Environment, Development and Sustainability
https://doi.org/10.1007/s10668-023-04116-w

Building composite indicators for the territorial quality of life assessment in European regions: combining data reduction and alternative weighting techniques

Eda Ustaoglu[1] · Gloria Ortega Lopez[2] · Alejandro Gutierrez-Alcoba[3]

Received: 15 February 2022 / Accepted: 27 October 2023

Abstract

Development of composite indicators is a challenging task given that sustainability indices are strongly dependent on how the sub-indicators are weighted. This is because relative indicator weights may significantly differ based on the chosen weighting methods used in the analysis. There is hardly any study that has paid attention to this issue so far. Therefore, this paper aims to fill this gap in the literature by searching the robustness of selected weighting methods, i.e. entropy-weight (EW), principal component analysis (PCA), machine learning approaches (random forest-RF), regression analysis (RA) and benefit-of-the-doubt (BOD) when constructing a composite indicator. To research the current sustainability performance of European regions, the present study focuses on the Territorial Quality of Life Index—initially proposed by the ESPON Programme—that are aligned with the specific targets of the Sustainable Development Goals of the 2030 Agenda. The methods to construct composite indicators include stages of data preparation (including the estimation of missing values with random forest method), normalization, statistical transformation of raw data, reduction of indicators in order to ease public communication (using the PCA method) and data interpretation, weighting of the sub-indicators using EW, PCA, RF, RA and BOD methods and their linear weighted aggregation, and checking for robustness and sensitivity. The results suggest that there are significant differences in the rank and spatial distribution of composite indicators based on the use of different weighting methods considered in the analysis. The results from sensitivity analysis support the robustness of entropy-weight method among others. The methodology used in the current analysis can be adapted to other study areas and regions internationally. The findings showed that Eastern European countries and some Mediterranean countries have relatively lower index values compared to other European regions; therefore, policy and planning actions are needed covering these regions specifically.

Keywords Sustainable urbanization · Territorial quality of life · Composite indicators · Weighting procedures · Europe

Extended author information available on the last page of the article

Published online: 24 November 2023

1 Introduction

European landscape has faced with substantial social, economic and environmental challenges due to uncontrolled growth of cities and regions during the past century. In 2016, more than 70% of European Union (EU) citizens lived in an urban area; and it is estimated that 80 percent of the total population will be living in cities by 2050 (Netherlands Presidency, 2016). The unprecedented demographic growth and the consequent changes in landscape do not only impact on the local environment but also have impacts on natural environment through exploitation of natural resources to support urban economy as well as social aspects such as urban poverty and social segregation (Michael et al., 2014; Wang et al., 2019). The urban activities raise socio-economic and environmental challenges in cities and regions of Europe (McPhearson et al., 2016). In 2016, EU launched the EU Urban Agenda, which covers a range of urban issues including housing, mobility, safety, poverty, migration, air quality, energy transition, and climate adaptation. The emphasis here is on the urgent need to enhance urban quality of life and ensure sustainable operations in cities and urban regions.

Indicators reduce complexity so that policy decisions can be framed, provide a linkage between science and policy, and further will assist in decision makers towards potential solutions to socio-economic and environmental problems (de Sherbinin et al., 2014). Towards this end, indicators have generally been developed for administrative units over which government authorities have responsibility. This requires aggregation of micro data to administrative units, normalization, statistical transformation of raw data and data reduction in order to ease public communication and data interpretation (OECD, 2008). There is an increasing number of countries attempted to the development of indicators that integrated three prominent pillars of sustainable development for different cities and regions, which vary according to their particular needs and goals (Brandon & Lombardi, 2005). Development of indicators for the assessment of urbanization quality has also gained attention from scholars. There are both qualitative and quantitative methods found in the literature used for the assessment of urban sustainability. Among the studies that utilized qualitative methods, Feng and Chen (2010), Chen et al. (2014), and Li et al., (2019a, 2019b) focused on quality of urbanization while Tannier and Thomas (2013), Gollin et al. (2016) and Zhao et al. (2017a) investigated theories of urbanization and morphological characteristics of cities. In this group, there are also studies examining dynamics of urbanization including Chen et al. (2013), Cao et al. (2014), Fang et al. (2015), Ye et al. (2017), Ustaoglu and Aydınoglu (2019), and Cai et al. (2020). There are more studies on quantitative assessment of urbanization quality, which consist of a complete assessment of urbanization and its drivers (Lopez et al., 2017; Ma & Sun, 2020; Ramachandra et al., 2015; Yang et al., 2019); development of index and indices systems (Merino-Saum et al., 2020; Mori & Yamashita, 2015; Phillis et al., 2017) and dynamic coupling of coordination between urbanization and socio-economic and environmental development (Li et al., 2019a, 2019b; Tomal, 2021; Zhao et al., 2017b). Two strands of literature aimed to develop indicators for the urbanization quality assessment: The first strand includes studies of single indicators that focused on socio-economic (Annoni & Bolsi, 2020; D'Adamo et al., 2020), land use (Bo et al., 2019; Chen et al., 2013) and environmental (Becker et al., 2017; De Sherbinin et al., 2014) aspects of urbanization. The second strand, on the other hand, covers studies of synthetic composite indicators which combines social-cultural, economic, technological and eco-environmental indicators for the development of a common index of sustainability. The latter consists of studies examining the urban sustainability indices at the local area

(Zhou et al., 2015), municipality/city (Zoeteman et al., 2016), regional (Yang et al., 2020) or at the continental (ESPON, 2020) levels.

Composite indicators are popular tools, especially due to their aims of summarizing, focusing and condensing complex information on technological development, social/socio-cultural, economic and environmental aspects. Composite indicators are a mathematical combination of a set of individual indicators representing different dimensions of a concept that have no common unit of measurement (Nardo et al., 2005). It is argued that composite indicators are simpler to interpret than researching to find a common trend in many single indicators (Paruolo et al., 2013). In fact, composite indicators are of significance for monitoring multidimensional processes where the use of individual indicators is notably difficult due to the challenges in linking trends across dimensions and capturing interactions between and within sub-systems (Dale & Beyeler, 2001; Munda, 2005). Indicator-based monitoring of spatial development can provide planning authorities and urban managers more flexibility in achieving sustainable development goals (Keiner, 2006). Further to this, rankings of composite indicators force institutions and government authorities to improve their standards (Kelley & Simmons, 2015). Composite indicators should aim to support needs of planning for sustainable development through integrating issues identified as important both from a scientific perspective and based on stakeholder concerns.

Even though rankings of indicators are carried out giving considerable attention, there are yet subjectivities in their construction, which raises the need for conducting uncertainty and sensitivity analysis on composite indicator assumptions (Becker et al., 2017; Saisana et al., 2005). The step discussed in this paper is the weighting process which can have a significant impact on the rankings of analysed regions/cities and subsequent policy making. Prior to the selection of indicator aggregation method, values of weights to apply to each sub-indicator should be specified. A wide range of composite indicators relies on equal weights implying that all variables are given the same weight (Blancas et al., 2010; Patias et al., 2021; UNDP, 1990; World Bank, 1999). However, not all the indicators have to be equally important in explaining the sustainability performance rankings. Larger weights are thus assigned to indicators which are more important than others. In such cases, weights are directly obtained from the data using, for instance, factor analysis (FA) (Salvati & Carlucci, 2014), principal component analysis (PCA) (Davino & Romano, 2014), entropy-weight method (Zhou et al., 2015), regression analysis (Porter & Stern, 2001) or Data Envelopment Analysis (DEA)/Benefit-of-the-Doubt (BOD) (Blancard et al., 2021) while others estimate the weights externally to the data using expert based systems such as conjoint analysis (Ülengin et al., 2001), budget allocation (Goedkoop & Spriensma, 2001), public opinion (Parker, 1991), Analytical Hierarchy Process (AHP) (D'Adamo et al., 2020), the Best–Worst method (Gomez-Limon et al., 2020) and others (see Freudenberg, 2003; Nardo et al., 2005 for a full review of the methods). In the current study, we aim to focus on data-based weighting methods and left the expert-based systems to be examined as a future research topic. Among the data-based methods, the most commonly used are entropy-weight method, PCA/FA, regression analysis, machine learning approaches (i.e. random forest), and DEA/BOD for the estimation of weights (Belu, 2009; Gan et al., 2017; Gerdessen & Pascucci, 2013; Hou et al., 2021; Huertas et al., 2020; Porter & Stern, 2001; Wang et al., 2015; Xavier et al., 2018). Therefore, we have decided to consider these key five methods in the current study and left comparison of statistical methods with those of expert-based approaches as a future research focus.

There is abundance of literature on composite indicators (Mascarenhas et al., 2015; Yang et al., 2020; Zoeteman et al., 2016) that are mainly based on single weighting schemes and hardly any of them researched the implications of different weighting methods on the results. For instance, at the pan-European level, there are examples of composite indicators on measuring the socio-economic performance of bio-economy sectors (D'Adamo et al., 2020), ranking innovativeness of EU member states complying with 2030 Agenda for sustainable development (Szopik-Depczynska et al., 2018), monitoring regional competitiveness performance at the regional level across EU member states (Annoni & Dijkstra, 2013, 2019), quality of life measurements at different regional levels for the studied European countries (ESPON, 2020), and assessing the quality of government (Fazekas, 2017), capacity of the creative economy (Boal-San Miguel & Herrero-Prieto, 2020) and social progress (Annoni & Bolsi, 2020) at the regional level in Europe. These studies have not paid particular attention to the impact of weighting procedure on composite indicator rankings. Therefore, the current study aims at filling this gap in the literature through constructing a composite indicator for quality of life with a specification for European regions, which is called the 'Territorial Quality of Life Index (TQLI)' as defined by the ESPON, 2020 Cooperation Programme. Given this framework, we compare different weighting methods that enable us to estimate weights internally to the data and examine the robustness and reliability of the composite indicators estimated in the study. In particular, we aim to compare the composite indicator values obtained from different weighting methods computed at the NUTS2 regional level for the EU member states.

2 Territorial quality of life indicators

As defined by ESPON (2020: 9), Territorial Quality of Life Index aims at "*measuring the capabilities of all living beings to survive and flourish in a place, thanks to the economic, social and ecological conditions that support life in that place*." According to this description, in the personal sphere, quality of human life corresponds to personal health and safety needs and needs for housing and basic utilities and in the socio-economic sphere, economic well-being, social and policy factors are being considered to support survival and flourishing of all people living in the place (ESPON, 2020). Finally, the ecological sphere concerns the quality of life of all living beings with a particular focus on the quality of environment as a key determinant of the territorial quality of life (ESPON, 2020). To measure the progress towards the Sustainable Development Goals (SDGs), the indicators in the current study were selected to cover the sustainability aspects specified under the ESPON's: (1) personal sphere, (2) socio-economic sphere, and (3) ecological sphere. Figure 1 presents the three dimensions of territorial quality of life and indicators and sub-indicators corresponding to each sphere that constructs the composite indicator. In the personal sphere, for instance, there are indicators of availability, accessibility and affordability of housing, basic utilities, education and health services, personal health and safety. In the socio-economic sphere, there are enablers of good mobility, digital connectivity, work and consumption choices, social and cultural life as well as indicators of inclusive economy and healthy society. Finally, ecological sphere comprises availability and maintenance of green infrastructure and protected areas, healthy environment and climate change indicators.

Besides social/cultural and ecological domains, there are other factors explaining sustainability (e.g. economic indicators) that are linked to regional competitiveness and prosperity. Dijkstra et al. (2011) defined regional competitiveness as '*the ability to offer an attractive and sustainable environment for firms and residents to live and work*'. It is noted (Annoni & Dijkstra, 2013) that regional competitiveness should be calculated for functional economic regions to take into account the functional economic links. As highlighted

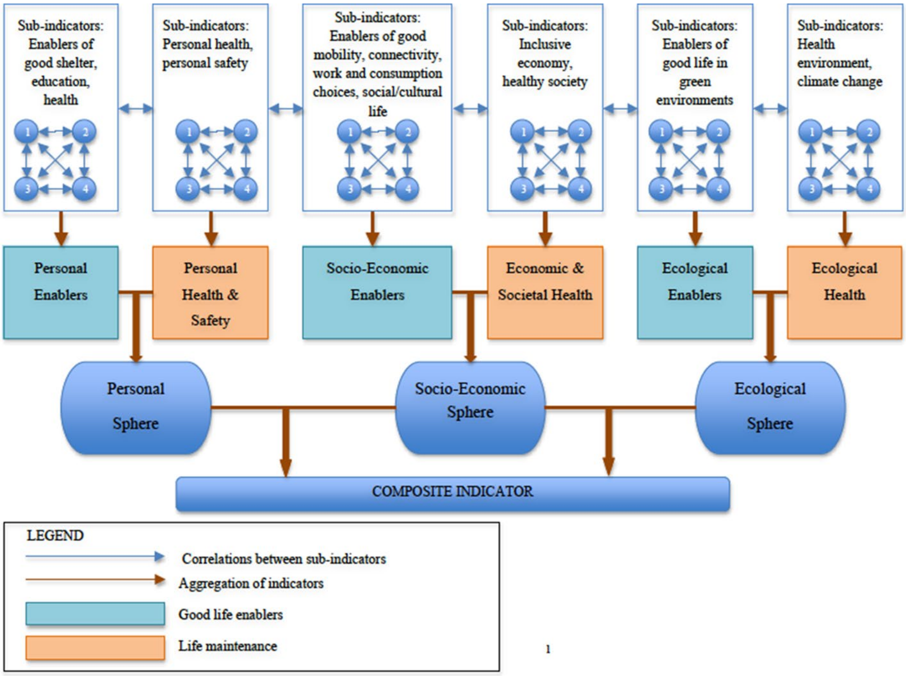

Fig. 1 Territorial Quality of Life Index (TQLI) development framework. *Note* Figure adapted from ESPON (2020) using the methodology given in Fig. 1 of Burgass et al. (2017)

by Dijkstra et al. (2011), NUTS2 regions can be considered as administrative and statistical regions which do not consider the functional economic linkages. For instance, in Europe, there are several cities including London, Brussels, Prague, Berlin, Amsterdam and Wien, which are administratively divided into NUTS2 regions where commuting zone of a city is not covered within the administrative boundaries. Annoni and Dijkstra (2013) suggested merging the region with the one containing the city if the region has at least 40% of its population inside the commuting zone. We followed the same procedure and merged one or more neighbouring regions with the region containing the main city for Brussels (BE), Prague (CZ), Berlin (DE), Amsterdam (NL) and Wien (AT) (Table 1). As an example, in Table 1, the NUTS2 boundary defining Brandenburg (DE40) was merged with the NUTS2 boundary of Berlin city (DE30), and the resulting NUTS2 polygon in Arc map is named as 'DE00'. The indicators associated with the newly formed NUTS2 region (e.g. DE00) were computed in such a way that for instance, regarding population or GDP, the corresponding values (e.g. values assigned to DE30 and DE40) were summed to get the unique value for the new NUTS2 region (e.g. DE00); and for instance, regarding accessibility or access to services, the population weighted average values were computed for the newly created NUTS2 region (e.g. DE00). All the values of indicators for these newly merged NUTS2 regions (the last column in Table 1) were computed either using summation or averaging the original values that were provided for official NUTS2 regions shown in the second column of Table 1.

