Environmental Modelling and Software 145 (2021) 105208

Contents lists available at ScienceDirect

Environmental Modelling and Software

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

A framework based on statistical analysis and stakeholders' preferences to inform weighting in composite indicators

David Lindén [a,b], Marco Cinelli [b,c,1,*], Matteo Spada [d], William Becker [e], Patrick Gasser [b], Peter Burgherr [d]

[a] *Department of Sustainable Development, Environmental Science and Engineering, KTH Royal Institute of Technology, Teknikringen 10 B, SE, 100 44, Stockholm, Sweden*
[b] *Future Resilient Systems (FRS), Swiss Federal Institute of Technology (ETH) Zürich, Singapore-ETH Centre (SEC), CREATE Tower #06-01, 1 Create Way, 138602, Singapore, Singapore*
[c] *Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965, Poznań, Poland*
[d] *Paul Scherrer Institut (PSI), Laboratory for Energy Systems Analysis, 5232, Villigen PSI, Switzerland*
[e] *European Commission, Joint Research Centre, Via E. Fermi 2749, 21027, Ispra, Italy*

ARTICLE INFO

Keywords:
Composite indicators
Index
Weights
Optimization
Resilience
Security of electricity supply
Sensitivity analysis

ABSTRACT

Composite Indicators (CIs, a.k.a. indices) are increasingly used as they can simplify interpretation of results by condensing the information of a plurality of underlying indicators in a single measure. This paper demonstrates that the strength of the correlations between the indicators is directly linked with their capacity to transfer information to the CI. A measure of information transfer from each indicator is proposed along with two weight-optimization methods, which allow the weights to be adjusted to achieve either a targeted or maximized information transfer. The tools presented in this paper are applied to a case study for resilience assessment of energy systems, demonstrating how they can support the tailored development of CIs. These findings enable analysts bridging the statistical properties of the index with the weighting preferences from the stakeholders. They can thus choose a weighting scheme and possibly modify the index while achieving a more consistent (by correlation) index.

1. Introduction

Composite Indicators (CIs), also called indices,[2] are widely used synthetic measures for ranking and benchmarking alternatives across complex concepts (Saisana and Tarantola 2002; Nardo et al., 2008). A recent review by Greco et al. (2019) identifies an almost exponential growth of CIs over the past 20 years, highlighting their popularity in all domains that require aggregation of information for decision-making. A CI is the result of a mathematical combination of individual indicators that together act as a proxy of the phenomena being measured (Mazziotta and Pareto 2013). By combining a plurality of variables, CIs are able to assess and evaluate the performance of alternatives across multidimensional concepts, which are not directly measurable or clearly defined. A broad range of studies can be found in the literature that address topics such as ecological and environmental quality (Reichert et al., 2015; Reale et al., 2017; Oţoiu and Grădinaru 2018), sustainability (Rowley et al., 2012; Cinelli et al., 2014; Eurostat 2015; Hirschberg and Burgherr 2015), human development (UNDP 2016; Biggeri and Mauro 2018), competitiveness (World Economic Forum 2017) and quality of governance (World Bank 2020). Thereby, they represent flexible tools for supporting decision-making when more than one criterion is being considered (Greco et al., 2016).

The purpose of constructing a CI is, among other things, to condense and summarise the information contained in a number of underlying indicators, in a way that accurately reflects the underlying concept. There are two key notions here: first, condensing information; and second, accurately representing the underlying concept. These two ideas will be revisited repeatedly in this work.

* Corresponding author. Future Resilient Systems (FRS), Swiss Federal Institute of Technology (ETH) Zürich, Singapore-ETH Centre (SEC), CREATE Tower #06-01, 1 Create Way, 138602, Singapore, Singapore.
E-mail address: m.cinelli@luc.leidenuniv.nl (M. Cinelli).
[1] Current address: Paul Scherrer Institut (PSI), Laboratory for Energy Systems Analysis, 5232, Villigen PSI, Switzerland.
[2] Composite Indicator (CI) and index are used interchangeably throughout the paper.

https://doi.org/10.1016/j.envsoft.2021.105208
Accepted 15 September 2021
Available online 21 September 2021

The rankings provided by a CI represent an invaluable tool for conveying complex and sometimes elusive phenomena to a larger audience (Freudenberg 2003), because it is easier to interpret a single figure than finding a common trend amongst a multitude of indicators (Singh et al., 2009; Paruolo et al., 2013). Furthermore, developers are often keen to stress that composite measures are complementary to the underlying indicators, and serve as a structured access point to a complex set of data (Becker et al., 2018). However, developing a CI is far from trivial, involving a number of steps where the developer is obliged to make compromises and subjective choices (Booysen 2002; Mazziotta and Pareto 2013; Cinelli et al., 2020). Hence, the complementary nature of a CI is largely contingent on its underlying construction scheme.

An important, but often overlooked, aspect in the construction of CIs is the correlation structure between the underlying indicators and its effect on the overall score (i.e., the CI). Ideally, there should be positive correlations between the indicators as this indicates that individual variables are linked to an overarching concept (Meyers et al., 2013). Negative (or weak) statistical relationships can have implications for the meaningfulness of the CI, as some of these might represent features different from the overarching target concept being measured (Furr 2011). It must however be noted that according to the area of application and scope of the analysis, there can be indicators that are not necessarily positively correlated, and their inclusion might be driven by stakeholders' choices. It is anyhow important to assess the statistical properties of CIs to judge their scoring and aid its interpretation (Nardo et al., 2008). An example of this can be found in the Sustainable Society Index – where aggregation was avoided due to negative correlations between sub-dimensions (Saisana and Philippas 2012).

Complex systems modelling and analysis is driven by indicators that in the majority of the cases are interwoven and interdependent (Allen et al., 2017). Information theory has been proposed as a prime solution to study and quantify such dependencies between indicators (Prokopenko et al., 2009). Dependencies mean that the information provided by one indicator can be partially or fully inferred from another one. According to the structure of the system under consideration, each indicator carries a certain level of information about its functioning and behaviour. Consequently, several measures have been advanced to study how much new information each indicator can add to characterize the system, such as the marginal utility of information (Allen et al., 2017). This type of measure can be characterized as carrying a variable weight or relevance in the description of the system, since the higher the utility of the information carried by one indicator, the higher its influence.

Even if there is a wide body of literature that demonstrates the need to account for dependencies and overlaps between indicators (Csiszár and Shields 2004; Prokopenko et al., 2009; Allen et al., 2017; Mao et al., 2019; Davoudabadi et al., 2020), CIs are often developed with limited attention to such interrelationships (Cinelli et al., 2020). In turn, this can have a nontrivial influence on subsequent stages of construction, such as the weighting (and aggregation) of indicators (Paruolo et al., 2013; Becker et al., 2017; Davoudabadi et al., 2020), as discussed below.

Recalling the objectives of constructing a CI, one key point is that the index should accurately reflect the underlying concept. This requires that each indicator contributes in a way that agrees with the decision maker(s)' views on its importance to the concept. In CI aggregation, weights are assigned to reflect the trade-offs[3] between the indicators, based on stakeholders' or decision-makers' preferences (Mazziotta and Pareto 2017; Greco et al., 2019). Consequently, it is usually assumed that the weight assigned can be directly interpreted as a measure of an indicator's importance, independent from the dataset under analysis (Munda and Nardo 2005). However, this assumption is rarely justified. In fact, in order to better understand the actual trade-offs (i.e., the influence that each indicator has on the CI) of each indicator on the CI, Paruolo et al. (2013) propose a methodology based on nonlinear regression. It compares the assigned weights with an *ex post* measure of importance – in this case Karl Pearson's correlation ratio (also known as the *first order sensitivity index*), which is a coefficient of nonlinear association. It is found that the structure of the dataset and correlations between the indicators often have a decisive effect on each indicator's influence in the index. In fact, their influence rarely coincides with the assigned weights.

In a more recent study, Becker et al. (2017) build on this research by extending the nonlinear regression approaches to include decomposing the correlation ratio to examine the "correlated" and "uncorrelated" contributions of each indicator, drawing on global sensitivity analysis literature (Xu and Gertner 2008; Da Veiga et al., 2009). Furthermore, the authors introduce a weight-optimization algorithm, which optimises (i.e., reallocates) the weights with the aim of achieving the indicators' pre-specified values of trade-offs. The authors thus propose an approach to adjust the value of each indicator's weight in relation to their desired trade-offs. However, adjusting indicator trade-offs is not the only issue/objective of CI aggregation. As previously stated, the other key aim of a CI is that it should be a good summary of its underlying indicators. One way to interpret this goal is that it should maximize the amount of information transferred from the underlying indicators to the CI.

The two issues above (adjusting indicators' influence on the index and maximizing information transfer from the indicators) are rarely considered in CI development and when they are, researchers and practitioners tend to focus on either one or the other in isolation. Moreover, work focusing on adjusting indicator influence misses a key point - that they are effectively balancing the *information transferred* by each indicator. In addition, as recently discussed in a review on CI construction, the weighting of indicators based on the statistical structure of the data has been widely criticized mostly because weights are assigned with these methods on the performance matrix and not using the preferences from the stakeholders (i.e., stakeholder-based weighting) (Greco et al., 2019). The available literature on CI development seems to neglect that the statistical properties of the dataset can be used to understand the actual contribution that each indicator is going to have on the index, independently from the weights assigned by the stakeholders. Identification of weights of indicators by means of statistical analysis of the data can be labelled as data-driven and it can be used to complement or even substitute the stakeholder-based weighting, whenever the latter is not available or it cannot be conducted with the relevant decision makers (Kojadinovic 2004).

Even if some approaches for combining stakeholder-based and data-driven methods to define the weights of the indicators have been proposed (Zardari et al., 2015; Davoudabadi et al., 2020), there is not yet a framework to guide the use of both types of methods in weighting CI indicators. Our research fills this gap by showing that stakeholder-based and data-driven weighting methods can be successfully combined to achieve a well-informed set of weights for the indicators of the CI. More specifically, our contribution consists in demonstrating how the desired weight of each indicator can be achieved by means of the statistical properties in the performance matrix. This work brings together the two objectives of CI construction, (I) reaching the desired indicator trade-offs and (II) maximizing information transfer, under a single framework built on information theory. It shows that the two objectives are (depending on the correlation structure) usually contradictory in the context of weighting. CIs developed with the aim of reaching the desired indicators' trade-offs may come at the cost of poor information transfer, while the CIs built via an information transfer maximization approach

[3] Algorithms used in CIs are frequently weighted sums and the weights of their indicators have the meaning of trade-offs (Munda 2008b, a). These indicate the level of compensation between the indicators. In other works, they define the improvement required in the performance on one indicator to compensate for the worsening in performance of another indicator. For example, if the weight of indicator 1 is half the weight of indicator 2, it means that the improvement of two units on indicator 1 are needed to compensate the worsening of one unit on indicator 2.

can potentially have a very unbalanced contribution from the underlying indicators. Hence, there is a pragmatic need for developing a deeper understanding on how statistical dependencies between indicators in the dataset affect the indicators' influence and information transfer in CIs and thus their outcomes.

The first objective (i.e., adjusting information transfer) is important as it relates to the essence of shaping a CI that reflects the desired trade-offs between the indicators. In fact, even if the DM desires equal trade-offs between the indicators, the correlation structure might not allow to reach it with equal weights. As an example, if the DM chooses that the weight of indicator 1 is the same as the weight of indicator 2, it conceptually means that the improvement of one unit on indicator 1 is needed to compensate the worsening of one unit on indicator 2. The conventional approach in CI construction is that the analyst then assigns equal weights to the indicators. However, our statistical tools that study the (nonlinear) dependence between each indicator and the index show that due to the correlations in the dataset, in order to achieve the same weights (i.e., equal trade-offs) the actual values of the weights for these indicators should for example be twice as high for indicator 1 when compared to indicator 2. This confirms the need for considering both the requirements from the DM (e.g., the desired trade-offs) and the statistical properties of the performance matrix.

The second objective (i.e., maximizing information transfer) is important as it accounts for a situation where the DM requests as much information transfer as possible, irrespective of a pre-defined value for the trade-offs on the indicators. In this situation, the trade-offs between the indicators are defined solely according to the maximization of information transfer.

