. (2008b) in which they inspected the robustness of the Human Development Index (HDI) through the effects of weighting and aggregation on eventual country rankings. They proposed several possible weight scenarios and using a dominance criterion conducted pairwise country comparisons. Several years later, Permanyer (2011) presented innovative ways to measure the sensitivity of a composite measure due to the choice of specific weights. Namely, he proposed an approach which aims to find a "neighbourhood system", a close, nested, and bounded neighbourhood of weights from which a decision maker can consider to choose the weighting scheme. He also proposed a robustness function which shows the percentage of rank shifts as the weighting scheme changes. Next, Foster et al. (2013) did a research in two directions and showed that a higher positive association between indicators leads to maximal robustness and that their novel robustness measure can be used to compare aggregate robustness properties of different composite indicators. In a more recent study, Sayed et al. (2018) wanted to avoid some of the drawbacks of the BoD and proposed a Goal Programming BoD (GP-BoD) model. The main benefit of the proposed method is that it overcomes the rank reversal issue by assigning data-driven weights which minimize rank reversal.

Robustness of the composite indicator in terms of capacity to produce stable measures should be conducted (Maggino 2009) to find the most stable model. Sayed et al. (2018) noted that the credibility of the composite indicator is limited if it produces inconsistent countries' rankings. Therefore, when creating a composite indicator, the issue of rank stability should be considered. Also, it is desirable that the chosen weighting method should be such that it avoids or minimizes rank reversal (Paracchini et al. 2008). Still, robustness is not a goal in itself, and the results of weighting schemes which are most stable should be analysed with caution (Permanyer 2011).

3 Background

In this Section, we provide insights into the methods used in our proposed approach for weight determination. First, we introduce the bootstrap CIDI, the methods it builds upon, and the rationale for its implementation. Second, we present the relative contributions and their role in the measurement of composite indicator stability. Finally, we give a brief overview of the enhanced Scatter Search (eSS), the metaheuristics used to solve the optimization model and provide the optimal weighting scheme.

Before introducing the above-mentioned methodologies, we provide notations used in the model development process and in the proposed approach.

Notation

i	Index for indicators within a composite indicator, $I = \{1, \ldots, k\}, i \in I$
x_{ie}	Value of indicator i for entity $e, i \in I, e \in E$
$d_i(e,f) = x_{ie} - x_{if}$	Distance between the values of indicator i of the entity e and fictive entity $f, i \in I, e \in E$
$r_{ji.12\ldots j-1}$	Coefficient of partial correlation between indicators i and j where $j < i$, $i \in I, j \in I$
$D^2(e,f)$	Value of the I-distance between entity e and fictive entity $f, e \in E$
w_{CIDI_i}	Weight assigned to indicator i using CIDI approach, $i \in I$
s	Index for bootstrap samples, $S = \{1, \ldots, p\}, s \in S$
$w_{\min i}$	Minimum weight of indicator i obtained after bootstrap CIDI, $i \in I$
w_i	Weight assigned to indicator $I, i \in I$
e	Index for entities ranked by a composite indicator, $E = \{1, \ldots, n\}$, $e \in E$
f	Fictive entity from which the distances are calculated
σ_i^2	Variance of indicator $i, i \in I$
$r_{ji.12\ldots j-1}^2$	Coefficient of partial determination between indicators i and j where $j < i, i \in I, j \in I$
r_i	Pearson's correlation coefficient between indicator i and the I-distance value, $i \in I$
v_{ie}	Relative contribution of indicator i to the overall composite indicator value of entity $e, i \in I, e \in E$
$w_{CIDI_{is}}$	Weight assigned to indicator i using CIDI approach based on the sample $s, i \in I, s \in S$
$w_{\max i}$	Maximum weight of indicator i obtained after bootstrap CIDI, $i \in I$

3.1 Bootstrap Composite I-Distance Indicator (Bootstrap CIDI)

This subsection provides insights to the Composite I-distance Indicator (CIDI) approach, a brief introduction to the bootstrap method, and the rationale for combining the bootstrap and the CIDI.

In the 1970s the growing need for a multivariate statistical analysis that could rank countries based on their level of socio-economic development led to the creation of the I-distance method (Ivanovic 1977). Since then the method has been applied for ranking entities in various fields, such as in education (Zornic et al. 2015), sustainable development (Savic et al. 2016), corporate social responsibility (Maricic and Kostic-Stankovic 2016) and ICT development (Dobrota et al. 2015).

In order to rank the entities (e.g. countries, regions, universities) using the I-distance method, we first select the entity with the minimal values for all indicators as a referent one in the observed dataset (Jeremic et al. 2013; Maricic and Kostic-Stankovic 2016). If such entity does not exist, a fictive entity, depicting minimal values for each observed indicator, is created.

For a selected set of indicators k chosen to characterize the entities, the square I-distance between the entity $e, e \in E$, and fictive entity f is defined as (Ivanovic 1977):

$$D^2(e,f) = \sum_{i=1}^{k} \frac{d_i^2(e,f)}{\sigma_i^2} \prod_{j=1}^{i-1} \left(1 - r_{ji.12\ldots j-1}^2\right) \tag{1}$$

The I-distance stands out for its ability to incorporate a number of variables into a single number without explicitly assigning weights. Therefore, it provides the possibility of ranking entities (Jeremic et al. 2011). Another benefit is that the importance of each variable for the ranking process can be obtained through the Pearson's correlation coefficients. The more the variable is correlated with the I-distance value, the more it contributes to the overall I-distance value and the ranking process (Maricic and Kostic-Stankovic 2016). Therefore, the obtained correlation coefficients can act as a foundation for determining weights. Accordingly, the new weights are formed using the Pearson's correlation coefficient between the variable and the obtained I-distance value. The weights suggested by the Composite I-distance Indicator (CIDI) approach are based on the ratio of the (I) correlation coefficient of a particular variable and (II) the I-distance value and the sum of correlation coefficients of variables with the I-distance value (Dobrota et al. 2015). The rationale of this approach is in line with the Hellwig's (1969) "weights based on the correlation matrix" approach. Namely, he proposed indicator weights as the ratio of correlation coefficient of a particular variable and the exogenous criterion and the sum of correlation coefficients of variables with the exogenous criterion (Ray 2007). Comparing the two approaches, in case of the CIDI methodology, the exogenous criterion is the I-distance value. The formula is given as:

$$w_{CIDI_i} = \frac{r_i}{\sum_{j=1}^{k} r_j} \tag{2}$$

where r_i is the Pearson's correlation coefficient between the indicator i, $i \in I$ and the obtained I-distance value and $\sum_{j=1}^{k} r_j$ is the sum of all Pearson's correlation coefficients between the indicators and the obtained I-distance value. The sum of weights obtained using CIDI is 1 (Dobrota et al. 2016) and the new weighting scheme is unbiased in the sense that it is data-driven and that no expert opinion has been included in the weighting process. The obtained CIDI weights can be used to evaluate a composite indicator or to create a novel one (Dobrota et al. 2015), but they can also be used to create weight constraints (Maricic et al. 2016; Radojicic et al. 2018).

Efron (1979) coined the term bootstrap and described both the parametric and nonparametric bootstrap. In his parametric approach, random samples are drawn for a specified probability density function. On the other hand, in the nonparametric approach, thousands of resamples are drawn with replacement from the original sample, and each resample is of the same size as the original sample (Kline 2005). The idea of both parametric and nonparametric bootstrap is to assess the quality of estimates based on finite data (Kleiner et al. 2014). The goal is to obtain confidence intervals, quality assessments that provide more information than a simple point estimate obtained using the Maximum Likelihood Method (Efron and Tibshirani 1993).

However, statistical shortcomings of the nonparametric bootstrap have been discussed (Bickel et al. 1997; Mammen 1992). These led to the development of related methods such as the *m out of n bootstrap with replacement* (Bickel et al. 1997) and the *m out of n bootstrap without replacement*, known as subsampling (Politis et al. 1999). The idea of reducing the bootstrap sample size came from Bretagnolle (1983) and Beran and Ducharme (1991). They showed that the failure of Efron's (1979) bootstrap with resampling size equal to the original sample size may be solved in some cases by undersampling. The idea of the *m out of n bootstrap* is to draw samples of size *m* with or without replacement, instead of resampling bootstrap samples of size *n*, where $m \to \infty$, and $m/n \to 0$.

The type of bootstrap which caught our attention is the subsampling, the *m out of n bootstrap without replacement*. Namely, subsampling is more general than the *m-bootstrap* since fewer assumptions are required (Bickel and Sakov 2008). It is based on two basic ideas – one being undersampling, the other the absence of replacement (Politis and Romano 1994). Such an approach to sampling, without replacement, results in deliberate avoidance of choosing any member of the population more than once. This process should be used when outcomes are mutually exclusive and when the researcher wants to eliminate the possibility of the same entity being repeated in the sample. Also, the bootstrap method without replacement improves the stability and accuracy of the estimate (Kawaguchi and Nishii 2007) and shows asymptotic consistency even in cases where the classical bootstrap fails (Davison et al. 2003).

On the other hand, bootstrap with replacement is easier to use and more widely used (Geyer 2013). We chose the *m out of n bootstrap without replacement* (subsampling) so as to avoid the possibility of choosing any entity more than once. Namely, composite indicators are used to rank different entities. Although it is always possible to get two or three entities with the same values of indicators, this cannot happen for a large percentage of the sample. Theoretically, if a bootstrap with replacement was used, no matter the sample size, we could observe a sample made out of just one or two entities. The possibility of such an occurrence is small, but we wanted to be sure no entity will repeat. Such an approach is used in medical researches, when there is a strict need that a subject is no longer eligible for consideration (Austin and Small 2014). Also, Strobl et al. (2007) showed that when bootstrap without replacement is conducted no bias occurs. Having these considerations in mind, we opted for the *m out of n bootstrap without replacement*.

Using a smaller bootstrap sample requires a choice of *m*. Several solutions have been suggested (see e.g. Arcones and Gine 1989; Bickel and Sakov 2008). An interesting approach to choosing *m* is to make it equal to $0.632 \cdot n$, namely to its nearest integer. This idea was first introduced by Efron and Tibshirani (1997) with the goal to reduce the bias of the leave-one-out bootstrap. The idea of this approach is to have a number of observations in the subsample equal to the average number of unique observations in the bootstrap sub-samples (Braga-Neto and Dougherty 2004; De Bin et al. 2016). If a bootstrap *n out of n with replacement* is conducted, in each of the subsamples, there will be $0.632 \cdot n$ original data points. Namely, the probability that each data point will not appear in the subsample is $(1 - 1/n)^n \approx e^{-1}$. If a subsample of size n is drawn then $(1 - 1/n)^n \cdot n \approx e^{-1} \cdot n \approx 0.632 \cdot n$. This approach to defining *m* has become popular because it shows low variability and moderate bias (Jiang and Simon 2007). In our study, we, as in Strobl et al. (2007), used the $0.632 \cdot n$ as *m*.

When it comes to bootstrap, one must decide on the number of replications. Again, several suggestions have been made. For example, Hedges (1992) recommended performing 400–2000 bootstrap replications, while a more recent study by Pattengale et al. (2009) suggest between 100 and 500 replicates. They also state that the stopping criteria can recommend very different numbers of replicates for different datasets of comparable size. In our study, we chose 1000 bootstrap replications.

The idea of implementing the bootstrap for various forms of multivariate analysis emerged in the 1990s. For example, Ferrier and Hirschberg (1999) combined bootstrapping methods to measure bank efficiency using DEA. Nevitt and Hancock (2001) showed that bootstrap methods could be applied in structural equations models. Kim et al. (2008) combined the bootstrap and discriminant analysis. More recently, Xu et al. (2013) conducted a parametric bootstrap approach for the two-way ANOVA, and Konietschke et al. (2015) applied parametric and nonparametric bootstrap to MANOVA. Lead by these positive

experiences and reputable results, we propose the application of the bootstrap to CIDI approach.

