Expert Systems With Applications 200 (2022) 116927

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Unsupervised machine learning approach for building composite indicators with fuzzy metrics

E. Jiménez-Fernández [a], A. Sánchez [b,*], E.A. Sánchez Pérez [c]

[a] Departament d' Economía, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain
[b] Departamento de Economía Aplicada, Universidad de Granada, Campus de Cartuja s/n, 18071 Granada, Spain
[c] Instituto Universitario de Matemática Pura Aplicada, Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, Spain

ARTICLE INFO

Keywords:
Machine learning
Fuzzy metric
Composite indicator
Benchmarking
Robustness and sensitivity analysis

ABSTRACT

This study aims at developing a new methodological approach for building composite indicators, focusing on the weight schemes through an unsupervised machine learning technique. The composite indicator proposed is based on fuzzy metrics to capture multidimensional concepts that do not have boundaries, such as competitiveness, development, corruption or vulnerability. This methodology is designed for formative measurement models using a set of indicators measured on different scales (quantitative, ordinal and binary) and it is partially compensatory. Under a benchmarking approach, the single indicators are synthesized. The optimization method applied manages to remove the overlapping information provided for the single indicators, so that the composite indicator provides a more realistic and faithful approximation to the concept which would be studied. It has been quantitatively and qualitatively validated with a set of randomized databases covering extreme and usual cases.

1. Introduction

Many socio-economic phenomena should be represented with a multiplicity of dimensions, each of them measured by single indicators with different singularities. People understand this kind of concepts more easily in the form of a sole measure or composite indicator that allows its interpretation, communication and comparison (Greco et al., 2019; Mazziotta & Pareto, 2017; Saltelli, 2007).

In this regard, two interconnected aspects have attracted attention in the composite indicator literature: the allocation of weights among the single indicators in the composite indicator and the method to aggregate the single indicators into a solely value. These two aspects are not minor since they will determine the values of the composite indicator and therefore the ranking of the units (countries, regions, companies, etc.) that are being analysed (Becker et al., 2017; Jiménez-Fernández & Ruiz-Martos, 2020; Keogh et al., 2021).

In order to establish the weights of the single indicators there are three possibilities: participatory weights based on public or expert opinion, equal weights, and statistical weighting techniques (Greco et al., 2019; Maggino, 2017; OECD, 2008). Participatory and equal weights are typically linked to arithmetic mean and geometric mean as form of aggregation, whereas statistically determined weights are mainly linked to a multivariate technique, being the principal component analysis (PCA) extensively used.

Of all of them, the no-weights or equal weights is the most widely used scheme for the aggregation of single indicators into a composite indicator. The arithmetic mean was used to calculate the Human Development Index of United Nations from 1990 until the 2010 edition, until it was substituted by the geometric mean (UNDP, 2018). Other initiatives of international institutions also use the arithmetic mean to aggregate indicators into a composite index, for example: the Better Life Index developed by the OECD (OECD, 2017), Canadian Index of Wellbeing (2016) developed by the Canadian Research Advisory Group (Canadian Index of Wellbeing, 2016), and the Sustainable Development Goals (SDG) Index (Sachs et al., 2018). Equal weights are favoured when simplicity is called upon; there is no consensus on the distribution of weights or/and lack, of theoretical basis or/and insufficient statistical knowledge (Jiménez-Fernández & Ruiz-Martos, 2020; Keogh et al., 2021). Their main inconvenience is that this approach is unlikely to be empirically or theoretically correct, since all variables are unlikely to have the same importance (Becker et al., 2017); that is to say, it does not permit to differentiate between crucial and non-crucial indicators (Greco et al., 2019).

Within the weighting schemes determined by the use of statistical techniques, PCA is a multivariate method to reduce the dimensionality of the set of data by explaining a high percentage of the variance or information of the data (Jolliffe, 2002). Usually, the first component is

* Corresponding author.
E-mail addresses: jimeneze@uji.es (E. Jiménez-Fernández), sancheza@ugr.es (A. Sánchez), easancpe@mat.upv.es (E.A. Sánchez Pérez).

https://doi.org/10.1016/j.eswa.2022.116927
Received 12 April 2020; Received in revised form 4 June 2021; Accepted 15 March 2022
Available online 4 April 2022

taken as the composite indicator. In this vein we find several applications in well-being studies (Cruz-Martínez, 2014; Madonia et al., 2013), the measurement of environment quality (Montero et al., 2010) and the socio-economic regional performance in the EU (Sánchez & Ruiz-Martos, 2013), among others. We can also find a combination of several methods. PCA is applied to verify whether the set of indicators within each dimension is internally consistent, which in a second step are aggregated by arithmetic mean, following a weighting-scheme decided by experts. This is the approach that follows the World Economic Forum to elaborate the Global Competitiveness Index from 2008 (Schwab & Porter, 2008) and the European Commission to elaborate the European Regional Competitiveness Index from 2009 until the edition of 2019 (Annoni & Dijkstr, 2019).

These aggregation methods are very popular, largely due to their simplicity. However, they have serious disadvantages in their weighting-schemes that should raise doubts about the reliability of their results (Becker et al., 2017). Neither arithmetic mean nor geometric mean avoid the redundant or overlapping information of single indicators, and PCA only removes the linear redundant information (Jiménez-Fernández & Ruiz-Martos, 2020; OECD, 2008). Additionally, and more importantly, none of these methods enables benchmarking because they do not provide a mathematical structure for the analysis of the results through a metric which allows comparisons between units. Why do we understand that the use of a metric is necessary in the context of the construction of composite indicators? This is due to the fact that it is the natural way to establish the proximity between the analysed observations and, therefore, to perform benchmarking. Without this construction, an ordering of the units studied is obtained that lacks structure and, thus, the notion of distance.

The aim of this paper is to provide a more advanced analytical framework which allows the generation of new composite indicators with mathematical and computational techniques. More specifically, (1) we use fuzzy metrics and (2) unsupervised machine learning with innovative applications that allow us to perform benchmarking, as well as overcoming the drawbacks in the construction of composite indicators.

Our starting point is that a large number of phenomena studied are fuzzy concepts. For instance, competitiveness is a fuzzy concept since we are not able to establish the boundary of this notion. That is to say, when we study competitiveness in different territories, we are measuring degrees of competitiveness, or we provide a measure of proximity to the concept, since we cannot affirm that a unit (country or region) has or does not have competitiveness.

The methodology proposed in this study makes it possible to address in a more realistic way the complexity of synthesizing a set of variables in a single indicator. Existing methodologies permit this operation to be carried out, in most cases in an elementary way. However, it is difficult to simplify a complex problem with the use of a trivial formulation. The intrinsic complexity of the method presented is at the same time its virtue. The proposed composite indicator is the result of the conjunction of as many fuzzy metrics as single indicators are synthesized. These metrics can all be the same, in other words, if all the single indicators have the same statistical nature (continuous, categorical, etc.); or different, if we combine single indicators of a different nature. From the theory of fuzzy metric spaces, it is possible to construct a metric as a result of joining through an appropriate t-norm the metrics used in each single indicator. For example, these metrics are being used in different fields of science. Gregori et al. (2011) propose an image filtering as a result of grouping a family of computationally attractive filters with good detail preservation capacity (FSVF) through the combination of fuzzy metrics. López-Ortega and Castro-Espinoza (2019) propose fuzzy metrics that capture the uncertainty and doubts of the experts to quantify the consensus in multi criteria group decision making.

To achieve our goal, throughout the article we follow the steps that are collected below. Section 2 introduces some technical formal concepts regarding the fuzzy metric, which are necessary to get a better understanding of the purposed methodology. In Section 3, two sections are distinguished. Section 3.1 presents the composite indicator calculation formula and analyses how to choose the baseline or targets of the single indicators. Section 3.2 focuses on the estimation of the weights or the relevance of each single indicator in the composite indicator by using machine learning. Specifically, we study how to determine the best polynomial approximation between the composite indicator and the set of single indicators and how to estimate the importance of the single indicators in the calculation of the composite indicator. Section 4 analyses the algorithm for building the composite indicator, that is to say, the iterative method of calculation and the statistical properties required for the iterative process to stop. Section 5 focuses on mathematical properties of composite indicators in order to study its goodness of fit. Section 5.1 illustrates the mathematical properties that our composite indicator satisfies and Section 5.2 performances a comparison with the widest methods used. Section 6 purposes three new strategies for checking the robustness of our composite indicator, since the traditional techniques (uncertainty and sensitivity analysis) are inappropriate to assess the validity of our method. Finally, conclusions and remarks are drawn in Section 7.