This paper provides a number of contributions to address these issues. In section 2, the concept of information transfer from indicators to the CI is formalised, by showing that the correlation ratio has a theoretical link with the concept of mutual information (a measure from information theory) under certain conditions. This formally demonstrates that the correlation ratio can be used as a tool to achieve both the objective of adjusting indicators' influence (e.g., balancing information contributions) and maximizing information transfer, by using an optimization approach with different objective functions. In section 3, the relationship between information transfer and the underlying correlation structure of CIs is explored with an analytical example, and it is shown that information transfer tends to a limit as more indicators are added to the framework. Then, in section 4, the tools proposed in this paper are applied to one version of the Electricity Supply Resilience Index (ESRI) developed at the Singapore-ETH Centre (Gasser et al., 2020), which was called Resilience Index for Analysis and Optimization (RIfAO). Discussion and conclusions complete the paper in section 5.

2. The concept of information transfer

This section proposes the use of the correlation ratio as a measure of the information transferred from each indicator to the CI. Its rationale is driven by the fact that the statistical relationships between the indicators in the dataset have an effect on how influential each indicator is in the overall system (Allen et al., 2017), which in this case is represented by the index.

The correlation ratio has been used in previous studies for adjusting the weights of CIs (Paruolo et al., 2013; Becker et al., 2017). Here, this idea is extended by linking it to the more intuitive concept of information transfer (or shared/mutual information), and by introducing two different objectives in weight adjustment: one based on balancing information transfer, and the other based on maximizing it.

Consider a CI y calculated as the *additive weighted average* (or weighted sum) – which is one of the most widely used methods for developing CIs (OECD 2008; Eisenfuhr et al., 2010; Bandura 2011; Langhans et al., 2014) – of n normalized variables x_i:

$$y_j = \sum_{i=1}^{n} w_i x_{ji},\ j = 1, 2, \ldots, m \tag{1}$$

where x_{ji} is the normalized score of alternative j (e.g., country) based on its raw value X_{ji} in the ith variable X_i, $i = 1, 2, \ldots, n$, and w_i is the weight (i.e., trade-off) assigned to the ith variable, such that $\sum_{i=1}^{n} w_i = 1$ and $w_i \geq 0$.

Fig. 1 illustrates this aggregation procedure. Now, after the aggregation, the objective is to understand the relationships between each indicator x_i and the aggregated CI y, and to see how it can be improved in terms of the two objectives mentioned above. In this work, the proposal is to measure the amount of information that is shared between the individual indicators and the CI, or the *information transferred* from each indicator to the CI (see again Fig. 1). Although equation (1) looks simple, correlations between indicators mean that the information transferred between y and x_i is not trivial to understand, and any of the three information transfer scenarios shown in Fig. 1 can occur, even with equal weighting.

The information transfer measure can be used as the basis for both the previously mentioned objectives of CI aggregation: (I) adjusting the influence of each indicator in relation to its assigned weight, and (II) maximizing the information transferred from the set of indicators to the CI. Information transfer is a more natural framework for assessing CIs than speaking directly in terms of correlations because CIs are effectively an information compaction problem: representing many indicators with one aggregated variable. In any case, this work will demonstrate that the two concepts are very similar and sometimes coincident. Building upon this logic, the concept of *information transfer* will, in this paper, be defined as: "*the (co-)dependence between the CI and each of its underlying indicators*". This could also be looked at as the information "shared" between the CI and each indicator, however since the CI is a product created by aggregating indicators, the term "transfer" will be used.

In the following sections, a measure of information transfer will be described, and two optimization problems, which satisfy the above-mentioned objectives, will be formulated.

2.1. Sensitivity index (S_i) as a measure of information transfer

One measure of information transfer is Mutual Information (I), which is an information theory measure that can be defined via entropy (Shannon 1948). Entropy is the foundational concept of information theory, which uses probability distributions to quantify the amount of information contained in a random variable (Cover and Thomas 2005). It can be used to measure the capacity of each variable to be used to predict the behaviour of the system in the next destination state, as well as to define the statistical complexity of a system (Prokopenko et al., 2009). With respect to the latter use, it is defined as Shannon's entropy and it defines the minimum amount of information required to statistically characterize the system. I can be understood as the amount of information that is shared between two random variables. The I between two continuous random variables $I(y, x_i)$, such as the CI y and one of its underlying indicators x_i, can be defined by:

$$I(y, x_i) = \iint f(y, x_i) \log \frac{f(y, x_i)}{f(y) f(x_i)} dy dx \tag{2}$$

where $f(y)$ and $f(x_i)$ are the marginal probability distributions and $f(y, x_i)$ is the joint probability distribution. Clearly, I allows us to directly measure a fundamental issue in composite indicators - how much information is passed from each indicator x_i to the CI y.

An intuitive way to think of information transfer in composite indicators is to consider: given the ranks of y, how well can one infer the ranks of the underlying indicators – in other words, how well is each

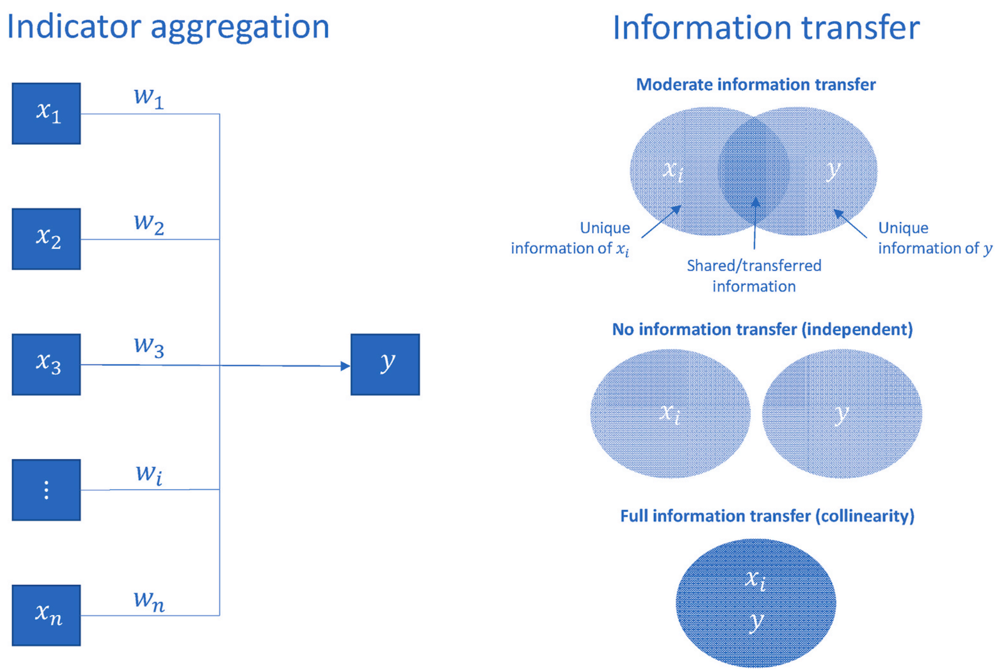

Fig. 1. Illustration of indicator aggregation and resulting information transfer, including examples of moderate/partial transfer, no information transfer, and full information transfer.

indicator represented in the final index ranking? If the mutual information between y and x_i is high, the ranks of x_i are very similar to those of y, therefore it can be considered as "well-represented". In the opposite case (low mutual information), the two ranks will differ markedly. Clearly, this is an important issue because a CI aims to summarise the information in its underlying indicators.

Although I is widely recognized within data analysis to possess ideal properties for measuring stochastic dependence – accounting for both linear and nonlinear dependencies – it has some drawbacks (Smith 2015). First, its interpretation is not straightforward. Unlike the well-known Pearson correlation coefficient (ρ), which has an absolute value in the range of 0 (complete linear independence) and 1 (complete linear dependence), the range of I is more open ended and can take on any value between 0 (complete independence) and infinity (complete dependence). Second, I is difficult to calculate from empirical data as it is based on probabilities and requires knowledge of the underlying marginal and joint distributions.

One way to alleviate these issues is to use a regression approach, which is simpler to estimate since the joint and marginal distributions do not need to be known (Kullback 1959). In fact, under restricted conditions it is possible to derive a direct link between I and coefficient of linear determination R^2 (Kullback 1959). When the joint probability distribution of both $\{x_i, y\}$ are normal, the expression for I in equation (2) reduces to:

$$I(y, x_i) = -\frac{1}{2}\ln\left(1 - R_i^2\right) \tag{3}$$

where R_i is the correlation between y and x_i. Thus, in the case of the multivariate Gaussian probability distribution, I between x_i and y can be fully represented by the coefficient of linear determination R_i^2. This is true because the dependence between two marginal distributions of a multivariate Gaussian distribution is by definition linear, hence the linear regression model is sufficient to capture the overall dependence (Dionisio et al., 2004).

In the nonlinear case, R_i^2 may still be used to approximate I, but becomes less accurate as associations start becoming nonlinear (Song et al., 2012; Smith 2015). To approximate I for more nonlinear cases, the proposal here is to use the *correlation ratio*, S_i, originally denoted η_i^2 (Pearson 1905). This is a coefficient of nonlinear association which can be estimated by a nonlinear regression model; see e.g., Paruolo et al. (2013) or Becker et al. (2017). Although this cannot be analytically linked to I, it is a direct nonlinear extension of R_i^2. In this respect, it should logically provide a good nonlinear approximation of I. Indeed, I has been shown to be directly related to the correlation ratio through Csiszár f-divergences (Da Veiga 2015).

The correlation ratio, also known as the *first order sensitivity index*, is a statistical measure of global 'variance-based' sensitivity (Saltelli et al., 2008). It is defined as:

$$S_i \equiv \eta_i^2 := \frac{V_{x_i}\left(E_{x_{\sim i}}(y|x_i)\right)}{V(y)} \tag{4}$$

where $V(y)$ is the unconditional variance of y, obtained when all factors x_i are allowed to vary and V_{x_i} is the variance of x_i as a function of the expected value $E_{x_{\sim i}}(y|x_i)$ for y given x_i. The expected value is the mean of y when only x_i is fixed, emphasised by the term $x_{\sim i}$, which is the vector containing all the variables $(x_1, \ldots, x_n)$ except variable x_i. Thus, $E_{x_{\sim i}}(y|x_i)$ is conditional on x_i and is, for that reason, also referred to as the *main effect* of x_i.

Notice that this definition, the ratio of the variance explained by x_i to the unconditional variance, is precisely a nonlinear generalisation of the well-known coefficient of determination R_i^2, such that S_i equals R_i^2 when the regression fit is linear (Wooldridge 2010). In fact, much like R_i^2, S_i can be interpreted as the expected reduction of variance in the CI scores if a given indicator could be fixed (Saisana and Saltelli 2011; Paruolo et al., 2013). S_i is also bounded within the range of 0–1, determining the degree of dependence between the CI and its underlying indicators. For instance, a value of 1 indicates complete dependence and a value of 0 implies complete independence. In information terms, a value of 1 means that all of the information contained in an indicator x_i has been transferred to the CI y, whereas a value of 0 implies that none of its information has been transferred. S_i is therefore a useful proxy of mutual information in more general nonlinear cases.

To estimate S_i, a regression approach is used. Since the main effect $E_{x_{\sim i}}(y|x_i)$ is a univariate function of x_i, it can be obtained by a nonlinear regression of y against x_i. In this study, a penalized cubic spline regression approach is used along the lines of Becker et al. (2017). To then obtain the first order sensitivity index S_i, the variance of the resulting curve is taken and standardised by the unconditional variance of y. Indeed, a comparative study by Song et al. (2012) showed that I can safely be replaced by a nonlinear regression model (based on splines), as it matches I for detecting nonlinear relationships.

The concept of entropy used in this study is an extension of the one presented in the work from Hwang and Yoon (1981). While these authors directly estimated the weights using the entropy method, in our study we make use of the results of the entropy method as input for the optimization models presented below. In fact, we defined the results of the entropy method as influence, or S_i, whose difference with respect to the initial weights (i.e., equal weights in our study) needs to be minimized using the optimization models.