In a typical application of the CIDI approach, the researcher would obtain new, data-driven CIDI weights (Dobrota et al. 2016). However, by performing bootstrap and repeating the CIDI approach multiple times on various samples, one obtains an interval for each weight. Such an interval can be used as constraint in GAR DEA method (Global assurance region DEA method) (Radojicic et al. 2018). Therefore, we propose the application of the CIDI approach to each of the bootstrap samples and the calculation the minimum and maximum weight assigned to each indicator. The obtained values can be used as weight constraints.

The procedure for the *m out of n bootstrap CIDI without replacement* with *p* replications is as follows:

1. From the observed *n* entities create *p* random samples S_s of size *m* ($m = 0.632 \cdot n$)
2. Perform the CIDI approach on each of the *p* created random samples
3. Obtain the minimum and maximum of the CIDI weights assigned to each of the *k* indicators.

The results of the bootstrap CIDI will be used in the next steps of the proposed hybrid weighting approach.

3.2 Relative Contributions

The weight assigned to the indicator does not guarantee its final contribution in the overall value (Saisana and D'Hombres 2008). One of the means to assess the share of an indicator in the overall value is using relative contributions. The relative contribution of each indicator to the overall composite indicator value is calculated as the proportion of the (I) product of the indicator score with the associated weight compared to the (II) total score (Murias et al. 2008). The formula is given as:

$$v_{ie} = \frac{x_{ie} w_i}{\sum_{j=1}^{k} x_{je} w_j} \tag{3}$$

where the relative contribution v_{ie} represents the relative contribution of indicator *i*, $i \in I$ to the overall index value of entity *e*, $e \in E$.

The relative contribution can be interpreted as percentual impact of the weighted indicator *i* to the overall composite indicator value of entity *e*. Therefore, we can say that as the value of the relative contribution of an indicator *i* of entity *e* rises, the impact of the indicator *i* in the overall composite indicator value of the entity *e* rises. The presented analysis can be quite useful when it comes to the assessment of composite indicators, especially when the values of relative contributions are analysed per indicator. Namely, such an approach can give insight into whether certain indicators are dominating the overall composite indicator values (Dobrota et al. 2016). It can also show that the relative contribution of an indicator to the overall composite indicator value might not necessarily be captured by the weight assigned (Saisana and D'Hombres 2008). This is measured through the mean relative contribution per indicator. It can deviate from the official weight assigned to the indicator thus indicating that an indicator has a higher or smaller impact. Therefore, relative contributions can indicate the level of composite indicator and/or rank stability. Accordingly, a higher standard deviation of relative contributions leads to a greater degree

of oscillation of overall composite indicator value subsequently leading to higher rank oscillation (Savic et al. 2016).

So far, relative contributions have been used with a great deal of success in the complex procedure of composite indicator creation. Besides being used to inspect the difference between nominal (explicit) and relative (implicit) weights, relative contributions can be used as input to the uncertainty and sensitivity analysis of weighting schemes such as in Dobrota and Dobrota (2016) and Dobrota et al. (2016). Another interesting application of relative contributions in uncertainty and sensitivity analysis is their use as constraints in the DEA model. Murias et al. (2008) aimed to create a composite indicator for the quality assessment of Spanish public universities using nine indicators and the DEA model. Since they wanted to allow every unit to choose its individual weights, but also to limit the flexibility of the weights, they imposed additional restrictions to the relative contributions of indicator weights. Similarly, Pérez et al. (2013) used linear programming to maximize the value of the composite indicator, while restraining the weights using relative contributions of indicators. Relative contributions can also be used as a stopping rule. Savic et al. (2016) suggested stopping post hoc I-distance (proposed by Marković et al. (2016)) when the sum of standard deviations of relative contributions starts to increase.

3.3 Enhanced Scatter Search (eSS)

Scatter Search (SS) is a population-based metaheuristics based on formulations proposed in the 1960s for combining decision rules and problem constraints. The SS was introduced by Glover (1977) for integer programming. However, the SS has recently produced promising results for solving combinatorial and nonlinear optimization problems (Glover et al. 2000). For that, it has seen several modifications of the original method such as the enhanced Scatter Search (eSS) (Egea et al. 2009), Cooperative enhanced Scatter Search (CeSS) (Villaverde et al. 2012), and the self-adaptive cooperative enhanced scatter search (saCeSS) (Penas et al. 2017). One of the reasons for the interest of academics and researchers in the SS is because it uses strategies for combining solution vectors that have proved effective in a variety of problem settings. As the modifications of SS have become more widely used, there was a growing need to implement them in the programming language R and make them more accessible (Egea et al. 2014).

The optimization algorithm used, the enhanced Scatter Search (eSS), can be considered as an evolutionary method similar to e.g. genetic algorithms but based on systematic combinations of the population members instead of recombination and mutation operators. Metaheuristic classification places Scatter Search in the group of population-based algorithms that constructs solutions by applying strategies of diversification, improvement, combination and population update (Penas et al. 2015). It is a naturally inspired and nonlinear metaheuristic which is suited for global dynamic optimization of nonlinear problems. The aim of the eSS is to enable the implementation of various solution strategies that can produce new solutions from combined elements to derive better solutions than strategies whose combinations are only based on a set of original elements. The SS algorithm uses different heuristics to choose suitable initial points for the local search which helps overcome the problem of switching from global to local search (Rodriguez-Fernandez et al. 2006). The pseudo code for the eSS can be found in Egea et al. (2009).

Although SS is similar to GAs, there are crucial differences which should be emphasized. Namely, GA approaches are based on the idea of choosing parents randomly and further on applying random crossover and mutation to create a new generation (Gen and

Cheng 2000). Contrarily, SS does not emphasize randomization. It does not use crossover and mutation but applies a solution combination method that operates among population members. The approach is designed to apply strategic responses, both deterministic and probabilistic, that take account of evaluations and history (Rodriguez-Fernandez et al. 2006). New solutions are generated using a systematic (partial random) combination rather than a fully random solution (Remli et al. 2017).

The eSS algorithm which we propose to implement in our study aims to find the balance between intensification (local search) and diversification (global search) using a small population size, more search directions, and an intensification mechanism in the global phase, which exploits the promising directions defined by a pair of solutions in the reference set (Egea et al. 2009). Therefore, eSS stands out for its good balance between robustness and efficiency in the global phase, and couples with a local search procedure to accelerate the convergence to optimal solutions (Penas et al. 2017). The eSS has proved to be an efficient metaheuristic in solving complex-process optimization problems from different fields, providing a good compromise between diversification (exploration by global search) and intensification (local search) (Otero-Muras and Banga 2014).

4 Proposed Approach: Hybrid Enhanced Scatter Search—Composite I-Distance Indicator (eSS-CIDI) Optimization Approach

This study aims to find a set of data-driven weights which creates the most stable composite indicator. The stability of a composite indicator, observed through the prism of its weighting scheme, is measured through the standard deviation of relative contributions of indicators. Therefore, we aim to minimize the sum of standard deviations of relative contributions of indicators which constitute the composite indicator. The goal is to adjust the weights assigned to composite indicator so as to achieve high overall composite indicator stability. However, if no constraints to each weight are imposed, the model would assign weight to just one indicator, the indicator with the minimum standard deviation. Accordingly, we impose weight restrictions using the previously introduced *m out of n bootstrap CIDI without replacement* with *p* replications. To solve the proposed optimization model, we propose the eSS.

The three-step algorithm which searches for the weights that create the most stable composite indicator via constrained optimization is as follows:

1. ***Conduct the m out of n bootstrap CIDI without replacement with p replications***. The suggested value of the parameter *m* is $0.632 \cdot n$, where *n* is the number of the entities ranked by the composite indicator. The suggested number of replications (*p*) is 1000.
2. ***Define the weight constraints***. Obtain the min and max CIDI weights of each indicator following the bootstrap CIDI procedure.
3. ***Solve the optimization model***. Use the min and max CIDI weights as constraints of the optimization model. To solve the model, we propose the eSS metaheuristic.

The issue which arises when applying optimization models relates to model constraints. Namely, as previously explained, if no constraints are assigned, the optimization model would assign weight to just one indicator (in case of BoD) or all entities would be efficient or close to efficient (in case of DEA). To conduct the eSS (Step 3), bound constraints for the decision variables must be defined.
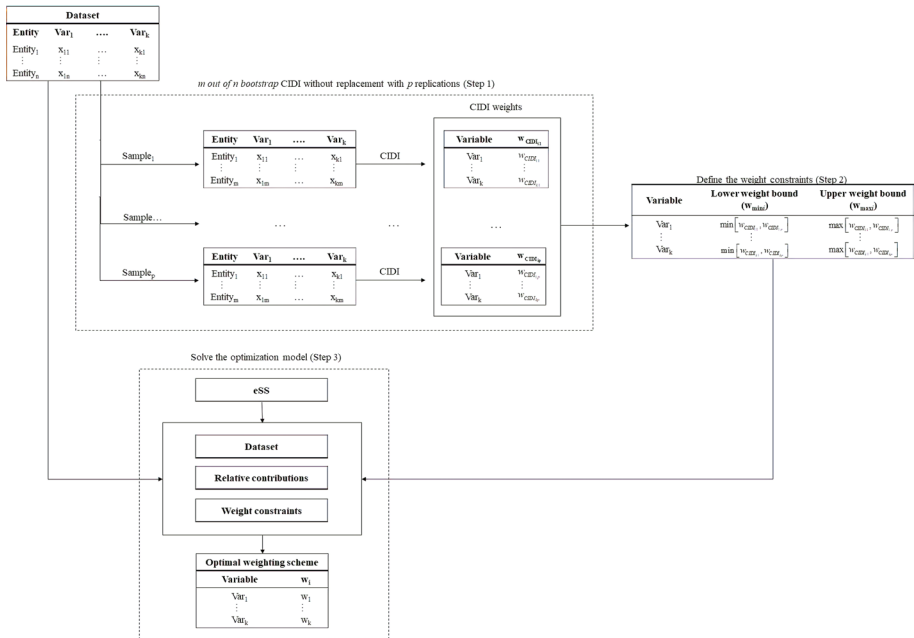
Fig. 1 An overview of the proposed eSS-CIDI weight optimization approach

Among various approaches for determining such bounds, the CIDI approach stood out as is was employed with success to constraint BoD and DEA models. Namely, Maricic et al. (2016) created weight constraints for the BoD model around CIDI weights. On the other side, Radojicic et al. (2018) used bootstrap CIDI to create constraints for the DEA model. Following the example of good practice, herein we use the *m out of n bootstrap CIDI without replacement* with *p* replications to create weight bounds for the optimization model (Step 1). These bounds are defined as the minimum and maximum weight assigned using the bootstrap CIDI (Step 2). Such an approach allows us to create a wide enough interval which will cover all bootstrap CIDI weights.

The proposed optimization approach is presented in Fig. 1.

The objective function of the optimization problem is given as follows:

$$\min_{w_i} \sum_{i=1}^{k} \sigma_{v_i} \tag{4}$$

Subject to

$$w_{\min i} \leq w_i \leq w_{\max i} \tag{5}$$

$$\sum_{i=1}^{k} w_i = 1 \tag{6}$$

$$w_i \geq 0 \quad \forall i \qquad (7)$$

where σ_{v_i} is the standard deviation of the relative contribution of indicator i, $i \in I$, in the composite indicator of entities e, $e \in E$, computed as:

$$\sigma_{v_i} = \sqrt{\left(\frac{1}{n} \sum_{e=1}^{n} \left(v_{ie} - \frac{1}{n} \sum_{e=1}^{n} v_{ie} \right)^2 \right)} \qquad (8)$$

Next, v_{ie} is the relative contribution of indicator i to the entity e, $w_{\max i}$ and $w_{\min i}$ are the maximum and minimum weight of indicator i obtained after bootstrap CIDI. Alongside the upper and lower constraints (Eq. 5), we added the constraint that the sum of assigned weights must be one (Eq. 6). Finally, the weights must be greater or equal to 0 (Eq. 7).