2. Preliminaries

In order to foster a better understanding of our methodological purpose, let us introduce some technical formal concepts and definitions regarding the fuzzy metric using the standard mathematical notation.

Definition 1. A function $*: [0,1]\times[0,1] \rightarrow [0,1]$ is said to be a continuous *t-norm* if the following conditions are satisfied for all elements $a,b,c,d \in [0,1]$:

1. Commutativity $a*b = b*a$;
2. Associativity $a*(b*c) = (a*b)*c$;
3. Monotonicity $a*b \leq c*d$ if $a \leq c$ and $b \leq d$;
4. Identity element, $a*1 = a$;
5. $*$ is continuous.

Some classical examples used in fuzzy metrics are, firstly, the usual product that we will denoted by $\cdot$ ($a*b = a\cdot b$), secondly, the *minimum t-norm* denoted by $\cdot_{min}$ ($a *_{min} b = min\{a,b\}$), and thirdly *Lukasievicz t-norm* $a *_L b = max\{a+b-1, 0\}$ denoted by ℓ. These t-norms satisfy $a *_{min} b \geq a\cdot b \geq a *_L b$.

Definition 2. Let X be a non-empty set. According to George and Veeramani (1994), a Fuzzy Metric Space is a triple $(X, M, *)$ where $*$ is a continuous t-norm and $M : X\times X\times[0,\infty) \rightarrow [0,1]$ is a mapping which satisfies the following properties for every $x,y,z \in X$ and two parameters $t,s > 0$

1. $M(x,y,t) > 0$;
2. $M(x,y,t) = 1$ if and only if $x = y$.
3. $M(x,y,t) = M(y,x,t)$;
4. $M(x,y,t) * M(y,z,t) \leq M(x,z,t+s)$;
5. $M(x,y,\cdot) :]0,+\infty] \rightarrow]0,1]$ is continuous.

M is called *fuzzy metric*.

$M(x,y,t)$ may be interpreted as the degree of proximity between x and y with respect to the parameter t that we call sensitivity[1] of the fuzzy metric M. When a fuzzy metric M on X does not depend on t is said to be stationary, i.e. if for each $x,y \in X$ the function $M_{x,y}(t) = M(x,y,t)$ is constant. Such a metric is enough for defining a topology τ_M by means of the basis of neighbours that is given by

[1] Notice that the parameter t represents an abstract parameter that allows modulating the fuzzy metric.

open balls $B_M(x,\varepsilon,t) := \{y \in X, 0 < \varepsilon < 1, t > 0, M(x,y,t) > 1-\varepsilon\}$. This topology allows us to establish a relation of proximity and therefore to provide an order to the calculated indicator. Thus $M(x,y) \simeq 1$ means that x is close to y (that is, there is high similarity or proximity) in the sense providing by the metric M (George & Veeramani, 1994).

In this paper, we use the fuzzy metrics approach for the construction of the composite indicator, although any of the classical metrics could have been used for this purpose. When simple indicators are represented by a continuous variable, we can use the following metrics, which are additionally used as a reference to show methodology.

Example 3. Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be an increasing continuous function and d is a distance on an nonempty set X

$$M(x,y,t) = \frac{g(t)}{g(t) + \alpha \cdot d(x,y)} \tag{1}$$

where $\alpha > 0$, then M is a fuzzy metric with $\cdot$ t-norm on X. A particular case to the previous (for instance $\alpha = 1$, $g(t) = k$) is the following

$$M(x,y) = \frac{k}{k + d(x,y)} \tag{2}$$

where k is a positive suitable constant, which provides a stationary fuzzy metric over X.

Notice that when the function $g(t)$ trends to infinity or k is sufficiently large, remaining constant $d(x,y)$, then the partial indicator is close to one, namely, the fuzzy distance between two pairs of vectors can be forced to the phenomenon to be measured, increasing the value of the function $g(t)$ or the parameter k. This property provides an advantage with respect to the classical metrics due to the fact that we can model the sensitivity of the metric. In our model, these functions are used in order to increase the sensitivity of the metric to avoid the duplication of information of the indicators that constitute the composite indicator, as well as differentially allocate the importance to each indicator.

Other fuzzy metrics that can be used depending on the nature of the indicators to be used are shown below.

Example 4. We assume that $X = \mathbb{R}^+$ and $t \in \mathbb{R}^+$, then

$$M(x,y,t) = \frac{\min\{x,y\} + t}{\max\{x,y\} + t}$$

is a fuzzy metric with $\cdot$ t-norm on X. Also

$$M(x,y) = \frac{\min\{x,y\}}{\max\{x,y\}}$$

is a stationary fuzzy metric with $\cdot$ t-norm on X.

When we use simple ordinals or binary simple indicators to make the composite indicators, we can use other metrics that require a specific t-norm. We assume the minimum t-norm ($a \cdot_{min} b = min\{a,b\}$), the following examples represent fuzzy metrics with the minimum t-norm on X.

Example 5.

$$M(x,y,t) = \begin{cases} 1, & \text{if } x = y \\ f(t), & \text{if } x \neq y \end{cases}$$

where $f : \mathbb{R}^+ \to [0,1[$ is an increasing continuous function. If f is constant $f(t) = k \in [0,1[$ then the fuzzy metric is stationary, also called discrete fuzzy metric.

In addition, we can assume the Lukasievicz t-norm $a *_L b = max\{a + b - 1, 0\}$. Let $g : X \times X \to [0, \frac{1}{2}[$ be a symmetric function. Then

$$M(x,y) = \begin{cases} 1, & \text{if } x = y \\ g(x,y), & \text{if } x \neq y \end{cases}$$

is a stationary fuzzy metric on X.

3. Fuzzy-machine learning approach

The methodology presented in this study provides the main tools in order to construct a composite indicator into the context of a formative measurement model, such as regional development, poverty, quality of life, well-being, etcetera, where the causality is from the single indicators to the composite indicator (see Coltman et al., 2008; Diamantopoulos et al., 2008; Jiménez-Fernández & Ruiz-Martos, 2020). Another consideration in the selection of the proper aggregation technique is whether compensability or substitutability among indicators should be permitted. The desired degree of compensation between the indicators will depend on the underlying theoretical framework of the phenomenon under investigation (Maggino, 2017; Mazziotta & Pareto, 2017). Our methodological proposal is partially compensatory which requires the use of non-linear functions or multiplicative methods. In what follows we present the calculation formula on the composite indicator and its required steps.

3.1. Composite indicator formula

The starting point is a $n \times m$-dimension matrix X, where the columns represent the normalized single indicators X_j $j \in \{1,\dots,m\}$, and the rows of X are referred to the units studied (regions, countries, etc.). Let $(X_j, M_j, *)$ be fuzzy metrics spaces $j = \{1,\dots,m\}$ according to George and Veeramani (1994) – for instance, any metric presented in the previous section – such that X_j is a (non-empty) *single indicator*, over the same t-norm $*$. Henceforth, we are going to use the t-norm of the usual product as a reference. We denote by $x_i = (x_{i1},\dots,x_{im})$ the m-dimension vector that groups all single indicators corresponding to the i-observation $i \in \{1,\dots,n\}$. Supposing that $x_* = (x_{*1},\dots,x_{*m})$ be a fictitious vector reference (target or baseline) that is composed by the results of a theoretical observation with the best-worst possible scenario for all the single indicators depending on the polarity[2] .

The *composite fuzzy indicator* (CFI) can be defined as

$$CFI_i = CFI(x_*, x_i) = \prod_{j=1}^{m} M_j(x_{*j}, x_{ij}) \tag{3}$$

Thus $(X_1 \times \cdots \times X_m, I, \cdot)$ is also a fuzzy metric defined with the same t-norm (Segi Rahmat & Noorani, 2008), namely, all metrics included in this product must be defined on the same t-norm. This metric computes the fuzzy distance between each observation $x_i = (x_{i1},\dots,x_{im})$ and the target $x_* = (x_{*1},\dots,x_{*m})$. As a result, if the composite indicator is equal to 1 for the unit i, then the i observation reaches the reference vector. Conversely, the closer to zero the composite indicator is, the observation i is further away from the reference vector and therefore, further from the concept we want to model. This property is fundamental, since it allows making a composite indicator that also is a fuzzy metric, namely, it satisfies Definition 2. Therefore, for each observation i, the metric provides a degree of proximity to the concept that the researcher wishes to model. The choice of the vector reference or baseline is a crucial step which must be done in connection with the theoretical framework of the phenomenon to be studied. If we were interested in measuring, for example, the technological development of a set of countries, the best values of the single indicators – within the set of countries – should be taken as a vector reference. On the contrary, the worse values would be taken as a reference vector when you want to measure concepts such as poverty or socio-economic vulnerability in

[2] The indicator's *polarity* is defined as the sign of the relation between the single indicator and the phenomenon to be measured. Some indicators may be positively correlated with the latent variable (positive polarity), whereas others may be negatively correlated with it (negative polarity). For instance, the investment in R&D would be positively associated with the economic development (latent variable), whereas the CO_2 emissions would be negatively associated.

which the higher the score of the composite indicator, the worse the real situation.