We did not include London in the analysis given that UK ended its membership to the European Union on 31 January 2020 and currently it is not a EU member state. Through

Table 1 NUTS2 classification adopted for TQLI. *Source* Annoni and Dijkstra (2013)

Merged regions based on commuting patterns	Official NUTS2 region	New merged region
Wien	AT12: Niederösterreich	AT00
	AT13: Wien	
Brussels	BE10: Reg. Bruxelles	BE00
	BE24: Prov. Vlaams-Brabant	
	BE31: Prov. Brabant Wallon	
Prague	CZ01: Praha	CZ00
	CZ02: Stredni Cechy	
Berlin	DE30: Berlin	DE00
	DE40: Brandenburg	
Amsterdam	NL23: Flevoland	NL00
	NL32: Noord-Holland	

excluding UK and merging the regions as explained in Table 1, we had a total of 223 NUTS2 regions that were considered in the study to construct the Territorial Quality of Life (TQL) composite indicators.[1]

3 Methodology

The methodological framework used for the construction of composite indicators is summarized in Fig. 2, which will be explained in detail in the following sub-sections. According to Fig. 2, first, data were compiled from various sources, data quality was checked, and missing values were estimated for the data. Next, data were selected and sub-indicators were constructed based on data quality and its suitability to be aligned with the aims of the study. This is followed by normalization of the sub-indicators so that indicators will be at a comparable scale. Due to existence of collinearity among sub-indicators, the number of sub-indicators was reduced to 30 using the principal component analysis (PCA), which is a method principally used to eliminate correlation across variables. Weighting was applied to final sub-indicators using the methods including entropy-weight, PCA, random forest (RF), regression and BOD. The sub-indicators were aggregated using weighted linear aggregation method. To check the robustness of the weighting methods used in the analysis, a sensitivity analysis was applied to each of the composite indicators that were constructed using different weighting methods. Finally, composite indicators were visualized for communication and dissemination purposes.

3.1 Data

In the current study, TQLIs were compiled based on the indicator construction framework presented in Fig. 1 and availability of data was also influential in the selection of indicators.

[1] The NUTS2 regions of the EU countries as of 2021 can be seen in: https://ec.europa.eu/eurostat/web/nuts/nuts-maps.

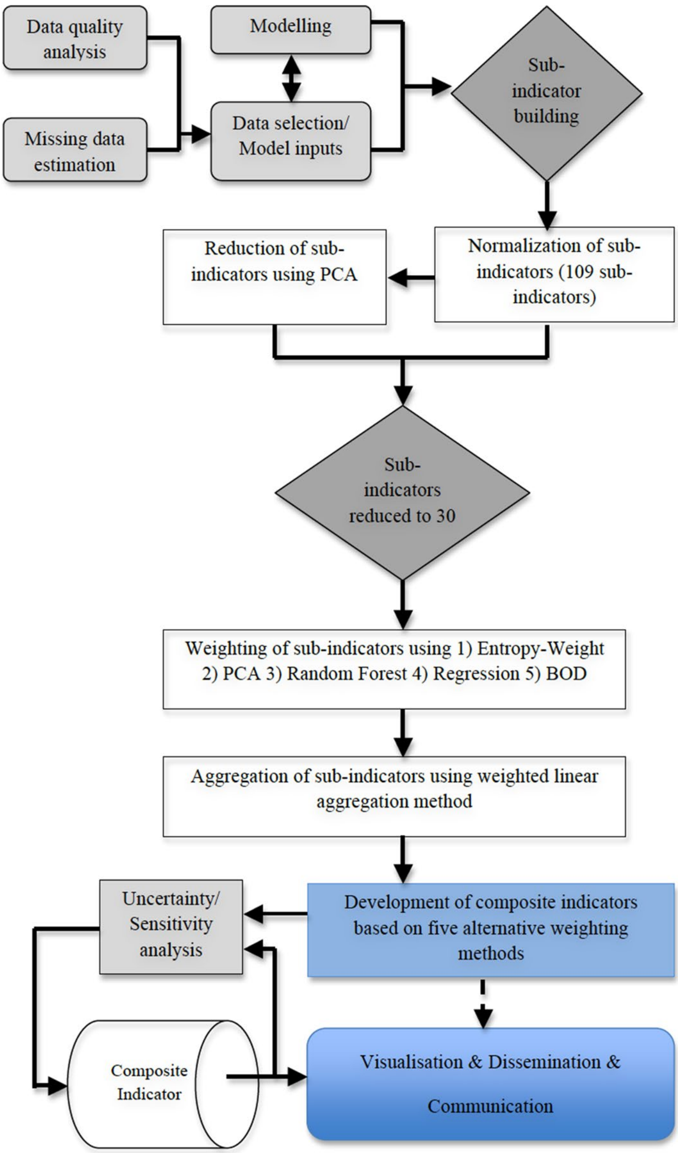

Fig. 2 Different stages of construction of the composite indicator in our study

The indicators included in the study and their data sources are given in Table 7 in Appendix. It is important to note that our key data sources are the databases obtained from Eurostat, European Environment Agency (EEA) and ESPON. There are also other data sources used in the study which are detailed in Table 7. As it can be seen in the table, the data are provided at the country (NUTS0), regional (NUTS1 or NUTS2) or more local (NUTS3) levels. Therefore, all the data were aggregated or disaggregated to the NUTS2 level, which is the statistical unit of our study that we aim to construct the composite indicators.

3.2 Estimation of the missing values in the dataset

The missing data are a vital issue in many statistical models and machine learning algorithms given that these models mainly rely on complete data sets. There are different methods to deal with the missing data problem ranging from simple approaches such as omitting the observations with missing values or mean imputation (Little & Rubin, 1987) for more sophisticated ones such as regression models (Bo et al., 2004), expectation–maximization (EM) approach (Nelwamondo et al., 2007) or nonparametric methods including classification and regression trees (CART) (Burgette & Reiter, 2010) and K-nearest neighbour (K-NN) (Tutz & Ramdan, 2014).

Because we had issues of multicollinearity and nonlinearity observed across parameters in the original dataset, we used classification with random forest (RF) approach, which is nonparametric and does not rely on distributional assumptions on the data (see Stekhoven & Bühlmann, 2012). According to the RF proximity imputation technique, the data are first roughly imputed using strawman imputation. Here, missing values for continuous variables are imputed using the median of non-missing values, and for missing categorical variables, the most frequently occurring non-missing value is used (Tang & Ishwaran, 2017). The RF is fitted to this imputed data. Using the resulting forest, the $n \times n$ (n is the sample size) symmetric proximity matrix is computed and used to impute the original missing values. For continuous predictors, the proximity weighted average of non-missing data is used; and for categorical predictors, the largest average proximity over non-missing data is used. The updated data are used to grow a new RF, and the procedure is iterated. The details on RF classification will be elaborated in the following sections.

3.3 Normalization of indicators

Prior to the PCA, all the indicators were standardized so that the indicators have values between 0 and 1. For a positive indicator, we used Eq. (1); and to represent a negative indicator, Eq. (2) being used.

$$z_{ij} = \frac{\left(x_{ij} - \min(x_i)\right)}{\left(\max(x_i) - \min(x_i)\right)} \tag{1}$$

$$z_{ij} = 1 - \frac{\left(x_{ij} - min(x_i)\right)}{\left(max(x_i) - min(x_i)\right)} \tag{2}$$

3.4 Principal component analysis (PCA) used for re-arrangement of indicators

Although indicator system can provide comprehensive information on the quality of life at the territorial or city scale, it is important to note that individual indicators, which are generally correlated with each other, may provide unnecessarily repetitive information. To set it differently, instead of using all the available indicators, the use of a set of comprehensive indicators is more manageable and handy from the perspective of policy makers and urban managers (Fraser et al., 2006; Liu et al., 2020; Newman, 1999). Therefore, the aim should be to minimize the overlapping information among individual indicators through constructing a comprehensive set of indices that represent the quality of urban liveability in regions and cities. To this aim, principal component analysis (PCA) provides the means to transform the correlated indicators into a new set of uncorrelated variables under the condition of maintaining the main information of the given sample. PCA quantifies how each indicator relates to the others in a given factorial space -where rotation of factors is performed (i.e. orthogonal vs. oblique rotations), which provides an understanding of whether different indicators correspond to the same factor; being therefore correlated. This allows us to identify the set of comprehensive indicators that better help explain differences in quality of living. PCA was applied to the 109 indicators which represent the total number of sub-indicators that are available at the regional level for the EU member states.

The principal components (PCs) were extracted according to the eigenvalue criterion where usually eigenvalues greater than 1.0 and in other cases greater than 0.5 are explanatory. The proportion of variance among the original variables that is explained by the corresponding PC was calculated next. Here, the values close to 1 indicated that common factors well represent the main information of the original indicators whereas values close to 0 indicated the opposite. This is followed by obtaining the PC matrix, also known as the PC loadings that showed weight of each PC in relation to each original indicator. The final matrix was rotated using the varimax criterion[2] to obtain loadings that would assist in calculating the final values for each PC. Regarding the PC loadings, PC score coefficients were examined and only one indicator associated with the highest score coefficient was selected from each PC to represent the other correlated indicators corresponding to the same PC loading. Therefore, we reduced the number of indicators to 30 and constructed a set of comprehensive indicators to analyse the territorial quality of living in Europe.

3.5 Weighting methods used in the study

As the literature suggests, there is no unique method for weighting the indicators. In fact, it is asserted by European Commission (EC) that "*no uniformly agreed methodology exists to weight individual indicators before aggregating them into a composite indicator*" (EC, 2021). For comparative purposes, the current study focuses on five different weighting methods which are: (1) entropy-weight method, (2) PCA, (3) random forest (RF) classification, (4) regression analysis, and (5) BOD.

3.5.1 Entropy-weight method

The entropy-weight method, which was initially developed by Shannon and Weaver (1947), is based on measuring the uncertainty in information, which relates to probability theory. Originally, the concept of entropy was measured by 'H', which satisfies some properties for all p_i within an estimated joint probability distribution P. It was shown that the only function which satisfies these properties is $H = -\sum_i^n p_i log(p_i)$(Shannon & Weaver, 1947). If there is large difference between the alternatives, the criterion will give large amount of information and is therefore considered as an important factor. Nevertheless, it can be

[2] Varimax searches for a rotation (a linear combination) of the original factors such that the variance of the loadings is maximized where we maximize the equation:$v = \sum \left(q_{j,l}^2 - \overline{q}_{j,l}^2\right)^2$ with $q_{j,l}^2$ being the squared loading of the jth variable on the l factor and $\overline{q}_{j,l}^2$ is the mean of the squared loadings. Following a varimax rotation, each original variable tends to be associated with one (or a small number) of factors, and each factor represents only a small number of variables.

applied for the assessment of significance of the factors in the multicriteria analysis (Delgado-Villanueva & Romero Gil, 2016). Assuming that there are m objects for evaluation, each having n evaluation criteria, the decision matrix can be formed as: $X=\{z_{ij};\ i=1,\ldots,m;\ j=1,\ldots,n\}$. The decision matrix '$X$' is normalized for each criterion C_j where P_{ij} represents the normalized values.

$$P_{ij} = \frac{z_{ij}}{\sum_{i=1}^{m} z_{ij}} \tag{3}$$

The entropy E_j of each criterion C_j is calculated as given in Eq. (4)

$$E_j = -g \sum_{i=1}^{m} P_{ij} ln(P_{ij}) \tag{4}$$

where g is a constant, i.e. $g=(\ln(m))^{-1}$. Given E_j, the degree of divergence D_j of the intrinsic information in each criterion C_j is calculated in Eq. (5). Finally, the entropy weight w_j of each criterion C_j is calculated by Eq. (6).

$$D_j = 1 - E_j \tag{5}$$

$$w_j = \frac{D_j}{\sum_{j=1}^{n} D_j} \tag{6}$$

3.5.2 PCA method

Intending to compute the weights of the indicators, we applied the PCA to the selected comprehensive indicator set having a total of 30 indicators, which were explained in Sect. 3.2. For the selection of factors, we followed the standard principles explained in Nicoletti et al. (1999), which include: (i) the associated eigenvalues need to be greater than one; (ii) the factor should contribute individually to the explanation of overall variance by more than 10%; (iii) the factor contributes cumulatively to the overall variance by more than 60%. In our case, the factors with eigenvalues larger than 1 are the first eight, some of which explained individually more than 10% of the total variance while some others are less than 10%, and overall they accounted for more than 65% of the variance. The weights were constructed from the matrix of factor loadings following the rotation (e.g. varimax). In the subject matrix, the square of factor loadings represents the proportion of the total unit variance of the indicator which is explained by the factor. This is aligned with the approach introduced by Nicoletti et al. (1999) in that the subject study is based on grouping the sub-indicators with the highest factor loadings in intermediate composite indicators. In our case, there are eight intermediate composite indicators.[3] After determination of the weights of all indicators, these were standardized so that all the weights sum up to 1.