4.1 Discussion on the Proposed Optimization Approach

A key 'managerial insight' of the eSS-CIDI approach is that it allows the composite indicator creators to be certain they have created a stable and robust composite indicator considering the chosen weight constraints. Namely, when choosing the weighting scheme, researchers are given a wide range of weighting approaches to choose from. Nevertheless, there is no guarantee that they will choose the most stable one. Furthermore, our approach provides an opportunity to improve the stability of a devised composite indicator. Besides suggesting new weights, the eSS-CIDI approach could be used to assess and inspect the official weighting scheme. The deviations in weights, as well as overall composite indicator values and ranks, could have valuable policy implications as composite indicators are often used to define strategies and policies and act as guiding lights (Dialga and Thi Hang Giang 2017). Therefore, the eSS-CIDI approach could be of use to policymakers, especially as the optimized weights are rounded to three decimal places which makes them easier to interpret and present to the wider public (Cole 2015a, b).

One of the prominent contributions of our method is that it is a data-driven weight optimization approach which builds upon methods that were verified in various applications (Dobrota et al. 2016; Egea et al. 2009; Nevitt and Hancock 2001). Furthermore, the algorithm provides a single solution, so the researcher does not need to choose between several solutions. Finally, it enables the index creators to be sure that, given the constraints, their metric will have the highest stability.

The question which arises is how we measure stability in the eSS-CIDI approach. To measure the stability of the composite indicator we used the standard deviation of relative contributions (σ_{v_i}). Namely, as shown in Sect. 3.1, relative contributions indicate the prevalence of an indicator in the overall composite indicator. If the values of relative contribution of indicator i for all observed entities e, $e \in E$, vary, therefore, the values of composite indicator of all observed entities are sensitive to the weight assigned to indicator i. To measure the stability of the impact of an indicator i we used the standard deviation of its relative contribution. The smaller σ_{v_i} of indicator i indicates that the overall composite indicator values are stable when it comes to the impact of indicator i. Therefore, our idea is to minimize the sum of standard deviations of relative contributions through weights and thus minimize the overall indicator volatility. The proposed approach leads to maximization of rank stability and robustness, all observed through the prism of weight uncertainty.

Table 2 Comparative analysis of the weighting approaches

Criteria	eSS-CIDI	CIDI	PCA	BoD	DEA
Entity-specific weighting scheme	No	No	No	Yes	Yes
Incorporates stability of the created composite indicator	Yes	No	No	No	No
Based on an objective function	Yes	No	No	Yes	Yes

Next, we provide comparison of the eSS-CIDI approach with related weighting approaches (Table 2). The eSS-CIDI approach like CIDI and PCA does not assign entity-specific weights. Namely, BoD and DEA assign a specific weighting scheme to each observed entity (Amado et al. 2016; Cherchye et al. 2007). The criteria for which the proposed approach stands out is because it takes into account the stability of the created composite indicator while assigning weights. For all other observed weighting approaches, the uncertainty analysis for the weighting scheme can be conducted, but after the weights have been assigned. The eSS-CIDI approach is an approach based on optimization, so it is, as BoD and DEA, based on an objective function.

Additionally, we compare the proposed eSS-CIDI to similar approaches which aim to improve or measure the stability of the composite indicator. Compared to the approach of Foster et al. (2013), the eSS-CIDI provides the decision maker with a vector of most stable weights, while their approach provides a measure of the stability of an initial vector of weights. Their approach gives a measure of robustness of an initial or official weighting scheme, while it does not suggest a vector of weights which will be more stable. The same accounts for the approach of Permanyer (2011), whereas his approach to measuring the robustness of an initial vector of weights differs as it observes the distance from equality. The approach of Cherchye et al. (2008b) is based on the rank robustness measured through Lorenz dominance. Therefore, their results do not produce an overall composite composite indicator value, whereas they provide pairwise dominance results. The eSS-CIDI, on the other hand, provides an overall composite indicator, but it does not take into account pairwise comparisons. Paruolo et al. (2013) proposed an approach based on the actual effect of the weights to the overall metric, measured through the 'main effect'. The eSS-CIDI is in a way similar as it measures the proportional impact of the weighted indicator in the overall composite indicator value through relative contributions. The theoretical approach of the weighting approach of Becker et al. (2017) resembles the eSS-CIDI. They aimed to optimize weights so as the impact of weights measured through the 'main effect' fits the pre-specified importance. In the eSS-CIDI also an optimization approach is used, but the weights are to fit the bootstrap-CIDI constraints. The provided comparison shows that the eSS-CIDI provides a continuation of the resent research in this area.

Lastly, the objective of the proposed eSS-CIDI weighting approach should be observed through the prism of policymakers. Although the objective of the eSS-CIDI is understandable to an analyst, the question rises will its objective be understood by policymakers or government representatives or field experts who are working on the process of creating a composite indicator. Namely, will they choose the eSS-CIDI weighting scheme over equal weighting or a simpler data-driven weighting approach? In an attempt to answer the question the pros and cons of this approach to the process of policymaking should be analysed. One of the first benefits of the proposed approach is that the uncertainty analysis is automatically performed. Namely, Freudenberg (2003) states that this step is usually skipped in

the process of composite indicator creation. By applying the eSS-CIDI the policymakers can be sure that this step will be conducted. Another benefit is that the eSS-CIDI proposes a weighting scheme which creates a most stable indicator considering predefined weight constraints. This also reduces the possibility of critique of the composite indicator methodology which the policymakers wish to evade. Namely, the weighting scheme is often criticized step in the process of composite indicator creation (e.g. Amado et al. 2016; Despotis 2005; Dobrota et al. 2016; Jeremic et al. 2011). On the other side, there are potential drawbacks of the proposed approach. First is that the approach is not so easy understandable for ones who have no basic knowledge of statistics and operational research, which might make it difficult to adopt by wider public. Second is that the obtained weights of indicators might significantly differ from the socially acceptable weights or weights proposed by experts. Nevertheless, according to Saltelli (2007) no matter how good the statistical and theoretical background of a composite indicator is, its acceptance relies on peer acceptance. The same can be said for the weighting methodology. We believe that this discussion might persuade the policymakers to put confidence in our approach.

5 Case Studies

This section shows the application of the hybrid eSS-CIDI approach to two existing composite indicators. Herein we chose the *Academic Ranking of World Universities (ARWU)* and the *Networked Readiness Index (NRI).* We chose two composite indicators on which to implement the suggested weighting scheme because they belong to two different fields of study; one is in the field of higher education, and the other in the field of information and communication technology (ICT). Another, more important reason, is the official weighting scheme. In case of *ARWU*, the weighting scheme is based on expert opinion and the assigned weights are not equal, while in case of *NRI*, the weights are equal. Finally, the chosen indicators have different structures. Namely, *NRI* has a three-level structure, while *ARWU* is based solely on indicators.

5.1 Academic Ranking of World Universities (ARWU) 2017

Higher education is just one of the spheres of life that has seen the introduction of quantitative metrics into its assessment through university rankings (Daraio and Bonaccorsi 2017). University rankings came into the spotlight after US News and World Report began providing rankings of US universities in 1983, while they proliferated after 2003 when the *Academic Ranking of World Universities (ARWU)* has been published (Moed 2017). Therefore, the general opinion is that the university rankings are here to stay (Hazelkorn 2007). The university ranking that attracted our attention was the *ARWU* ranking published by the Institute of Higher Education of the Jiao Tong University in Shanghai which ranks the world's top 1000 universities (ShanghaiRanking 2018).

The ranking itself is based on six indicators which aim to rank institutions according to academic and research performance (Liu and Cheng 2005). The first two indicators, *Alumni* and *Award*, are related to the number of alumni/staff who won the Nobel Prize and/or Fields Medal. The following three indicators are bibliometric. The indicator *HiCi*, aims to measure the number of staff who are classified as highly cited researchers by *Clarivate Analytics*. The next two indicators measure the research output. The *N&S*, on one

Table 3 Official *ARWU* weights, min and max bootstrap CIDI weights and optimal weights suggested by the hybrid eSS-CIDI approach

Indicator	Official weight (%)	$w_{\min_i}$ (%)	$w_{\max_i}$ (%)	w_i (%)
Alumni	10	15.2	19.1	15.2
Award	20	15.2	21.5	15.2
HiCi	20	14.8	18.3	14.8
N&S	20	18.0	21.9	21.9
PUB	20	7.5	15.4	13.3
PCP	10	15.0	19.6	19.6

hand, indicates the number of papers published in *Nature* and *Science* in the last 5 years, while the *PUB* indicates the total number of papers indexed by the Science Citation Index-Expanded (SCIe) and the Social Science Citation Index (Social SCI) in the previous year. Both indicators take into account only publications of 'Article' type (ShanghaiRanking 2018). Finally, the *PCP* attempts to measure the academic performance of an institution by taking into account the number of full-time equivalent academic staff.

Taking a closer look at the *ARWU* methodology, to calculate the score of the *ARWU* ranking first raw indicator values are normalized. The normalization method used is the distance to a referent entity (Nardo et al. 2005), which is in this case the university with the highest indicator value. The best performing university is given a score of 100 and becomes the benchmark against which the scores of other universities are measured. In the next step, the normalized scores are weighted accordingly and aggregated using the linear sum (Dehon et al. 2010). The list of the *ARWU* indicators and their respective weights are given in Table 3.

Since it first appeared in 2003, the *ARWU* ranking has attracted both positive feedback and rigorous critique. Academics specialized in data analysis, bibliometrics, and composite indicators have tried to attract the attention of the ranking creators and the broader public to some of the methodological flaws the *ARWU* ranking faces. The methodological issue which is particularly significant for this research is the weighting scheme of the *ARWU* indicators. Namely, Dehon et al. (2010) scrutinized the *ARWU* and showed that its results are sensitive to the relative weight attributed to each of the indicators. Their research is also notable because it provided insight that the university rankings, as well as *ARWU*, are sensitive to weighting scheme alterations.

In the presented case study, we initially chose to observe the top 100 universities for the year 2017. The data was publicly available on the official site of the *ARWU* ranking (ShanghaiRanking 2017). As the indicators *Award* and *Alumni* are related to Nobel Prize and Fields medals winners, which not many universities can boast, many universities have zero values for these indicators. The high percentage of zero values of these indicators is seen an issue of the ARWU ranking (Docampo and Cram 2015; Maricic et al. 2017). In general, high per cent of zero values in the data can greatly complicate any statistical analysis or interpretation of results (Zuur et al. 2010). In some cases, the high percentage of zero values is inherent (e.g. murder rate), but otherwise it indicates poor data quality (Yang et al. 2015). Therefore, to improve the quality of the observed ranking, all universities within the top 100 which had zero value for any indicator were removed. In our analysis, we included 74 world-class universities within the top 100 for which all six indicator values are above 0.

According to the suggested algorithm, the first step was to perform the bootstrap CIDI and determine the allowed weight intervals needed for the eSS optimization. We performed

Table 4 Standard deviation of relative contribution per *ARWU* indicator for eSS-CIDI, official, CIDI, and equal weighting scheme

	Indicator						Sum
	Alumni	Award	HiCi	N&S	PUB	PCP	
eSS-CIDI weighting scheme	0.038803	0.045714	0.039835	0.037612	0.048795	0.041819	0.252578
Official weighting scheme	0.025635	0.060233	0.047017	0.032840	0.064685	0.022975	0.253385
CIDI weighting scheme	0.044214	0.049619	0.045404	0.034081	0.050273	0.035956	0.259547
Equal weighting scheme	0.041631	0.049941	0.043262	0.029554	0.058052	0.036127	0.258567

1000 iterations with a sample size of 47 ($0.632 \cdot 74 = 46.768 \approx 47$). The obtained intervals are also presented in Table 3. The original weights of indicators *Award* and *N&S* are covered by the interval, while the suggested weight intervals for *HiCi* and *PUB* are below the officially assigned weight. For example, *PUB* was assigned 20% while the bootstrap CIDI suggested weights from 7.5 to 15.4%. On the other hand, the latter two weight intervals were above the original weight. For example, the weight of the indicator *Alumni*, officially assigned a weight of 10%, has a suggested weight from the interval ranging from 15.2 to 19.1%. The obtained intervals give more significance to prestigious awards of alumni (*Alumni*) and per capita performance (*PCP*), while relegating the number of highly cited researchers (*HiCi*) and the level of publications (*PUB*). Bootstrap CIDI intervals are more oriented towards excellence measured through Nobel Prizes and Fields Medals.