2.2. Adjusting the weights to optimize information transfer

Given the information transfer measure proposed in the previous section, how can a CI be modified to either (I) adjust the relative information contribution of each indicator according to the desired trade-offs by the DM, or (II) maximize the overall information transfer? As hinted in the introduction, these objectives are often contradictory. Moreover, it is assumed that the input data for the indicators (i.e., normalized set) cannot be altered, and the aggregation method (e.g., arithmetic or geometric mean) is kept constant. In this case, the adjustments can be made by altering the weights. However, it is far from obvious which weight values will lead to the best properties in terms of objectives (I) and (II). The solution is found by framing the issue as a computational optimization problem. The first step is to build an "objective function", which, for any given weight values, calculates a score representing either (I) how "adjusted" the mean information transferred is, or (II) how much information is overall transferred to the composite index, by calculating correlation ratio (S_i) values for each indicator. The best set of weights are then found by an iterative optimization search algorithm, in this case the Nelder-Mead simplex search method (Lagarias et al., 1998; McKinnon 1998), which tries to find the highest value of the objective function. The two objective functions for (I) and (II) are described in detail in the following sections.

2.2.1. Objective I – Adjusting information transfer

Adjusting the relative information transfer (i.e., the influence) from the indicators to the CI in relation to their assigned weight is achieved in two steps – see details in Becker et al. (2017). First, to render the correlation ratios comparable to the weights, a normalization step is needed:

$$\widetilde{S}_i = S_i \Big/ \sum_{i=1}^{n} S_i \tag{5}$$

where $\widetilde{S}_i$ is the normalized correlation ratio of x_i, and $\sum_{i=1}^{n} \widetilde{S}_i = 1$. This allows the normalized correlation ratios to be directly compared to their target, the weights w_i (since the w_i also sum to 1).

Second, the problem of adjusting the contribution of the indicators can be formulated by defining an objective function as the sum of squared differences between the $\widetilde{S}_i$ at a given set of weights and the target $\widetilde{S}_i^*$, accordingly:

$$\boldsymbol{w}_{opt} = argmin_w \sum_{i=1}^{n} \left(\widetilde{S}_i^* - \widetilde{S}_i(\boldsymbol{w}) \right)^2 \tag{6}$$

where $\boldsymbol{w} = \{w_i\}_{i=1}^{n}$ and $\boldsymbol{w}_{opt} \geq 0$. Here it is assumed that the initially assigned weights represent the relative information transfer that is desired from each indicator, i.e., $\widetilde{S}_i^* = w_i$. Hence, the optimization problem in equation (6) tries to find a set of weights that minimises the discrepancy between the normalized correlation ratios ($\widetilde{S}_i$) and the initially assigned weights (w_i). From the perspective of information transfer, this equates to adjust the relative information transfer of each indicator in relation to the assigned weights by the DM.

2.2.2. Objective II – Maximizing information transfer

Mathematically, this problem is formulated by defining an objective function as the difference between a vector of all ones, $\vec{1}$ (i.e., the maximum information transfer, $S_i = 1$) and the S_i obtained at a given set of weights, accordingly:

$$\mathbf{w}_{opt} = argmin_w \sum_{i=1}^{n} \left(\vec{1}_i - S_i(\mathbf{w}) \right) \tag{7}$$

where the weights must sum to one $\mathbf{w} = \{w_i\}_{i=1}^{n}$, and are constrained to be positive $\mathbf{w}_{opt} \geq 0$. By minimising this objective function, the weights $\mathbf{w}_{opt}$ that maximize the total sum of information transferred from the indicators to the index can be found.

3. Relation between information transfer and average correlation

This section gives an analytical exploration of CI aggregation. It discusses how correlations between a set of indicators, $x_{i,\ldots,n}$, influence the information that is transferred from those indicators to the CI y. Here, R_i^2 (or "linear S_i") captures the linear dependence between x_i and y, as shown in equation (3). Consider the definition of R_i^2:

$$R_i^2 = \mathrm{corr}^2(y, x_i) = \frac{\mathrm{cov}^2(y, x_i)}{\mathrm{var}(y)\mathrm{var}(x_i)} \tag{8}$$

Now, assume a set of n variables with correlation matrix $\sum$. For this set of variables, the weighted mean is explored, such that $\mathbf{y} = \mathbf{X}\mathbf{w}$, where $\mathbf{X}$ is the $m \times n$ sample matrix, $\mathbf{w}$ is the $n \times 1$ vector of weights, and $\mathbf{y}$ is the vector of output values. By letting $\mathbf{e_i}$ be a $n \times 1$ vector where all elements are zero except the ith element, which is set to one, this linear combination gives (Johnson and Wichern 2007):

$$R_i^2 = \frac{(\mathbf{w}' \sum \mathbf{e}_i)^2}{(\mathbf{w}' \sum \mathbf{w})(\mathbf{e}' \sum \mathbf{e})} \tag{9}$$

Using the expression in equation (9) to obtain R_i^2, Fig. 2 shows its convergence as the number of indicators (n) changes from 2 to 100, for correlation matrices with average correlation coefficients (ρ) ranging from 0 to 1 with an interval of 0.1. It can be seen that $R^2(y, x_i)$ converges to ρ for large n, with faster convergence the closer ρ is to 1. This convergence is also mathematically derived in Appendix A in the Electronic Supplementary Information (ESI), where it is shown that, for indicators with equal weights and equal variance, R_i^2 tends to the average correlation coefficient (between indicators) as n tends to infinity.

From this analysis, it can be concluded that the strength of the correlations between the indicators is directly linked with their capacity to transfer information to the CI. A linear combination of poorly correlated indicators will, on average, have a weaker dependence (i.e., information transfer) between the indicators and the CI than a linear combination of highly correlated indicators. Although here information transfer has been framed via R_i^2, the fact that S_i is a nonlinear generalisation of R_i^2 allows these conclusions to be extended to the nonlinear case. Thus, the average correlation coefficient ρ of a given correlation matrix can provide a useful rule of thumb on how the information transfer capacity of a CI will be affected, when considering adding/subtracting indicators to a framework. This relationship will be further examined in the following section by applying the proposed measure to a case study.

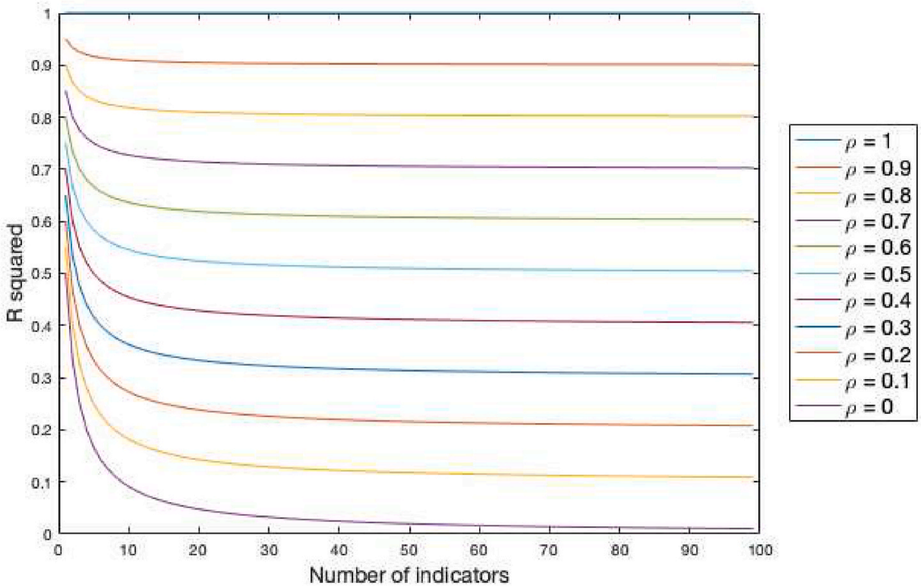

Fig. 2. $\boldsymbol{R}^2$as a function of the number of indicators n with different values of average correlation coefficient $\boldsymbol{\rho}$. The lines represent the different correlation scenarios, ranging from 0 to 1 with an interval of 0.1.

4. Case study: Electricity Supply Resilience Index

The management of complex socio-technical systems that are also embedded in environmental ones requires a dedicate array of tools to lead (i) the conception of their structure, (ii) the identification of their key variables and functions, (iii) the development of their underlying model, and (iv) the assessment of their integrated performance, as well as the effect of uncertainty in the input variables on the model output. One of the premier concepts proposed to conduct integrated assessment and management of systems is the one of resilience. It empowers analysts to consider technical, biophysical and socio-economic factors under one framework to support the understanding of the systems (Roostaie et al., 2019). A main example of complex socio-technical systems that requires a dedicated evaluation from a resilience perspective is the one of energy. The pervasive nature of this type of systems is such that it encompasses multiple others, including the biophysical ones at multiple scales (Fernandes Torres et al., 2019). In fact, energy systems have direct and indirect implications on the environmental systems, including water, land and air. Given the importance of this topic, the tools presented in Section 2 are tested with one CI developed to assess energy systems resilience. More specifically, they are used with one CI out of the 38 that constitute the Electricity Supply Resilience Index (ESRI), a CI developed within the Future Resilient Systems (FRS) program, at the Singapore-ETH Centre (SEC). It is based on 12 indicators evaluating countries' security of electricity supply from a resilience perspective (Gasser et al., 2020). The targets of the evaluation are 140 countries that represent a wide spectrum of nations from all around the world. ESRI uses data compiled from the International Energy Agency (IEA), the International Renewable ENergy Agency (IRENA), Paul Scherrer Institute's (PSI) ENergy-related Severe Accidents Database (ENSAD), the World Bank, the Swiss Reinsurance Company (Swiss Re) and the U.S. Energy Information Administration (EIA). The underlying data has been treated for outperformers, identified with the Interquartile range (IQR) method. Values are considered as outperformers if they lay outside 1.5 times the IQR from the first and third quartiles (Q1 and Q3 respectively). These were trimmed to the nearest value that is not an outperformer.[4] After trimming, missing values have been replaced by the average indicator values using an unconditional mean imputation,[5] as one of the common methods to deal with missing data (Nardo et al., 2008). The final scoring and ranking of ESRI is obtained by 38 different combinations of normalization methods and aggregation functions (Gasser et al., 2020). Normalization methods are used to render the raw data comparable and suitable for aggregation. In the cited study, eight of these approaches were selected. Ordinal, linear and non-linear normalizations were chosen to account for the variability of approaches that can be selected by the analysts. In CI development, once the indicators are normalized, they have to be aggregated to provide a final score and ranking. Gasser et al. (2020) considered six aggregation functions, in order to include different preferences of the decision maker in the form of compensation between the indicators.

The research in Gasser et al. (2020) is an extensive exploration of how different combinations of normalization methods and aggregation functions can affect the final score and ranking of the countries. However, the correlation analysis is limited to the assessment of the positive and negative trends between the indicators, as well as the coherence of the set of indicators (i.e., reliability of the scale). As shown in this paper in Section 2, correlation analysis can be used to do much more, including the exploration of the correlations between the indicators by assessing the information transferred from each indicator to the CI and study the effect that different weighting schemes have on each of them. Consequently, the tools proposed in Section 2 are used in this case study to extend the understanding of the effect of the data structure on the weighting stage in the CI. It must be noted that the CI resulting from the proposed weighting scheme is not more nor less valid compared to the ESRI proposed in Gasser et al. (2020). Given that CIs cannot be validated with objective measures as they model a concept that is not directly

[4] Note that the trimming is based on the actual data for the chosen 140 countries, not the theoretical min and max values. Across the 12 indicators, 88 values were identified as outperformers and trimmed to the nearest value within the IQR range.

[5] Across the 12 indicators, 65 instances of missing values were identified and replaced. It must be noted that the use of the indicator mean can result in a decrease of the correlations.

measurable, the value of the research resides in refining the learning about the implications of different data structure on the influence that indicators have in CIs.

In this paper, the tools presented in Section 2 are applied to one CI, developed with the combination of one normalization method (i.e., *min-max normalization)* and one aggregation function (i.e., *additive weighted sum)* to develop ESRI. The reason for this choice is that these are among the most commonly used approaches in their respective discipline (Carrino 2017; El Gibari et al., 2019; Greco et al., 2019), so the results are of interest to a large audience of analysts and decision makers. The index used in this paper and obtained with this combination of normalization method and aggregation function is called Resilience Index for Analysis and Optimization (RIfAO). The software called Composite Indicator Analysis and Optimization (CIAO) (Lindén et al., 2021), developed by some of the authors of this paper too, was used to perform the statistical analysis. Appendix B in the ESI provides more details on the framework and the indicators that constitute RIfAO, while Appendix C in the ESI includes the raw and normalized dataset used to construct RIfAO. It must be pointed out that no final scores of RIfAO are actually presented and discussed, since the objective of this case study is not to focus on the rankings obtained with this index, but rather to apply the optimization algorithms according to the objectives (I) and (II) presented in section 2.2 to achieve the desired information transfer from each indicator to the CI. Furthermore, Appendix D in the ESI presents the results of the same analysis by using the raw dataset, i.e., the dataset without trimming the outperformers, which shows that similar trends have been found as with the application of CIAO tool with the RIfAO dataset with the trimmed outperformers.