[3] As an example, the first composite includes X1 (weight: 0.162), X23 (weight: 0.093), X29 (weight: 0.119), X42 (weight: 0.067), X43 (weight: 0.107). The second is formed by X71 (weight: 0.096), X76 (weight: 0.064), X87 (weight: 0.172), X90 (weight: 0.148), X94 (weight: 0.126), and the third only by X14 (weight: 0.096) (see Table 7 in Appendix for the indicator codes and their explanation).

3.5.3 Machine learning approaches (random forest)

Random forests (RF) (Breiman, 2001) are examples of nonparametric techniques consisting of a combination of tree predictors where each tree is generated from different subsets from the training dataset using a random vector sampled independently from the input vector. Bagging is a method to generate a training data set by randomly drawing data from the original dataset or a randomly selected part of the training set be used for the construction of individual trees for each feature combination (Breiman, 1996). In the RF applications, training set generally consists about 70% of the data from the original dataset, leaving about one-third of the data out from every tree grown. The latter is the so-called 'out-of-bag' (OOB) (out of the bootstrap sampling) data. The number of variables used (n) at each node to generate a tree and the number of trees to be grown (k) are user-defined parameters considered in RF classifications. At each node of the tree, a subset of randomly chosen parameters is selected and the predictor variable resulting in the best split is used to do a binary split on that node (Liaw & Wiener, 2002). Each tree gives a classification which is validated using the OOB data and the RF predictor is formed by averaging of generalization error over k trees. In comparison to parametric techniques, this method provides higher prediction accuracy and provides information about the underlying modelling mechanism as well as reporting the importance of the predictor variables.

Gini Index (Breiman et al., 1984) is one of the commonly used variable selection measures in tree induction, which measures impurity of a variable with respect to output. Gini importance score provides a relative ranking of the features, which is a by-product in the training of the random forest classifier: at each node τ within the binary trees T of the random forest, the optimal split is sought using the Gini impurity $i(\tau)$ measuring how well a potential split is separating the samples of the two classes in this particular node (Breiman, 2001).[4] In fact, Gini importance can be used as a general indicator of feature relevance (Menze et al., 2009). The RF classification approach applied in this study used the Gini Index as a proxy for measuring the importance of each indicator, which were used as weights. These weights were normalized to sum up to one prior to the application to the subject indicators.

3.5.4 Regression approach

Linear regression models explicate the relationship between a large number of indicators X_{1i}, X_{2i},…,X_{ni} and an output measure Y_i, being the objective to be attained. A multiple (linear) regression model is estimated to find out the relative weights of sub-indicators, which is given as $\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_n X_{ni}$ where $1 = 1, \ldots, M$ where Y_i indicates the objective that sub-indicators aim to measure, α is the constant, and β_1,…, β_n are the regression coefficients (weights) of the associated sub-indicators represented by X_{1i},…,X_{ni}. The equation was estimated with OLS technique using our comprehensive set of indicators being independent variables; and EU territorial quality index constructed in the EU ESPON, 2020 Cooperation

4 Gini impurity $i(\tau)$ can be calculated as: $i(\tau) = 1 - \rho_1^2 - \rho_0^2$ where $\rho_k = \frac{n_k}{n}$ being the fraction of the n_k samples from the class $k = \{0, 1\}$ out of the total of n samples at node τ. Its increase Δi that results from splitting and sending the samples to two sub-nodes τ_l and τ_r by a threshold t_0 on variable θ is defined as: $\Delta i(\tau) = i(\tau) - \rho_l i(\tau_l) - \rho_r i(\tau_r)$. Concerning all variables θ available at the node and all possible thresholds t_0, the pair $\{\theta, t_0\}$ leading to a maximal Δi is determined. The decrease in Gini impurity resulting from this optimal split $\Delta i_0(\tau, T)$ is recorded and accumulated for all nodes τ in all trees T in the forest, individually for all variables θ: $I_G(\theta) = \sum_T \sum_\tau \Delta i_\theta(\tau, T)$. The resulting Gini importance index (I_G) indicates how often a particular feature θ was selected for a split and how large its overall discriminative value was for the classification problem (see Menze et al., 2009 for the formulas).

Programme (see ESPON, 2020) was considered as the dependent variable. The coefficient estimated from the regression model quantifies the relative effect of each indicator on the measured output; therefore, we used the estimates of regression coefficients as weights in order to construct the composite indicator at the regional level in Europe.

3.5.5 Benefit-of-the-doubt

The data envelopment analysis (DEA) centres on measuring relative efficiency of homogeneous set of decision making units (DMUs) through applying linear programming tools (Charnes et al., 1978). Benefit-of-the-doubt (BOD) approach, originally developed for the macroeconomic performance assessment and recently adopted to the index theory, is an application of the DEA (Cherchye & Kuosmanen, 2002). Following Melyn and Moesen (1991) and Cherchye et al. (2007), the BOD model has been increasingly used for constructing composite indicators (CIs). The composite indicator is defined as the ratio of a country's (or region's) performance over its benchmark performance:

$$\mathrm{CI}_i = \frac{\sum_{i=1}^{M} I_{qi} w_{qi}}{\sum_{i=1}^{M} I_{qi}^{*} w_{qi}} \tag{7}$$

where I_{qi} is the normalized score of q$^{\text{th}}$ sub-indicator ($q = 1,\ldots,Q$) for country i and w_{qi} is the corresponding weight. Cherchye et al. (2004) suggested the method of solving a maximization problem for obtaining the benchmark:

$$I^{*} = I^{*}(w) = \underset{I_k, k \in \{1,\ldots,M\}}{argmax} \left(\sum_{q=1}^{Q} I_{qk} w_q \right) \tag{8}$$

I^* is the score of the hypothetical country that maximizes the overall performance given a set of weights w. The next step is the determination of optimum set of weights for each country, which is obtained by solving the following problem in Eq. (9) subject to non-negativity constraints on weights.

$$\mathrm{CI}_i^{*} = \underset{w_{qi}, q=1,\ldots,Q}{argmax} \frac{\sum_{q=1}^{Q} I_{qi} w_{qi}}{\underset{I_k, k \in \{1,\ldots,M\}}{max} \left(\sum_{q=1}^{Q} I_{qk} w_{qi} \right)} \quad \text{where } i = 1, \ldots, M \tag{9}$$

The resulting composite index takes values between 0 and 1 where 0 is the lowest possible performance and 1 is the benchmark value. Equation (9) can be reduced to a linear programming problem through multiplying all weights by a common factor and solving the following optimization problem.

$$\begin{aligned}
&\mathrm{CI}_i^* = \underset{wqi}{argmax} \sum_{q=1}^{Q} I_{qi} w_{qi} \\
&\text{st} \\
&\sum\nolimits_{q=1}^{Q} I_{qk} w_{qk} \le 1 \\
&w_{qk} \ge 0 \\
&\forall k = 1, \dots, M; \forall q = 1, \dots, Q
\end{aligned} \tag{10}$$

Regarding the BOD model, initially a 5% restriction was imposed to the lower weight bound value so that each sub-indicator should have a relative contribution of at least 5% to the CI value. The reason for imposing such a restriction is that CIs cannot be constructed by excluding some of its constituent sub-indicators through assigning a corresponding zero weight value (Rogge, 2018). In fact, all dimensions contributing to the CI value should be considered in the analysis aiming to provide at least some valuable information.

3.6 Aggregation of indicators

To calculate a region's quality of life index, we multiply our comprehensive set of indicators (of 30) with the corresponding weights obtained from the first four approaches (entropy-weight, PCA, RF and regression model) and sum them up separately to represent the impacts of weights obtained from four different methods on the Territorial Quality of Life Index (TQLI) for each of the NUTS2 regions (Eq. 11).

$$\begin{aligned}
&\mathrm{TQLI}_i = \sum_{q=1}^{Q} CI_{qi} w_q \ \text{ where } q = 1, \dots, Q;\ i = 1, \dots, M \\
&\text{with } \sum\nolimits_q w_q = 1;\ 0 \le w_q \le 1
\end{aligned} \tag{11}$$

In Eq. (11), CI_{qi} is the qth composite indicator assigned to region i that was normalized and selected by the PCA model, and w_q is the corresponding weight of the qth indicator. BOD model assigns virtual weights in the process of computing CIs, and the subject method provides TQLIs computed as the model outcome. Therefore, Eq. (11) does not apply to BOD model and we present the estimations from the subject model directly provided as TQLIs.

3.7 Sensitivity analysis

For conducting sensitivity analysis, the model can be written as follows:

$$y = f(x_1, x_2, \dots, x_n) \tag{12}$$

where the x_i are the input variables and y is named as the model response. Through applying sensitivity analysis, we aim to quantify the influence of input factors on the model outcome. The application of sensitivity analysis techniques enables the exploration of the entire interval for each input variable and is not based on any assumption on the model's structure (e.g. linearity or additivity). As highlighted by Chen et al. (2010), the application of sensitivity analysis increases the compatibility of results. In the case of our model, we consider the influence of change in indicator weights on the model outcome. As in all sensitivity analysis methods, we will examine the relative influence of each x_i on y through searching the value of $\partial y/\partial x_i$, which is known as the *i*-th elemental effect (EE_i) employing an one factor at a time (OAT) calculation scheme which fixes at each step all input variables except the one the effect of which is being calculated.

In the current study, we conducted a sensitivity analysis for different ranges to all criteria, which are the composite indicators that were computed through the application of alternative weighting methods. Following Romano et al. (2015) and Ustaoglu and Aydınoglu (2020), sensitivity testing was applied to the highest weights which were computed using entropy-weight, PCA, RF, and regression analysis approaches. Regarding the BOD method, the weights of each criterion were computed virtually in the model; therefore it is not possible to change the weights manually to search for its impact on the model outcomes. In the BOD model, initially a 5% restriction was imposed to the lower weight bound value. For the purpose of sensitivity analysis, the model outcomes corresponding to 5% restriction were compared to the outcomes where a 2.5% restriction was applied to the lower weight bound value.

Regarding the weights obtained from entropy-weight, PCA, RF and regression analysis, a series of sensitivity analysis was performed where a quarter percent changed each criterion weight (i.e. the highest weight) (e.g. $\pm 25, \pm 50, \pm 75\%$) (see Chen et al., 2010; Perpina et al., 2013). Changing the value of x_i in Eq. (12), it should be noted that the weights of the remaining variables were adjusted proportionally so that it satisfies the additivity constraint where all weights sum up to one. The summary of the sensitivity analysis applied in this study is given in Table 2. The results on sensitivity analysis will be elaborated later on in Sect. 4.3.

4 Results

4.1 Selection of the indicators from PCA

Table 3 shows the individual variance explained from each PC and the indicators selected from each PC based on their high score coefficients. Because thirty PC were retained, the total number of selected indicators from each PC sums up to thirty. These are the highlighted indicators in Table 3 with bold colour; also including those given in the table notes. From Table 3, we note that the first eighteen PCs account to more than 80% of data variance: 26.6, 10.8 and 8.4% for the first, second and third factors, respectively. Correlations between indicators and PCs indicate that the first factor is mainly correlated to governance indicators, while the second and the third factors are correlated to economic development and health & work indicators, respectively. PCs correlations to other factors can be seen in Table 3.