In the following step, the eSS was performed and the results are given in Table 3. As it can be observed, for the first three indicators, the suggested weights are the lower limits of the bootstrap CIDI intervals. The weights of *N&S* and *PCP* are set as the upper limit of the bootstrap CIDI intervals. Interestingly, only one weight did not achieve the upper or lower bound of the constraint interval, the weight of *PUB*.

To evaluate the obtained weighting scheme, we compared the sum of standard deviations of relative contributions of the optimized, official, equal, and the CIDI weighting scheme[1] (Table 4). The sum of standard deviations of relative contributions of the official weighting scheme is 0.253385, while the sum of standard deviation of relative contributions of the eSS-CIDI weighting scheme is 0.252578. Compared with the equal weighting scheme and with the CIDI weighting scheme, the optimized weighting scheme again showed more stable results. Also, it is of interest to analyse the standard deviation of relative contributions. Comparing the official and the optimized weighting schemes, the indicators which were assigned more weight than the official have a greater standard deviation of relative contributions. Bearing in mind that university rankings, both international and national, should be stable and not subject to sharp fluctuations in the overall scores (Shattock 2017), the new eSS-CIDI weighting scheme might be very useful.

---

[1] The obtained CIDI weighting scheme is: *Alumni* 17.6%, *Award* 16.7%, *HiCi* 16.9%, N&S 19.4%, *PUB* 13.4%, and *PCP* 16.0%.
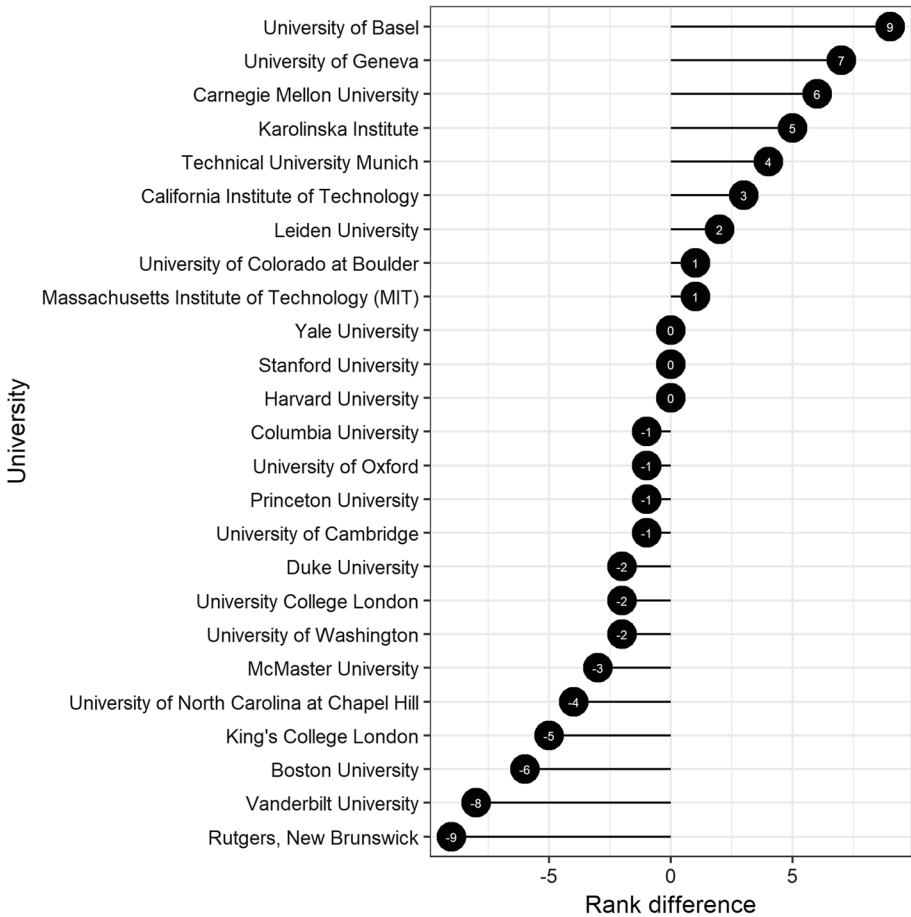
Fig. 2 Rank change between official *ARWU* and optimized *ARWU* for a chosen number of universities

Finally, we apply the novel weighting scheme. The changes in the *ARWU* ranks as the result of the application of optimal weights for a selected number of universities are given in Fig. 2. We can observe that the rank difference ranges from + 9 places (University of Basel) to − 9 places (Rutgers, The State University of New Jersey—New Brunswick). Most commonly universities dropped one ranking position (23.0% of universities), while the second most common case was no rank change (17.6% of universities). For example, the Karolinska Institute and University of Munich improved their rank by 5 places. One of the reasons for such rank advancement is their high values for the indicator *PCP*, 53.3 and 51.7, respectively, whose importance for the ranking procedure rose. On the other side of Fig. 2 we can see that King's College London and University of British Columbia dropped by 5 ranking places. Their overall score did not change dramatically, but other universities benefited more from the new weighting scheme and increased their overall score.

Looking at the top 10 universities according to the official ranking and comparing their ranks with the eSS-CIDI ranks we see that the universities within the top 10 did not change, but that there was a change in their ranks (Table 5). The biggest rank change was for the California Institute of Technology which advanced from 9th to 6th place. On the other

Table 5 Top and bottom ten universities by the official *ARWU* weighting scheme and their ranks according to the eSS-CIDI approach

University	Official rank	eSS-CIDI rank	University	Official rank	eSS-CIDI rank
Harvard University	1	1	Rutgers, The State University of New Jersey—New Brunswick	65	74
Stanford University	2	2	Boston University	66	72
University of Cambridge	3	4	Carnegie Mellon University	67	61
Massachusetts Institute of Technology (MIT)	4	3	Nagoya University	68	68
University of California, Berkeley	5	5	Leiden University	69	67
Princeton University	6	7	The University of Western Australia	70	71
University of Oxford	7	8	Technion-Israel Institute of Technology	71	69
Columbia University	8	9	University of Goettingen	72	65
California Institute of Technology	9	6	University of Basel	73	64
University of Chicago	10	10	The Australian National University	74	73

Table 6 Official *NRI* weights, min and max bootstrap CIDI weights and optimal weights suggested by the hybrid eSS-CIDI approach

Sub-index	Official weight (%)	$w_{\min_i}$ (%)	$w_{\max_i}$ (%)	w_i (%)
Environment	25	24.2	25.4	24.5
Readiness	25	21.7	23.1	21.7
Usage	25	26.0	26.9	26.9
Impacts	25	25.9	26.9	26.9

hand, there was a notable shift in the bottom 10 universities. The biggest rank changes occurred for the universities which were in the bottom 10 according to the official ranking. The higher sensitivity in the lower part of the ranks can be explained by the fact that those entities have neither good nor bad indicator values and therefore, they are highly dependent on the chosen weighting scheme (Saisana et al. 2011).

5.2 Networked Readiness Index 2016

The *Networked Readiness Index (NRI)* is a composite indicator which aims to measure the level up to which countries leverage ICT to increase their competitiveness. It was developed by the World Economic Forum and ranks and compares 151 countries using 53 indicators divided into 10 sub-categories and later into four sub-indexes: *Environment* of ICTs, *Readiness* of a society to use ICTs, the actual *Usage* of ICTs by all primary stakeholders, and *Impacts* that ICTs generate in the economy and society (Bilbao-Osorio et al. 2014). Such a complex framework provides broad and systematic comparison of the ICT development of countries around the globe (Kirkman et al. 2002).

The *NRI* is computed successively. Indicators are aggregated into sub-categories, which are later aggregated into sub-indexes, which are finally combined to obtain the overall composite indicator value. Prior to aggregation, scores of each indicator are normalized onto a common scale ranging from 1 to 7 (World Economic Forum 2016b). At each step of the aggregation equal weights are applied (Table 6). Such an approach to weighting remains a stumbling block in the process of constructing a composite indicator (Saisana and Tarantola 2002). Namely, the *NRI* weighting scheme was scrutinized using the Composite I-distance Indicator approach (CIDI) which showed that the current weighting scheme could be altered (Jovanovic-Milenkovic et al. 2016). Herein, we will place our attention on the weighting scheme on the level of sub-indexes, whereas we presumed that the weighting schemes on the level of indicators and sub-categories should not be changed.

In our case study, we observed 138 countries which had no missing data for the sub-indexes of the newest version of the *NRI* for the year 2016. The data on which the analysis was performed is publicly available (World Economic Forum 2016a). The bounds obtained using the bootstrap CIDI after 1000 replications with a subsample size of 87 are presented in Table 6. It can be observed that the bootstrap CIDI suggests a lower weight to the sub-index *Readiness*, in the range 21.7–23.1%, while the indicator *Usage* was given more importance, in the range 26.0–26.9%. The only indicator whose official weight is covered in the obtained interval is *Environment*, while the proposed weight interval for *Impacts* is slightly above the official 25%.

The next step was to perform the eSS and find the optimal set of weights. The results are also given in Table 6. Analysis of the optimized weights provides some interesting insights. Namely, while the weight of *Readiness* decreased to 21.7% from 25%, the importance of

Table 7 Standard deviation of relative contribution per *NRI* sub-index for eSS-CIDI, official and CIDI weighting scheme

	Sub-index				Sum
	Environment	Readiness	Usage	Impacts	
eSS-CIDI weighting scheme	0.025104	0.030313	0.013074	0.014000	0.082491
Official weighting scheme	0.026168	0.033376	0.012757	0.013960	0.086261
CIDI weighting scheme	0.025491	0.031094	0.013047	0.013896	0.083528

Usage and *Impacts* increased to 26.9%. We must also note that the weight of *Environment* exhibited almost no change. The weights of *Usage* and *Impacts* are set as the upper bounds of the bootstrap CIDI interval, while the weight of *Readiness* is set as the lower bound of the determined interval. Interestingly, as in the *ARWU* case study, only one weight did not achieve the upper or lower bound of the constraint interval, the weight of *Environment.*

The value of the objective function using the optimized weights is 0.082491, while the value of the objective function using the original weights is 0.086261 (Table 7). Interestingly, in this case CIDI[2] produced better results than the official weighting scheme, but not better than the eSS-CIDI. Again, the same pattern occurs: the indicators which were assigned lower weight than the official, have a smaller sum of standard deviations of relative contributions.

The changes in the *NRI* ranks as the result of the application of optimal weights for a selected number of countries are given in Fig. 3. We can observe that the rank difference ranges from + 3 places (for example China) to − 3 places (for example Ukraine). Most commonly countries did not change ranks (48.6% of countries), while the second most frequent shift was − 1 place (20.3% of countries).

5.3 Discussion on the Obtained Results

When it comes to the comparison of the obtained weighting schemes for the *ARWU* and *NRI* there are differences which should be emphasized. Both proposed weighting schemes are unequal. In the case of *ARWU* that goes in line with the official weighting scheme which is unequal, but the same does not account for *NRI* as its weighting scheme on the level of sub-index is equal. The eSS-CIDI weighting scheme of *ARWU* significantly differs from the official. The greatest difference is for the indicator *PCP* which gained 9.6 weight points compared to the official weight. On the other hand, the indicator *PUB* lost the most importance, from official 20 to 13.3%. The proposed weighitng scheme of *NRI* is unequal, but the proposed weights slightly deviate from equal. The biggest difference is in the case of sub-indexes *Usage* and *Impacts*, just 1.9 weight points.

