sustainability

MDPI

Article

Building a Composite Indicator to Measure Environmental Sustainability Using Alternative Weighting Methods

José A. Gómez-Limón [1], Manuel Arriaza [2] and M. Dolores Guerrero-Baena [1,*]

[1] WEARE: Water, Environmental and Agricultural Resources Economics Research Group, Faculty of Law and Business Sciences, University of Córdoba, 14071 Córdoba, Spain; jglimon@uco.es

[2] Department of Agricultural Economics, Institute of Agricultural Research and Training (IFAPA), 14080 Córdoba, Spain; manuel.arriaza@juntadeandalucia.es

* Correspondence: dolores.guerrero@uco.es; Tel.: +34-957-212-205

Received: 30 April 2020; Accepted: 23 May 2020; Published: 27 May 2020

Abstract: Environmental sustainability in agriculture can be measured through the construction of composite indicators. However, this is a challenging task because these indexes are heavily dependent on how the individual base indicators are weighted. The main aim of this paper is to contribute to the existing literature regarding the robustness of subjective (based on experts' opinions) weighting methods when constructing a composite indicator for measuring environmental sustainability at the farm level. In particular, the study analyzes two multi-criteria techniques, the analytic hierarchy process and the recently developed best-worst method, as well as the more straightforward point allocation method. These alternative methods have been implemented to empirically assess the environmental performance of irrigated olive farms in Spain. Data for this case study were collected from a panel of 22 experts and a survey of 99 farms. The results obtained suggest that there are no statistically significant differences in the weights of the individual base indicators derived from the three weighting methods considered. Moreover, the ranking of the sampled farms, in terms of their level of environmental sustainability measured through the composite indicators proposed, is not dependent on the use of the different weighting methods. Thus, the results support the robustness of the three weighting methods considered.

Keywords: agricultural sustainability; environmental performance; sustainability indices; multi-criteria analysis; analytic hierarchy process; best-worst method; irrigated olive groves; Spain

1. Introduction

There is a broad consensus about the definition of 'sustainable agriculture' as an activity that satisfies the following requirements for an indefinite period of time [1,2]: a) it protects biodiversity and natural resources and prevents environmental degradation, b) it is economically viable, and c) it is socially acceptable. Taking these requirements into consideration, agricultural sustainability can be defined as a concept that encompasses three main dimensions:

- *Environmental sustainability*. Sustainable agriculture must preserve biological biodiversity and the provision of ecosystem services. Thus, environmental sustainability can be defined as the ability to ensure greater agricultural productivity while simultaneously conserving natural resources and preventing the degradation of ecosystems.
- *Economic sustainability*. Sustainable agriculture must be economically viable, ensuring not only adequate profitability for farmers (the microeconomic level) but also a positive contribution to national/regional income (the macroeconomic level).

- *Socio-cultural sustainability*. Sustainable agriculture must be socially and culturally beneficial, i.e., it should ensure food security and the fair and equitable distribution of the wealth it generates, as well as contribute to the viability of rural communities.

This paper is focused on the measurement of environmental sustainability. To date, various indicator-based methods have been developed for this purpose, constructed using a wide range of agri-environmental indicators (AEIs) (for a review of the approaches proposed, interested readers can consult Bockstaller et al. [3]). All these proposals are based on large sets of AEIs aimed at assessing the multidimensional environmental impacts of agricultural activity and the provision of ecosystem services (natural resources, biological processes, biodiversity, etc.) (e.g., [4–6]). However, the quantification of environmental sustainability through AEIs has been criticized for several reasons. On the one hand, there are technical problems related to data and measurability (qualitative aspects that are hard to quantify) issues and a lack of sound ecological models that enable the interpretation of the indicators (e.g., a lack of reference levels or thresholds for reversibility processes), the multiple spatial scales needed for an overall assessment or the appropriate time horizon required (extended monitoring is needed for long-term environmental changes) [7,8]. On the other hand, there are also operational concerns related to the interpretation of the whole set of indicators required for such analyses, which is an obstacle to their use as a practical decision-support tool. In order to deal with the latter problem, composite indicators or indexes have been proposed as a means of summarizing the information provided by multiple indicators into an overall assessment of environmental performance (see, for instance, [9,10]).