4.2 Main results from TQLIs based on different weighting methods

The weights that were computed based on the application of four different methods (entropy-weight, PCA, RF, regression analysis) are presented in Fig. 3. In the figure, the most notable result is that the two highest weights associated with X23 (higher education attainment rate) and X41 (life expectancy) indicators were obtained from the regression

Table 2 Criteria for sensitivity analysis

Weights of the sub-indicators subject to sensitivity testing	Method used for the weight calculation	Sensitivity Test
Development of urban use per capita (X79)	Entropy-weight	±25%, ±50%, ±75% change in the value of the highest weights
Emissions of CO2 per capita (X58)	PCA	±25%, ±50%, ±75% change in the value of the highest weights
Control of corruption (X1)	Random forest	±25%, ±50%, ±75% change in the value of the highest weights
Life expectancy (X41)	Regression analysis	±25%, ±50%, ±75% change in the value of the highest weights
NA	BOD	5% and 2.5% restrictions were imposed to lower weight bound value

analysis approach. Other methods, i.e. RF, entropy-weight and PCA, have also provided considerably high weights regarding X1 (control of corruption), X41 (life expectancy) and X58 (CO2 emissions per capita) indicators, respectively. The weights of the remaining indicators are relatively smaller where the regression analysis method produced the smallest weights particularly for the X67 (supply–demand balance from water retention index), X69 (fresh water consumption), X85 (total employment in CBM sectors) and X25 (early school leavers) indicators. Based on entropy-weight and RF approaches, the influence of X14 (crime rates), X25 (early school leavers), and X103 (change in a number of days with snow cover) are fairly equal, whereas it is the case for X87 (poor access to primary schools), X90 (ratio of closest doctors), and X95 (poor accessibility to shops) indicators when PCA and regression analysis methods were used. The top five indicators that retained the highest weighting based on the weighting method used in the analysis are summarized in Table 4. We note from the table that the highly weighted sub-indicators are mainly associated with environment, governance, health, education and work classifications. It is clear from these findings that the use of different methods for the calculation of weights resulted in significant differences in the distribution of weights associated with each of the sub-indicator specified in the study.

Figure 4 presents the values of TQLIs resulting from the use of five different methods where the warm colours show the lower index values with cold colours indicating the highest values. From the figure, the distribution of index values across the EU countries seems to be similar when entropy-weight and PCA were used as the indicator weighting schemes in the construction of composite indicators. There are also similarities in the index values that were computed from RF classification and regression analysis methods. Although there are similarities of the index values computed for particular NUTS2 regions, the distribution of index values is considerably different in the case of BOD in comparison with index values developed from other weighting methods. Further differences in the distribution of index values can be seen in Fig. 5, which presents the frequency distribution of TQLIs based on the use of alternative weighting methods. Overall, though there are similarities, the distribution and ranking of index values vary significantly for each of the alternative weighting method used in the analysis.

It appeared from the index values computed from entropy-weight, PCA, RF and regression approaches that Sweden and Finland are the two countries reported the highest value

Table 3 Variance explained by principal components and selected indicators

PC number	PC name	PC loadings > 1.0 Original variables	Code	Variance explained (%)
1	Governance	Control of Corruption	X1	26.6
		Government Effectiveness	X2	
		Regulatory Quality	X4	
		Rule of Law	X5	
		Voice and Accountability	X6	
		Public service quality	X8	
		Impartiality	X9	
		Corruption in public services	X10	
		Trust in the legal system	X12	
		Not employed persons (% pop15_29)	X30	
		Internet at home	X31	
2	Economic Development	Road accident fatalities	X40	10.8
		Disposable income of households	X44	
		Total employment (btw 15–64)	X46	
		Unemployment	X47	
		Employment in high-tech sectors	X49	
		Employment in Science & Technology	X50	
		Total waste production (metric tones)	X60	
		Total turnover generated by material providers' activities (million euro)	X82	
		Total employment in technology providers sectors	X83	
		Total turnover generated by technology providers sectors (million euro)	X84	
3	Health & Work	Lower secondary education completion (%pop25-64)	X24	8.4
		Involuntary part-time/temporary employment	X27	
		Hearth diseases death rate	X36	
		Premature mortality (< 65)	X39	
		Life expectancy	X41	
		MEAN_PM2.5 emissions	X56	
		MEAN_PM10 emissions	X57	
4	Access to Services	Poor access to primary schools	X87	8.1
		Poor access to secondary schools	X88	
		Poor access to hospitals	X89	
		Poor access to pharmacies	X91	
		Poor access to bank offices	X92	
		Poor access to train station	X93	
		Poor access to shops	X96	
		Poor access to regional centres	X97	
5	Environmental Quality	Coverage percentage of GI within a region (%)	X66	5.7
		Share of agricultural area in protected areas	X75	
		Percentage share of agricultural use areas	X78	
		Employment in material providers for the circular economy	X81	
		Change in annual mean number of days with snow cover	X103	

Table 3 (continued)

PC number	PC name	PC loadings > 1.0 Original variables	Code	Variance explained (%)
6	Soil Protection	MEAN_Ozone3 emissions	X55	3.7
		Soil retention	X71	
		Soil sealing in a region	X74	
		Percentage share of agricultural use areas	X78	
		Annual amount of raw material extracted from natural environment (1000 tones)	X80	
7	Quality and Affordability of Housing	Burdensome cost of housing	X18	3.2
		Lack of adequate heating	X22	
		Early school leavers	X25	
		MEAN_PM10 emissions	X57	
		Uncollected sewage	X62	
		Development of urban use per capita 2000–2018	X79	
8	Green Land	Green infrastructure initiatives	X7	2.8
		Higher education attainment rate (%) pop_25_64	X23	
		Cancer death rate	X35	
		EU trademark applications per million inhabitant	X52	
		Forest area within a region between 2012_2018	X72	
9	Security of Citizens	Political Stability and Absence of Violence/Terrorism	X3	2.1
		Assaulted/Mugged (%)	X17	
		Higher education attainment rate (%) pop_25_64	X23	
		Early school leavers (%)	X25	
10	Quality of Public Services	Internet access (%)	X34	2
		Unmet medical needs (%)	X42	
		CO_2 emissions per capita	X58	
		Sewage treatment	X63	
12	Public Health	Cancer death rate	X35	1.5
		EU trademark applications per million inhabitant	X52	
		Uncollected sewage	X62	
13	Education & Environment	Early school leavers	X25	1.4
		CO_2 emissions per capita	X58	
		Electricity generated by renewable energy sources (PJ)	X59	
		Area of NATURA2000 areas relative to NUTS2 region area	X73	
14	Quality of Communities	Crime rates	X14	1.3
		Higher education attainment rate (%) pop_25_64	X23	
		Community design applications per million inhabitants	X53	
		Area of NATURA2000 areas relative to NUTS2 region area	X73	
		Crime rates	X14	1.3
		Higher education attainment rate (%) pop_25_64	X23	
		Community design applications per million inhabitants	X53	
		Area of NATURA2000 areas relative to NUTS2 region area	X73	
15	Security of Urban Areas	Money stolen	X16	1.1
		Lifelong learning	X29	
		Cancer death rate	X35	
		Number of vulnerability causes	X65	

Table 3 (continued)

PC number	PC name	PC loadings > 1.0 Original variables	Code	Variance explained (%)
16	Environmental Sustainability	Number of hazards	X64	1
		Total employment in Circular Business Models (CBM) sectors	X85	
		Total turnover generated by CBM sectors	X86	
		Poor accessibility to shops	X95	
17	Social Well-being	Housing quality	X19	1
		Suicide death rate	X37	
18	Basic Human Needs	Assaulted/Mugged	X17	0.9
		Unmet medical needs	X42	
		Insufficient food	X43	
Total variance explained		0.83		

Other than these, the selected indicators from PC loadings > 0.5 are: PC-19: Share of urban area in protected areas (X76); PC-20: Freshwater consumption (Lt/day/capita) (X69); PC-21: Supply–demand balance quantified by Water Retention Index (X67); PC-22: Early school leavers (X25); PC-23: Electricity generated by renewable energy sources (PJ) (X59); PC-24: Area of NATURA2000 land relative to NUTS2 region area (X73); PC-25: Crime rates (X14); PC-26: Municipal solid waste recycling rate (X61); PC-27: Total employment in CBM sectors (X85); PC-28: Soil retention (X71); PC-29: Ratio of closest doctors (X90); PC-30: Poor access to urban morphological zones (jobs) (X94)

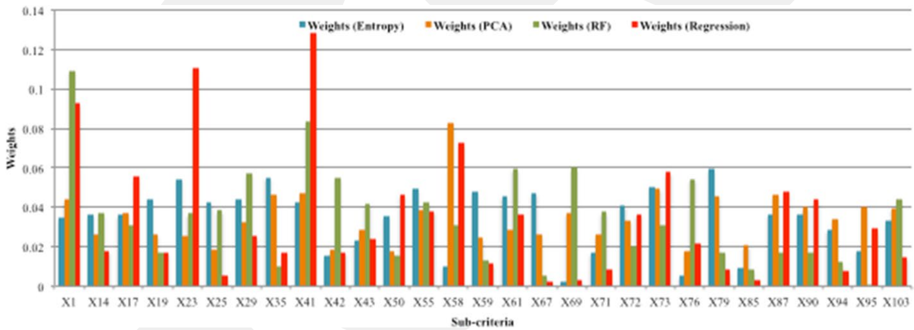

Fig. 3 Weights obtained from entropy-weight, PCA, RF and regression methods

of territorial quality of life in Europe. This is also true for some other NUTS2 regions that are located mostly in western EU countries but they are small in number. In general, these cover capital regions and regions comprising large cities indicating that high quality of life indices were observed in Paris, Brussels but also in regions like Warsaw in Poland, Lisbon and Porto in Portugal, Wien in Austria, Prague in Czech Republic and others. The lowest values were reported for the regions in Eastern Europe, particularly for Bulgaria and Romania. According to BOD method, the regions in Sweden, Denmark, Netherlands, Belgium, western Germany, western France and northern Spain account for the highest values of territorial quality of life whereas Eastern Europe and some Baltic countries (i.e. Lithuania and Latvia) reported the lowest values. This also includes some wider parts of

Table 4 The top weighted (sub) indicators across four different weighting methods

Weighting method	Main indicator	Top weighted sub-indicators	Weights
Entropy-weight	Environment (Land)	1) X79 (Development of urban use per capita)	0.059
	Health	2) X35 (Cancer diseases death rate)	0.055
	Education & Work	3) X23 (Higher education attainment rate)	0.054
	Environment (Green Infrastructure)	4) X73 (Area of NATURA2000 areas)	0.05
	Environment (Air quality)	5) X55 (Mean Ozone3 concentration)	0.049
PCA	Environment (Air quality)	1) X58 (CO_2 emissions per capita)	0.083
	Environment (Green Infrastructure)	2) X73 (Area of NATURA2000 areas)	0.049
	Health	3) X35 (Cancer diseases death rate)	0.047
	Health	4) X41 (Life expectancy)	0.047
	Environment (Land)	5) X79 (Development of urban use per capita)	0.046
RF	Governance	1) X1 (Control of corruption)	0.109
	Health	2) X41 (Life expectancy)	0.083
	Environment (Waste)	3) X61 (Municipal solid waste recycling rate)	0.06
	Environment (Water)	4) X69 (Freshwater consumption)	0.06
	Education & Work	5) X29 (Lifelong learning)	0.057
Regression	Health	1) X41 (Life expectancy)	0.128
	Education & Work	2) X23 (Higher education attainment rate)	0.111
	Governance	3) X1 (Control of corruption)	0.093
	Environment (Air quality)	4) X58 (CO_2 emissions per capita)	0.072
	Environment (Green Infrastructure)	5) X73 (Area of NATURA2000 areas)	0.058

Mediterranean countries where southern Italy and Greece are such examples that are associated with low performance values in terms of territorial quality of living.

Table 5 presents the descriptive statistics of TQLIs with respect to the five different weighting methods used in the analysis. As shown in the table, territorial quality of life in European regions has a mean value ranging from 0.499 ± 0.064 SD to 0.622 ± 0.187 SD; the former represents the value from entropy-weight method, while the latter is the value from the BOD. The min and max index values range between 0.06 and 0.96, both reported from the BOD method. The intervals of min and max values are quite similar for the regression and RF classification approaches. These intervals are quite different regarding entropy-weight and PCA methods, the former ranges between 0.29 and 0.66 and the latter between 0.41 and 0.71. Examining the correlations among territorial quality of life indices developed from different weighting methods (Table 6) indicated that TQLIs from entropy-weight method have a high and significant correlation coefficient with those developed from the PCA. The TQLIs from RF classification method also correlate well with the indices obtained from regression analysis approach, and have the lowest correlation with those of the BOD method. We note that TQLIs from BOD method in general have the lowest correlations with the rest of the indices developed from entropy-weight, PCA, RF and regression approaches. This is mainly due to the fact that different from other methods, the indices constructed by BOD are ratio based where the relative performance of a particular