More specifically, given the fuzzy metric shown by Eq. (2), the composite indicator can be written as follows

$$CFI(x_*, x_i) = \prod_{j=1}^{m} M_j(x_{*j}, x_{ij}) = \prod_{j=1}^{m} \frac{k_j}{k_j + |x_{*j} - x_{ij}|} \tag{4}$$

where k_j is the sensitivity constant associated to each fuzzy metric M_j.

Notice that the function $f(k) = k/(k + d)$ is an increasing function, due to the fact that d is positive. Therefore, we can use this parameter in order to modify each fuzzy metric according to its relevance. For example, taking $k_1 = \cdots = k_m$, then the composite indicator only considers a theoretical situation of equal importance of each single indicator or, equivalently, there is not overlapping or redundant information between them (influence). To check if there exists any influence, according to Section 3.2.2, we must check if there are statistical function relationships among the single indicators and then attribute them to their corresponding fuzzy metrics through the sensitivity constant k_j. The next step is to estimate the more suitable value of the sensitivity constant k_j for each single indicator X_j.

3.2. Computing the sensitivity of fuzzy metric

One of the crucial stages in this methodology focuses on the computation of the sensitivity constant k_j, namely, the importance of each single indicator in the CFI. Taking as a reference the initial CFI, for which all single indicators have the same relevance ($k_1 = \cdots = k_m$), the procedure must check if there exist statistical relationships between them. To carry out this task, the best approximation polynomial between the composite indicator and the set of single indicators is obtained through Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991). Unlike to other approaches to indicator construction such a Distance P2 (Pena Trapero, 1977; Sánchez et al., 2018; Sánchez & Ruiz-Martos, 2018) which use ordinary linear regression as a hinge to quantify the weights of the relationships between the CFI and the single indicators, MARS is a non-parametric modelling method which extends the linear model by incorporating non-linearities and the indicators interactions without the assumptions that traditional regression models must meet. It is insensitive to not normally distributed predictors and response. Likewise, it is insensitive to irrelevant variables and unscaled variables. These tools are used to understand data through new improved software packages Kuhn (2008).

3.2.1. Selecting an efficient approach to the data set

In this section, we select the best functional relationship among singles indicators and CFI. For each observation $i \in \{1, \ldots, n\}$, MARS model can be written as follows:

$$CFI_i = \beta_0 + \sum_{j=1}^{m} \beta_i B(x_{ij}) + \epsilon_i \tag{5}$$

where CFI_i is the composite indicator, x_{ij} is the observation of the j-normalized single indicator $j \in \{1, \ldots, m\}$, β_0 is the intercept, $B(x_{ij})$ is a basis of disjoint functions and finally ϵ is the error term. The procedure is implemented by constructing on the previous suitable basis of disjoint functions (polynomials of degree q) tied by knots, where a final model is constituted as a combination of this generated base functions that can be fitted by ordinary least-squares (Friedman, 1991).

To carry out this approach, two crucial tuning hyperparameters are used in order to minimize mean-square error (MSE). Firstly, the degree of the basis functions $B(x_{ij})$ used to perform Eq. (5), and secondly, the number of knots used to link the disjoint polynomials. In order to identify the optimal combination of the previous hyperparameters, the procedure performs a grid of two dimensions search. MARS procedure will assess all the potential combinations between these two hyperparameters and will discard them until obtaining the optimal selection.

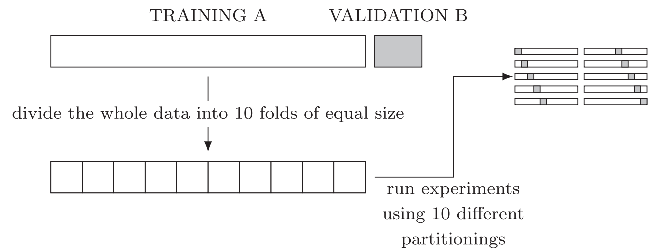

Fig. 1. 10-cross validation partition.

To perform this optimization process, we use k-fold cross-validation as a resampling procedure (Craven & Wahba, 1971). We split the data set into 10 groups of equal size. The first fold (B) is treated as a validation set, and the model (5) is fit on the remaining 9 folds (A) for each combination of two hyperparameters which are tested using B to adjust the coefficients values to best fit the data, as shown in a Fig. 1 . More precisely, let $\{(x_{i1}, \ldots, x_{im}, CFI_i)\}_{i=1}^{n}$ be n observations where each i-observation is composed by m+1 coordinates, the m first represented by the single indicators and the last by the CFI. We generate a partition of the whole data, the training subset $A = \{(x_{i1}, \ldots, x_{im}, CFI_i)\}_{i=1}^{\ell}$ and test subset $B = \{(x_{i1}, \ldots, x_{im}, CFI_i)\}_{i=\ell+1}^{n}$ where $\ell > n - \ell$. For each two hyperparameters belonging to the performed grid, model (5) is fit with respect to the A training subset. We assess $\{(x_{i1}, \ldots, x_{im})\}_{i=\ell+1}^{n}$ and then $\{\widehat{CFI}\}_{i=\ell+1}^{n}$ is obtained. Due to the fact that B was not used in the previous process, the Mean Square Error (MSE) can be estimated as

$$MSE = \frac{1}{n-\ell} \sum_{i=\ell+1}^{n} (CFI_i - \widehat{CFI}_i)^2 \tag{6}$$

We can repeat the procedure 10 times by selecting new partitions of data and obtaining 10 square errors $\{MSE_k\}_{k=1}^{10}$. The cross-validation error is defined as follows

$$CV_{10} = \frac{1}{10} \sum_{i=k}^{10} MSE_k \tag{7}$$

3.2.2. Single indicators importance scores (k_j). The sensitivity scores of the fuzzy metrics

Once we know which is the best polynomial basis that relates the single indicators with respect to the composite indicator, we determine the sensitivity scores (k_j). These parameters provide a rank of importance of the single indicators. In other words, we want to quantify the strength of the relationship between the single indicators and the composite indicator. One of the primary reasons to measure the strength or relevance of the single indicators is to select the suitable sensitivity k_j, which should be used as inputs in the fuzzy metric. To do this, we select sensitivity scores through a function variable importance, using Partial Dependence Plots (PDP) (Greenwell et al., 2018). In turn, PDP provides the metric with the suitable constants (k_j) in order to reduce the overlapping information from the m-single indicators. Sensitivity scores are computed as follows. Let $X = \{X_1, \ldots, X_m\}$ be the set of m indicators in the chosen model (5) whose prediction function computed using MARS technique is denoted by $\hat{f}$. Let $Z_s = \{X_1, \ldots, X_s\}$, $s < m$ be an indicator subset of X and let $Z_{p-s} = \{X_{s+1}, \ldots, X_p\}$ be the complementary subset into X. The partial dependence of the response on Z_s is defined by

$$f_s(Z_s) = E[\hat{f}(Z_s, Z_{p-s})]_{z_{p-s}} = \int \hat{f}(Z_s, Z_{p-s}) \mathbf{p}_{p-s}(Z_{p-s}) dZ_{p-s} \tag{8}$$

where the function $\mathbf{p}_{p-s}$ represents the marginal density of the subset Z_{p-s}. Let n be the number of observations in the training data $Z_{i,p-s} \in \{Z_{1,s+1}, \ldots, Z_{n,p}\}$ for each single indicator, then the model (8) can be estimated by

$$\overline{f}_s(Z_s) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(Z_s, Z_{i,p-s}) \tag{9}$$