The methodology used to develop RIfAO, conduct the statistical analysis with the tools from Section 2, and elaborate the resulting recommendations for weighting scenario choice and index revision is shown in Fig. 3. Step 1 refers to the normalization of the dataset with the min-max normalization. In step 2, the correlations are analysed by means of Pearson correlation coefficient ρ to study the interrelations between the indicators. The normalized indicators are then aggregated with the additive weighted sum in step 3. Step 4 studies the information transferred (S_i) at equal weights and discusses the average correlation measured with respect to the step-wise addition of indicators. Lastly, step 5 provides recommendations for the choice of a weighting scheme according to a set of conditions that the DM might be interested to set for the index development. This leads to three scenarios (i.e., scenario A, B, C) which represent different combinations of three main features of the problem: (i) the variability of the information transferred (S_i) from each indicator to the index; (ii) the possible removal of one or more indicators from the index; and (iii) the possible loss of mean information transfer (S_i^{mean}). Each scenario is described in detail in section 4.2 and 4.3.

Step 1 in RIfAO development leads to the normalization of the dataset. For indicators with a positive polarity - meaning that the higher the value the better for the evaluation - the chosen normalization method is given by the formula $[X_{ji} - \min(X_i)]/[\max(X_i) - \min(X_i)]$. Indicators with a negative polarity - meaning that the lower the value the better for the evaluation - are transformed via $[1 - [X_{ji} - \min(X_i)]/[\max(X_i) - \min(X_i)]]$, where X_{ji} is the raw country value in the ith indicator X_i, $i = 1, 2, \ldots, n$. This procedure results in a linear transformation of the data, ranging from 0 (min) to 1 (max), and is performed on all indicators to render them comparable. Table 1 gives an overview of each of the 12 indicators that are included in the RIfAO framework, and Fig. 4 shows the Pearson correlation coefficients (ρ) between them (step 2 in Fig. 3). For conciseness, the indicators are labelled according to their ID number (e.g., IND 1), as defined in Table 1, in all graphs and figures.

By examining the correlation structure of RIfAO, it can be noticed that there is a large variation in the correlation strength between the indicators, with values ranging from −0.44 to 0.94. Although many indicators show a positive correlation between them – the highest (ρ = 0.94) being between IND_3 (Control of corruption) and IND_{10} (Government effectiveness) – there are also a number of negative trends visible. IND_6 (Electricity import dependence) showcases negative correlations with all the other indicators. This finding shows that IND_6 is mostly capturing a trend which is opposite to the other indicators in the dataset. Also, a few non-significant correlations[6] can be seen. Four out of the eleven negative correlations displayed by IND_6 are non-significant. IND_7 (Equivalent availability factor), except for a high positive correlation with IND_2 (Severe accident risks), presents non-significant correlations, all close to 0. This finding confirms how IND_7 is mostly disconnected from the trends of the other indicators in the dataset. These last two indicators proved to be of high interest in the subsequent stages of the analysis, especially when discussing the possible re-structuring of RIfAO.

4.1. Information transfer at equal weights

As far as weighting is concerned, equal weights are assigned to each indicator, with the modelling assumption that the trade-offs between each one included in the conceptual framework should be equal. This section explores information transfer in RIfAO at equal weights and it is performed in two steps. First, the RIfAO indicators are aggregated with equal weights (step 3 in Fig. 3) and an ex-post assessment of information transfer is performed by estimating the correlation ratios, via regression analysis, between the indicators and the index (step 4 in Fig. 3). The resulting regression fits are shown in Fig. 5, where both a linear (R_i^2) and nonlinear (S_i) regression model are fitted to the data. Second, the resulting correlation ratios (S_i) are then normalized and assessed in comparison to the vector of equal weights. This comparison is shown in Table 2.

From observing the resulting regression fits and the estimated R_i^2 and S_i values in Fig. 5, it can be noted that the indicators showing a linear trend towards the index (e.g., IND_3 – *Control of corruption* or IND_4 – *Political stability*) also have a low discrepancy between their R_i^2 and S_i measure. In these cases, linear estimates are sufficient to capture their dependence. However, there are also indicators that display nonlinear tendencies towards the index (e.g., IND_1 – *SAIDI* or IND_2 – *Severe accident risks*). In these cases, the linear regression model underestimates their dependence (see e.g., IND_2 which has an R_i^2 of 0.48 but an S_i of 0.66). This highlights the importance of also considering nonlinearities between the indicators and the CI when estimating dependence.

What is further evident from Fig. 5 is that not all indicators are transferring an equal amount of information, hence they do not have the same influence on the index, even though they are assigned equal weights. Thus, they are not equally influential in representing countries across the concept measured by RIfAO. The normalized correlation ratios ($\widetilde{S}_i$) in Table 2 further showcase this discrepancy (see "Deviation ratio" column), with values ranging from 64% overrepresentation (IND_{10}) to − 77% underrepresentation (IND_7). By re-examining the correlation matrix in Fig. 4, a connection between correlation strength and information transfer is evident: the information in the highly correlated indicators (e.g., $IND_{3,8,10,12}$) tends to be overrepresented, whereas the opposite holds true for the poorly, non- or negatively correlated indicators (e.g., $IND_{5,6,7,11}$). These findings are especially relevant in relation to the previously defined link between correlation and information transfer under restricted conditions (see Section 3). Indeed, even when distributions are not strictly linear, an indicator's correlation with the other aggregated indicators provides a strong indication of its capacity to transfer information to the CI.

Based on this statistical analysis, it is possible to assign the indicators to three groups (Table 2):

[6] Defined according to significance level p = 0.05.

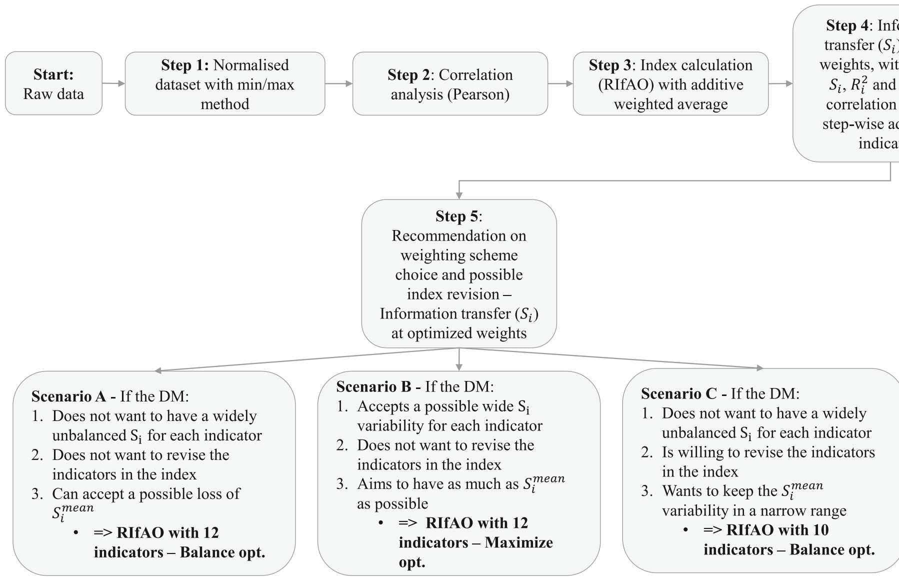

Fig. 3. The methodology used to develop RIfAO and the resulting recommendation for weighting scenario choice and index revision (w.r.t. = with respect to).

- Group 1: $IND_{3,8,10,12}$ high correlations and S_i, and also high positive deviation ratios. This characterises indicators that are overrepresented.
- Group 2: $IND_{5,6,7,11}$: low correlations and S_i, and also the highest negative, as well as absolute, deviation ratios. This characterises indicators that are underrepresented.
- Group 3: $IND_{1,2,4,9}$: intermediate correlations and deviation ratios, leading to moderate over- or under-representation.

The analytical analysis presented in Section 3 was adapted to RIfAO to study the effect of each indicator on the average correlations of the index (step 4 in Fig. 3). The results are presented in Fig. 6, showing how the average R_i^2, S_i and Pearson correlation (ρ) perform when indicators are added incrementally one-by-one to develop RIfAO. The measures show a common trend. Nonetheless, it can be seen how notable divergence emerges between S_i and Pearson correlation (ρ) when IND_6 and IND_7 are added. This analysis also shows that there is a significant "drop-off" in information transfer when IND_6 and IND_7 are added to the framework, which confirms that low correlated indicators result in low information transfer. In addition to the findings in Section 3, these results show that the average correlation can provide a useful, albeit not perfect, rule of thumb with respect to how much information (on average) is transferred from a set of indicators to the CI – even for a smaller sample size and when distributions are not strictly linear.

4.2. Information transfer at optimized weights

The variance-based analysis of RIfAO shows that the information transfer from the indicators to the CI is not equal, even though equal weights are applied, and strongly driven by the correlation structure. In addition, the information transfer from each indicator to the CI is not maximized. This section explores two avenues of weighting that a decision-maker might be interested in case he/she wants to achieve a balanced information transfer or a maximized one, while the framework of indicators has to remain the same. They are contextualized as two different scenarios, Scenario A and Scenario B, with different conditions that a DM might require to be met (step 5 in Fig. 3).

Scenario A considers a DM who:

1. Does not want to have a widely unbalanced S_i for each indicator;
2. Does not want to revise the indicators in the index;
3. Can accept a possible loss of S_i^{mean}.

This scenario results in RifAO with 12 indicators, where the main objective is to equally balance the information transfer from each indicator (*Balance opt.*).

Scenario B considers a DM who:

1. Accepts a possible wide S_i variability for each indicator;
2. Does not want to revise the indicators in the index;
3. Aims to have as much as S_i^{mean} as possible.

This scenario results in RifAO with 12 indicators, where the main objective is to maximize the total information transferred from each indicator (*Maximize opt.*).

The scenarios are modelled by optimizing the weights in line with the objective functions (equations (6) and (7), respectively) defined in Section 2. The next sections describe the results of each scenario.

4.2.1. Scenario A – Equally balancing the information transfer from each indicator (Balance opt.)

Scenario A results in the most unbalanced set of weights, as shown in Fig. 7. Most notably, the negatively correlated indicator (IND_6 - *Electricity import dependence*) receives the highest weight (35%) and also the non-correlated indicator (IND_7 - *Equivalent availability factor*) receives a substantial share of the weight (10%). Furthermore, five indicators (IND_2 - *Severe accident risk*, IND_3 - *Control of corruption*, IND_8 - *GDP* per capita, IND_{10} - *Government effectiveness* and IND_{12} - *Ease of doing business*) receive zero weight and two more (IND_1 - *SAIDI* and IND_9 - *Insurance penetration*) obtain a weight close to zero (i.e., 0.01). Even though only five indicators receive a weight greater than 0.01, as shown by the correlation ratios in Fig. 8, the information contained within the zero-weighted indicators is still captured by the CI simply through correlation. Judging from previous observations, it can be assumed that these indicators (excluding IND_2) are sufficiently represented by the inclusion of IND_4, with which they are all highly positively correlated (see Fig. 4).

The error bars in Fig. 8, representing the 5–95% percentiles, show that the resulting weighting vector from the *Balance opt.* objective would achieve the most well-balanced information transfer from each indicator, ranging from $S_i^{min} = 0.14$ to $S_i^{max} = 0.25$. However, the average contribution is relatively low ($S_i^{mean} = 0.19$). The correlation ratios in Fig. 8 show that only two indicators (IND_6 and IND_7) measure an increased information transfer, compared to the case of equal weights. Hence, this weighting scheme does practically not improve the total information transfer but rather reduces the information transfer from

Table 1
Descriptive statistics (prior to normalization) for the 12 indicators used to develop RIfAO. Min and Max values refer to the studied countries indicator scores, but not necessarily the whole value range that a country can take.