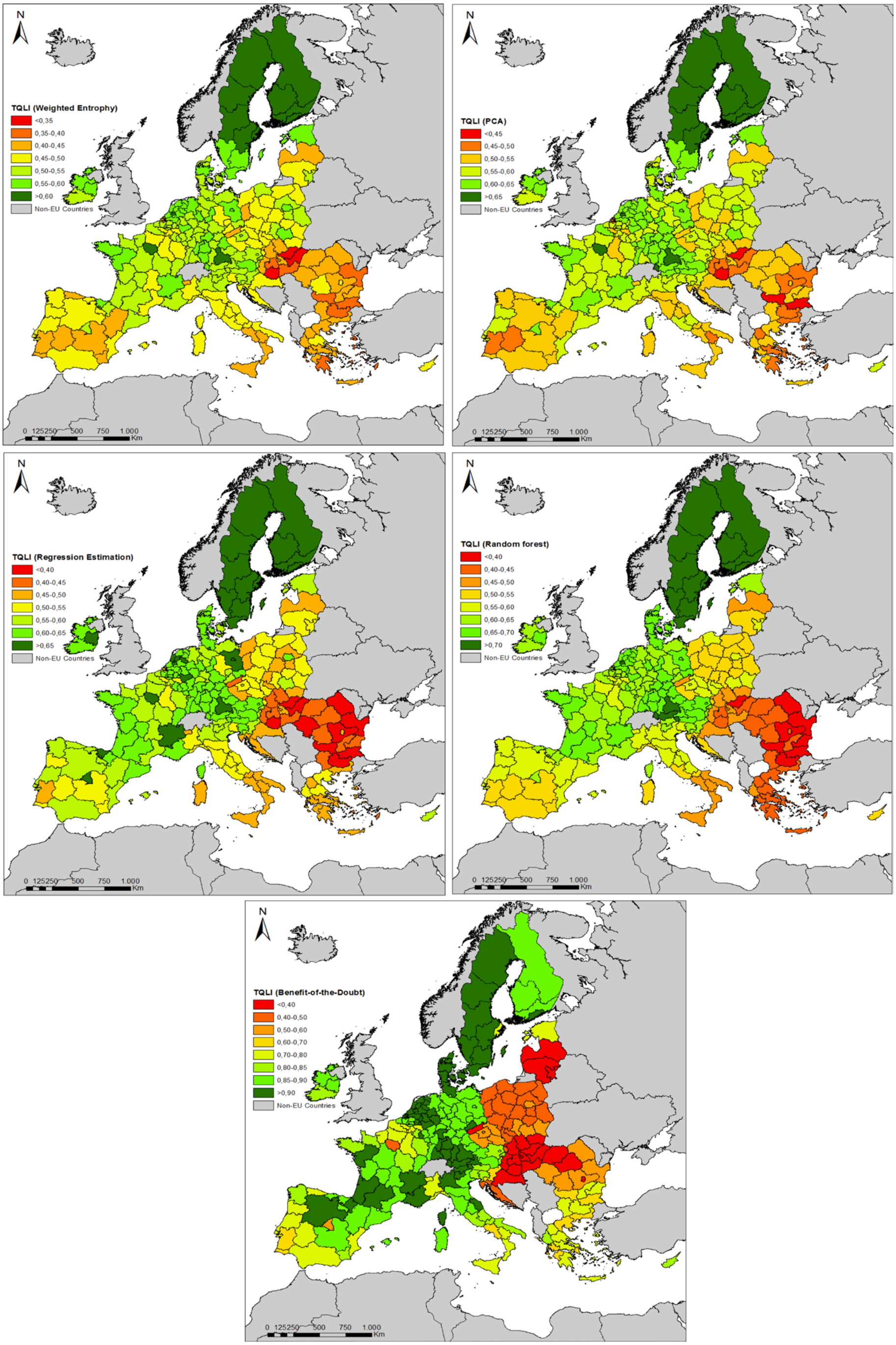

Fig. 4 Spatial distribution of the TQLI based on the results of different weighting methods applied to EU countries

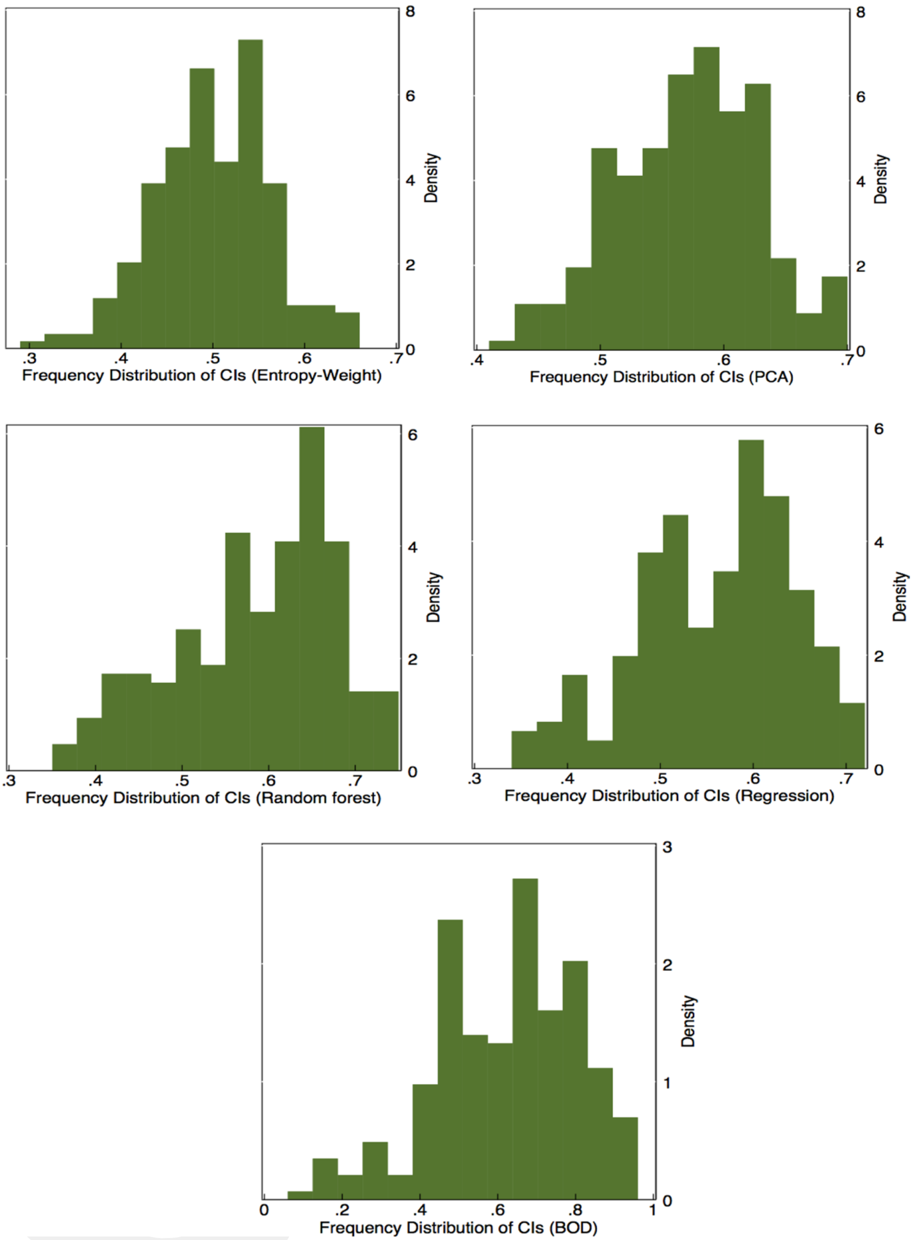

Fig. 5 Frequency distribution of Composite Indicators (CIs) (i.e. TQLIs) based on alternative weighting methods

unit is compared with a benchmark performance value, which is acquired endogenously through linear optimization. Here the benchmark is based on the maximum weighted sum of a decision unit which applies the same weight as the one to be assessed in the analysis. Therefore, the benchmark can be unit dependent if no decision unit has the highest score

Table 5 Descriptive statistics of the territorial quality of life in EU countries based on different weighting methods

Method	Mean	SD	Min	Max
TQLI (entropy-weight)	0.499	0.064	0.29	0.66
TQLI (PCA)	0.569	0.055	0.41	0.71
TQLI (regression)	0.561	0.083	0.34	0.72
TQLI (RF)	0.584	0.092	0.35	0.75
TQLI (BOD)	0.622	0.187	0.06	0.96

Table 6 Correlation among indices of territorial quality of life from different weighting methods

	TQLI (entropy-weight)	TQLI (PCA)	TQLI (RF)	TQLI (regression)	TQLI (BOD)
TQLI (entropy-weight)	1				
TQLI (PCA)	0.968*	1			
TQLI (RF)	0.902*	0.908*	1		
TQLI (regression)	0.917*	0.921*	0.942*	1	
TQLI (BOD)	0.547*	0.556*	0.628*	0.541*	1

TQLI Territorial Quality of Life Indicators; *significance at 10% level

in all indicators. This implies that there is no unique benchmark applicable in the subject method for the decision units considered in the study.

4.3 Findings from sensitivity analysis

Figure 6 presents the one factor at a time (OAT) sensitivity analysis applied to the weights of the sub-indicators where the highest weights were changed $\pm 25, \pm 50$, and $\pm 75\%$ (Table 2) and the remaining weights were adjusted accordingly so that the weights sum up to one. The results in Fig. 6 represent each of the four weighting schemes used in the analysis. The percentage change in the composite indicator value when there are quarter percent changes in the value of weights varied between -0.28 and 0.32 in entropy-weight method; -0.75 and 1.50 in PCA; -0.50 and 1.20 in RF; and -0.50 and 0.75 in regression analysis approaches. Therefore, entropy-weight method showed the least sensitivity of the changes in the values of composite indicators to the changes in value of weights, which is followed by regression and RF methods, respectively. The highest sensitivities were observed in the model which applied the PCA method for the calculation of weights.

Because OAT sensitivity analysis is not applicable to the BOD method, the BOD model outcomes corresponding to 5% restriction were compared to the outcomes where a 2.5% restriction was applied to the lower weight bound value. The resulting percentage changes in the value of composite indicators are presented in Fig. 7. From the figure, the composite indicator values are highly sensitive to the changes in the lower weight bound value. More than 100% changes in the value of composite indicators were observed for the NUTS2 regions located in Greece, Germany, Bulgaria, Italy and Netherlands. The smallest changes were observed for the NUTS2 regions in Poland, France, Finland, Belgium and Austria. The changes reported for other regions range between -75% and $+75\%$.

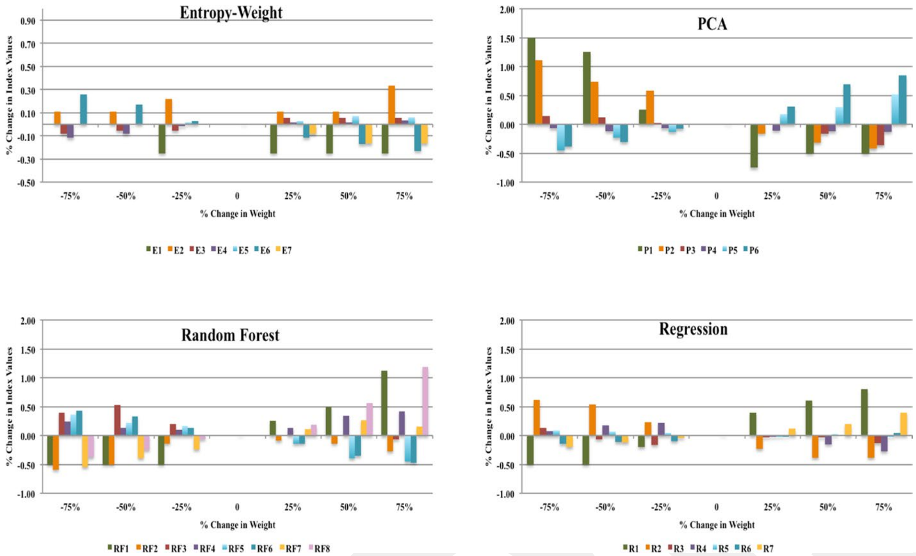

Fig. 6 Summary of the results of sensitivity analysis applied to weights obtained from: **a** entropy-weight method, **b** PCA, **c** random forest, **d** regression analysis. *Note* The figures demonstrate the changes in Composite Index values when there are ± 25%, ± 50%, ± 75% changes in the corresponding weights having the maximum value. In the entropy-weight, E1: < 0.35 E2: 0.35–0.40 E3: 0.40–0.45

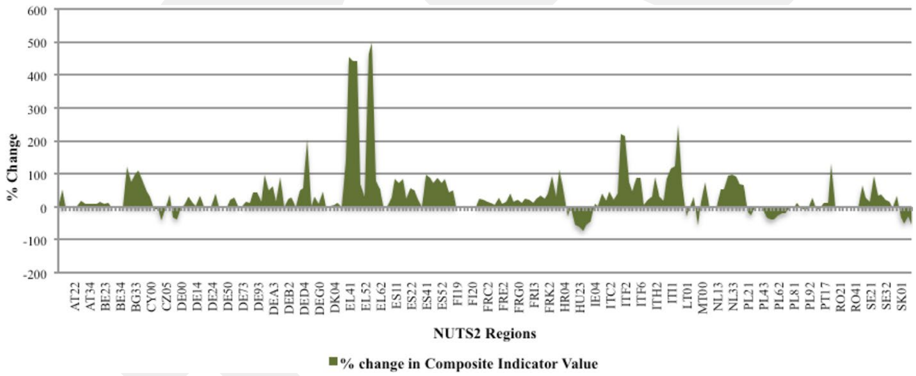

Fig. 7 Results of sensitivity analysis applied to BOD model. *Note* The figure demonstrates % change in the value of composite indicators for each NUTS2 region following the sensitivity analysis applied to BOD model

5 Discussion

Although there has been rising awareness by government and public on urban quality of life particularly in the recent decades, quality of urban living is still understudied, especially at the NUTS regional level in the EU member states. Based on our constructed set of indicators aligned with the social, economic and environmental aspects of sustainability, this study aimed to fill the gap in the literature through searching the impact of using five different weighting methods in the construction of composite indicators on the ranking and value of Territorial Quality of Life (TQL) indices in Europe. The findings from this study not only adds further evidence to support policy making and strategy development regarding the NUTS2 regions in Europe but contribute to the literature from two perspectives. First, this study presents an enhanced methodological framework in the construction of composite indicators (Booysen, 2002; Nardo et al., 2005) that is based on data preparation, normalization, statistical transformation of raw data, reduction of indicators in order to ease public communication and data interpretation, weighting of the sub-indicators and their aggregation, and checking for robustness and sensitivity applied to the TQLIs at the regional level in Europe as specified in the study. This methodological framework can be applied in any study aimed to construct composite indicators to assess urban and regional quality of living (Mori & Yamashita, 2015; Pravitasari et al., 2018) or to research other sustainability-related topics such as social progress (Jitmaneeroj, 2017), competitiveness and innovativeness (Annoni & Dijkstra, 2013; Myszczyszyn et al., 2021), transportation (Reisi et al., 2014), agriculture (Xavier et al., 2018), ecosystems and environment (Smith et al., 2013) or energy (Mainali et al., 2014). Second, we showed that the values and rankings of composite indicators are dependent on how the sub-indicators are weighted through focusing on five different data-based weighting methods in the construction of composite indicators. We mainly considered assessing the data-based weighting methods in the study and excluded the expert-based approaches for two reasons: First, there is a high cost of organizing stakeholders that should evaluate the sub (indicators) at the pan-European level. Second, weighting methods based on expert views are more powerful at finer scales than coarser scales (Gan et al., 2017) given that public opinions on controversial issues are spatially auto correlated (Van de Kerk & Manuel, 2008). This implies that expert-based evaluations reflect the local conditions and they cannot be adopted to larger scales (Gan et al., 2017). In fact, statistical approaches are more effective in coarse spatial scales (Mayer, 2008) which is the case in our study which rely on political boundaries of the NUTS2 regions in Europe.