Notice that this function depends on Z_s indicators and all observations are assessed in the complementary set of indicators. In fact, $\overline{f}_s(Z_s)$ is an average over the set $\{Z_{p-1}, \dots, Z_{p-s}\}$. Therefore, for each indicator – or subset of indicators – we can compute the average out of the effects of all the other indicators in the model (5). Without loss of generality, we select $Z_j = X_j$ for each $j \in \{1, \dots, m\}$, thus the score of the indicator X_j is defined by

$$i(X_j) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[\overline{f}_j(x_{ij}) - \frac{1}{n}\sum_{i=1}^{n}\overline{f}(x_{ij})\right]^2} \tag{10}$$

Zhou et al. (2010) provide a novel methodology to minimize the loss of information in the aggregation of single indicators to build the composite indicator. In our study, the weighs are defined as scores in the fuzzy metric with respect to each single indicator, namely, assigning its respective scores to the parameters k_j, $j \in \{1, \dots, m\}$ which generates a specific importance for each single indicator. The final composite indicator is constituted by full/partial information depending on the sensitivity computed in the previous step of all fuzzy metrics.[3]

In brief, the methodology presented in this study incorporates the non-linearities and the machine learning tool in order to obtain a composite indicator as close as possible to the nature of the data and their interactions with the analysed model. Notice that PDP could generate erroneous results if the single indicators are strongly correlated. In this case we will recommend using the Accumulated Local Effects (ALE) plots in order to minimize these interactions (Apley & Zhu, 2016).

4. Algorithm for building the composite indicator

Once the method for constructing the composite indicator has been studied, the next step focuses on its practical implementation.[4] In this section we analyse the iterative method of calculation, and the statistical properties required for the iterative process to stop.

Let Ω be the set of single indicators $\Omega = \{1, \dots, m\}$, where m denotes the number of single indicators selected and let $(X_j, M_j, *)$ be fuzzy metric spaces under the same t-norm $*$ and $j \in \Omega$. Let x_* be the m-dimension reference vector performed by the baseline of each indicator and let x_i be the m-dimension observation where each component $x_{ij} \in X_j$, and $i \in \{1, \dots, n\}$ being n be the number of observations considered. In this study, all single indicators X_j are performed by real numbers and normalized. For all $x_i \in X_j$ we select the fuzzy metric as we pointed out in a previous section

$$M_j(x_{*j}, x_{ij}) = \frac{k_j}{k_j + d(x_{*j}, x_{ij})} \tag{11}$$

where k_j, $j \in \Omega$ represents the sensitivity scores for each indicator according to PDP described in the previous section. Finally, d is the absolute value of the difference between the target x_{*j} and the i-observation x_{ij} of the j-observation.

4.1. Iterations

We start computing the composite indicator with identical sensitivity $k_1^{(0)} = \cdots = k_m^{(0)} = 0.5$, namely all indicators have the same importance, for each indicator which is defined by the composite fuzzy indicator $CFI^{(0)}$. Assuming $CFI^{(0)}$ as an output variable, MARS and CV are used to select in Eq. (5) the optimal disjoint basis functions to approximate the function relation between the $CFI^{(0)}$ and the single indicators. As presented in a previous section, we determine the sensitivity score of single indicators through PDP. Those single indicators whose sensitivity is close to zero contract its metric by subtracting relevance, those whose sensitivity is maximum will approximate the metric to one endowing with more importance. To impute these relevance to the new composite indicator, we assign the sensitivity scores to each fuzzy metric Eq. (11) through $k_j^{(1)}$ for each $j \in \{1, \dots, m\}$, obtaining a new composite indicator $CFI^{(1)}$. In turn, this new composite indicator generates new sensitivity scores for each of the single indicators. This iterative process generates a sequence of sensitivity scores $(k_j^{(\ell)})_{(\ell)}$ and composite indicators $(CFI^{(\ell)})_{(\ell)}$ for all $j \in \{1, \dots, m\}$ and $\ell = \{0, 1, 2 \dots\}$.

$$CFI(x_*, x_i)^{(\ell)} = \prod_{k=1}^{m} \frac{k_j^{(\ell)}}{k_j^{(\ell)} + d(x_{*j}, x_{ij})} \tag{12}$$

4.2. When will the iterative process stop?

The iterative process will stop when a similar rank correlation between two composite indicators $CFI(x_*, x_i)^{(\ell)}$ and $CFI(x_*, x_i)^{(\ell-1)}$ is met. In order to measure this rank correlation, the three most commonly measures used are Pearson's r, Spearman's ρ and Kendall's τ coefficients for monotone association in continuous data. The normality of the variables analysed is the main assumption for the first one. However, this assumption is not necessary for the other tests. The second one takes account of the Pearson's coefficient used for continuous non-normal data. Nonetheless, Kendall's τ provides the difference between the probability that the composite indicators $CFI(x_*, x_i)^{(\ell)}$, $CFI(x_*, x_i)^{(\ell-1)}$ are in the same order versus the probability that the composite indicators $CFI(x_*, x_i)^{(\ell)}, I(x_*, x_i)^{(\ell-1)}$ are not in the same order (Hauze & Kossowski, 2011).

Another rank correlation measure designed for ordinal data is the γ rank correlation measure (Goodman & Kruskal, 1954). In all these tests, the null hypothesis H_0 assumes that the two composite indicators are independent, whereas the alternative hypothesis H_1 assumes that they are dependent. In order to compute the p-value and therefore computing the test statistic on the given data, the knowledge of the statistic's distribution associated to the null hypothesis is required. However, this information is not known for the classical gamma and Kendall's test. The robust γ rank correlation measures by using some fuzzy ordering with smooth transitions is a new approach that does not require these assumptions and generalizes the classical γ rank correlation (Bodenhofer & Klawonn, 2008; Bodenhofer et al., 2013). The robust γ rank correlation coefficient is defined as:

$$\gamma = \frac{C - D}{C + D} \tag{13}$$

where C represents the number of concordant pairs and D the number of discordant pairs. For instance, if we consider an extremal case where all observations of two data set are equal ($D = 0$), we will obtain a gamma rank correlation equal to one.

One advantage of this test is that it solves the difficulty of unknown distribution through permutation testing, which assumes that the two concatenated composite indicators are independent, namely, any combination of two observations is likely to be equal. In this setting, fixed ℓ, if we compute the test statistic for $CFI(x_*, x_i)^{(\ell)}$ and all possible permutations of $CFI(x_*, x_i)^{(\ell-1)}$ varying in the set $\{1, \dots, n,\}$, the p-value corresponding to the two-sided tails means the relative frequency of how many times the absolute value of the test statistics for the

[3] Notice that one can impute scores to the fuzzy metric according to the importance that the researcher decides to assign to each single indicator. However, the subjectivity of this choice may, in some cases, bias the information of the composite indicator towards the indicators with the most weight. Partial dependence plots assign the scores from the data information avoiding this subjectivity, although it is worth noting that from the observed data (samples) a partial information can be derived that is far from what we are really looking for. Therefore, these two approaches must be taken into account before using the methodology presented in this paper.

[4] The algorithm was developed with R software. Caret package has been used to estimate the scores of the single indicators (Kuhn, 2008).

permuted observations is at least as large as the absolute value of the statistic test for the no permuted observations.

We use the p-value and the robust gamma rank correlation coefficient as control parameters in our algorithm. Fixed a level of significance α and a threshold γ, the iterative method is stopped when two properties are satisfied. The first one is reached when the p-value is lower than the fixed level of significance, therefore we will have empirical evidence that the two composite indicators are not independent. The second one has to fulfil that the calculated robust gamma statistic is higher than the figurative threshold. In this case, we obtain that the compared composite indicators are significantly similar and then the iterative process stops. The pseudo-code of the purposed algorithm can be described as Algorithm 1.

Algorithm 1: Algorithm of the Composite Fuzzy Indicator $CFI(x_*, x_i)$ with respect to a reference vector.