The next point which should be more closely observed is the difference in the sums of standard deviations of relative contributions of indicators between the official and the eSS-CIDI weighting scheme. In the first case study for the *ARWU* ranking, the difference is 0.000807, while in the second case study, the difference is 0.003770. It can be observed that the difference does not implicate the rank differences. For *ARWU*, the differences in

---

[2] The CIDI weighting scheme is: *Environment* 24.8%, *Readiness* 22.5%, *Usage* 26.5%, and *Impacts* 26.2%.
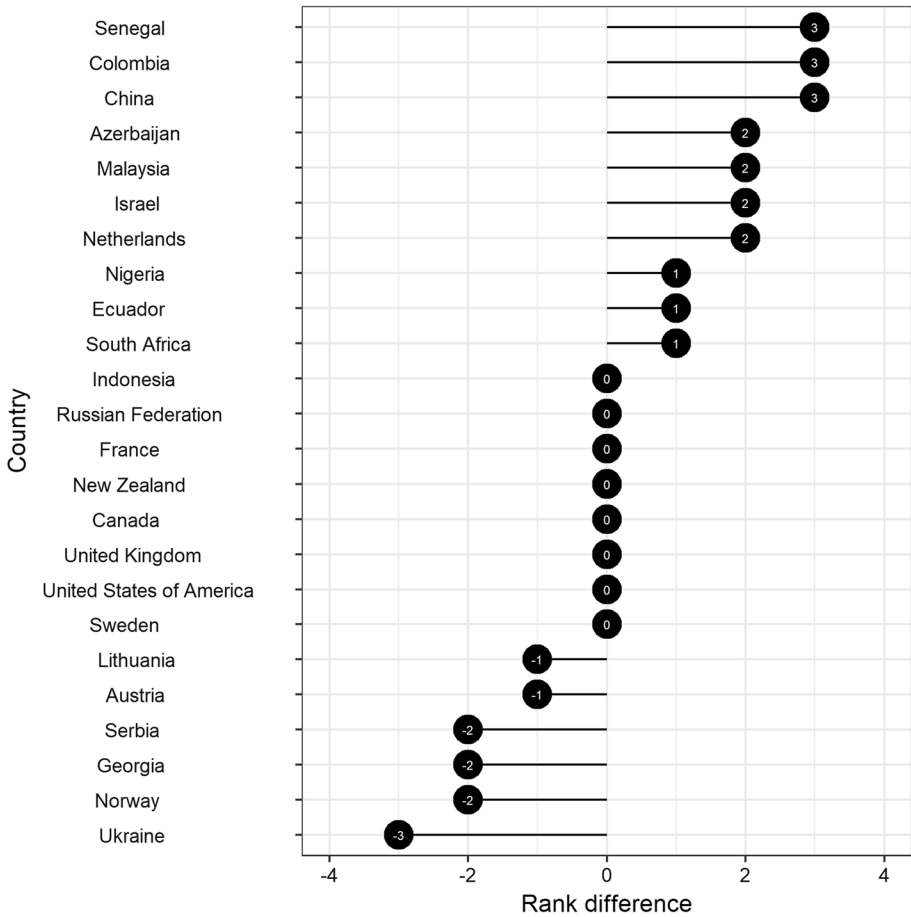
Fig. 3 Rank change between official *NRI* and optimized *NRI* for a chosen number of countries

the sums are smaller while the rank differences varied from + 9 to − 9 places. For *NRI* the rank differences varied from + 3 to − 3 places.

In both case studies, the stability of ranks of entities has been improved if the sole observed source of uncertainty was the weighting scheme. Our results showed that, in the two conducted case studies, the proposed eSS-CIDI weighting approach provided more stable ranks. However, this should be taken with caution. Namely, weighting schemes in both case studies are based on expert opinion, while the proposed eSS-CIDI weighting scheme is entirely data-driven. Therefore, the viewpoint of the interpretation of the results should be "what would happen if the weights were devised using a data-driven approach", rather than "the official weighting scheme is improper and should be corrected". Namely, there is no correct weighting scheme and no ambiguous solution. The proposed approach adds to the growing body of papers addressing the issue of establishing weighting schemes.

6 Prospects for Future Research

Assigning weights to composite indicators is one of the vital steps in the process of composite indicator creation (Nardo et al. 2005; Saisana et al. 2005) therefore novel approaches and insights are needed. Herein, we aimed to propose a novel approach which envelops two steps in the composite indicator creation: weighting and robustness analysis of the weighting scheme. The results presented in this study provide some interesting findings regarding the possibility to creating a data-driven weighting scheme which will propose the most stable composite indicator considering the given weight constraints.

Although the eSS-CIDI is easily adaptable to a wide range of composite indicator frameworks, there is place for further improvements of the approach. During our research, we identified several future directions of study. First modification of the proposed approach could be related to the application of another metaheuristics. Herein we used a variant of Scatter Search as it has demonstrated promising results (Egea et al. 2009; Remli et al. 2017; Rodriguez-Fernandez et al. 2006). This is especially true in the case of highly nonlinear and multimodal problems as is the case presented in this study. However, not many studies have conducted Scatter Search and its variants and the algorithm is not very popular and widespread among researchers compared to other metaheuristics. Therefore, it would be of interest to compare the results of various metaheuristics with the results obtained using eSS. For example, genetic algorithms and naturally inspired metaheuristics showed promising results in the field (Grupp and Schubert 2010; Jain et al. 2015). Another plausible drawback of the approach related to the applied metaheuristics is that it uses a global optimization method whose performance deteriorates when it is used to solve high dimension problems (Remli et al. 2017). This indicates that the process of assigning hundreds of weights in a large-scale composite indicator cannot be handled efficiently by eSS.

The second future research could tackle the issue of the type of aggregation. The proposed eSS-CIDI approach is based on the weighted sum type of aggregation. However, this type of aggregation has been criticised as it allows trade-off and does not penalize underperformance (Zhou et al. 2006). Also, as there is usually collinearity among indicators, the linear aggregation might generate biased metrics (Fusco 2015). Therefore, multiplicative aggregation, especially weighted geometric average, is being more and more used in the process of creation of composite indicators (Blancas et al. 2013; Giambona and Vassallo 2013; Rogge 2018a; Van Puyenbroeck and Rogge 2017; Verbunt and Rogge 2018). Namely, geometric average does not imply perfect substitutability and constant trade-offs between indicators/dimensions compared to arithmetic mean whereas it penalizes inequality among indicators/dimensions (Rogge 2018b). Thus, the possible direction of the study could be the modification of the approach to weighted geometric average.

In the presented case studies, the eSS-CIDI approach was used to assess an official weighting scheme or the weighting scheme suggested by the experts. However, the eSS-CIDI approach could be used for other purposes. It could be used to propose a weighting scheme of a novel composite indicator. Also, it could be employed to reduce the number of indicators which make the composite indicator framework. Namely, as the number of indicators rises, the significance of some indicators for the ranking process can decrease due to fact that other indicators cover the same and/or more variability of the observed multidimensional phenomenon (Maricic and Kostic-Stankovic 2016). Therefore, the correlation of these indicators with the final I-distance value can be 0 or close to zero. Consequently, the bootstrap CIDI weights can be close to zero. In future studies, the eSS-CIDI could see its application beyond weighting scheme assessment.

Finally, different weight constraints could be employed. Namely, other approaches could have easily been implemented. For example, an interesting approach to assigning weight constraints was employed by Saisana and Saltelli (2014). They used a wide enough interval around the official indicator weight in the sensitivity analysis of the Rule of Law Index. Herein, such an interval could be determined around the mean bootstrap CIDI weights. The $\pm\, n\sigma$ constraint has been used with success in robust optimization, where n denotes the desired sigma level (Koch et al. 2004). Radojicic et al. (2015) used $\pm 3\sigma$ and $\pm 6\sigma$ intervals around CIDI weights as constraints for DEA weights. Antony et al. (2006) conducted a Six Sigma constrained optimization using GA.

Some limitations of the given study could be stressed out which could be solved in some of the future studies. One of the characteristics of our approach is that it does not include expert opinion and thus constructs rigid metrics. However, this can easily be overcome, depending on the composite indicator creators' attitude towards the inclusion of expert opinion. Namely, if composite indicator creators want to create a strictly data-driven weighting scheme, they can use the model as proposed in this study. On the other hand, if they want to include expert opinion, they could use a hybrid eSS-CIDI weighting approach where weights at one indicator level could be expert-driven or equal, while on the other they could be obtained using the eSS-CIDI algorithm. The rationale for such an approach is that expert opinion can be quite useful in the process of weighting as it can improve the acceptability of the final composite indicator (Cherchye et al. 2007). Completely data-driven weighting approaches can propose weights which significantly deviate from public opinion on the importance of particular indicator, thus making it socially unacceptable (Banerjee 2018). Sometimes, expert opinion on indicator importance is available (Cherchye et al. 2008a) and should be included in the process of composite indicator creation process. Therefore, a hybrid expert opinion data-driven model could be a solution which would encompass the pros of both approaches (Edirisinghe and Zhang 2010; Joro and Viitala 2004; Reggi et al. 2014) and find a balance between the rigid data-driven approaches and the subjective expert-driven ones (Wang et al. 2009).

The next potential issue is that our model does not guarantee a more stable solution compared to the official composite index. Namely, it can occur that considering the bootstrap CIDI constraints, the approach suggests a weighting scheme which creates a less stable composite indicator. In that case, the official weighting scheme is proven to be more stable and according to our proposed approach it should not be altered. In such occurrence, the results of our proposed method can be interpreted as a mean of verification of the official weighting scheme. However, if one aims to find another weighting scheme other than the official, and the proposed bootstrap CIDI procedure with 1000 replications suggests rigid weight constraints, the number of replications can be reduced or increased. Also, other approaches to defining m can easily be implemented. For some suggestions on the size of m see Arcones and Gine (1989) or Bickel and Sakov (2008).

7 Conclusion

Composite indicators are used in various types of public debate such as those surrounding sustainable development policy, tourism policy, or higher education. Also, they have various policy applications such as aiding benchmark, performance assessment and resource allocation (Sevigny and Saisana 2016). This is exactly why it is crucial to create

statistically sound and robust composite indicators and unravel their construction methods to a wider audience (Dialga and Thi Hang Giang 2017). We have attempted here to open the black box in the segment of the weighting procedures by proposing a novel data-driven, optimization approach.

In this paper we introduced the enhanced Scatter Search—Composite I-distance Indicator (eSS-CIDI) optimization approach for determining weights within composite indicators. The weights obtained using our approach aim to increase the stability of the composite indicator. The goal function of the model is based on relative contributions of indicators to the overall indicator value (Sect. 3.1). The upper and lower weight constraints are defined using the here proposed *m out of n bootstrap CIDI without replacement* with *p* replications. To solve the optimization model the eSS metaheuristic was used (Sect. 3.2).

The proposed approach has two main contributions which should be pointed out. First is the proposed bootstrap CIDI as an approach for devising weight bounds. Namely, the suggested optimization could be carried out without bounding the weights via bootstrap CIDI. However, the bootstrap CIDI provides reliable bounds for the constraints which facilitate the optimization process by reducing the search space. This could be particularly crucial in high dimensional problems with large search space in which the performance of metaheuristics can deteriorate (Remli et al. 2017). Second is that it facilitates the usually cited stumbling block in the creation process of composite indicators: the weighting process. It provides completely data-driven weights which do not depend on the experts' opinions. Also, it helps the composite indicator creators as it guarantees that, taking into account the constraints, they will create a metric whose values are least susceptible to change due to weighting. On the other hand, the novel weighting scheme is rounded to three decimal places and the sum of weights is one, what makes it easier for stakeholders to understand and interpret.