Through the application of sensitivity analysis, we analysed the robustness of five alternative weighting methods. Therefore, the results from sensitivity analysis can serve as an input to other studies in their selection of the most appropriate weighting method to be used for their construction of composite indicators. Sensitivity analysis showed that TQLIs are less sensitive to the changes in weights computed from entropy-weight method, which can be considered as a robust and flexible method in the construction of composite indicators. The sensitivities of TQLIs to the changes in weights are relatively higher in the PCA and RF classification methods in comparison to entropy-weight method, and we found modest sensitivities in the case of regression analysis approach. From the sensitivity analysis applied to BOD, we found very high sensitivities in the model outcomes particularly for some NUTS2 regions located in Greece, and there are also some others in Italy, Germany and Bulgaria accounting for 100–500% changes in the index values when there is 2.5% change in the restriction applied to the lower weight bound value. We conclude that the highest sensitivities were observed in the BOD method; and therefore there are increasing input uncertainties in the construction of composite indicators resulting from the BOD.

In the literature, there is no clear justification in the choice for one weighting method rather than another. That is because each method has its advantages and disadvantages. For instance, the main advantage of the entropy-weight method is that it calculates the weights avoiding the influence of subjective factors. But this way, it does not account the decision makers' preferences on the specified indicators. To encounter the preferences of decision makers into the analysis, the entropy method is now widely integrated to well-known multicriteria analysis approaches such as AHP (Zhao & Wang, 2019), Fuzzy Best–Worst

(Tian et al., 2018), VIKOR (Fei et al., 2019) and TOPSIS (Li et al., 2014) models. Regarding PCA, one advantage using the subject method is that the model does not introduce any manipulation of weights through ad-hoc restrictions (Chen et al., 2020; Nardo et al., 2005) and it reduces the risk of double weighting which may occur in the equal weighting method (Islam et al., 2020; Yeheyis et al., 2013). There are also a number of drawbacks with the use of the PCA model: First, the meaning of the factors which are extracted using the PCA may be difficult to interpret given that some unrelated indicators may be grouped into the same factor due to high correlations with the subject factor (Gan et al., 2017). Second, the method is sensitive to modifications of the original data which may change the set of weights used in the composite (Nardo et al., 2005). Further, the optimum number of dimensions can change if different methods for factor extraction and rotation are used (OECD, 2008). Another disadvantage of the method is that it is sensitive to the presence of outliers and small sample problems that may result in difficulties in the statistical identification and economic interpretation (Nardo et al., 2005). Regression analysis, on the other hand, is a useful technique that uncovers the relationship between a set of indicators and a single output. The method performs well even if the component indicators are not correlated and it is useful to update and validate the applied set of weights (Gan et al., 2017). One limitation of the method is the multicollinearity issue which is common in sustainability assessments. A possible solution to this problem is to associate the PCA with regression analysis (Nardo et al., 2005; Ul-Saufie et al., 2013). A further limitation is the need of large amount of data to produce estimates with specific statistical properties. With the use of the data, it is expected that selected indicators can properly explain the variation of the objective variable, which is usually not the case. It is therefore important to select an appropriate response variable which can reflect the main objective and be well explained by the selected indicators (Gan et al., 2017).

RF models show good practical performance, particularly in high-dimensional settings. In fact, the RF model can work well with large dimensional databases aimed at ranking predictors through retaining many benefits of decision trees but achieving better results through the use of bagging on samples. Through introducing randomness to the data at the modelling level, RFs have been shown to produce accurate and robust results. One further advantage of RFs is that the procedure automatically predicts the probability with which an item belongs to a certain class (Guns & Rousseau, 2014). This is not the case, for instance, for support vector machines and neural networks which require a time consuming cross-validation procedure to obtain probability estimates (Liaw & Wiener, 2002; Platt, 1999). Another advantage of RF is that it allows to model nonlinear relationships and the technique is robust against overfitting (Breiman, 2001). A disadvantage of applying the RFs is that there are growing computational intensity with the increase of calibration data and covariates as well as the high sensitivity of predictions to input data quality. Different from the explained four methods, BOD is an application of the DEA which optimizes the relative performance of a given set of indicators through integrating the processes of weighting, aggregation and index formation. Because weights are retrieved virtually in the model, this would help to identify trade-offs that aim to mitigate some of the difficulties arising from linear aggregation methods (Nardo et al., 2005). The weights are selected to maximize the index for each decision unit; therefore policy makers could not complain about unfair weighting. However, impossible cross-country comparisons, multiplicity of potential results, incomparability among the results, and lack of transparency in the weighting procedure are the major disadvantages of the subject method (Nardo et al., 2005; Yeheyis et al., 2013). From the sensitivity analysis, we found that entropy-weight method is the most robust one that has the least sensitivity to the changes in weight values. To our knowledge, there is no published case study focusing on the assessment of the different statistical weighting methods, therefore we are unable to compare our findings on the influence of weights on regions performances with those of international findings. However, this can be specified as a future research topic to provide cross-city/region or cross-country comparisons.

The results from the four different weighting methods (other than BOD) revealed that territorial quality of living in Europe had a strong positive relationship particularly with the environment, health, education and work, and governance dimensions. This is supported in many studies arguing that urban quality of life is a multidimensional index that represents the importance of each dimension of urban sustainability (Shen et al., 2011; Zhou et al., 2015; Zoeteman et al., 2016; Yang et al., 2020). Environment and health have been assigned a primary role given that land consumption, green infrastructure, air pollution, waste and water, and environmental health can have a significant role in the perceptions of territorial quality of life of the European citizens. Education, institutional trust and good governance are the other priorities given that high quality education and trust in governmental institutions are key factors for the quality of community life that can improve the quality of living in the EU member states. Which factor contributes most to the satisfaction of territorial quality of living tend to vary significantly based on the weighting method used in the analysis. From entropy-weight and PCA, environment is the most significant factor associated with the highest weight among other criteria, whereas it is the governance and health when RF and regression approaches are used. Therefore, it is recommended that one needs to be cautious in the selection of the weighting method for the construction of composite indicators and a sensitivity analysis is certainly required that will assist in the method decisions.

6 Conclusion

The NUTS2 regions of the EU member states were used as our case study. We assessed the territorial quality of life through focusing on sustainability indicators that were normalized and reduced, weighted and aggregated to obtain the final composite indicator values aimed at assessing the performance of the European regions. The results and rankings of composite indicators rely on how the subindicators are weighted. To this aim, this paper analysed the robustness of five alternative weighting methods (entropy-weight, PCA, RF, regression and BOD) and compared the results obtained for TQLIs related to each NUTS2 region. Our findings confirmed that the value and ranking of TQLIs differ significantly with the use of different weighting methods. Therefore, the end users need to understand how the indices were calculated and how the weighting methodology may influence their performance.

From sensitivity analysis, we found that entropy-weight method is the most robust one resulting in the lowest sensitivities to the changes in weights among others. By contrast, the uncertainties are the highest in the case of the BOD method. Because we did not cover the expert-based methods for the calculation of weights, these can be considered in the future research where our findings can be compared with those that will be obtained from expert-based approaches. However, it is important to mention that a future construction of composite indicators based on expert-based methods should consider NUTS2 regions in Europe as the unit of analysis for more accurate comparison. The methodological framework of the study can be adapted to other study areas or regions to construct quality of life composite indicators. The uncertainties in the construction of indicators are not only limited to the weighting methods. There are other methodological issues including the data (i.e. the indicators) included in the analysis, normalization and aggregation methods that influence the value and ranking of the composite indicators. However, these can be specified as future research topics.

The rankings of the TQLIs (Fig. 4) indicated that the regions in the Eastern Europe, particularly Romania and Bulgaria, have resulted in the lowest indicator values. Low performance values were also observed for Mediterranean regions including southern Italy and Greece, and southern parts of Spain and Portugal are also included. Compared to northern and western EU countries, the eastern countries and some regions of the southern countries have relatively less economic development; also, they are facing with problems regarding the natural environment, health, education, governance, and institutional quality among others.

These findings have several policy implications. First, to improve territorial quality of living in the poor performance regions, EU and government authorities should monitor the TQLIs and provide the means for regional and local authorities regarding public investments and new regulations to improve the quality of life. Special efforts need to be devoted to various kinds of environmental problems including air pollution, land management, green infrastructure, waste treatment and clean water provision along with public health, safety, education, institutional quality and governance. Second, private sector, citizens and other stakeholders should be included in policy co-creation processes during the planning and construction processes for high quality cities. It is expected that they contribute to the definition of quality of life as well as contribute to a better understanding of citizens' needs for provision of goods and services to improve quality of living in collaboration with regional and local authorities. Lastly, the poor performance regions identified in this study should be considered in the post-2020 EU Cohesion Policy Framework to support engaging investments for facilitating economic recovery as well as providing high-quality environment, convenient transportation, improved social services, good governance and institutional quality.

Appendix

See Table 7.

Table 7 Indicators and data sources

Main indicator	Sub-indicator name	Code	Scale	Source
Governance	Control of corruption	X1	Country	Kaufmann et al. (2010) (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130)
	Government effectiveness	X2	Country	
	Political stability and absence of violence/terrorism	X3	Country	
	Regularity quality	X4	Country	
	Rule of law	X5	Country	
	Voice and accountability	X6	Country	
	Green infrastructure initiatives	X7	Country	European Environment Agency
	Public service quality	X8	Country	Charron et al. (2019)
	Impartiality: all treated equally or some get special advantages in education, health care, law	X9	Country	
	Corruption in public service provision	X10	Country	
	Trust in the national government	X11	NUTS1, NUTS2	Gallup World Poll (WP139)
	Trust in the legal system	X12	NUTS1, NUTS2	Gallup World Poll (WP138)
	Trust in the police	X13	NUTS1, NUTS2	Gallup World Poll (WP112)
Safety	Crime	X14	NUTS0, NUTS1, NUTS2	EU-Statistics on income and living conditions
	Safety at night	X15	NUTS1, NUTS2	Gallup World Poll (WP113)
	Money stolen in the household	X16	NUTS1, NUTS2	Gallup World Poll (WP117)
	Assaulted/mugged	X17	NUTS1, NUTS2	Gallup World Poll (WP118)
Shelter	Burdensome cost of housing	X18	NUTS0, NUTS1, NUTS2	EU-Statistics on income and living conditions
	Housing quality	X19	NUTS0, NUTS1, NUTS2	
	Overcrowded housing	X20	NUTS0, NUTS1, NUTS2	
	Lack of adequate heating in the dwelling	X21	NUTS0, NUTS1, NUTS2	
	Lack of toilet in the dwelling	X22	NUTS0, NUTS1, NUTS2	

Table 7 (continued)

Main indicator	Sub-indicator name	Code	Scale	Source
Education & Work	Higher education attainment rate	X23	NUTS2	Eurostat regional education statistics
	Lower secondary education completion rate	X24	NUTS2	
	Early school leavers	X25	NUTS2	
	Job opportunities	X26	NUTS1, NUTS2	Gallup World Poll (WP89)
	Involuntary part-time/temporary employment	X27	NUTS2	Eurostat
	Young people not in education, employment or training	X28	NUTS2	
	Lifelong learning	X29	NUTS2	
	Not employed persons	X30	NUTS2	
Access to ICT	Internet at home	X31	NUTS1, NUTS2	
	Broadband at home	X32	NUTS1, NUTS2	
	Online interaction with public authorities	X33	NUTS1, NUTS2	
	Internet access	X34	NUTS1, NUTS2	Gallup World Poll (WP16056)
Health	Cancer diseases death rate	X35	NUTS2	Eurostat regional health statistics
	Hearth diseases death rate	X36	NUTS2	
	Suicide death rate	X37	NUTS2	
	Infant mortality rate	X38	NUTS2	
	Premature mortality rate	X39	NUTS2	
	Road accident fatalities	X40	NUTS2	
	Life expectancy	X41	NUTS2	
	Unmet medical needs	X42	NUTS0, NUTS1, NUTS2	EU-Statistics on income and living conditions
	Insufficient food	X43	NUTS0, NUTS1, NUTS2	