Data: (Inputs) $error = \alpha$,
$\ell = 1$, $iterations = p$, $p - value_1 = 1$, ρ;
Result: (Output) Composite indicator $CFI(x_*, x_i)^{(\ell)}$
Initialization: Compute $CFI(x_*, x_i)^{(0)}$ and the sensitivity scores $\{k_1^{(0)}, ..., k_m^{(0)}\}$;
repeat
  Compute $CFI(x_*, x_i)^{(\ell)}$ using the sensitivity scores $\{k_1^{(\ell-1)}, ..., k_m^{(\ell-1)}\}$
  Compute the sensitivity scores $\{k_1^{(\ell)}, ..., k_m^{(\ell)}\}$ of the composite indicator $CFI(x_*, x_i)^{(\ell)}$
  Apply gamma robust correlation to $CFI(x_*, x_i)^{(\ell)}$ and $CFI(x_*, x_i)^{(\ell-1)}$
  Compute $p - value_\ell$
  Compute $\hat{\rho}_{(\ell)}$
  $\ell = \ell + 1$;
until $\ell \leq iterations$ *or* $(\hat{\rho}_\ell \leq \rho$ *and* $p - value_\ell < error)$;

5. Properties of the aggregation methods

As far as we know, the analysis of mathematical properties that a composite indicator must fulfil in order to allow us to study its goodness of fit is scarcely treated in the literature. In this section we present the mathematical properties of our methodological proposal according to what was pointed out by Pena Trapero (1977, 2009), Zarzosa Espina (1996). In addition, we review whether the traditional methods to build composite indicator fulfil these properties.

5.1. Mathematical properties of the composite Fuzzy indicator

The main properties that our proposal CFI satisfies are: existence, transitivity, invariance, exhaustiveness, monotony and symmetry.

- 1th. Existence and determination. The aggregation method (Eq. (3)) is well-defined for all $i \in \{i, \dots, n\}$ due to Definition 2.
- 2th. Transitivity. Due to the range of the composite indicator belongs into $]0, 1]$, the transitivity property (if $I_i > I_j$ and $I_j > I_k$ implies $I_i > I_k$) is satisfied.
- 3th

Proposition 6. *The aggregation method is invariant by origin and scale changes.*

Proof. Let X_j be a single indicator such that $\|X_j\|_2 = 1$. For each $j \in \{1, \dots, m\}$, we define $Z_j = (\alpha X_j + B_j)/\|\alpha X_j + B_j\|_2$ where α is a constant and $B_j = (\beta_j, \dots, \beta_j)$ is a n-dimension vector.

$$M_j(z_{*j}, z_{ij}) = \frac{k_j}{k_j + |z_{*j} - z_{ij}|} = \frac{k_j}{k_j + \frac{|\alpha x_{*j} + \beta_j - \alpha x_{ij} - \beta_j|}{\|\alpha X_j + B_j\|_2}}$$

$$= \frac{k_j}{k_j + \frac{|\alpha x_{*j} - \alpha x_{*j}|}{\|\alpha X_j\|_2}} = M_j(x_{*j}, x_{ij})$$

Therefore the fuzzy metric is invariant by origin and scale changes, in turn, the composite indicator too. □

- 4th. Exhaustiveness. The weighs of the single indicators are introduced according to their relevance through variable importance scores, therefore, the composite indicator presented in this study compiles the information of the indicators in a hierarchical way through the variable function importance. This property is called *exhaustiveness* by some authors Pena Trapero (1977, 2009), Zarzosa Espina (1996).
- 5th.

Proposition 7. *Let X_j be the single indicator, the CFI defined as Eq. (3) is monotone.*

Proof. If we are measuring the proximity to a phenomenon in which the closer 1 the CFI indicates the better real situation is (i.e. competitiveness, economic development, technological level, socio-economic status,etc.), an increase in a single indicator with positive polarity, it must also generate an increase in the composite indicator.

Suppose that some X_j single indicator has positive polarity $j \in \{1, \dots, m\}$, $x_{ij} < x_{\ell j} < x_{*j}$ where x_{*j} is the best scenario. If we are measuring the proximity to a phenomenon in which the closer 1 the CFI indicates the worst real situation (i.e. poverty, vulnerability, corruption, etc.), the reference vector shows the worse real situation is in each single indicator.

$$x_{ij} - x_{*j} < x_{\ell j} - x_{*j};$$

$$|x_{*j} - x_{ij}| > |x_{*j} - x_{\ell j}|;$$

$$\frac{k_j}{k_j + |x_{*j} - x_{ij}|} < \frac{k_j}{k_j + |x_{*j} - x_{\ell j}|};$$

$$M_j(x_{*j}, x_{ij}) < M_j(x_{*j}, x_{\ell j}).$$

Conversely, supposing that some X_j single indicator has negative polarity $j \in \{1, \dots, m\}$, an increase in a single indicator must generate a decrease in the composite indicator. $x_{*j} < x_{ij} < x_{\ell j}$ where x_{*j} is the best scenario.

$$x_{ij} - x_{*j} < x_{\ell j} - x_{*j};$$

$$|x_{ij} - x_{*j}| < |x_{\ell j} - x_{*j}|;$$

$$\frac{k_j}{k_j + |x_{*j} - x_{ij}|} > \frac{k_j}{k_j + |x_{*j} - x_{\ell j}|};$$

$$M_j(x_{*j}, x_{ij}) > M_j(x_{*j}, x_{\ell j}).$$

Taking as a reference an indicator with positive polarity, if one single indicator shows a better situation, whereas the rest of indicators remain constant, the composite indicator must reflect this improvement, and vice versa. We suppose that $M_p(x_{*p}, x_{ip}) < M_p(x_{*p}, x_{\ell p})$ for any $i \neq \ell$,where $i, \ell \in \{1, \dots, n\}$ and $M_j(x_{*j}, x_{ij}) = M_j(x_{*j}, x_{\ell j})$ for all $j \neq p$, then

$$\begin{aligned} CFI(x_*, x_i) &= \prod_{j=1}^{p-1} M_j(x_{*j}, x_{ij}) \cdot M_p(x_{*p}, x_{ip}) \cdot \prod_{j=p+1}^{m} M_j(x_{*j}, x_{ij}) \\ &< \prod_{j=1}^{p-1} M_j(x_{*j}, x_{\ell j}) \cdot M_p(x_{*p}, x_{\ell p}) \cdot \prod_{j=p+1}^{m} M_j(x_{*j}, x_{\ell j}) \\ &= CFI(x_*, x_\ell) \end{aligned}$$

The proof taking an indicator with negative polarity is symmetrical. □

Table 1
Comparison of aggregation methods.

	Arithmetic mean	Geometric mean	PCA	CFI
Existence	✔	✖	✔	✔
Transitivity	✔	✔	✔	✔
Invariance	✔	✔	✖	✔
Exhaustiveness	✖	✖	✖	✔
Monotonicity	✔	✔	✖	✔
Symmetry	✔	✔	✔	✔
Homogeneity	✔	✖	✖	✖

PCA = Principal components analysis. CFI = composite fuzzy indicator.

- 6th. The composite indicator defined as Eq. (3) is symmetric. The value of the composite indicator, as result of method presented in this study does not depend on the order of the single indicators. According to Grabisch et al. (2011), Proposition 5), we test several permutations in the order in which the single indicators have been introduced, obtaining the same result. Proof of this property will be shown in the following section.

5.2. Comparison with other aggregation methods

In this section we compare the widest applied methods to build composite indicators with our proposal CFM. Table 1 illustrates that the composite indicators built with the arithmetic mean, the geometric mean and PCA do not fulfil all the six properties analysed in the previous section, and also that our CFI proposal does not observe the property of homogeneity. Next, these cases are stressed. A more detailed analysis about the mathematical properties can be consulted in Herrero et al. (2012) (for arithmetic and geometric means) and Jiménez-Fernández and Ruiz-Martos (2020), Jolliffe (2002) and (OECD, 2008) (for PCA). Firstly, the geometric mean is indeterminate in the case that one of the indicators, in any unit, takes the value zero or negative. In the case of PCA, the values of the composite indicators depend on the extraction method and the rotation of axes used, so that the results are more easily interpretable.

Secondly, composite indicators built with the arithmetic mean, the geometric mean and PCA are not unique to scale changes (invariance property), hence the results are affected by the choice of normalization values.

Thirdly, regarding exhaustiveness, arithmetic mean and geometric mean do not avoid the duplication of information provided by the single indicators, so that an increase in information does not necessarily translate into a better composite indicator. Although PCA avoids the duplicity of information provided by the single indicators, it only removes the linear redundant information. Additionally, the PCA composite indicators derived from just the first component do not take full advantage because they ignore any useful non-redundant information present in the data.

Fourthly, the PCA composite indicators verify the monotonicity as long as single indicators with equal polarities are positively correlated; otherwise inconsistent results would be reached (Mazziotta & Pareto, 2019).