ID – Indicator	Unit	Polarity	Mean	SD	Min	Max
IND 1 – System Average Interruption Duration Index (SAIDI)	Hours per year and customer	–	6.8	7.5	0	21.4
IND 2 – Severe accident risks	Fatalities/GWeyr	–	1.7	2.1	0	7
IND 3 – Control of corruption	Percentile rank[a]	+	49.1	29.6	0.5	100
IND 4 – Political stability and absence of violence/terrorism	Percentile rank	+	45.4	28.1	0	99.1
IND 5 – Electricity mix diversity	Normalized Shannon index	+	0.4	0.2	0	0.8
IND 6 – Electricity import dependence	Ratio (cons/prod)	–	0.9	0.1	0.6	1.2
IND 7 – Equivalent availability factor	%	+	70.3	14.2	37.3	85.2
IND 8 – GDP per capita	2010 USD per capita	+	14582	16348	332	50107
IND 9 – Insurance penetration	premiums paid in % of GDP	+	1.6	0.9	0.1	3.9
IND 10 – Government effectiveness	Percentile rank	+	53.3	29	0.5	100
IND 11 – Average outage time	Hours	–	1.7	1	0	4
IND 12 – Ease of doing business	Distance to frontier	+	62.7	13.1	32.8	86.4

[a] Percentile rank is the proportion of scores in its frequency distribution that are equal to or lower than it. For example, if country A has a percentile rank of 88%, it means that 88% of the other countries have a score below the one of country A.

the highly correlated indicators, to target a balanced contribution. In other words, the *Balance opt.* weighting scheme focuses mostly on the indicators which are underrepresented ($IND_{5,6,7,11}$, see Group 2 in Table 2), at the cost of reduced mean information transfer (S_i^{mean}).

4.2.2. Scenario B - Maximize the total information transferred from each indicator (Maximize opt.)

This scenario results in a slightly less unbalanced set of weights than in scenario A (see Fig. 7). In this setting, the weights are mostly assigned to the highly correlated indicators (e.g., IND_4 - Political stability (12%), IND_8 - GDP per capita (19%) and IND_{10} - Government effectiveness (16%)) whereas the two non- or negatively correlated indicators (IND_6 and IND_7) receive zero weight. Interestingly, the correlation ratios in Fig. 8 reveal that the information in these two indicators is, albeit only slightly for IND_7, still represented by the CI through correlation. Most notably, IND_6 shows an increased information transfer compared to the *equal weights* and *Balance opt.* weighting scenario, even though it is receiving zero weight.

In line with its objective, most indicators show an increased information transfer to the CI when the *Maximize opt.* weighting scheme is applied. Only three indicators (IND_1, IND_7 and IND_{11}) show a decline in relation to the equal weighting scenario. When comparing the average correlation ratios, Fig. 8 shows that this weighting vector does achieve the highest total information transfer ($S_i^{mean} = 0.54$). However, the large error bars (even higher than for the equal weights case) suggest that it is unevenly distributed amongst the indicators, ranging from $S_i^{min} = 0.04$ to $S_i^{max} = 0.93$. It can thus be concluded that the pursuit of maximizing total information transfer comes at the expense of certain poorly correlated indicators (especially IND_7), which are barely represented by the CI.

4.3. Revising the CI based on the information transfer analysis

For both optimized (i.e., *Balance* and *Maximize opt.*) weighting schemes in RIfAO with 12 indicators, the poorly correlated indicators (especially IND_6 and IND_7) revealed to be problematic from a perspective of information transfer. When the *Balance opt.* weighting scheme is employed, these indicators receive a substantial share of the weights. The result is a balanced information transfer from the indicators to the CI, but with a low total information transfer. When the *Maximize opt.* weighting scheme is deployed, these indicators receive low or zero weights. This results in a high total information transfer, but with a large discrepancy between the individual indicators. A third scenario (**Scenario C**, step 5 in Fig. 3) has thus been developed, where the DM:

1. Wants to keep the S_i variability in a narrow range;
2. Is willing to revise the indicators included in the index;
3. Does not want to have an excessive (compared to equal weights and maximize weighting schemes) loss of S_i^{mean}.

This is mainly performed for exploratory reasons. The previous analysis shows that these indicators are not transferring much information to the index and their inclusion does not allow achieving a balanced information transfer from each indicator. Hence, we explore if we can achieve this by omitting them from the CI. A key drawback/consequence of omitting low correlated indicators is that these can contain a high information content of that indicator dimension. This information would then be “lost”. However, what we have shown is that this information is not really represented by the index in the first place, so removing them will have a low effect on the index scores and resulting rankings.

This problem framing leads to what is called RIfAO 10, an index with 10 indicators where IND_6 and IND_7 are removed from the CI (see above discussion) and the balance optimization is used (i.e., RIfAO with 10 indicators with *Balance opt.*). The resulting weights and information transfer measures are shown in Fig. 9 and Fig. 10, respectively.

Similarly to the case of RIfAO with 12 indicators, Fig. 9 shows that the *Balance opt.* still results in an unbalanced set of weights, even though IND_6 and IND_7 are removed. The same five highly correlated indicators (IND_2 - *Severe accident risk*, IND_3 - *Control of corruption*, IND_8 - *GDP* per capita, IND_{10} - *Government effectiveness* and IND_{12} - *Ease of doing business*) receive zero weight. However, the distribution of the remaining weights is not the same as for RIfAO with 12 indicators. In the absence of IND_6 and IND_7, IND_{11} now receive the most substantial share of the weights; followed by IND_5, IND_9, IND_4 and IND_1 (in decreasing order). Again, it is important to note that the information in the zero-weighted indicators would still be captured by the CI simply through correlation by the inclusion of IND_4 and IND_9. This is shown by the resulting correlation ratios in Fig. 10.

The key difference compared to the previous case of RIfAO 12, however, is the magnitude of information transfer achieved at *Balance opt.* weights. Contrary to the case of 12 indicators, it is now possible to achieve a rather well-balanced information transfer, ranging from $S_i^{min} = 0.41$ and $S_i^{max} = 0.52$ (see Fig. 10), without reducing the total information transfer to the same extent ($S_i^{mean} = 0.46$ compared to $S_i^{mean} = 0.19$ in the case of 12 indicators). For comparative purposes, Fig. 10 also includes the S_i^{mean} for the *Maximize opt.*, for RIfAO with 10

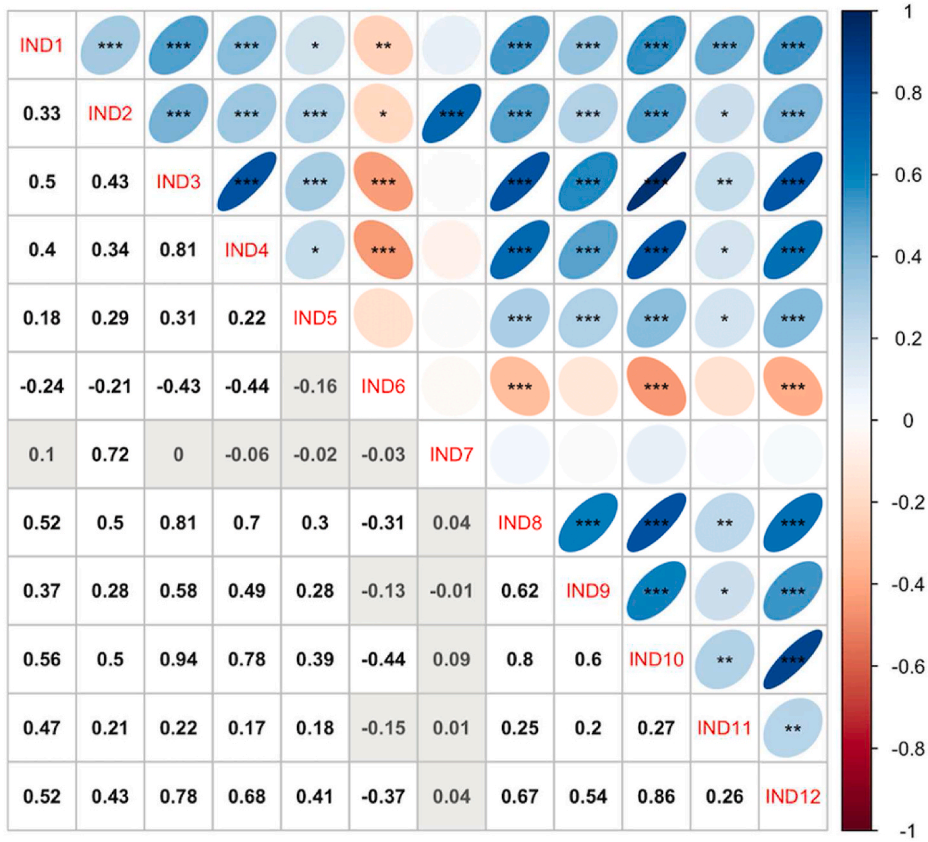

Fig. 4. Pearson correlation coefficients (ρ) (significance level = 0.05) between the 12 indicators of the RIfAO. Colours and ellipses represent strength and direction of the correlation. Numbers in grey background represents non-significant correlations. Asterisks represent significance levels, accordingly: * = 0.05, ** = 0.01, *** = 0.001.

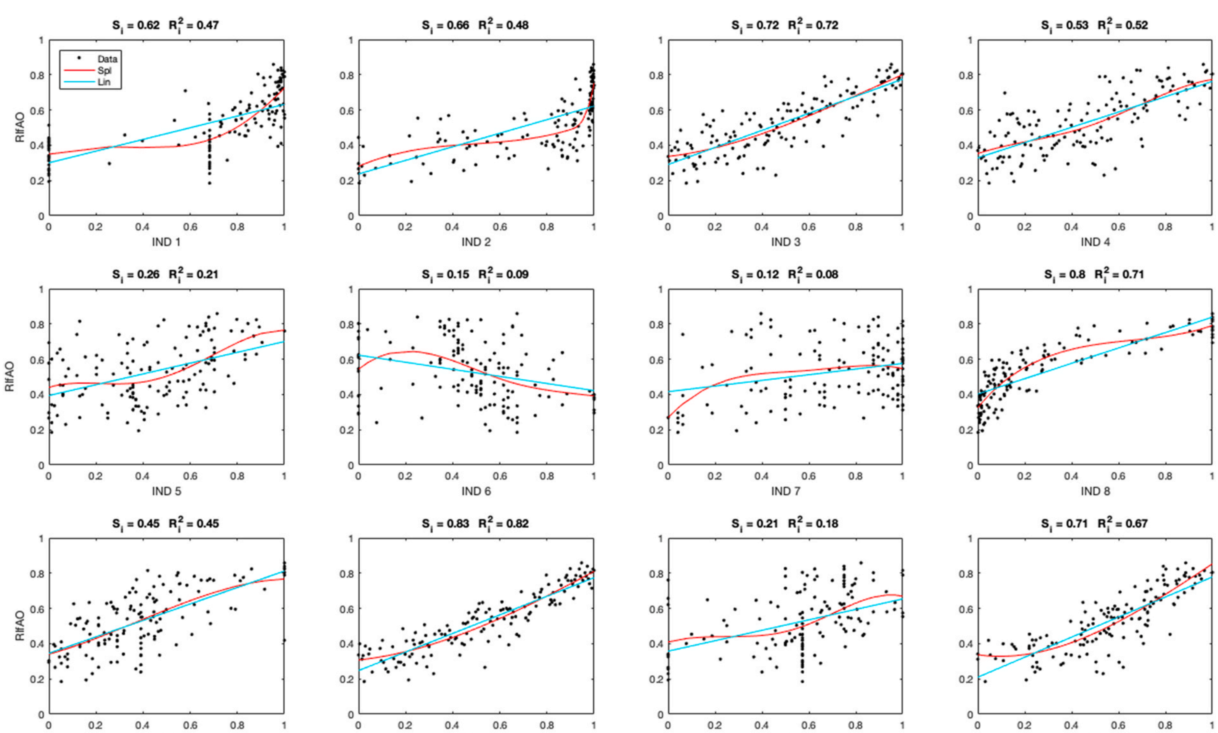

Fig. 5. Regression fits of RIfAO (y-axis), obtained with equal weights, against each indicator (x-axis), using two different regression approaches: linear (cyan) and splines (red). Above each plot is the estimated dependence, both linear (R_i^2) and nonlinear (S_i), between each of the 12 indicators and RIfAO.