In order to construct a composite environmental indicator, specific methodological approaches for the normalization, weighting, and aggregation of base AEIs must be selected from several alternatives (see further details in Section 3). All these choices have a significant effect on the overall composite indicator built. The inherent subjectivity of the choice of these approaches is behind most of the criticisms leveled at the different sustainability indexes proposed in the empirical studies carried out to date [11]. This issue has prompted an academic debate on the robustness and sensitivity of the methodological approaches used in the construction of composite indicators [12,13].

The objective of this paper is to contribute to the existing literature regarding the robustness of three alternative weighting methods [14]. For this purpose, we build a composite environmental indicator using three different methods to assess the relative importance of the individual base AEIs; then, by comparing the results obtained from a real-world case study, we provide further insights into the robustness of the weighting methods implemented.

In this context, it is worth mentioning that two of the weighting methods used in this paper are at the core of the multi-criteria decision-making (MCDM) paradigm, since they are used as key tools to provide information about the relative importance of the different criteria considered in these kinds of problems. For this reason, MCDM weighting procedures have commonly been used to build composite indicators [15], with the analytic hierarchy process (AHP) being the most popular one (e.g., [16–18]). Moreover, other more consistent and less time-consuming weighting methods have recently been developed, with the notable among them being the best-worst method (BWM) [19]. As this is a new technique, there have only been a few applications to date, and it has not yet been used in the construction of composite indicators. In this paper, we use the two aforementioned multi-criteria methods, in addition to the more straightforward point allocation (PA) method, to weight the base AEIs that are to be included in an index. We then compare the results obtained from the three methods in an empirical case study. To the best of our knowledge, this is the first study focused on the analysis of the robustness of weighting methods when constructing a composite environmental indicator, although there have been several empirical studies comparing some alternative weighting methods in different types of composite indicators (e.g., [20,21]). The comparative analysis proposed will enable us to draw useful conclusions about the construction of sound composite indicators.

The empirical case study considered is the assessment of the environmental sustainability of olive groves in Spain. For this purpose, the composite environmental indicator built (*ENV_SUST*) relies on

previous research focused on the selection of relevant AEIs for this particular agricultural system [22], and uses primary data gathered from Spanish olive farms to calculate the whole set of selected AEIs at the farm level [23]. In any case, it is worth noting that, as far as the authors are aware, this paper is the first to develop a single composite environmental index for an overall assessment of environmental performance in olive growing. Thus, this paper also contributes to the existing literature by providing a sound instrument that is particularly useful for designing targeted policy interventions aimed at promoting sustainable olive farming.

2. Olive Farming in Spain: Environmental Sustainability Assessment

2.1. Recent Developments of Olive Groves in Spain

The current area of olive groves in Spain has reached its historically highest level, with around 2.5 million hectares (14% of the country's utilized agricultural area). In fact, Spain is the world's leading olive-producing country, accounting for one-third of the total olive grove area worldwide and half of the total olive oil production.

Spain's accession to the EU in 1986 allowed the Spanish olive sector to benefit from the implementation of the Common Agricultural Policy (CAP), which granted olive growers coupled subsidies that made olive farming more profitable than other types of agriculture (i.e., extensive herbaceous crops). The CAP subsidies encouraged these farmers to increase their olive grove area and olive production. This process of expansion has been also supported by new growing techniques aimed at production intensification, such as irrigation (olive has traditionally been a non-irrigated crop) and higher tree density (increasing from traditional densities of around 100 olive trees per hectare to 300–500 trees per hectare in 'intensive' orchards, or even more than 1000 trees per hectare in 'super-intensive' ones). As a result of these changes, Spain has increased its olive grove area by 25% between 1990 and 2020, and doubled its production of olive oil.

However, this expansion and intensification of Spanish olive groves has given rise to a number of environmental problems [23,24]:

1. *Soil erosion*. Erosion is the main environmental problem caused by this crop. The high soil erosion rates are due to the fact that more than 40% of all olive groves are located on land with unfavorable soil conditions for agricultural production (steep slopes, land particularly sensitive to erosion, or affected by frequent torrential rain), and that poor soil management by farmers has damaged natural vegetation cover (leading to farms with uncovered soils) [25]. This environmental impact has been aggravated in recent years by the expansion of olive groves into areas with especially adverse characteristics (steep slopes, extreme torrential rainfall, high erodibility of soils) [26].
2. *Loss of biodiversity*. One of the main characteristics of olive groves in the 1980s (under traditional farming) was the high biodiversity associated with the crop, with olive being an example of a 'high natural value' agricultural system. The low-intensity olive farming (minimum use of agrochemicals) and the existence of old olive trees with semi-natural herbaceous vegetation located in areas with different land uses (vineyards, cereals, pastures, and Mediterranean forest) provided a varied habitat, where a large number of insects, birds, reptiles, and mammals found refuge. However, the extension (large olive monoculture areas where hedgerows, stone walls, and islands of shrubs and trees have been eliminated) and intensification of olive groves (disappearance of vegetable cover, intensive use of biocides, fertilizers and machinery, water pollution, and soil erosion) has changed this situation, leading to a reduction in both the number and diversity of animal species in olive grove systems [27,28].
3. *Non-point source water pollution*. Modern olive growing has contributed to a decline in water quality due to the intense use of agrochemical products (mainly herbicides and fertilizers). This has resulted in non-point source water pollution problems in rivers, reservoirs and aquifers. Although in recent years some of the most polluting products widely used in olive farming (e.g., simazine

and diuron) have been banned, water quality could be further improved by modifying some of the current olive farming practices [29].

4. *Overexploitation of water resources*. Before the 1990s, most olive trees in Spain were rain-fed, but the intensification of the crop has seen the emergence of 800,000 hectares of irrigated olive groves. Although olive trees have low water requirements and are usually irrigated using highly efficient irrigation systems (water consumption using drip irrigation is around 1500 m^3/ha·year), there is substantial pressure on water resources [30]. Increasing water extraction not only causes the overexploitation of water resources but also jeopardizes the ability to meet other water demands in basins with a higher degree of water scarcity [31].

It is worth noting that several policy initiatives have been implemented in recent years to (partially) solve all these relevant environmental problems related to olive groves, by encouraging farmers to adopt various biodiversity-friendly and resource conservation practices. Along with the rational use of agrochemicals and water, these include some compatible soil conservation practices, such as disposing of olive-desuckering debris without burning, which helps mitigate climate change; the shredding of olive-pruning debris for use as soil cover, improving soil texture and reducing the impact of rain and water run-offs; and the use of cover crops under mower control as a sustainable practice in terms of soil protection [26,32,33]. In addition, some practices related to functional elements (hedgerows, riparian vegetation, plots margins, etc.) have proven effective in enhancing biodiversity as well as having a positive effect on other ecosystem services, such as landscape aesthetics [34]. Yet there is still plenty of room for further improvement. To effectively guide the future development of olive groves, more in-depth analyses are required, especially those that provide a quantitative assessment of the environmental performance of individual olive farms.

2.2. Indicators Measuring Environmental Sustainability of Olive Farming

Although there are several methodological frameworks for the quantitative assessment of the environmental sustainability of agricultural systems, there is widespread scientific agreement that constructing and calculating environmental indicators is the most suitable approach for this purpose [35,36]. Thus, the present study relies on the set of AEIs proposed by Gómez-Limón and Riesgo [22] to evaluate the sustainability of olive farms in Spain.

In order to evaluate the environmental sustainability of agriculture, the approach followed is founded on the SAFE (*Sustainability Assessment of Farming and the Environment*) analytical framework [37]. SAFE is based on a hierarchical structure with three levels: i) principles, ii) criteria and iii) indicators. *Principles* are general conditions for achieving environmental sustainability related to the ecological functions of the agro-ecosystems. In this sense, the environmental sustainability of agricultural systems centers on two principles regarding the protection of (a) biodiversity and (b) natural resources. *Criteria* are the resulting states of agricultural systems when their related principles are respected. For the particular case of olive groves, biodiversity is considered to be protected when the following elements are guaranteed: a1) olive grove genetic diversity (trees in the orchard), a2) biological diversity (a range of different species within the farm boundaries, from 0.1 km^2 to 10 km^2), and a3) habitat diversity (a range of different habitats within the landscape unit, from 10 km^2 to 1000 km^2). In addition, natural resources conservation is achieved when: b1) soil erosion is minimized, b2) soil fertility is protected or enhanced, b3) soil and water quality are maintained or improved, b4) water extraction is minimized, and b5) the energy balance (primary energy supply minus primary energy used per cultivation unit) is optimized. Lastly, *indicators* are variables that can be assessed to measure compliance with a criterion, thus producing a representative picture of the environmental sustainability of the agricultural system under analysis. Taking the former criteria into account, a set of 11 AEIs was selected on the basis of analytical soundness, measurability and policy relevance [38,39], as shown in Table 1. Technical details on why each indicator was chosen, how it was calculated at farm level and how its value should be interpreted can be found in Gómez-Limón and Arriaza [23].