To test our approach, we conducted two case studies on composite indicators from two spheres of life and with completely different frameworks. Namely, *ARWU* has six indicators and unequal weights (Sect. 5.1), while the *NRI* has four indicators and equal weights (Sect. 5.2). The optimization model had a unique solution for both cases which satisfied all the imposed constraints. The results show that the application of weights suggested by the eSS-CIDI approach leads to a smaller sum of standard deviations of relative contributions of the observed indicators compared to other tested weighting schemes. Namely, the proposed approach outperforms the existing tested approaches. The stability of the overall metrics could be improved using the proposed weights and weighting approach.

The current study contributes to the research by providing a comprehensive framework for determining data-driven weighting schemes. We believe that the proposed eSS-CIDI approach could serve as a catalyst for further research on the topic of the weighting approaches, especially on the topic of optimization through weights.

References

Ács, Z. J., Autio, E., & Szerb, L. (2014). National systems of entrepreneurship: Measurement issues and policy implications. *Research Policy, 43*(3), 476–494. https://doi.org/10.1016/j.respol.2013.08.016.

Amado, C. A. F., São José, J. M. S., & Santos, S. P. (2016). Measuring active ageing: A data envelopment analysis approach. *European Journal of Operational Research, 255,* 207–223. https://doi.org/10.1016/j.ejor.2016.04.048.

Angilella, S., Catalfo, P., Corrente, S., Giarlotta, A., Greco, S., & 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. https://doi.org/10.1016/j.knosys.2018.05.041.

Antony, J., Swarnkar, R., Kumar, M., & Tiwari, M. K. (2006). Design of synchronised supply chain: A genetic algorithm based six sigma constrained approach. *International Journal of Logistics Systems and Management, 2,* 120–140.

Arcones, M. A., & Gine, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l'I.H.P. Probabilités et Statistiques, 25*(4), 457–481.

Austin, P. C., & Small, D. S. (2014). The use of bootstrapping when using propensity-score matching without replacement: A simulation study. *Statistics in Medicine, 33*(24), 4306–4319. https://doi.org/10.1002/sim.6276.

Banerjee, A. K. (2018). Multidimensional indices with data-driven dimensional weights: A multidimensional coefficient of variation. *Arthaniti: Journal of Economic Theory and Practice.* https://doi.org/10.1177/0976747918792644.

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators, 80,* 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056.

Beran, R., & Ducharme, G. (1991). *Asymptotic theory for bootstrap methods in statistics.* Montréal: Centre de Recherches Mathematiques.

Bickel, P. J., Gotze, F., & van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses and remedies for losses. *Statistica Sinica, 7,* 1–31.

Bickel, P. J., & Sakov, A. (2008). On the choice of m in the m out of n bootstrap and its application to confidence bounds for extreme percentiles. *Statistica Sinica, 18,* 967–985.

Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014). *The global information technology report 2014. Rewards and risks of big data.*

Birkmann, J. (2007). Risk and vulnerability indicators at different scales: Applicability, usefulness and policy implications. *Environmental Hazards, 7*(1), 20–31. https://doi.org/10.1016/j.envhaz.2007.04.002.

Blancas, F. J., Contreras, I., & Ramírez-Hurtado, J. M. (2013). Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. *Journal of the Operational Research Society, 64*(5), 668–678. https://doi.org/10.1057/jors.2012.90.

Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys, 35*(3), 189–213. https://doi.org/10.1007/s10479-005-3971-7.

Bojórquez-Tapia, L. A., Sánchez-Colon, S., & Florez, A. (2005). Building consensus in environmental impact assessment through multicriteria modeling and sensitivity analysis. *Environmental Management, 36*(3), 469–481. https://doi.org/10.1007/s00267-004-0127-5.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research, 59*(2), 115–151. https://doi.org/10.1023/A:1016275505152.

Braga-Neto, U. M., & Dougherty, E. R. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics, 20*(3), 374–380. https://doi.org/10.1093/bioinformatics/btg419.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Annales de l'IHP Probabilités et Statistiques, 19*(3), 281–296.

Burgass, M. J., Halpern, B. S., Nicholson, E., & Milner-Gulland, E. J. (2017). Navigating uncertainty in environmental composite indicators. *Ecological Indicators, 75,* 268–278. https://doi.org/10.1016/j.ecolind.2016.12.034.

Casadio Tarabusi, E., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research, 112*(1), 19–45. https://doi.org/10.1007/s11205-012-0070-4.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to "benefit of the doubt" composite indicators. *Social Indicators Research, 82*(1), 111–145. https://doi.org/10.1007/s11205-006-9029-7.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., et al. (2008a). Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *Journal of the Operational Research Society, 59*(2), 239–251. https://doi.org/10.1057/palgrave.jors.2602445.

Cherchye, L., Ooghe, E., & Van Puyenbroeck, T. (2008b). Robust human development rankings. *The Journal of Economic Inequality, 6*(4), 287–321. https://doi.org/10.1007/s10888-007-9058-8.

Choquet, G. (1953). Theory of capacities. *Annales de l'Institut Fourier, 5*(54), 131–295.

Chowdhury, S., Moral, R. J., & Dulikravich, G. S. (2008). Predator-prey evolutionary multi-objective optimization algorithm: Performance and improvements. In *Proceedings of 7th ASMO-UK/ISSMO international conference on engineering design optimization* (pp. 1–10). Bath, UK. Retrieved from https://maidroc.fiu.edu/wp-content/uploads/2012/05/CP16pcp30.pdf.

Cole, T. J. (2015a). Setting number of decimal places for reporting risk ratios: Rule of four. *The BMJ, 350,* h1845. https://doi.org/10.1136/BMJ.H1845.

Cole, T. J. (2015b). Too many digits: The presentation of numerical data. *Archives of Disease in Childhood, 100*(7), 608–609. https://doi.org/10.1136/archdischild-2014-307149.

Corrente, S., Greco, S., Nicotra, M., Romano, M., & Schillaci, C. E. (2018). Evaluating and comparing entrepreneurial ecosystems using SMAA and SMAA-S. *The Journal of Technology Transfer*. https://doi.org/10.1007/s10961-018-9684-2.

Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J., & Jones, D. R. (1992). Quality-of-life assessment: Can we keep it simple? *Journal of the Royal Statistical Society. Series A, 155*(3), 353–393. https://doi.org/10.2307/2982889.

Daraio, C., & Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology, 68*(2), 508–529. https://doi.org/10.1002/asi.23679.

Davison, A. C., Hinkley, D. V., & Young, G. A. (2003). Recent developments in bootstrap methodology. *Statistical Science, 18*(2), 141–157. https://doi.org/10.1214/ss/1063994969.

De Bin, R., Janitza, S., Sauerbrei, W., & Boulesteix, A. L. (2016). Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. *Biometrics, 72*(1), 272–280. https://doi.org/10.1111/biom.12381.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews, 32*(1), 7–34. https://doi.org/10.1080/07474938.2012.690641.

Dehon, C., McCathie, A., & Verardi, V. (2010). Uncovering excellence in academic rankings: a closer look at the Shanghai ranking. *Scientometrics, 83*(2), 515–524. https://doi.org/10.1007/s11192-009-0076-0.

Despotis, D. K. (2005). A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society, 56*(8), 969–980. https://doi.org/10.1057/palgrave.jors.2601927.

Dialga, I., & Thi Hang Giang, L. (2017). Highlighting methodological limitations in the steps of composite indicators construction. *Social Indicators Research, 131*(2), 441–465. https://doi.org/10.1007/s11205-016-1263-z.

Dijkstra, L., Annoni, P., & Kozovska, K. (2011). *A series of short papers on regional research and indicators produced by the Directorate-General for Regional Policy*.

Dobbie, M. J., & Dail, D. (2013). Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators, 29,* 270–277. https://doi.org/10.1016/j.ecolind.2012.12.025.

Dobrota, M., Bulajic, M., Bornmann, L., & Jeremic, V. (2016). A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses. *Journal of the Association for Information Science and Technology, 67*(1), 200–211. https://doi.org/10.1002/asi.23355.

Dobrota, M., & Dobrota, M. (2016). ARWU ranking uncertainty and sensitivity: What if the award factor was excluded? *Journal of the Association for Information Science and Technology, 67*(2), 480–482. https://doi.org/10.1002/asi.23527.

Dobrota, M., & Jeremic, V. (2017). Shedding the light on the stability of university rankings in the ICT field. *IETE Technical Review, 34*(1), 75–82. https://doi.org/10.1080/02564602.2016.1144487.

Dobrota, M., Martic, M., Bulajic, M., & Jeremic, V. (2015). Two-phased composite I-distance indicator approach for evaluation of countries' information development. *Telecommunications Policy, 39*(5), 406–420. https://doi.org/10.1016/j.telpol.2015.03.003.

Docampo, D., & Cram, L. (2015). On the effects of institutional size in university classifications: The case of the Shanghai ranking. *Scientometrics, 102*(2), 1325–1346. https://doi.org/10.1007/s11192-014-1488-z.

Dubey, H. M., Pandit, M., & Panigrahi, B. K. (2016). Hydro-thermal-wind scheduling employing novel ant lion optimization technique with composite ranking index. *Renewable Energy, 99,* 18–34. https://doi.org/10.1016/j.renene.2016.06.039.

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management, 47*(2), 270–283. https://doi.org/10.1016/j.jeem.2003.09.001.

Edirisinghe, N. C. P., & Zhang, X. (2010). Input/output selection in DEA under expert information, with application to financial markets. *European Journal of Operational Research, 207*(3), 1669–1678. https://doi.org/10.1016/j.ejor.2010.06.027.

Efron, B. (1979). Bootstrap methods: Another look at the jacknife. *The Annals of Statistics, 7*(1), 1–26. https://doi.org/10.1214/aoms/1177692541.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC.

Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The 632 + bootstrap method. *Journal of the American Statistical Association, 92*(438), 548–560. https://doi.org/10.1080/01621459.1997.10474007.

Egea, J. A., Balsa-Canto, E., García, M.-S. G., & Banga, J. R. (2009). Dynamic optimization of nonlinear processes with an enhanced scatter search method. *Industrial and Engineering Chemistry Research, 48*(9), 4388–4401. https://doi.org/10.1021/ie801717t.

Egea, J. A., Henriques, D., Cokelaer, T., Villaverde, A., MacNamara, A., Danciu, D.-P., et al. (2014). MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinformatics, 15*(1), 136. https://doi.org/10.1186/1471-2105-15-136.

Ferrier, G. D., & Hirschberg, J. G. (1999). Can we bootstrap DEA scores? *Journal of Productivity Analysis, 11*(1), 81–92. https://doi.org/10.1023/A:1007787406099.

Foster, J. E., McGillivray, M., & Seth, S. (2013). Composite indices: Rank robustness, statistical association, and redundancy. *Econometric Reviews, 32*(1), 35–56. https://doi.org/10.1080/07474938.2012.690647.

Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment. *OECD Science, Technology and Industry Working Papers, 16*, 35. https://doi.org/10.1787/405566708255.

Fusco, E. (2015). Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research, 242*(2), 620–630. https://doi.org/10.1016/j.ejor.2014.10.017.

Gen, M., & Cheng, R. (2000). *Genetic algorithms and engineering optimization*. Hoboken: Wiley.

Gendreau, M., & Potvin, J. Y. (2010). *Handbook of metaheuristics* (Vol. 146). Berlin: Springer. https://doi.org/10.1007/978-1-4419-1665-5.

Geyer, C. (2013). *5601 Notes: The subsampling bootstrap*. Retrieved February 10, 2018 from http://www.stat.umn.edu/geyer/5601/notes/sub.pdf.

Giambona, F., & Vassallo, E. (2013). Composite indicator of financial development in a benefit-of-doubt approach. *Economic Notes, 42*(2), 171–202. https://doi.org/10.1111/j.1468-0300.2013.12005.x.

Giannetti, B. F., Bonilla, S. H., Silva, C. C., & Almeida, C. M. V. B. (2009). The reliability of experts' opinions in constructing a composite environmental index: The case of ESI 2005. *Journal of Environmental Management, 90*(8), 2448–2459. https://doi.org/10.1016/j.jenvman.2008.12.018.

Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision Sciences, 8*(1), 156–166. https://doi.org/10.1111/j.1540-5915.1977.tb01074.x.

Glover, F., & Kochenberger, G. A. (2003). *Handbook of metaheuristics*. Dordrecht: Kluwer Academic Publishers.

Glover, F., Laguna, M., & Martí, R. (2000). Fundamentals of scatter search and path relinking. *Control and Cybernetics, 29*(3), 652–684. https://doi.org/10.1007/0-306-48056-5_1.

Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. (2018a). Stochastic multi-attribute acceptability analysis (SMAA): An application to the ranking of Italian regions. *Regional Studies, 52*(4), 585–600. https://doi.org/10.1080/00343404.2017.1347612.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2018b). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1832-9.

Grupp, H., & Mogee, M. E. (2004). Indicators for national science and technology policy: How robust are composite indicators? *Research Policy, 33*(9), 1373–1384. https://doi.org/10.1016/j.respol.2004.09.007.

Grupp, H., & Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy, 39*(1), 67–78. https://doi.org/10.1016/j.respol.2009.10.002.

Gupta, J. N., & Sexton, R. S. (1999). Comparing backpropagation with a genetic algorithm for neural network training. *Omega*. https://doi.org/10.1016/S0305-0483(99)00027-4.

Hatefi, S. M., & Torabi, S. A. (2010). A common weight MCDA-DEA approach to construct composite indicators. *Ecological Economics, 70*(1), 114–120. https://doi.org/10.1016/j.ecolecon.2010.08.014.

Hazelkorn, E. (2007). How do rankings impact on higher education? *OECD Institutional Management in Higher Education*. December 2007, 1–4. Available at http://www.oecd.org/education/imhe/39802910.pdf. Accessed 21 Mar 2018.

Hedges, S. (1992). The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Molecular Biology and Evolution, 9*(2), 366–369. https://doi.org/10.1093/oxfordjournals.molbev.a040725.

Hellwig, Z. (1969). *On the problem of weighting in international comparisons*. Paris: UNESCO, Methods and Analysis Unit, Department of Social Sciences.

Ivanovic, B. (1977). *Teorija klasifikacije*. Beograd: Institut za ekonomiku industrije.

Jain, V., Kumar, A., Kumar, S., & Chandra, C. (2015). Weight restrictions in data envelopment analysis: A comprehensive genetic algorithm based approach for incorporating value judgments. *Expert Systems with Applications, 42*(3), 1503–1512. https://doi.org/10.1016/j.eswa.2014.09.034.

Jeremic, V., Bulajic, M., Martic, M., & Radojicic, Z. (2011). A fresh approach to evaluating the academic ranking of world universities. *Scientometrics, 87*(3), 587–596. https://doi.org/10.1007/s11192-011-0361-6.

Jeremic, V., Jovanovic Milenkovic, M., Radojicic, Z., & Martic, M. (2013). Excellence with leadership: The crown indicator of Scimago Institutions Rankings Iber report. *El Profesional de La Información, 22,* 474–480. https://doi.org/10.3145/epi.2013.sep.13.

Jiang, W., & Simon, R. (2007). A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification. *Statistics in Medicine, 26*(29), 5320–5334. https://doi.org/10.1002/sim.2968.

Joro, T., & Viitala, E.-J. (2004). Weight-restricted DEA in action: From expert opinions to mathematical models. *Journal of the Operational Research Society, 55*(8), 814–821. https://doi.org/10.1057/palgrave.jors.2601752.

Jovanovic, M., Jeremic, V., Savic, G., Bulajic, M., & Martic, M. (2012). How does the normalization of data affect the ARWU ranking? *Scientometrics, 93*(2), 319–327. https://doi.org/10.1007/s11192-012-0674-0.

Karaboga, D., Akay, B., & Ozturk, C. (2007). Artificial Bee Colony (ABC) optimization algorithm for training feed-forward neural networks. *Modeling Decisions for Artificial Intelligence, 4617,* 318–329. https://doi.org/10.1007/978-3-540-73729-2_30.

Kawaguchi, S., & Nishii, R. (2007). Hyperspectral image classification by bootstrap adaboost with random decision stumps. *IEEE Transactions on Geoscience and Remote Sensing, 45*(11), 3845–3851. https://doi.org/10.1109/TGRS.2007.903708.

Kim, D. Y., Han, S. H., & Kim, H. (2008). Discriminant analysis for predicting ranges of cost variance in international construction projects. *Journal of Construction Engineering and Management, 134*(6), 398–410. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:6(398).

Kim, K., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications, 19*(2), 125–132. https://doi.org/10.1016/S0957-4174(00)00027-0.

Kirkman, S., Osorio, A., & Sachs, D. (2002). The networked readiness index: Measuring the preparedness of nations for the networked world. In Kirkman (Ed.), *The global information technology report 2001–2002 readiness for the networked world* (pp. 10–29). New York: Oxford University Press.

Kleiner, A., Talwalkar, A., Sarkar, P., & Jordan, M. I. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society. Series B: Statistical Methodology, 76*(4), 795–816. https://doi.org/10.1111/rssb.12050.

Kline, R. B. (2005). *Principles and practice of structural equation modeling. Principles and practice of strucutral equation modeling* (2nd ed., Vol. 156). New York: The Guliford press. https://doi.org/10.1038/156278a0.

Koch, P. N., Yang, R. J., & Gu, L. (2004). Design for six sigma through robust optimization. *Structural and Multidisciplinary Optimization, 26*(3–4), 235–248. https://doi.org/10.1007/s00158-003-0337-0.

Konietschke, F., Bathke, A. C., Harrar, S. W., & Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis, 140,* 291–301. https://doi.org/10.1016/j.jmva.2015.05.001.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization, 9*(1), 112–147. https://doi.org/10.1137/S1052623496303470.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA—Stochastic multiobjective acceptability analysis. *European Journal of Operational Research, 106*(1), 137–143. https://doi.org/10.1016/S0377-2217(97)00163-X.

Ligmann-Zielinska, A., & Jankowski, P. (2008). A framework for sensitivity analysis in spatial multiple criteria evaluation. In T. J. Cova, H. J. Miller, K. Beard, A. U. Frank & M. F. Goodchild (Eds.),

Geographic information science (pp. 217–233). Berlin: Springer. https://doi.org/10.1007/978-3-540-87473-7_14.

Liu, N. C., & Cheng, Y. (2005). The academic ranking of world universities. *Higher Education in Europe, 30*(2), 127–136. https://doi.org/10.1080/03797720500260116.

Macedo, L. L., Godinho, P., & Alves, M. J. (2017). Mean-semivariance portfolio optimization with multiobjective evolutionary algorithms and technical analysis rules. *Expert Systems with Applications, 79,* 33–43. https://doi.org/10.1016/j.eswa.2017.02.033.

Macharis, C., Springael, J., De Brucker, K., & Verbeke, A. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis. *European Journal of Operational Research, 153*(2), 307–317. https://doi.org/10.1016/S0377-2217(03)00153-X.

Maggino, F. (2009). Towards more participative methods in the construction of social indicators: Survey techniques aimed at determining importance weights. In *62nd Annual conference of the world association for public opinion research on public opinion and survey research in a changing world*. Lausanne, Switzerland. Retrieved from https://flore.unifi.it/retrieve/handle/2158/370603/11441/Maggino_-_paper.pdf.

Mammen, E. (1992). *When does bootstrap work?* (Vol. 77). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-2950-6.

Mariano, E. B., Sobreiro, V. A., & do Nascimento Rebelatto, D. A. (2015). Human development and data envelopment analysis: A structured literature review. *Omega, 54,* 33–49. https://doi.org/10.1016/j.omega.2015.01.002.

Maricic, M., Bulajic, M., Martic, M., & Dobrota, M. (2015). Measuring the ict development: The fusion of biased and objective approach. *Naval Academy Scientific Bulletin, 18*(2), 326–334.

Maricic, M., Bulajic, M., Radojicic, Z., & Jeremic, V. (2016). Multivariate approach to imposing additional constraints on the benefit-of-the-doubt model: The case of QS world university rankings by subject. *Croatian Review of Economic, Business and Social Statistics, 2*(1), 1–14. https://doi.org/10.1515/crebss-2016-0005.

Maricic, M., & Kostic-Stankovic, M. (2016). Towards an impartial Responsible Competitiveness Index: A twofold multivariate I-distance approach. *Quality & Quantity, 50*(1), 103–120. https://doi.org/10.1007/s11135-014-0139-z.

Maricic, M., Zornic, N., Pilcevic, I., & Dacic-Pilcevic, A. (2017). ARWU vs. alternative ARWU ranking: What are the consequences for lower ranked universities? *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies, 22*(1), 1–14. https://doi.org/10.7595/management.fon.2017.0002.

Marković, M., Zdravković, S., Mitrović, M., & Radojičić, A. (2016). An iterative multivariate post hoc I-distance approach in evaluating OECD Better Life Index. *Social Indicators Research, 126*(1), 1–19. https://doi.org/10.1007/s11205-015-0879-8.

Mazziotta, M., & Pareto, A. (2007). Un indicatore sintetico di dotazione infrastrutturale: il metodo delle penalità per coefciente di variazione. In *Lo sviluppo regionale nell'Unione Europea-Obiettivi, strategie, politiche. Atti della XXVIII Conferenza Italiana di Scienze Regionali*. Bolzano.

Mazziotta, M., & Pareto, A. (2012). Methods for constructing composite indices: One for all or all for one. *Rivista Italiana Di Economia Demografia e Statistica, 67*(2), 67–80.

Melyn, W., & Moesen, W. (1991). *Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available*. Leuven University, Working paper ID 26175691.

Milenkovic, M. J., Brajovic, B., Milenkovic, D., Vukmirovic, D., & Jeremic, V. (2016). Beyond the equal-weight framework of the Networked Readiness Index: A multilevel I-distance methodology. *Information Development, 32*(4), 1120–1136. https://doi.org/10.1177/0266666915593136.

Mirjalili, S., Mohd Hashim, S. Z., & Moradian Sardroudi, H. (2012). Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. *Applied Mathematics and Computation, 218*(22), 11125–11137. https://doi.org/10.1016/j.amc.2012.04.069.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics, 110*(2), 967–990. https://doi.org/10.1007/s11192-016-2212-y.

Mousavi-Nasab, S. H., & Sotoudeh-Anvari, A. (2018). A new multi-criteria decision making approach for sustainable material selection problem: A critical study on rank reversal problem. *Journal of Cleaner Production, 182,* 466–484. https://doi.org/10.1016/j.jclepro.2018.02.062.

Munda, G. (2005a). "Measuring sustainability": A multi-criterion framework. *Environment, Development and Sustainability, 7*(1), 117–134. https://doi.org/10.1007/s10668-003-4713-0.

Munda, G. (2005b). Multiple criteria decision analysis and sustainable development. In S. Greco, M. Ehrgott, & J. Figueira (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 953–986). New York: Springer.

Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics, 41*(12), 1513–1523. https://doi.org/10.1080/0003684060 1019364.

Murias, P., de Miguel, J. C., & Rodríguez, D. (2008). A composite indicator for university quality assessment: The case of Spanish higher education system. *Social Indicators Research, 89*(1), 129–146. https://doi.org/10.1007/s11205-007-9226-z.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators. *OECD Statistics Working Papers*. https://doi.org/10.1787/53341 1815016.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling—Google Académico. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 353–377. https://doi.org/10.1207/S1532 8007SEM0803_2.

OECD. (2004). *The OECD-JRC handbook on practices for developing composite indicators*. Paris: OECD.

Otero-Muras, I., & Banga, J. R. (2014). Multicriteria global optimization for biocircuit design. *BMC Systems Biology, 8*(1), 113. https://doi.org/10.1186/s12918-014-0113-3.