Table 2

A comparison of the normalized correlation ratios $\tilde{S}_i$, obtained by nonlinear regression, and the assigned weights w_i (in this case equal). The deviation refers to the difference between the two and for the description of the groups, see the text.

Indicator	$\tilde{S}_i$	w_i	Deviation	Deviation ratio (%)	Group
IND 1 – SAIDI[a]	0.103	0.083	−0.020	−24%	3
IND 2 – Severe accident risks	0.108	0.083	−0.025	−30%	3
IND 3 – Control of corruption	0.120	0.083	−0.037	−44%	1
IND 4 – Political stability	0.088	0.083	−0.005	−6%	3
IND 5 – Electricity mix diversity	0.043	0.083	−0.040	−48%	2
IND 6 – Electricity import dependence	0.024	0.083	−0.059	−71%	2
IND 7 – Equivalent availability factor	0.019	0.083	−0.064	−77%	2
IND 8 – GDP per capita	0.132	0.083	−0.049	−59%	1
IND 9 – Insurance penetration	0.074	0.083	−0.009	−11%	3
IND 10 – Government effectiveness	0.137	0.083	−0.054	−64%	1
IND 11 – Average outage time	0.035	0.083	−0.048	−58%	2
IND 12 – Ease of doing business	0.117	0.083	−0.034	−40%	1

[a] System Average Interruption Duration Index.

indicators. It can be seen that the discrepancy between the two S_i^{mean} is considerably reduced with respect to the case of the CI based on 12 indicators. Most importantly, the wide variability in the S_i^{mean} shows that there is still a considerable imbalance of information transfer from each indicator in this RIfAO with *Maximize opt,* though the mean value is higher than in RIfAO 12, and the lower bound increases from about 0.1 to 0.2, whereas the upper bound remains at about 0.9. Maintaining the S_i variability in a narrow range was a binding condition to be met for Scenario C, and for this reason, only the *Balance opt.* is considered as a viable option, in the case of RIfAO with 10 indicators.

5. Discussion

Information transfer and correlations are intricately related in the construction of CIs. In this paper, it was confirmed that correlations lead the indicators to transfer information differently and hence have a different influence/impact on the CI as compared to the assigned weight. In order to deal with this discrepancy between desired influence of indicators (i.e., weights) and their actual influence driven by correlations, we provide tools that allow a deep-dive into this complex interrelationship and study the information transfer in relation to both weights and correlations. The main contributions of this research consist in:

1. Proposing a measure of information transfer based on correlations between the indicators along with two weight-optimization methods. The analyst can now adjust the weights to achieve either a targeted or maximized information transfer from a set of indicators.
2. Showing that while targeting indicator contributions is important, it is also relevant to consider the overall information conveyed by the index, thereby introducing the second optimization objective (maximizing information transfer).
3. Showing how the number of indicators, and the average correlation, can inform the analyst about the overall information transfer. More specifically, we demonstrate the convergence of information transfer towards the average correlation coefficient. The resulting analysis indicates that the strength of the correlations between the indicators is directly linked with their capacity to transfer information to the CI. In fact, correlations can be a good rule of thumb of how information transfer from a set of indicators will behave in the aggregation of a CI.
4. Applying these tools to a case study on electricity supply resilience assessment.

Regarding the case study, we apply the proposed tools to one version

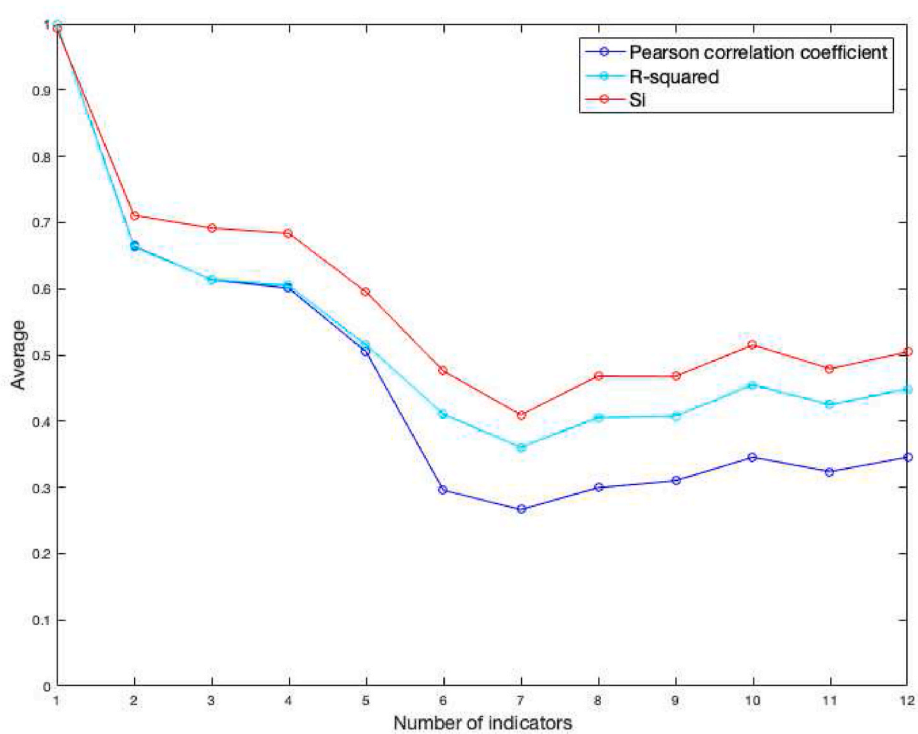

Fig. 6. Average S_i, R_i^2 and Pearson correlation (ρ) with respect to ordinal addition of indicators.

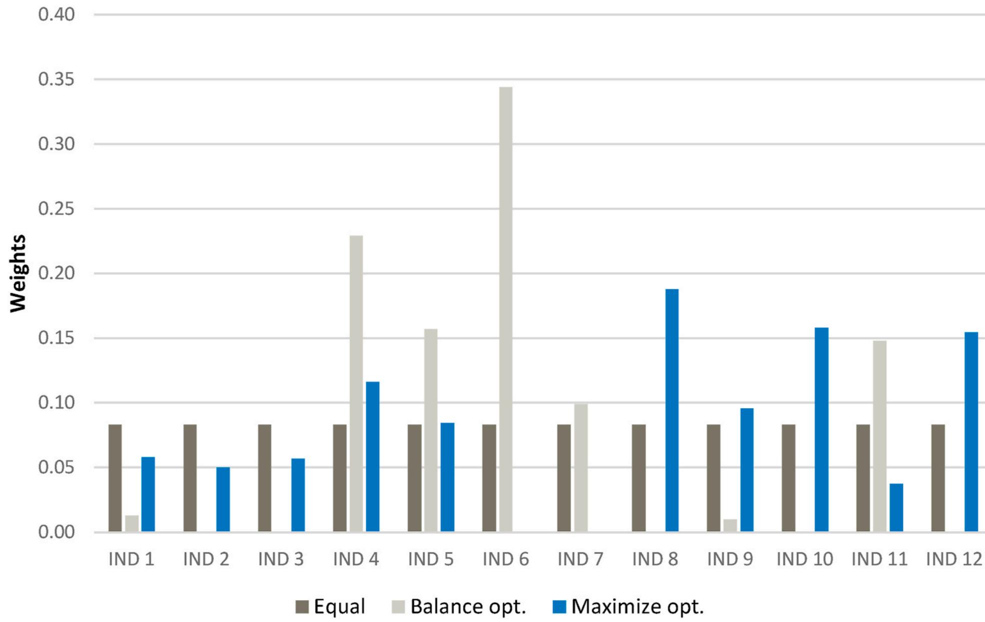

Fig. 7. The weights obtained from the two optimization problems, Balance (grey) and Maximize (blue), compared to the vector of equal weights (dark grey).

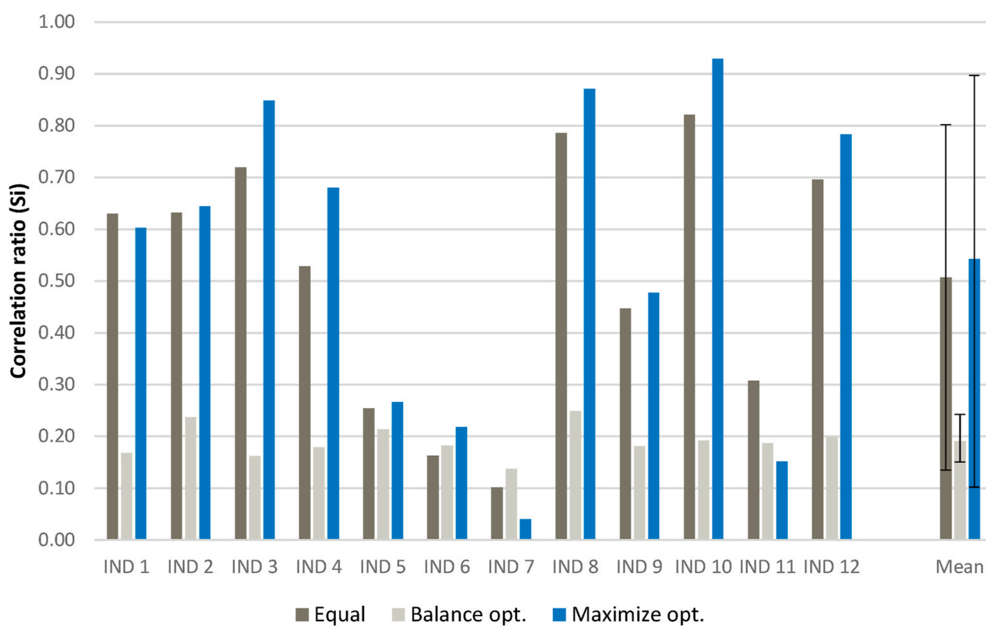

Fig. 8. The resulting correlation ratios (S_i), obtained at each weighting scenario: Equal (dark grey), Balance (light grey) and Maximize (blue). To the right, the mean values for each weighting scenario are presented along with error bars, indicating the 5th and 95th percentiles.

of the Electricity Supply Resilience Index (ESRI) developed at the Singapore-ETH Centre, which was called Resilience Index for Analysis and Optimization (RIfAO). The resulting analysis shows that correlations between RIfAO's underlying indicators have a direct influence on the index, preventing the equal weights assigned to correspond to an equal information transfer from each indicator. Different weighting schemes and index revision scenarios are also proposed according to specific requests that the DM might have with respect to possible loss and balance of information transfer, as well as indicators' inclusion in the index. When the weighting scheme used to distribute influence equally between indicators (i.e., *Balance opt.*) is employed, highly correlated indicators are poorly weighted, and less correlated indicators receive a substantial share of the weights. The outcome is a balanced, but low information transfer from the indicators to the CI. When the weighting scheme proposed to maximize the information transfer from the indicators (i.e., *Maximize opt.*) is applied, it is instead the less correlated indicators that are poorly weighted in favour of the more highly correlated indicators. The result is a high total information transfer, but with a large discrepancy between the individual indicators. However, when the two poorly correlated indicators are removed from RIfAO, the results indicate a less evident trade-off between the two weighting schemes, with comparable average information transfer though well-balanced with the *Balance opt.* scenario compared to the *Maximize opt.* scenario. Thus, if there is a large inconsistency (variation) in correlation strength between the indicators, it is probable that there will be an unbalanced information transfer from each indicator even though equal weights are applied. This phenomenon is not possible to counterbalance by adjusting the weights without compromising the

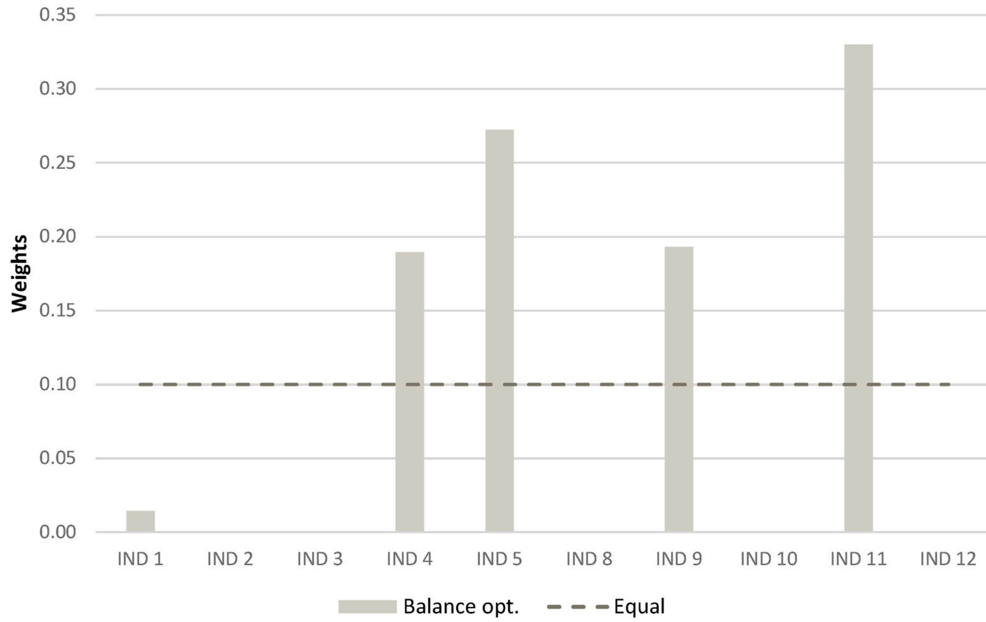

Fig. 9. The weights obtained from the Balance opt. compared to equal weights (dotted line), for RIfAO with 10 indicators.