Finally, the property of homogeneity referred to the composite indicator is a degree 1 homogeneous function with respect to the single indicators. This would mean that if all the single indicators are multiplied by a constant, the composite indicator is also multiplied by the same constant. Our CFI proposal does not verify this property due to the multiplicative nature of its calculation formula. The same happens to the geometric mean. For PCA composite indicators, if all the single indicators are multiplied by a constant, the values of the composite indicator will not change, because the correlation matrix is the same in both cases (namely, the matrix of data and the matrix of data multiplied by a constant).

6. Robustness of the composite indicator

Checking the robustness of a composite indicator is a crucial step to increase its transparency and also to prevent drawing misleading implications from it. In spite of the importance of robustness check, little attention has been paid to this step in the empirical applications (see Greco et al., 2019). Uncertainty and sensitivity analyses are the widest used techniques for checking the robustness under the framework of the traditional methods for building composite indicators (OECD, 2008; Saisana et al., 2005).

In order to analyse the robustness of the composite indicator introduced in this study, three strategies are used. In all of them, we start from an input set artificially generated by using a random uniform sample where four hundred observations ($i \in \{1, \ldots, 400\}$) are ranked according to the composite indicator presented in this study. We assume that 10 single indicators have been considered ($j \in \{1, \ldots, 10\}$) to carry out the experiment. Likewise, in the three strategies, the goodness of the model will be referred to changes observed in the final outcome of the composite indicator, that is the changes in the ranking of the units studied.

The first strategy analyses the degree of importance of each single indicator in the composite indicator. As described above, we determine the importance of the single indicators identifying the subset of the p single indicators that best explain the composite indicator through PDP. According to the 6th property of the aggregation method, we perform two permutations of the single indicators. We exchange the first indicator with the last one, and the first indicator with the second one. We obtain the same rank in the observations for all permutations. Therefore, the order of the composite indicator observations is invariant by permutations of the single indicators.

In order to visualize the geometric behaviour of each single indicator with respect to the composite indicator while remains constant all the other single indicators, we use PDP. Fig. 2 displays the projection of that particular single indicator on the model's predictions for the random sample considered in this study. As shown, in the database studied and for all the single indicators, a truncated association is observed between the simple indicator and the composite indicator. In this case, this result gives the intuition that it would not be correct to apply a unique type of adjustment between a single indicator and a composite indicator for the entire data sample. Accordingly, the use of the MARS procedure seems a suitable option.

In addition, Table 2 shows a statistical description of the averages of the residues of all the samples used in the procedure carried out through the Monte Carlo method. In the same way that Fig. 2, MARS model is the best model according to cross-validated horizontal axis, Root-Mean-Square-Error (RMSE) for our Monte Carlo random samples.

The second strategy analyses variability of the composite indicator scores when some units are randomly deleted. To carry out this test, we perform a Monte Carlo procedure. To perform this analysis, firstly, 10 random observations or units are deleted of the database. The composite indicator (CFI) is calculated on the remaining 390 observations. Secondly, we use the original database with the 400 units to calculate the CFI. Once computed, we remove the 10 observations CFI results corresponding to the same units that were eliminated in the previous step. The two performed CFI have been compared using Spearman's, Kendall and robust rank correlation statistics. Assuming a type I error $\alpha = 0.05$, and also a coefficient γ, as parameters of robust gamma rank correlation the algorithm stops when the p-value is smaller than α and γ greater than 0.9. One hundred random sample data set were analysed, for which 10 observations were randomly deleted for each sample following the previous steps. The results show γ computed as significance. As shown in Fig. 3, except in the case of the Kendall test, in which correlation is also evident, the others present strong evidence of correlation between the two composite indices analysed in each iteration. Table 3 summarizes the statistical outputs for each test.

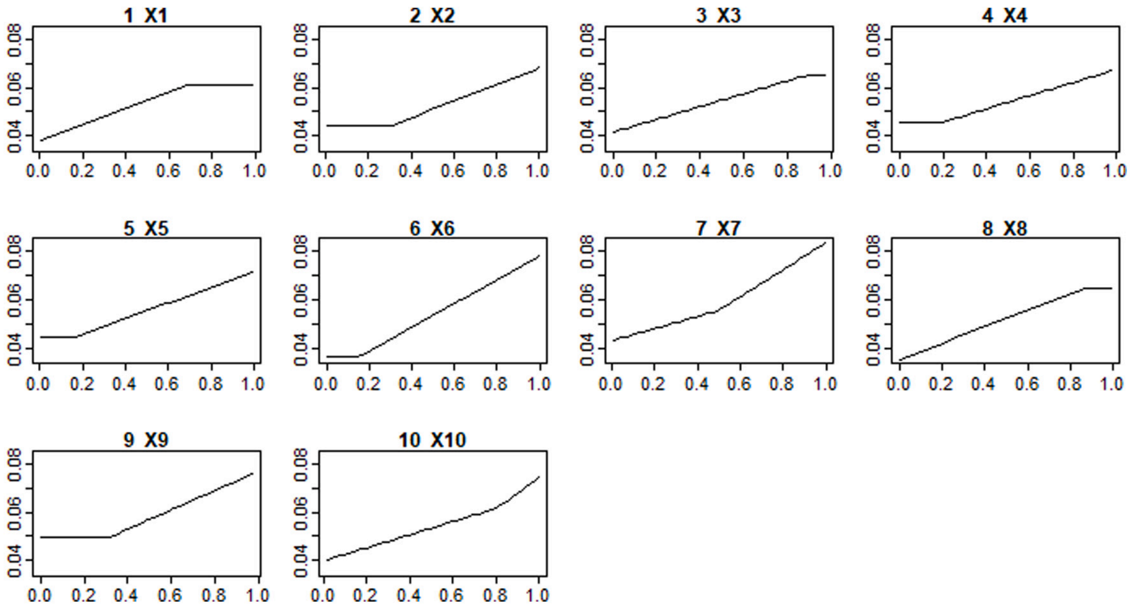

Fig. 2. Projection over single indicators. Single indicators horizontal axis and composite indicator in vertical axis.

Table 2
Statistical outputs of the residuals of estimates of fit between CFI and single indicators.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
OLS	0.000823	0.000864	0.001030	0.001144	0.001121	0.002335
PCR	0.000883	0.000985	0.001139	0.0011300	0.001424	0.002432
PLS	0.000833	0.000867	0.001025	0.001143	0.001120	0.002320
Elastic net	0.000748	0.000811	0.001004	0.001135	0.001167	0.002412
MARS	0.000625	0.000771	0.001018	0.001046	0.001324	0.001446

Note. CFI: Composite fuzzy indicator. OLS: Ordinary less squares. PCR: Principal component regression. PLS: Partial least squares. Elastic net: Regularized regression. MARS: Multivariate adaptive regression splines. Residual calculated by cross-validated Root-Mean-Square-Error (RMSE).

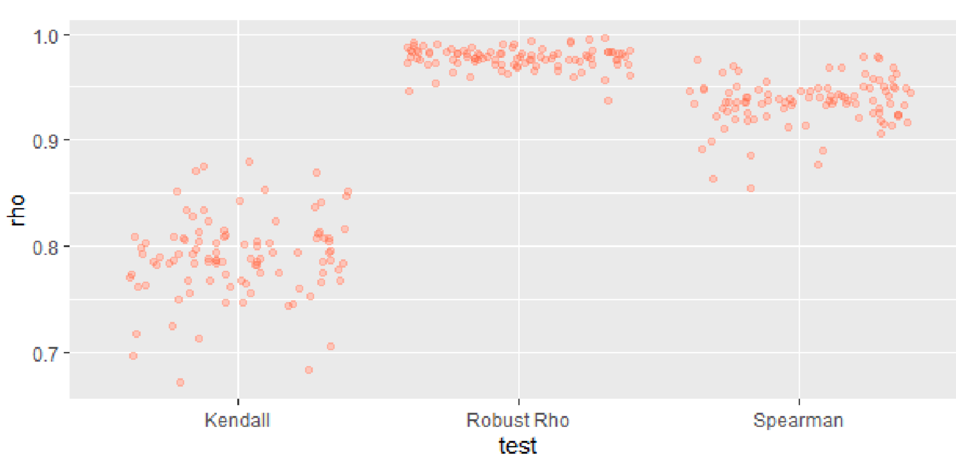

Fig. 3. Results of simulations of composite fuzzy indicators when ten observations are randomly deleted.