Paracchini, M. L., Pacini, C., Calvo, S., & Vogt, J. (2008). Weighting and aggregation of indicators for sustainability impact assessment in the SENSOR context. In K. Helming, M. Pérez-Soba & P. Tabbush (Eds.), *Sustainability impact assessment of land use changes* (pp. 349–372). Berlin: Springer. https://doi.org/10.1007/978-3-540-78648-1_18.

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society), 176*(3), 609–634. https://doi.org/10.1111/ j.1467-985X.2012.01059.x.

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2009). How many bootstrap replicates are necessary? In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, C. Pandu Rangan, B. Steffen, D. Terzopoulos & D. Tygar (Eds.), *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 5541 LNBI, pp. 184–200). https://doi. org/10.1007/978-3-642-02008-7_13.

Penas, D. R., González, P., Egea, J. A., Banga, J. R., & Doallo, R. (2015). Parallel metaheuristics in computational biology: An asynchronous cooperative enhanced scatter search method. *Procedia Computer Science*, *51*, 630–639. https://doi.org/10.1016/j.procs.2015.05.331.

Penas, D. R., González, P., Egea, J. A., Doallo, R., & Banga, J. R. (2017). Parameter estimation in large-scale systems biology models: A parallel and self-adaptive cooperative strategy. *BMC Bioinformatics, 18*(1), 52. https://doi.org/10.1186/s12859-016-1452-4.

Pérez, V., Guerrero, F., González, M., Pérez, F., & Caballero, R. (2013). Composite indicator for the assessment of sustainability: The case of Cuban nature-based tourism destinations. *Ecological Indicators, 29,* 316–324. https://doi.org/10.1016/j.ecolind.2012.12.027.

Permanyer, I. (2011). Assessing the robustness of composite indices rankings. *Review of Income and Wealth, 57*(2), 306–326. https://doi.org/10.1111/j.1475-4991.2011.00442.x.

Politis, D. N., & Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics, 22*(4), 2031–2050. https://doi.org/10.1214/ aos/1176325770.

Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York, NY: Springer. https://doi. org/10.1007/978-1-4612-1554-7.

Pollesch, N. L., & Dale, V. H. (2016). Normalization in sustainability assessment: Methods and implications. *Ecological Economics, 130,* 195–208. https://doi.org/10.1016/j.ecolecon.2016.06.018.

Radojicic, M., Savic, G., & Jeremic, V. (2018). Measuring the efficiency of banks: The bootstrapped I-distance gar dea approach. *Technological and Economic Development of Economy, 24*(4), 1581–1605. https://doi.org/10.3846/tede.2018.3699.

Radojicic, M., Savic, G., Radovanovic, S., & Jeremic, V. (2015). A novel bootstrap dba-dea approach in evaluating efficiency of banks. *Naval Academy Scientific Bulletin, 18*(2), 375–384.

Ramón, N., Ruiz, J. L., & Sirvent, I. (2012). Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications, 39*(5), 4882–4889. https://doi.org/10.1016/j.eswa.2011.10.004.

Ray, A. K. (2007). Measurement of social development: An international comparison. *Social Indicators Research, 86*(1), 1–46. https://doi.org/10.1007/s11205-007-9097-3.

Reggi, L., Arduini, D., Biagetti, M., & Zanfei, A. (2014). How advanced are Italian regions in terms of public e-services? The construction of a composite indicator to analyze patterns of

innovation diffusion in the public sector. *Telecommunications Policy, 38*(5–6), 514–529. https://doi.org/10.1016/j.telpol.2013.12.005.

Remli, M. A., Deris, S., Mohamad, M. S., Omatu, S., & Corchado, J. M. (2017). An enhanced scatter search with combined opposition-based learning for parameter estimation in large-scale kinetic models of biochemical systems. *Engineering Applications of Artificial Intelligence, 62,* 164–180. https://doi.org/10.1016/j.engappai.2017.04.004.

Renzi, G., Canfora, L., Salvati, L., & Benedetti, A. (2017). Validation of the soil Biological Fertility Index (BFI) using a multidimensional statistical approach: A country-scale exercise. *CATENA, 149,* 294–299. https://doi.org/10.1016/j.catena.2016.10.002.

Rodriguez-Fernandez, M., Egea, J. A., & Banga, J. R. (2006). Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics, 7*(1), 483. https://doi.org/10.1186/1471-2105-7-483.

Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. *Ecological Indicators, 23,* 143–154. https://doi.org/10.1016/j.ecolind.2012.03.020.

Rogge, N. (2018a). Composite indicators as generalized benefit-of-the-doubt weighted averages. *European Journal of Operational Research, 267*(1), 381–392. https://doi.org/10.1016/j.ejor.2017.11.048.

Rogge, N. (2018b). On aggregating benefit of the doubt composite indicators. *European Journal of Operational Research, 264*(1), 364–369. https://doi.org/10.1016/j.ejor.2017.06.035.

Saisana, M., & D'Hombres, B. (2008). *Higher education rankings: Robustness issues and critical assessment. How much confidence can we have in higher education rankings? JRC.* https://doi.org/10.2788/92295.

Saisana, M., D'Hombres, B., & Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research Policy, 40*(1), 165–177. https://doi.org/10.1016/j.respol.2010.09.003.

Saisana, M., & Saltelli, A. (2014). JCR statistical audit of the WJP Rule of Law index 2014. In *World justice project: The world justice project Rule of Law index 2014* (pp. 188–197). Washington, DC: World Justice Project.

Saisana, M., Saltelli, A., & 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. https://doi.org/10.1111/j.1467-985X.2005.00350.x.

Saisana, M., & Tarantola, S. (2002). *State-of-the-art re port on current methodologies and practices for composite indicator development.* Retrieved December 20, 2017 from http://bookshop.europa.eu/en/state-of-the-art-report-on-current-methodologies-and-practices-for-composite-indicator-development-pbEUNA20408/%2520.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*(1), 65–77. https://doi.org/10.1007/s11205-006-0024-9.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis. Wiley series in probability and statistics.* Hoboken: Wiley.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2007). *Global sensitivity analysis. The primer.* Chichester: Wiley. https://doi.org/10.1002/9780470725184.

Savic, D., Jeremic, V., & Petrovic, N. (2016). Rebuilding the pillars of sustainable society index: A multivariate post hoc I-distance approach. *Problemy Ekorozwoju – Problems of Sustainable Development, 12*(1), 125–134.

Sayed, H., Hamed, R., Hosny, S. H., & Abdelhamid, A. H. (2018). Avoiding ranking contradictions in human development index using goal programming. *Social Indicators Research, 138*(2), 405–442. https://doi.org/10.1007/s11205-017-1663-8.

Sayed, H., Hamed, R., Ramadan, M. A.-G., & Hosny, S. (2015). Using meta-goal programming for a new human development indicator with distinguishable country ranks. *Social Indicators Research, 123*(1), 1–27. https://doi.org/10.1007/s11205-014-0723-6.

Sébastien, L., Bauler, T., & Lehtonen, M. (2014). Can indicators bridge the gap between science and policy? An exploration into the (non)use and (non)influence of indicators in EU and UK policy making. *Nature and Culture.* https://doi.org/10.3167/nc.2014.090305.

Sevigny, E. L., & Saisana, M. (2016). Measuring interstate variations in the consequences of illegal drugs: A composite indicator approach. *Social Indicators Research, 128*(2), 501–529. https://doi.org/10.1007/s11205-015-1042-2.

ShanghaiRanking. (2017). *ARWU world university rankings 2017. Academic ranking of world universities 2017. Top 500 universities. Shanghai ranking—2017.* Retrieved February 14, 2018, from http://www.shanghairanking.com/ARWU2017.html.

ShanghaiRanking. (2018). *Academic ranking of world universities 2018*. Retreived December 24, 2018 from http://www.shanhairanking.com/ARWU2018.html.

Shattock, M. (2017). The 'world class' university and international ranking systems: what are the policy implications for governments and institutions? *Policy Reviews in Higher Education, 1*(1), 4–21. https://doi.org/10.1080/23322969.2016.1236669.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2007). Development of composite sustainability performance index for steel industry. *Ecological Indicators, 7,* 565–588. https://doi.org/10.1016/j.ecolind.2006.06.004.

Socha, K., & Blum, C. (2007). An ant colony optimization algorithm for continuous optimization: Application to feed-forward neural network training. *Neural Computing and Applications, 16*(3), 235–247. https://doi.org/10.1007/s00521-007-0084-z.

Sörensen, K., Sevaux, M., & Glover, F. (2018). A history of metaheuristics. In R. Martí, P. Panos & M. G. C. Resende (Eds.), *Handbook of heuristics* (pp. 1–18). Cham: Springer. https://doi.org/10.1007/978-3-319-07153-4_4-1.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8,* 25.

Taghdisian, H., Pishvaie, M. R., & Farhadi, F. (2015). Multi-objective optimization approach for green design of methanol plant based on CO2-efficeincy indicator. *Journal of Cleaner Production, 103,* 640–650. https://doi.org/10.1016/j.jclepro.2014.05.032.

Talukder, B., Hipel, K., & VanLoon, G. (2017). Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques. *Resources, 6*(4), 66. https://doi.org/10.3390/resources6040066.

Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with benefit-of-the-doubt weights. *European Journal of Operational Research, 256*(3), 1004–1014. https://doi.org/10.1016/j.ejor.2016.07.038.

Verbunt, P., & Rogge, N. (2018). Geometric composite indicators with compromise benefit-of-the-doubt weights. *European Journal of Operational Research, 264*(1), 388–401. https://doi.org/10.1016/j.ejor.2017.06.061.

Villaverde, A., Egea, J. A., & Banga, J. R. (2012). A cooperative strategy for parameter estimation in large scale systems biology models. *BMC Systems Biology, 6*(1), 75. https://doi.org/10.1186/1752-0509-6-75.

Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2009.06.021.

Wang, X., & Triantaphyllou, E. (2008). Ranking irregularities when evaluating alternatives by using some ELECTRE methods. *Omega, 36*(1), 45–63. https://doi.org/10.1016/j.omega.2005.12.003.

Wang, Y.-M., & Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling, 49*(5–6), 1221–1229. https://doi.org/10.1016/j.mcm.2008.06.019.

World Economic Forum. (2016a). *Global information technology report 2016—Reports—World economic forum*. Retrieved December 23, 2017, from http://reports.weforum.org/global-information-technology-report-2016/.

World Economic Forum. (2016b). *Structure and methodology*. Retrieved September 10, 2017, from http://reports.weforum.org/global-information-technology-report-2015/structure-and-methodology/.

Xu, L. W., Yang, F. Q., Abula, A., & Qin, S. (2013). A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances. *Journal of Multivariate Analysis, 115,* 172–180. https://doi.org/10.1016/j.jmva.2012.10.008.

Yang, W., Mckinnon, M. C., & Turner, W. R. (2015). Quantifying human well-being for sustainability research and policy. *Ecosystem Health and Sustainability, 1*(4), 1–13. https://doi.org/10.1890/EHS15-0004.1.

Zanakis, S. H., Newburry, W., & Taras, V. (2016). Global social tolerance index and multi-method country rankings sensitivity. *Journal of International Business Studies, 47*(4), 480–497. https://doi.org/10.1057/jibs.2016.5.

Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research, 245*(2), 517–530. https://doi.org/10.1016/j.ejor.2015.03.036.

Zhou, P., Ang, B. W., & Poh, K. L. (2006). Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics, 59*(3), 305–311. https://doi.org/10.1016/j.ecolecon.2005.10.018.

Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research, 96*(1), 169–181. https://doi.org/10.1007/s11205-009-9472-3.

Zornic, N., Bornmann, L., Maricic, M., Markovic, A., Martic, M., & Jeremic, V. (2015). Ranking institutions within a university based on their scientific performance: A percentile-based approach. *El Profesional de La Información, 24*(5), 551–566. https://doi.org/10.3145/epi.2015.sep.05.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x.