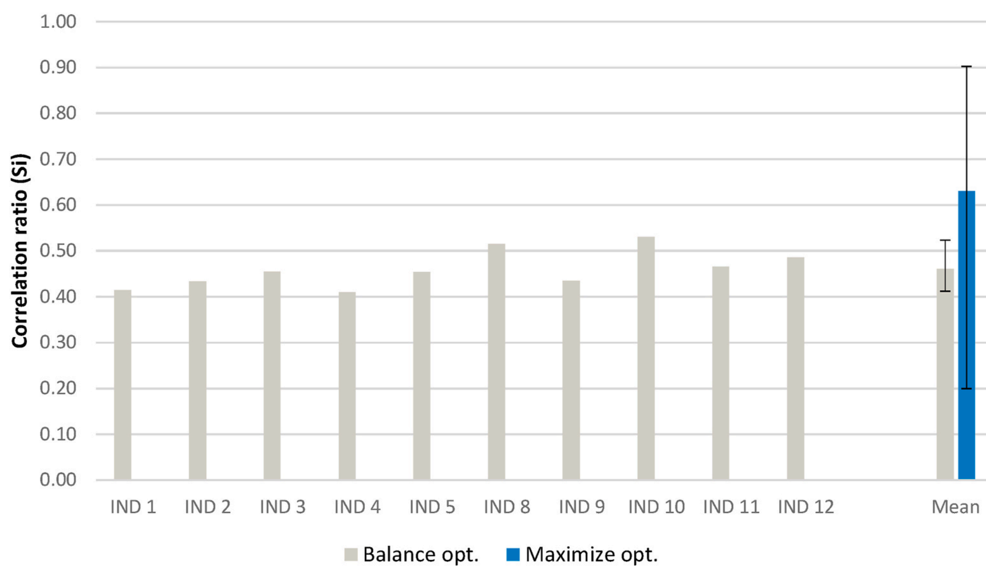

Fig. 10. (Grey) The resulting correlation ratios (S_i) obtained at Balance opt. weights, for RIfAO aggregated with 10 indicators. To the right, the mean value presented along with error bars, indicating the 5th and 95th percentiles. (Blue) The mean S_i for the Maximize opt., for RIfAO with 10 indicators.

information transferred to the CI, and its overall capacity to convey a representation of its underlying components, the indicators.

Our research also contributes to an ongoing debate on the inclusion of positively and/or negatively correlated indicators in CIs. On the one hand, there are authors like Marttunen et al. (2019) who advocate for the inclusion of not or negatively correlated indicators as they can be more informative for a decision since they bring unique perspectives on the aspects under evaluation. On the other hand, there are other authors like Munda et al. (2020) who warn about the risk of including indicators with low or negative correlations as their information might not be represented in the CI. Our research advocates for a balanced reasoning between these perspectives as follows.

When correlation exists between indicators, it means that information is shared between the two indicators. To take extreme cases, if (nonlinear) correlation is zero, that means that there is no shared information, and the two indicators are bringing completely unique information contributions. If correlation is one, the indicators are collinear and encode effectively the same information. Clearly, the second case is not useful because it implies double counting.[7] However, the first case comes with some pros and cons. On the one hand, as pointed out by Marttunen et al. (2019), zero correlation between indicators means that there is no overlap, and that can be seen as a good. But this comes at a

[7] This reasoning applies to a decision-making problem with a flat structure for the indicators. It would nonetheless be possible to keep the same indicator in two different dimensions if there would a hierarchy of indicators where the same indicator is present in more than one dimension. In this case, it would be possible keep the same indicator twice and use for example value functions to transform/normalize the data, so that e.g., value X of indicator A in dimension 1 means a 0.2, while the same value X of indicator 1 in dimension 2 means a 0.4, assuming the transformation is between 0 and 1 with an increasing order of preference.

cost, as we show in Fig. 2, since if one combines several indicators with zero correlation this will result in a CI that contains relatively little information from any one of the indicators. Therefore, in our opinion, if a concept can be summarised by some very few indicators with low correlations, this can still be acceptable as it is still possible to have a moderate information transfer. However, as Fig. 2 shows, above 10 indicators with an average correlation coefficient of zero, $R^2$is less than 0.1 between indicators and index, which contrasts with the fundamental objective of CI development itself, being the condensation of information of many indicators into one. Consequently, we recommend that when only looking at the correlations, if they are low, only a few indicators should be aggregated together, but if they are high, more indicators can be aggregated. However, the whole development of the CI should in the ideal case be embedded in a stakeholder consultation process, i.e., decisions on indicators will not just be driven by correlations but influenced by the priorities of the stakeholders. Additionally, potential interactions between the indicators might also be included in the development of the CI, which are not necessarily equal to correlations.

The authors also think that it is relevant to separate two different concepts: "Information transfer" and "Information content". It is true that a low correlated indicator can imply a high information content of that indicator dimension. However, what we show is that because of its low correlation with the other indicators, it will not transfer much of that information to the index, i.e., the index will not contain much of the information of that indicator dimension. Hence, a low correlated indicator will have a *low information transfer* to the index but can still, by itself, have a *high information content* of that specific indicator dimension.

This research also comes with a number of limitations that are presented below, together with options for future research to tackle them. This study has not considered the effects of changing aggregation methods and input data, which can be considered as one of the inherent uncertainties in composite indicators. In order to understand the effects of changing input data and aggregation method, one would have to perform an uncertainty analysis, e.g., a Monte Carlo sampling, along the lines of Saisana et al. (2005). What we propose in this research is not to investigate the uncertainties in weights, but more to calibrate them to a desired objective (i.e., target or maximize information transfer). Any uncertainty analysis is thus an avenue for future research. The same reasoning applies to the assessment of the effect that each source of uncertainty can have on the index variance. A possible option is this respect would be fuzzy MCDA methods (Kaya et al., 2019).

The application of the CIAO tool to the case study is based on the fully compensatory additive weighted sum, which means that its results are meaningful only for this type of aggregation function. However, the CIAO tool can be used with aggregation functions that have lower compensation levels than the additive weighted sum, such as the geometric and harmonic ones. Like the additive weighted sum, also the geometric and harmonic weighted sums are already included in the CIAO tool, and they can surely be a very interesting opportunity for future testing of our tool. There are however aggregation functions which would not be suitable for the CIAO tool, like extreme "aggregation" operators, such as the minimum and maximum operators. The reason is that since only one indicator would determine the final score (the worst with the minimum and the best with the maximum operators), there would be no optimization of weights to be performed as only one indicator would be defining the overall performance.

This research has not accounted for a DM who is willing to accept a compromise between the two objectives proposed for the weight optimization. This is because the goal of our research is to offer the users the CIAO tool to exactly achieve the desired target behind each optimization objective. In case the DM would like a compromise between these two objectives, the option of applying a multi-objective model could be explored.

Finally, the $S_i$measure proposed in this research has been developed for a decision-making challenge with a flat structure of the indicators, meaning that there is only one level between the constructed concept and the variables used to measure it with the CI. It can however be noted that, for a hierarchical index with multiple pillars and based on an additive weighted average, it would be also possible to calculate the effective weight of each indicator in the index by multiplying the indicator weight by its pillar weight, or by optimizing one level at a time.

The statistical analysis presented in this paper surely adds a layer of complexity for the well-informed development of composite indicators. The weighting of the indicators in fact results as a combination of data-driven (i.e., statistical) and stakeholder-based (i.e., value choices of the DM) input, which might be difficult to communicate, especially if the index is developed for advocacy purposes. Nonetheless, these types of advanced statistical analyses can be used to assess and enhance the robustness of the models that are developed, ultimately leading to more sound decision-making. This is in line with the recent call for such type of research as presented, for example, by Moallemi et al. (2020) and Saltelli et al. (2019).

6. Conclusions

The tools introduced in this study allow developers of CIs to explore in detail the effect of weighting choices, in an easily interpretable framework based on the concept of information transfer. For the first time, this work has shown that trying to balance the contributions of indicators may often come at the expense of reducing the overall information transferred from each indicator to the index. Most likely, developers will wish to find a compromise point between balancing and maximizing information transfer, and the optimization algorithms here give the means to assign selected weights in the perspective of these two competing criteria. As demonstrated with the RIfAO case study, this can sometimes be achieved by re-structuring the index.

This research also relates to an existing discussion on the use of supervised (DM-driven) and unsupervised (machine-driven) methods for studying and defining the complexities and interdependencies of a certain decision problem. When the complexity is such that the required knowledge cannot be easily given or the decision maker is not knowledgeable enough, the unsupervised method can be useful in at least providing an initial mapping of the decision problem (Kojadinovic 2008). Consequently, unsupervised methods are not to be seen as competitors to the methods that employ active interaction with the decision makers to define these dependencies and the resulting weights. Rather, they should be viewed as aiding tools to navigate the difficulties embedded in shaping the understanding of complex systems evaluated by means of multiple criteria.

Furthermore, it is important to note that the users of the tools proposed in this research are envisioned to be analysts with a mathematical background in statistical analysis and development of CI. A key distinctive feature of this type of users is their desire of providing a bridge between two scientific communities, on the one side data analysis without stakeholders' involvement, and on the other side decision aiding based on inclusion of stakeholders' preferences. The users can in fact use the tools provided by this research to achieve the desired contributions of the underlying indicators in the CI.

The tools proposed here are intended to provide "goalposts", between which developers can pick a desired target, and are not meant to supersede the conceptual relevance of the indicators, communication issues, and methodological choices in other stages of the CI construction, which are other highly relevant factors. More specifically, the DM can define the conditions for the index development with respect to (i) the possible loss of mean information transfer, (ii) the possible variability range of the information transferred from each indicator to the index and (iii) the willingness to discuss the possible removal of one or more indicators from the index. Once these conditions are defined, the weighting scheme can be obtained with the proposed tools and their results discussed among the stakeholders to decide how to proceed in the development of the index.

Finally, the current findings should not be simply generalized and applied, but the wider applicability of the proposed tools requires further testing with different datasets, with a varying number of indicators and alternatives, and with further normalization and aggregation functions. The difference of this research with respect to other sensitivity analyses is that the proposed framework does not aim to study the variability of the results according to the choices involved in its construction, such as the selection of the indicators, the normalization methods or the aggregation algorithms (Saltelli et al., 2019; Douglas-Smith et al., 2020; Zhang et al., 2020). It instead focuses on the effect of the correlation structure on the influence that each indicator has in the CI. When foreseeing a link with the other uses of sensitivity analyses, the proposed framework could also be applied to different conceptualizations of the CI to study how the recommended weighting would change based on e.g., different normalization methods and/or aggregation functions.