Table 3
Statistical summary of the ranking test.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Spearman	0.8551	0.9252	0.9370	0.9359	0.9487	0.9786
Kendall	0.6720	0.7697	0.7884	0.7898	0.8084	0.8794
Robust γ	0.9372	0.9727	0.9785	0.9774	0.9831	0.9966

This allows us to deduce that the ranking associated to the composite indicators are significantly the same when we eliminate observations.

The third strategy to check the robustness of the method proposed focuses on influence analysis. We generate an input as random convex linear combinations of the original single indicators called *Xadded*, which is included in the underlying model. Monte Carlo procedure assesses one hundred random convex linear combinations. Variable importance scores provided by the algorithm indicates in all the cases the irrelevance in the underline model of the added indicator (Fig. 4).

Fig. 4 shows that on average the most important single indicators in the underlying model are 2 and 4, the added single indicator being irrelevant because it provides redundant information. If we compare the composite indicators of the original model with respect to the

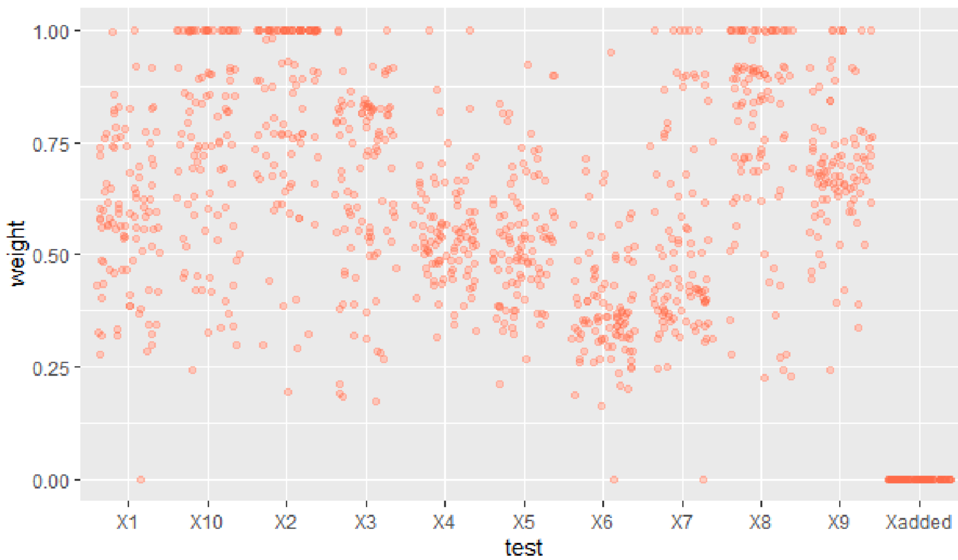

Fig. 4. Importance of single indicators in one hundred iterations.

Table 4
Influence analysis test.

Spearman	Kendall	Robust rho
Min. :0.8823	Min. :0.7038	Min. :0.9609
1st Qu. :0.9389	1st Qu. :0.7922	1st Qu. :0.9849
Median :0.9517	Median :0.8125	Median :0.9888
Mean :0.9479	Mean :0.8124	Mean :0.9868
3rd Qu. :0.9649	3rd Qu. :0.8420	3rd Qu. :0.9927
Max. :0.9896	Max. :0.9163	Max. :0.9989

model with the added linearly dependent indicator, we observe that the correlation in range is significantly equal, being especially similar when using the Spearman tests and robust rank γ correlation test (Table 4).

Therefore, it does not seem to be significant differences in the range of observations when adding single indicators that are linear combinations of the rest, that is the algorithm discriminates the information linearly redundant. This technique commonly referred to as influence analysis gives us a 0 weight for *Xadded* single indicator, and therefore minimizes their influence on the resulting composite indicator (Table 5).

7. Conclusions and remarks

There is an increasing popularity of rankings obtained with composite indicators based on simple methods, but these are not reliable in some cases (Becker et al., 2017; Jiménez-Fernández & Ruiz-Martos, 2020). This paper aims at developing a new method for building composite indicators which allows making more accurate and reliable estimations of the concept being studied.

The methodology is designed for formative measurement models using a set of indicators measured on different scales (quantitative, ordinal and binary) and it is partially compensatory. The CFI composite indicator is a fuzzy metric that is obtained from the grouping of a set of fuzzy metrics (George & Veeramani, 1994). The use of this mathematical approach is very appropriate for the study of phenomena in which their boundaries are not well defined. In this regard, as a limitation of our proposal, it is worth noting that the method we present is sensitive to the change of metric and t-norm. An appropriate metric must be used in each case and, therefore, an appropriate choice of the t-norm that makes its use compatible too (Gregori et al., 2011). The CFI will be more realistic, or its goodness of fit will be better, to the extent that the researcher has the ability to provide the most suitable metric to apply to each single indicator in the context of the concept studied.

In this study we have been concerned with the allocation of the weights of the single indicators and the method of aggregation. From the comparison made of the mathematical properties that verify the composite indicators built with arithmetic mean, geometric mean, PCA and CFI, we found that one of the major disadvantages of the most popular methods is the lack of exhaustiveness. To be exhaustive, a composite indicator should take full advantage, and in a useful way, of the information provided by the single indicators (Zarzosa Espina, 1996). That is, a composite indicator is better than another if it provides more useful information about the phenomenon studied. In turn, this implies that it should be able to eliminate duplicate information. These aspects are key to performing a rigorous benchmarking. In our proposed method for building composite indicators, we use unsupervised machine learning techniques. Most of the current methodologies for the construction of composite indicators deal with this issue by using algorithms that are intrinsically linear or models that adapt to the nonlinear patterns of the data a priori. In contrast, in the methodology presented in this study it is not necessary to explicitly know or specify the exact form of non-linearity before the model's training. In this vein, the algorithms provided will look for, and discover, non-linearity in the data that will help to optimize the relationships. Once the functional relationship is detected, the relationships between the composite indicator and the individual indicators are quantified through partial dependency plots (PDP). The weight attributed to each single indicator will show the relevance of each single indicator in the constructed metric. The resulting composite indicator will be the result of a metric structure that allows the comparison of observations and contains the precise information provided by each single indicator.

A common practice is to make comparisons between the most popular methodologies for the construction of composite indicators and the proposed methodology grounded on the outcomes. However, this type of analysis is omitted for the following three reasons. Firstly, unlike the more traditional methods, the proposed methodology is a metric and, in addition, it satisfies the properties of existence, transitivity, invariance, exhaustiveness, monotony and symmetry. Secondly, to eliminate the implicit subjective choices of other methodologies, unsupervised machine-learning techniques are used. Thirdly, the concept of fuzzy metrics is used to provide a more realistic approach to the concepts that will be modelled. Henceforth, the methodology presented is diametrically different from the traditional ones, so that, from our point of view, the comparison of the results that would be obtained with the same database would lack conceptual solidity. Instead, we have studied and compared the mathematical properties that satisfy the most

Table 5
Sensitivity scores by single indicator.

X1	X2	X3	X4	X5	X6
Min. :0.0000	Min. :0.1950	Min. :0.1748	Min. :0.3175	Min. :0.2107	Min. :0.0000
1st Qu. :0.4788	1st Qu. :0.7245	1st Qu. :0.5396	1st Qu. :0.4810	1st Qu. :0.4460	1st Qu. :0.3136
Median :0.5846	Median :0.8931	Median :0.7263	Median :0.5357	Median :0.5217	Median :0.3587
Mean :0.5878	Mean :0.8354	Mean :0.6636	Mean :0.5517	Mean :0.5340	Mean :0.3842
3rd Qu. :0.7169	3rd Qu. :1.0000	3rd Qu. :0.8241	3rd Qu. :0.5964	3rd Qu. :0.5983	3rd Qu. :0.4358
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :0.9232	Max. :0.9516
X7	X8	X9	X10	Xadded	
Min. :0.0000	Min. :0.2267	Min. :0.2437	Min. :0.2447	Min. :0	
1st Qu. :0.3754	1st Qu. :0.6959	1st Qu. :0.6290	1st Qu. :0.6063	1st Qu. :0	
Median :0.4337	Median :0.8652	Median :0.6841	Median :0.7812	Median :0	
Mean :0.5301	Mean :0.7919	Mean :0.6943	Mean :0.7565	Mean :0	
3rd Qu. :0.6642	3rd Qu. :0.9200	3rd Qu. :0.7521	3rd Qu. :0.9966	3rd Qu. :0	
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :0	

commonly used methods for constructing composite indicators and our proposal. We have found that, except for the property of homogeneity, our method verifies all the properties indicated in the literature of composite indicators, whereas the most traditional methods presented a worse balance.