Table 7 (continued)

Main indicator	Sub-indicator name	Code	Scale	Source
Economy	Disposable income of households	X44	NUTS2	Eurostat
	GDP per capita	X45	NUTS2	
	Total employment	X46	NUTS2	
	Unemployment	X47	NUTS2	
	Share of full-time employment	X48	NUTS2	
	Employment in high-tech sectors	X49	NUTS2	
	Employment in science & technology	X50	NUTS2	
	EU patent applications	X51	NUTS2	
	EU trademark applications	X52	NUTS2	
	EU community design applications	X53	NUTS2	
Environment				
(a) Air quality	Average NO_2 concentration	X54	NUTS2	European Environment Agency
	Average Ozone3 concentration	X55	NUTS2	
	Average PM2.5 concentrations	X56	NUTS2	
	Average PM10 concentrations	X57	NUTS2	
	Emissions of CO_2 per capita	X58	NUTS2	ESPON TIA (https://database.espon.eu/project-archives/#/archives)
	Electricity produced by renewable energy sources	X59	NUTS2	Baranzelli et al. (2016)
(b) Waste	Total waste production	X60	NUTS2	ESPON CIRCTER Project (https://database.espon.eu/project-archives/#/archives)
	Municipal solid waste recycling rate	X61	NUTS2	European Environment Agency Country profiles on Solid Waste Management
	Uncollected sewage	X62	NUTS2	European Environment Agency (Waterbase_UWWTD_v7)
	Sewage treatment	X63	NUTS2	
(c) Hazards	Number of hazards	X64	NUTS2	European Environment Agency
	Number of vulnerability causes	X65	NUTS2	

Table 7 (continued)

Main indicator	Sub-indicator name	Code	Scale	Source
(d) Water	Supply–demand balance quantified by Water Retention Index	X67	NUTS2, NUTS3	ESPON GRETA Project (https://database.espon.eu/project-archives/#/archives)
	Satisfaction with water quality	X68	NUTS1, NUTS2	Gallup World Poll (WP95)
	Freshwater consumption per capita	X69	NUTS2	ESPON TIA (https://database.espon.eu/project-archives/#/archives)
(e) Soil	Capacity of ecosystems to avoid soil erosion	X70	NUTS2	
	Soil retention	X71	NUTS2	
(f) Green infrastructure	Coverage percentage of Green Infrastructure (GI)	X66	NUTS2, NUTS3	ESPON GRETA Project (https://database.espon.eu/project-archives/#/archives)
	Forest area within a region	X72	NUTS3	European Environment Agency (Corine Land Cover)
	Area of NATURA2000 areas relative to regional area	X73	NUTS3	European Environment Agency
(g) Land	Soil sealing within a region	X74	NUTS3	
	Share of agricultural area in protected areas	X75	NUTS3	ESPON SUPER Project (https://database.espon.eu/project-archives/#/archives)
	Share of urban area in protected areas	X76	NUTS3	
	Percentage share of urban use areas within a region	X77	NUTS3	
	Percentage share of agricultural use areas within a region	X78	NUTS3	
	Development of urban use per capita	X79	NUTS3	
Circular Economy	Annual amount of raw material extracted from natural environment	X80	NUTS2	ESPON CIRCTER Project (2017–2019): Circular Economy and Territorial Consequences (https://database.espon.eu/project-archives/#/archives)
	Total employment in material providers	X81	NUTS2	
	Total turnover generated by material providers' activities	X82	NUTS2	
	Total employment in technology providers' sectors	X83	NUTS2	
	Total turnover generated by technology providers' sector	X84	NUTS2	
	Total employment in Circular Business Models (CBM) sectors	X85	NUTS2	
	Total turnover generated by Circular Business Models (CBM) sectors	X86	NUTS2	

Table 7 (continued)

Main indicator	Sub-indicator name	Code	Scale	Source
Access to services	Share of regions that have poor access to primary schools	X87	NUTS3	ESPON PROFECY Project (https://database.espon.eu/project-archives/#/archives)
	Share of regions that have poor access to secondary schools	X88	NUTS3	
	Share of regions that have poor access to hospitals	X89	NUTS3	
	Share of regions that have poor access to closest doctors	X90	NUTS3	
	Share of regions that have poor access to pharmacies	X91	NUTS3	
	Share of regions that have poor access to bank office	X92	NUTS3	
	Share of regions that have poor access to train station	X93	NUTS3	
	Share of regions that have poor access to urban morphological zone (jobs)	X94	NUTS3	
	Share of regions that have poor access to cinemas	X95	NUTS3	
	Share of regions that have poor access to shops	X96	NUTS3	
	Share of regions that have poor access to regional centres	X97	NUTS3	
Access to transport network	Potential accessibility to roads	X98	NUTS3	ESPON TIA (https://database.espon.eu/project-archives/#/archives)
	Potential accessibility to rail	X99	NUTS3	
	Potential accessibility to air	X100	NUTS3	
	Potential accessibility to multimodal transport	X101	NUTS3	

Table 7 (continued)

Main indicator	Sub-indicator name	Code	Scale	Source
Climate	Change in annual mean number of days with heavy rainfall	X102	NUTS3	ESPON CLIMATE Project (https://database.espon.eu/project-archives/#/archives)
	Change in annual mean number of days with snow cover	X103	NUTS3	
	Change in annual mean number of summer days	X104	NUTS3	
	Potential vulnerability to climate change	X105	NUTS3	
	Combined adaptive capacity to climate change	X106	NUTS3	
	Relative change in annual mean evaporation	X107	NUTS3	
	Relative change in annual mean precipitation in summer months	X108	NUTS3	
	Relative change in annual mean precipitation in winter months	X109	NUTS3	

Acknowledgements We thank two anonymous reviewers for their helpful comments and suggestions that have significantly improved this paper.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Annoni, P., & Bolsi, P. (2020). The regional dimension of social progress in Europe: Presenting the new Social Progress Index. Working Paper No. 06/2020. EC Directorate-General for Regional and Urban Policy.

Annoni, P., & Dijkstra, L. (2013). EU Regional Competitiveness Index RCI 2013. EUR26060 EN, European Union, Luxembourg.

Annoni, P., & Dijkstra, L. (2019). *The EU regional competitiveness index 2019*. Publication Office of the European Union.

Baranzelli, C., Lavalle, C., Sgobbi, A., Aurambout, J-P., Trombetti, M., Jacobs-Crisioni, C., Cristobal, J., Kancs, D. & Kavalov, B. (2016). Regional patterns of energy production and consumption factors in europe-explaratory project EREBILAND-European Regional Energy Balance and Innovation Landscape. EUR27697, European Union, Luxembourg.

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

Belu, C. (2009). Ranking corporations based on sustainable and socially responsible practices. A data envelopment analysis (DEA) approach. *Sustainable Development, 17*(4), 257–268.

Blancard, S., Bonnet, M., & Hoarau, J.-F. (2021). The influence of agriculture on the structural economic vulnerability of small island spaces: Assessment using DEA based composite indicators. *Applied Economics, 53*(1), 79–97.

Blancas, F. J., Caballero, R., Gonzalez, M., Lozano-Oyola, M., & Perez, F. (2010). Goal programming synthetic indicators: An application for sustainable tourism in Andalusian coastal counties. *Ecological Economics, 69*(11), 2158–2172.

Bo, F., Danlin, Y., & Yaojun, Z. (2019). The liveable urban landscape: GIS and remote sensing extracted land use assessment for urban livability in Changchun Proper, China. *Land Use Policy, 87*, 104048.

Bo, T. H., Dysvik, B., & Jonassen, I. (2004). LSimpute: Accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Research, 32*(3), e34.

Boal-San Miguel, I., & Herrero-Prieto, L. C. (2020). Reliability of creative composite indicators with territorial specification in the EU. *Sustainability, 12*, 3070.

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

Brandon, P. S., & Lombardi, P. (2005). *Evaluating sustainable development in the built environment*. Blackwell.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

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

Burgette, L. F., & Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology, 172*, 1070–1076.

Cai, Z., Liu, Q., & Cao, S. (2020). Real estate supports rapid development of China's urbanization. *Land Use Policy, 95*, 104582.

Cao, S., Lv, Y., Zheng, H., & Wang, X. (2014). Challenges facing China's unbalanced urbanization strategy. *Land Use Policy, 39*, 412–415.

Charron, N., Lapuente, V., & Annoni, P. (2019). Measuring quality of government in EU regions across space and time. *Papers in Regional Science, 98*(5), 1925–1953.

Charnes, A., Cooper, W., & Rhodes, W. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(4), 429–444.

Chen, Y., Yu, J., & Khan, S. (2010). Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environmental Modelling & Software, 25*(12), 1582–1591.
Chen, M., Liu, W., & Tao, X. (2013). Evolution and assessment on China's urbanization 1960–2010: Under-urbanization or over-urbanization? *Habitat International, 38*, 25–33.
Chen, M., Huang, Y., Tang, Z., Lu, D., Liu, H., & Ma, L. (2014). The provincial pattern of the relationship between urbanization and economic development in China. *Journal of Geographical Sciences, 24*(1), 33–45.
Chen, J., Ji, J., Wang, H., Deng, M., & Yu, C. (2020). Risk assessment of urban rainstorm disaster based on multi-layer weighted principal component analysis: A case study of Nanjing, China. *International Journal of Environmental Research and Public Health, 17*(15), 5523.
Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research, 82*, 111–145.
Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the EU. *Journal of Common Market Studies, 42*, 919–955.
Cherchye, L., & Kuosmanen, T. (2002). Benchmarking sustainable development: A synthetic meta-index approach. WIDER Working Paper Series RP2004-28. World Institute for Development Economic Research (UNU-WIDER).
D'Adamo, I., Falcone, P. M., & Morone, P. (2020). A new socio-economic indicator to measure the performance of bioeconomy sectors in Europe. *Ecological Economics, 176*, 106724.
Dale, V. H., & Beyeler, S. C. (2001). Challenges in the development and use of ecological indicators. *Ecological Indicators, 1*, 3–10.
Davino, C., & Romano, R. (2014). Assessment of composite indicators using the ANOVA model combined with multivariate methods. *Social Indicators Research, 119*, 627–646.
De Sherbinin, A., Levy, M. A., Zell, E., Weber, S., & Jaiteh, M. (2014). Using satellite data to develop environmental indicators. *Environmental Research Letters, 9*, 084013.
Delgado-Villanueva, K. A., & Romero Gil, I. (2016). Environmental conflict analysis using an integrated grey clustering and entropy-weight method: A case study of a mining project in Peru. *Environmental Modelling & Software, 77*, 108–121.
Dijkstra, L., Annoni, P., & Kozovska, K. (2011). A new regional competitiveness index: Theory, methods and findings. European Union Regional Policy Working Papers No. 02/2011. EC Directorate General of Regional Policy.
EC. (2021). Online Document. Available at: https://knowledge4policy.ec.europa.eu/composite-indicators/10-step-guide/step-6-weighting_en
ESPON (2020). ESPON QoL-Quality of life measurements and methodology-Final Report. ESPON.
Fang, C., Ma, H., & Wang, J. (2015). A regional categorization for "New-Type Urbanization" in China. *PLoS ONE, 10*(8), e0134253.
Fazekas, M. (2017). Assessing the quality of government at the regional level using public procurement data. Working Paper No. 12/2017. Publications Office of the European Union.
Fei, L., Deng, Y., & Hu, Y. (2019). DS-VIKOR: A new multi-criteria decision-making method for supplier selection. *International Journal of Fuzzy Systems, 21*, 157–175.
Feng, J. A., & Chen, Y. G. (2010). Spatiotemporal evolution of urban form and land-use structure in Hangzhou, China: Evidence from fractals. *Environmental Planning B, 37*(5), 838–856.
Fraser, E. D. G., Dougill, A. J., Mabee, W. E., Reed, M., & McAlpine, P. (2006). Bottom up and top down: Analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. *Journal of Environmental Management, 78*, 114–127.
Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment. DSTI/IND Report No. 5. OECD.
Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators, 81*, 491–502.
Gerdessen, J. C., & Pascucci, S. (2013). Data Envelopment Analysis of sustainability indicators of European agricultural systems at regional level. *Agricultural Systems, 118*, 78–90.
Goedkoop, M., & Spriensma, R. (2001). The Eco-indicator99: A damage oriented method for life cycle impact assessment: Methodology report. Pre Consultants b.v.
Gollin, D., Jedwab, R., & Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth, 21*, 35–70.
Gomez-Limon, J. A., Arriaza, M., & Guerrero-Baena, M. D. (2020). Building a composite indicator to measure environmental sustainability using alternative weighting methods. *Sustainability, 12*, 4398.