7. Software

The calculations for the case study on electricity supply resilience were performed with the software Composite Indicator Analysis and Optimization (CIAO) (Lindén et al., 2021), which was specifically developed for this research, and it is now freely available at the link: https://bitbucket.org/ensadpsi/ciao-tool/src/master/.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was conducted at the Future Resilient Systems (FRS) at the Singapore-ETH Centre (SEC), which was established collaboratively between ETH Zürich and Singapore's National Research Foundation (FI 370074011) under its Campus for Research Excellence And Technological Enterprise (CREATE) program. Matteo Spada and Peter Burgherr also received support from the Swiss Competence Center for Energy Research (SCCER) Supply of Electricity (SoE). Marco Cinelli acknowledges that this project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 743553. The authors also thank Paolo Paruolo from the European Commission's Joint Research Centre, for helpful input on analytical correlation analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2021.105208.

References

Allen, B., Stacey, C.B., Bar-Yam, Y., 2017. Multiscale information theory and the marginal utility of information. Entropy 19.

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

Becker, W., Domingeuz-Torreiro, M., Neves, A.R., Tacao Moura, C.J., Saisana, M., 2018. Exploring ASEM Sustainable Connectivity – what Brings Asia and Europe Together? European Commission. ISBN 978-92-79-92901-4.

Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I., 2017. Weights and importance in composite indicators: closing the gap. Ecol. Indicat. 80, 12–22.

Biggeri, M., Mauro, V., 2018. Towards a more 'sustainable' human development index: integrating the environment and freedom. Ecol. Indicat. 91, 220–231.

Booysen, F., 2002. An overview and evaluation of composite indices of development. Soc. Indicat. Res. 59, 115–151.

Carrino, L., 2017. The role of normalisation in building composite indicators. Rationale and consequences of different strategies, applied to social inclusion. In: Maggino, F. (Ed.), Complexity in Society: from Indicators Construction to Their Synthesis. Springer International Publishing, Cham, pp. 251–289.

Cinelli, M., Coles, S.R., Kirwan, K., 2014. Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. Ecol. Indicat. 46, 138–148.

Cinelli, M., Kadziński, M., Gonzalez, M., Słowiński, R., 2020. How to support the application of multiple criteria decision analysis? Let us start with a comprehensive taxonomy. Omega 96, 102261.

Cover, T.M., Thomas, J.A., 2005. Entropy, relative entropy, and mutual information. In: Cover, T.M., Thomas, J.A. (Eds.), Elements of Information Theory. Wiley, pp. 13–55.

Csiszár, I., Shields, P.C., 2004. Information theory and statistics: a tutorial. Found. Trends™ Commun. Inf. Theory 1, 417–528.

Da Veiga, S., 2015. Global sensitivity analysis with dependence measures. J. Stat. Comput. Simulat. 85, 1283–1305.

Da Veiga, S., Wahl, F., Gamboa, F., 2009. Local polynomial estimation for sensitivity analysis on models with correlated inputs. Technometrics 51, 452–463.

Davoudabadi, R., Mousavi, S.M., Sharifi, E., 2020. An integrated weighting and ranking model based on entropy, DEA and PCA considering two aggregation approaches for resilient supplier selection problem. Journal of Computational Science 40, 101074.

Dionisio, A., Menezes, R., Mendes, D.A., 2004. Mutual information: a measure of dependency for nonlinear time series. Phys. Stat. Mech. Appl. 344, 326–329.

Douglas-Smith, D., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: an overview of software tools and techniques. Environ. Model. Software 124. https://doi.org/10.1016/j.envsoft.2019.104588.

Eisenfuhr, F., Weber, M., Langer, T., 2010. Rational Decision Making. Springer Berlin.

El Gibari, S., Gómez, T., Ruiz, F., 2019. Building composite indicators using multicriteria methods: a review. J. Bus. Econ. 89, 1–24.

Eurostat, 2015. 2015 Monitoring Report of the EU Sustainable Development Strategy. European Union. ISSN 2443-8480.

Fernandes Torres, C.J., Peixoto de Lima, C.H., Suzart de Almeida Goodwin, B., Rebello de Aguiar Junior, T., Sousa Fontes, A., Veras Ribeiro, D., Saldanha Xavier da Silva, R., Dantas Pinto Medeiros, Y., 2019. A literature review to propose a systematic procedure to develop "nexus thinking" considering the water–energy–food nexus. Sustainability 11.

Freudenberg, M., 2003. Composite Indicators of Country Performance. OECD Publishing, OECD Science, Technology and Industry Working Papers, Paris.

Furr, M.R., 2011. Evaluating Psychometric Properties: Dimensionality and Reliability, pp. 25–51. Scale Construction and Psychometrics for Social and Personality Psychology. SAGE Publications Ltd.

Gasser, P., Suter, J., Cinelli, M., Spada, M., Burgherr, P., Hirschberg, S., Kadziński, M., Stojadinović, B., 2020. Comprehensive resilience assessment of electricity supply security for 140 countries. Ecol. Indicat. 110, 105731.

Greco, S., Ehrgott, M., Figueira, J., 2016. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer-Verlag, New York.

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G., 2019. On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. Soc. Indicat. Res. 141, 61–94.

Hirschberg, S., Burgherr, P., 2015. Sustainability Assessment for Energy Technologies. In: Handbook of Clean Energy Systems. John Wiley & Sons, Ltd, pp. 1–22.

Hwang, C.L., Yoon, K., 1981. Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, New York.

Johnson, R.A., Wichern, D.W., 2007. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, New Jersey.

Kaya, İ., Çolak, M., Terzi, F., 2019. A comprehensive review of fuzzy multi criteria decision making methodologies for energy policy making. Energy Strategy Reviews 24, 207–228.

Kojadinovic, I., 2004. Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functionals. Eur. J. Oper. Res. 155, 741–751.

Kojadinovic, I., 2008. Unsupervized aggregation of commensurate correlated attributes by means of the choquet integral and entropy functionals. Int. J. Intell. Syst. 23, 128–154.

Kullback, S., 1959. Information Theory and Statistics. Wiley, New York.

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 J. Optim. 9, 112–147.

Langhans, S.D., Reichert, P., Schuwirth, N., 2014. The method matters: a guide for indicator aggregation in ecological assessments. Ecol. Indicat. 45, 494–507.

Lindén, D., Cinelli, M., Spada, M., Becker, W., Burgherr, P., 2021. Composite Indicator Analysis and Optimization (CIAO) Tool, v.2. Singapore-ETH Centre. https://doi.org/10.13140/RG.2.2.14408.75520.

Mao, F., Zhao, X., Ma, P., Chi, S., Richards, K., Clark, J., Hannah, D.M., Krause, S., 2019. Developing composite indicators for ecological water quality assessment based on network interactions and expert judgment. Environ. Model. Software 115, 51–62.

Marttunen, M., Haag, F., Belton, V., Mustajoki, J., Lienert, J., 2019. Methods to inform the development of concise objectives hierarchies in multi-criteria decision analysis. Eur. J. Oper. Res. 277, 604–620.

Mazziotta, M., Pareto, A., 2013. Methods for constructing composite indices: one for all or all for one. Riv. Ital. Econ. Demogr. Stat. 67, 67–80.

Mazziotta, M., Pareto, A., 2017. Synthesis of indicators: the composite indicators approach. In: Maggino, F. (Ed.), Complexity in Society: from Indicators Construction to Their Synthesis. Springer International Publishing, Cham, pp. 159–191.

McKinnon, K.I., 1998. Convergence of the nelder–mead simplex method to a nonstationary point. SIAM J. Optim. 9, 148–158.

Meyers, L.S., Gamst, G.C., Guarino, A., 2013. Reliability analysis: internal consistency. In: 311-318 Performing Data Analysis Using IBM SPSS. John Wiley & Sons, pp. 311–318.
Moallemi, E.A., Zare, F., Reed, P.M., Elsawah, S., Ryan, M.J., Bryan, B.A., 2020. Structuring and evaluating decision support processes to enhance the robustness of complex human–natural systems. Environ. Model. Software 123, 104551.
Munda, G., 2008a. The issue of consistency: basic discrete multi-criteria "methods. In: Munda, G. (Ed.), Social Multi-Criteria Evaluation for a Sustainable Economy. Springer Berlin Heidelberg, Berlin, pp. 85–109.
Munda, G., 2008b. The issue of consistency: basic methodological concepts. In: Munda, G. (Ed.), Social Multi-Criteria Evaluation for a Sustainable Economy. Springer Berlin Heidelberg, Berlin, pp. 57–84.
Munda, G., Albrecht, D., Becker, W., Havari, E., Listorti, G., Ostlaender, N., Paruolo, P., Saisana, M., 2020. Chapter 18 - the use of quantitative methods in the policy cycle. In: Šucha, V., Sienkiewicz, M. (Eds.), Science for Policy Handbook. Elsevier, pp. 206–222.
Munda, G., Nardo, M., 2005. Constructing consistent composite indicators: the issue of weights. EUR 21834 EN.
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2008. Handbook on Constructing Composite Indicators. Methodology and User Guide. OECD, Paris.
OECD, 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD publishing, Paris.
Oţoiu, A., Grădinaru, G., 2018. Proposing a composite environmental index to account for the actual state and changes in environmental dimensions, as a critique to EPI. Ecol. Indicat. 93, 1209–1221.
Paruolo, P., Saisana, M., Saltelli, A., 2013. Ratings and rankings: voodoo or science? J. Roy. Stat. Soc. 176, 609–634.
Pearson, K., 1905. Mathematical Contributions to the Theory of Evolution, vol. XIV. Drapers' Company Research Memoirs (On the General Theory of Skew Correlation and Non-linear Regression. Dulau and Company, London).
Prokopenko, M., Boschetti, F., Ryan, A.J., 2009. An information-theoretic primer on complexity, self-organization, and emergence. Complexity 15, 11–28.
Reale, F., Cinelli, M., Sala, S., 2017. Towards a research agenda for the use of LCA in the impact assessment of policies. Int. J. Life Cycle Assess. 22, 1477–1481.
Reichert, P., Langhans, S.D., Lienert, J., Schuwirth, N., 2015. The conceptual foundation of environmental decision support. J. Environ. Manag. 154, 316–332.
Roostaie, S., Nawari, N., Kibert, C.J., 2019. Sustainability and Resilience: A Review of Definitions, Relationships, and Their Integration into a Combined Building Assessment Framework. Building and Environment.
Rowley, H.V., Peters, G.M., Lundie, S., Moore, S.J., 2012. Aggregating sustainability indicators: beyond the weighted sum. J. Environ. Manag. 111, 24–33.
Saisana, M., Philippas, D., 2012. Sustainable Society Index (SSI): Taking Societies' Pulse along Social, Environmental and Economic Issues. JRC.
Saisana, M., Saltelli, A., 2011. Rankings and ratings: instructions for use. Hague Journal on the Rule of Law 3, 247–268.
Saisana, M., Saltelli, A., Tarantola, S., 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. J. Roy. Stat. Soc. 168, 307–323.
Saisana, M., Tarantola, S., 2002. State-of-the-art report on current methodologies and practices for composite indicator development. Cités.
Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. Environ. Model. Software 114, 29–39.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: the Primer. John Wiley & Sons.
Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423.
Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An overview of sustainability assessment methodologies. Ecol. Indicat. 9, 189–212.
Smith, R., 2015. A mutual information approach to calculating nonlinearity. Stat 4, 291–303.
Song, L., Langfelder, P., Horvath, S., 2012. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinf. 13, 328.
UNDP, 2016. The United Nations Development Programme: Human Development Report 2016. Technical Report. UNDP, New York.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.
World Bank, 2020. Worldwide governance indicators. Accessed on 2 Febraury 2020 at: https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators.
World Economic Forum, 2017. The Global Competitiveness Report 2017–2018. Technical Report, WEF, Geneva.
Xu, C., Gertner, G.Z., 2008. Uncertainty and sensitivity analysis for models with correlated parameters. Reliab. Eng. Syst. Saf. 93, 1563–1573.
Zardari, N.H., Ahmed, K., Shirazi, S.M., Yusop, Z.B., 2015. Weighting Methods and Their Effects on Multi-Criteria Decision Making Model Outcomes in Water Resources Management. Springer International Publishing, Cham.
Zhang, Y., Spada, M., Cinelli, M., Kim, W., Burgherr, P., 2020. MCDA index tool. An interactive software to develop indices and rankings - user manual. Future resilient systems (FRS) team at Singapore-ETH Centre and laboratory for energy systems analysis (LEA) at Paul scherrer Institute, Switzerland. Cluster 2.1: assessing and measuring energy systems resilience. http://www.frs.ethz.ch/research/energy-and-comparative-system/energy-systems-resilience.html.