Additionally, simulation methods have been employed to investigate the robustness of composite indicators and the conclusions based on them. More specifically, we developed specific strategies to check the robustness in three directions. Firstly, we checked that the composite indicator is invariant regarding the permutations of single indicators. Secondly, it has been tested that the method is not sensitive to the elimination of observations. The CFI composite indicator is stable when observations are removed. In other words, the ranking of the units does not register significant alterations when some of the units are eliminated. Finally, with the addition of noise to the model for the detection of multicollinearity problems, we showed that our method is able to solve problems of this nature. In other words, the multicollinearity problem generates composite indicators with redundant information, thus giving to some indicators a greater weight than they should have. This anomaly must be addressed. From a Monte-Carlo test it is found that our methodology is capable of correcting multicollinearity.

To sum up, and without seeking to be presumptuous, we strongly believe that our methodology to build composite indicators solves to a great extent those aspects that have not been resolved by classical methodologies, providing a more realistic and faithful vision of the phenomenon studied.

CRediT authorship contribution statement

E. Jiménez-Fernández: Conceptualization, Methodology, Software, Validation, Writing – original draft, Funding acquisition. **A. Sánchez:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Funding acquisition. **E.A. Sánchez Pérez:** Conceptualization, Writing – original draft, Supervision.

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.

Acknowledgment

This work was supported by the project FEDER-University of Granada (B-SEJ-242.UGR20), 2021-2023: An innovative methodological approach for measuring multidimensional poverty in Andalusia (COMPOSITE). Eduardo Jiménez-Fernández would also like to thank the support received from Universitat Jaume I under the grant E-2018-03. Funding for open access charge: Universidad de Granada / CBUA.

References

Annoni, P., & Dijkstr, L. (2019). *The EU regional competitiveness index 2019*. Luxembourg: Publications Office of the European Union.

Apley, D. W., & Zhu, J. (2016). Visualizing the effects of predictor variables in black box supervised learning models. arXiv:1612.08468 [stat.ME].

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22.

Bodenhofer, U., & Klawonn, F. (2008). Robust rank correlation coefficients on the basis of fuzzy orderings: initial steps. *Mathware & Soft Computing*, *15*, 5–20.

Bodenhofer, U., Krone, M., & Klawonn, F. (2013). Testing noisy numerical data for monotonic association. *Information Sciences*, *245*, 21–37.

Canadian Index of Wellbeing (2016). *How are canadians really doing? The 2016 CIW national report*. Waterloo, ON: Canadian indicator of Wellbeing and University of Waterloo.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262.

Craven, P., & Wahba, G. (1971). Smoothing noisy data with spline functions. *Numerische Mathematik*, *31*, 377–403.

Cruz-Martínez, G. (2014). Welfare state development in latin america and the caribbean (1970s–2000s): Multidimensional welfare index, its methodology and results. *Social Indicators Research*, *119*(3), 1295–1317.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*, 1203–1218.

Friedman, J. H. (1991). Multivariate adaptive regresion splines. *The Annals of Statistics*, *19*(1), 1–141.

George, A., & Veeramani, P. (1994). On some results in fuzzy metric spaces. *Fuzzy Sets and Systems*, *64*, 395–399.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764.

Grabisch, M., Marichal, J. L., Mesiar, R., & E, Pap. (2011). Aggregation functions: Means. *Information Sciences*, *181*(1), 1–22.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of cmposite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, *141*, 61–94. http://dx.doi.org/10.1007/s11205-017-1832-9.

Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). A simple and effective model-based variable importance measure. arXiv:1805.04755v1 [stat.ML].

Gregori, V., Morillas, S., & Sapena, A. (2011). Examples of fuzzy metrics and applications. *Fuzzy Sets and Systems*, *170*, 995-111.

Hauze, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data. *Questiones Geographicae*, *30*(2), 87–93.

Herrero, C., Martínez, R., & Villar, A. (2012). A newer human development index. *Journal of Human Development and Capabilities*, *13*, 247–268.

Jiménez-Fernández, E., & Ruiz-Martos, M. J. (2020). Review of some statistical methods for constructing composite indicators. *Estudios de Economía Aplicada-Studies of Applied Economy*, *38*(1), 1–15.

Jolliffe, I. (2002). *Principal component analysis*. New York: Springer.

Keogh, S., O'Neill, S., & Walsh, K. (2021). Composite measures for assessing multidimensional social exclusion in later life: Conceptual and methodological challenges. *Social Indicators Research*, http://dx.doi.org/10.1007/s11205-021-02617-7.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. http://dx.doi.org/10.18637/jss.v028.i05.

López-Ortega, O., & Castro-Espinoza, F. (2019). Fuzzy similarity metrics and their application to consensus reaching in group decision making. *Journal of Intelligent & Fuzzy Systems*, *36*, 3095–3104.

Madonia, G., Cracolici, M. F., & Cuffaro, M. (2013). Exploring wider well-being in the EU-15 countries: An empirical application of the Stiglitz report. *Social Indicators Research*, *111*(1), 117–140.

Maggino, F. (2017). Developing indicators and managing the complexity (Chapter 4). In F. Maggino (Ed.), *Complexity in society: From indicators construction to their synthesis* (pp. 87–114). Cham, Switzerland: Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-60595-1.

Mazziotta, M., & Pareto, A. (2017). Synthesis of indicators: The composite indicators approach (Chapter 7). In F. Maggino (Ed.), *Complexity in society: From indicators construction to their synthesis* (pp. 159–192).

Mazziotta, M., & Pareto, A. (2019). Use and misuse of PCA for measuring well-being. *Social Indicators Research*, *1422*, 451–476. http://dx.doi.org/10.1007/s11205-018-1933-0.

Montero, J. M., Chasco, C., & Larraz, B. (2010). Building an environmental quality index for a big city: a spatial interpolation approach combined with a distance indicator. *Journal of Geographical Systems*, *12*(4), 435–459.

OECD (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD publishing, http://dx.doi.org/10.1787/9789264043466-en.

OECD (2017). *How's Life? 2017: Measuring Well-Being*. OECD Publishing, http://dx.doi.org/10.1787/how-life-2017-en.

Pena Trapero, J. B. (1977). *Problemas de la Medición Del Bienestar Y Conceptos Afines (Problems of Welfare Measurement and Related Concepts)*. Madrid: INE.

Pena Trapero, J. B. (2009). La medición del bienestar social: una revisión crítica (measuring social welfare: a critical review). *Estudios de Economía Aplicada/Studies of Applied Economy*, *27*(2), 299–324.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., & Fuller, G. (2018). *SDG indicator and dashboards report 2018*. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

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

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, *81*(1), 65–77.

Sánchez, A., Chica-Olmo, J., & Jiménez-Aguilera, J. D. (2018). A space–time study for mapping quality of life in andalusia during the crisis. *Social Indicators Research*, *135*(2), 699–728. http://dx.doi.org/10.1007/s11205-016-1497-9.

Sánchez, A., & Ruiz-Martos, M. J. (2013). A multidimensional regional development index as an alternative allocation mechanism of EU structural funds remittances. In *39 international conference on regional studies*. https://old.reunionesdeestudiosregionales.org/Oviedo2013/htdocs/pdf/p683.pdf.

Sánchez, A., & Ruiz-Martos, M. (2018). Europe 2020 strategy and citizens' life satisfaction. *Journal of Happiness Studies*, *19*, 2315–2338. http://dx.doi.org/10.1007/s10902-017-9928-0.

Schwab, K., & Porter, M. (2008). *The global competitiveness report 2008–2009*. Geneva, Switzerland: World Economic Forum.

Segi Rahmat, M. R., & Noorani, M. S. (2008). Product of fuzzy metric spaces and fixed point theorems. *International Journal of Contemporary Mathematical Sciences*, *3*(15), 703–712.

UNDP (2018). *Human development indices and indicators*. New York: United Nations Development Programme.

Zarzosa Espina, P. (1996). *Aproximación a la Medición Del Bienestar Social*. Valladolid (Approach to the measurement of social welfare)- University of Valladolid.

Zhou, P., Fan, L. W., & Zhou, D. Q. (2010). Data aggregation in constructing composite indicators: A perspective of information loss. *Expert Systems with Applications*, *37*, 360–365.