Guns, R., & Rousseau, R. (2014). Recommending research collaborations using link prediction and random forest classifiers. *Scientometrics, 101*(2), 1461–1473.
Hou, X., Lv, T., Xu, J., Deng, X., Liu, F., & Pi, D. (2021). Energy sustainability evaluation of 30 provinces in China using the improved entropy weight-cloud model. *Ecological Indicators, 126*, 107657.
Huertas, J. D. B., Farinha, F., Oliveira, M. J., Silva, E., & Lanca, R. (2020). Comparison of artificial intelligence algorithms to estimate sustainability indicators. *Sustainable Cities and Society, 63*, 102430.
Islam, A. R. M. T., Al Mamun, A., Rahman, M. M., & Zahid, A. (2020). Simultaneous comparison of modified-integrated water quality and entropy weighted indices: Implication for safe drinking water in the coastal region of Bangladesh. *Ecological Indicators, 113*, 106229.
Jitmaneeroj, B. (2017). Beyond the equal-weight framework of the Social Progress Index: Identifying causal relationships for policy reforms. *International Journal of Social Economics, 44*(12), 2336–2350.
Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). The worldwide governance indicators: Methodology and analytical issues. World Bank Policy Research Working Paper No. 5430. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130
Keiner, M. (2006). Advancing sustainable development and its implementation through spatial planning. In M. Keiner (Ed.), *The future of sustainability*. Springer.
Kelley, J. G., & Simmons, B. A. (2015). Politics by number: Indicators as social pressure in international relations. *American Journal of Political Science, 59*(1), 55–70.
Li, Y., Zhao, L., & Suo, J. (2014). Comprehensive assessment on sustainable development of highway transportation capacity based on Entropy Weight and TOPSIS. *Sustainability, 6*(7), 4685–4693.
Li, D., Ma, J., Cheng, T., van Genderen, J. L., & Shao, Z. (2019a). Challenges and opportunities for the development of MEGACITIES. *International Journal of Digital Earth, 12*(12), 1382–1395.
Li, S., Ying, Z., Zhang, H., Ge, G., & Liu, Q. (2019b). Comprehensive assessment of urbanization coordination: A case study of Jiangxi province, China. *Chinese Geographical Science, 29*, 488–502.
Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2*, 18–22.
Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
Liu, Z., Ren, Y., Shen, L., Liao, X., Wei, X., & Wang, J. (2020). Analysis on the effectiveness of indicators for evaluating urban carrying capacity: A popularity-suitability perspective. *Journal of Cleaner Production, 246*, 119019.
Lopez, J. M. R., Heider, K., & Scheffran, J. (2017). Frontiers of urbanization: Identifying and explaining urbanization hot spots in the south of Mexico City using human and remote sensing. *Applied Geography, 79*, 1–10.
Ma, H., & Sun, Z. (2020). Comprehensive urbanization level and its dynamic factors for five Central Asian countries. *Journal of Geographical Sciences, 30*, 1761–1780.
Mainali, B., Pachauri, S., Rao, N. D., & Silveira, S. (2014). Assessing rural energy sustainability in developing countries. *Energy for Sustainable Development, 19*(15), 28.
Mascarenhas, A., Nunes, L. M., & Ramos, T. B. (2015). Selection of sustainability indicators for planning: Combining stakeholders' participation and data reduction techniques. *Journal of Cleaner Production, 92*, 295–307.
Mayer, A. L. (2008). Strengths and weaknesses of common sustainability indices for multidimensional systems. *Environment International, 34*(2), 277–291.
McPhearson, T., Haase, D., Kabisch, N., & Gren, Å. (2016). Advancing understanding of the complex nature of urban systems. *Ecological Indicators, 70*, 566–573.
Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available. Public Economics Research Paper No. 17. CES KU Leuven
Menze, B. H., Kelm, B. M., Masuch, R., et al. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10*, 213.
Merino-Saum, A., Halla, P., Superti, V., Boesch, A., & Binder, C. R. (2020). Indicators for urban sustainability: Key lessons from a systematic analysis of 67 measurement initiatives. *Ecological Indicators, 119*, 106879.
Michael, F. L., Noor, Z. Z., & Figueroa, M. J. (2014). Review of urban sustainability indicators assessment-Case study between Asian countries. *Habitat International, 44*, 491–500.
Mori, K., & Yamashita, T. (2015). Methodological framework of sustainability assessment in City Sustainability Index (CSI): A concept of constraint and maximisation indicators. *Habitat International, 45*(1), 10–14.

Munda, G. (2005). 'Measuring sustainability': A multi-criterion framework. *Environment, Development and Sustainability, 7*, 117–134.
Myszczyszyn, J., Lacka, I., Golab, S., Bedzik, B., & Supron, B. (2021). A preliminary analysis of the relationship between economic growth and selected innovativeness indicators on the example of four European Union countries. *European Research Studies Journal, 24*(3), 402–418.
Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. EUR 21682EN. EC Joint Research Centre.
Nelwamondo, F. V., Mohamed, S., & Marwala, T. (2007). Missing data: A comparison of neural network and expectation maximization techniques. *Current Science, 93*(11), 1514–1521.
Netherlands Presidency. (2016). Urban Agenda for the EU: Pact of Amsterdam. Factsheet. Available online at: https://agendastad.nl/wp-content/uploads/2015/02/EU-Urban-Agenda-factsheet.pdf. Accessed 9 July 2021.
Newman, P. W. G. (1999). Sustainability and cities: Extending the metabolism model. *Landscape and Urban Planning, 44*, 219–226.
Nicoletti, G., Scarpetta, S., & Boylaud, O. (1999). Summary indicators of product market regulation with an extension to employment protection legislation. OECD Economics Department Working Papers No. 226, ECO-WKP (99)18. OECD
OECD. (2008). *Handbook on constructing composite indicators*. OECD.
Parker, J. (1991). Environmental reporting and environmental indices. PhD Dissertation. University of Cambridge.
Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science. *Journal of the Royal Statistical Society, 176*, 609–634.
Patias, N., Rowe, F., Cavazzi, S., & Arribas-Bel, D. (2021). Sustainable urban development indicators in Great Britain from 2001 to 2016. *Landscape and Urban Planning, 214*, 104148.
Perpina, C., Martinez-Llario, J. C., & Perez-Navarro, A. (2013). Multicriteria assessment in GIS environments for sitting biomass plants. *Land Use Policy, 31*, 326–335.
Phillis, Y. A., Kouikoglou, V. S., & Verdugo, C. (2017). Urban sustainability assessment and ranking of cities. *Computers, Environment and Urban Systems, 64*, 254–265.
Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers*. MIT Press.
Porter, M. E., & Stern, S. (2001). *National innovative capacity. The global competitiveness report*. Oxford University Press.
Pravitasari, A. E., Rustiadi, E., Mulya, S. P., & Fuadina, L. N. (2018). Developing regional sustainability index as a new approach for evaluating sustainability performance in Indonesia. *Environment and Ecology Research, 6*(3), 157–168.
Ramachandra, T. V., Bharath, A. H., & Sowmyashree, M. V. (2015). Monitoring urbanization and its implications in a mega city from space: Spatiotemporal patterns and its indicators. *Journal of Environmental Management, 148*, 67–81.
Reisi, M., Aye, L., Rajabifard, A., & Ngo, T. (2014). Transport sustainability index: Melbourne case study. *Ecological Indicators, 43*, 288–296.
Rogge, N. (2018). On aggregating Benefit of the Doubt composite indicators. *European Journal of Operational Research, 264*, 364–369.
Romano, G., Dal Sasso, P., Trisorio Liuzzi, G., & Gentile, F. (2015). Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy. *Land Use Policy, 48*, 131–143.
Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society Series A, 168*(2), 307–323.
Salvati, L., & Carlucci, M. (2014). A composite index of sustainable development at the local scale: Italy as a case study. *Ecological Indicators, 43*, 162–171.
Shannon, C. E., & Weaver, W. (1947). *The mathematical theory of communication*. The University of Illinois Press.
Shen, L.-Y., Ochoa, J. J., Shah, M. N., & Zhang, X. (2011). The application of urban sustainability indicators-A comparison between various practices. *Habitat International, 35*, 17–29.
Smith, L. M., Case, J. L., Smith, H. M., Harwell, L. C., & Summers, J. K. (2013). Relating ecosystem services to domains of human well-being: Foundation for a U.S. index. *Ecological Indicators, 28*, 79–90.
Stekhoven, D. J., & Bühlmann, P. (2012). Missforest: Non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112–118.

Szopik-Depczynska, K., Kedzierska-Szczepaniak, A., Szczepaniak, K., Cheba, K., Gajda, W., & Ioppolo, G. (2018). Innovation in sustainable development: An investigation of the EU context using 2030 agenda indicators. *Land Use Policy, 79*, 251–262.

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining, 10*(6), 363–377.

Tannier, C., & Thomas, I. (2013). Defining and characterising urban boundaries: A fractal analysis of theoretical cities and Belgian cities. *Computers, Environment and Urban Systems, 41*, 234–248.

Tian, Z.-P., Wang, J.-Q., & Zhang, H.-Y. (2018). An integrated approach for failure mode and effects analysis based on fuzzy best-worst, relative entropy, and VIKOR methods. *Applied Soft Computing, 72*, 636–646.

Tomal, M. (2021). Analysing the coupling coordination degree of socio-economic-infrastructural development and its obstacles: The case study of Polish rural municipalities. *Applied Economics Letters, 28*(13), 1098–1103.

Tutz, G., & Ramdan, S. (2014). Improved methods for the imputation of the missing data by Nearest Neighbour methods. Technical Report No. 172. University of Munich, Germany.

UNDP. (1990). *Human development report 1990*. Oxford University Press.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., Rosaida, N., & Hamid, H. A. (2013). Future daily PM10 concentrations prediction by combining regression models and feedforward back propagation models with principle component analysis (PCA). *Atmospheric Environment, 77*, 621–630.

Ustaoglu, E., & Aydınoglu, A. C. (2019). Regional variations of land-use development and land-use/cover change dynamics: A case study of Turkey. *Remote Sensing, 11*, 885.

Ustaoglu, E., & Aydınoglu, A. C. (2020). Suitability evaluation of urban construction land in Pendik district of Istanbul, Turkey. *Land Use Policy, 99*, 104783.

Ülengin, B., Ülengin, F., & Güvenç, Ü. (2001). A multidimensional approach to urban quality of life: The case of Istanbul. *European Journal of Operational Research, 130*(2), 361–374.

Van de Kerk, G., & Manuel, A. R. (2008). A comprehensive index for a sustainable society: The SSI-the sustainable society index. *Ecological Economics, 66*(2), 228–242.

Wang, Q., Yuan, X., Zhang, J., Gao, Y., Hong, J., Zuo, J., & Liu, W. (2015). Assessment of the sustainable development capacity with the entropy weight coefficient method. *Sustainability, 7*(10), 13542–13563.

Wang, L., Omrani, H., Zhao, Z., et al. (2019). Analysis on urban densification dynamics and future models in southeastern Wisconsin, USA. *PLoS ONE, 14*(3), e0211964.

World Bank. (1999). World Development Indicators 1999. World Bank, Washington D.C.

Xavier, A., Freitas, M. B. C., Fragoso, R., & Rosario, M. S. (2018). A regional composite indicator for analyzing agricultural sustainability in Portugal: A goal programming approach. *Ecological Indicators, 89*, 84–100.

Yang, Z., Lei, J., & Li, J.-G. (2019). Identifying the determinants of urbanization in prefecture-level cities in China: A quantitative analysis based on spatial production theory. *Sustainability, 11*(4), 1204.

Yang, Z., Yang, H., & Wang, H. (2020). Evaluating urban sustainability under different development pathways: A case study of the Beijing-Tianjin-Hebei region. *Sustainable Cities and Society, 61*, 102226.

Ye, C., Chen, M., Duan, J., & Yang, D. (2017). Uneven development, urbanization and production of space in the middle-scale region based on the case of Jiangsu province, China. *Habitat International, 66*, 106–116.

Yeheyis, M., Hewage, K., Alam, M. S., Eskicioglu, C., & Sadiq, R. (2013). An overview of construction and demolition waste management in Canada: A lifecycle analysis approach to sustainability. *Clean Technology and Environmental Policy, 15*(1), 81–91.

Zhao, M., Cheng, W., Zhou, C., Li, M., Wang, N., & Liu, Q. (2017a). Spatial differentiation and morphological characteristics of China's urban core zones based on geomorphologic partition. *Journal of Applied Remote Sensing, 11*(1), 016041.

Zhao, Y., Wang, S., Ge, Y., Liu, Q., & Liu, X. (2017b). The spatial differentiation of the coupling relationship between urbanization and the eco-environment in countries globally: A comprehensive assessment. *Ecological Modelling, 360*, 313–327.

Zhao, G., & Wang, D. (2019). Comprehensive evaluation of AC/DC hybrid microgrid planning based on Analytic Hierarchy Process and Entropy Weight method. *Applied Sciences, 9*(18), 3843.

Zhou, D., Xu, J., Wang, L., & Lin, Z. (2015). Assessing urbanization quality using structure and function analysis: A case study of the urban agglomeration around Hangzhou Bay (UAHB), China. *Habitat International, 49*, 165–176.

Zoeteman, K., Mommaas, H., & Dagevos, J. (2016). Are larger cities more sustainable? Lessons from integrated sustainability monitoring in 403 Dutch municipalities. *Environmental Development, 17*, 57–72.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Eda Ustaoglu[1] · Gloria Ortega Lopez[2] · Alejandro Gutierrez-Alcoba[3]

✉ Eda Ustaoglu
eda.ustaoglu@agu.edu.tr

Gloria Ortega Lopez
gol315@ual.es

Alejandro Gutierrez-Alcoba
gutialco@gmail.com

1 Department of Economics, Abdullah Gul University, Kayseri, Turkey

2 Department of Computer Architecture and Electronics, University of Almeria, Almeria, Spain

3 Business School, University of Edinburgh, Edinburgh, UK