Table 1. Principles, criteria and indicators of the environmental sustainability of olive groves.

Principles	Criteria	Agri-environmental Indicators (*ACRONYM*) [Measurement Unit]
Biodiversity protection	Ensuring olive grove genetic diversity	Number of olive grove varieties (*NUMVAR*) [olive grove varieties number]
	Enhancing or protecting biological diversity	Index of biological diversity (*DIVERSIND*) [dimensionless] bounded [0,1]
		Pesticide risk (*PESTRISK*) [kg live organism/ha·year]
	Enhancing or protecting habitat diversity (ecosystem)	Percentage of land with other crops (*OTHERCROP*) [%]
		Percentage of non-cultivated land (*NONCULTIV*) [%]
Natural resources conservation (soil and water)	Minimizing soil erosion	Soil erosion (*EROSION*) [t/ha·year]
	Enhancing or protecting soil fertility	Soil organic matter (*ORGMAT*) [dimensionless] bounded [0,1]
	Enhancing or protecting soil and water quality	Nitrogen balance (*NITROGBAL*) [N kg/ha·year]
		Residual herbicide use (*RESHERB*) [kg active matter/ha·year]
	Minimizing water extraction	Irrigation water use (*WATERUSE*) [m^3/ha·year]
	Optimizing energy balance	Energy balance (*ENERGYBAL*) [MJ/ha·year]

Source: Gómez-Limón and Riesgo [22] and Gómez-Limón and Arriaza [23].

3. Building a Composite Indicator to Measure Environmental Sustainability

3.1. The Methodological Approach for Building Composite Indicators

The literature contains a plethora of techniques for building environmental sustainability indices. In any case, the Organization for Economic Co-operation and Development (OECD) and the Joint Research Centre of the European Commission (JRC) [11] provide guidance on the transparent construction of composite indicators, identifying the steps that analysts should follow:

1. *Indicator selection and data gathering*. As explained in Section 2, an essential element of this kind of study is the selection of relevant AEIs based on strict quality criteria, and accurate data gathering to calculate the empirical values of these indicators. Given the huge number of possible indicators, the use of a solid theoretical framework is recommended; in this paper, the SAFE approach is applied.
2. *Normalization of indicators*. Transforming indicators into dimensionless variables (normalization) is essential before they are weighted and aggregated, as they have usually been calculated using different units of measurement. To be able to compare them and perform arithmetic operations on them, they need to be expressed in homogeneous units within the same range. In our case, selecting from among the various normalization techniques available [40,41], we applied the min-max or re-scaling normalization, taking the reference values for each of the AEI considered as sustainability thresholds. Thus, the values of all the normalized indicators vary within a dimensionless range [0,1], where 0 is assigned to all cases where the AEI value is worse than or equal to an 'unacceptable level of sustainability' (i.e., the worst environmental performance)

and 1 is assigned to all cases where the AEI value is better than or equal to a 'desired level of sustainability' (i.e., the best environmental performance).

3. Weighting of indicators. Assigning weights enables us to identify the relative importance of the individual indicators. There are several valid procedures for weighting indicators, but the composite indicator may yield different results depending on the procedure used [42,43]. Therefore, the selection of a particular technique is a challenging task. The weighting techniques for constructing indices can be divided into 'objective' and 'subjective' ones [44]. With the former, weights are derived endogenously using statistical or mathematical procedures, such as principal components analysis (PCA), data envelopment analysis (DEA), the benefit of the doubt (BOD) approach or regression analysis (RA). With the latter, weights are determined exogenously on the basis of value judgments expressed by experts or decision-makers, as is the case with AHP, BWM, PA, budget allocation process (BAP) or conjoint analysis (CA). It is worth mentioning that environmental sustainability is a technical concept that requires scientific knowledge to define and measure, especially when it is applied to a specific ecosystem (in this case, olive groves in Spain). This justifies the use of exogenously determined weights in our case study, based on the opinion of experts on the environmental performance of olive groves. In particular, we have chosen AHP because it is the most commonly-applied technique among the subjective weighting methods available, BWM has been selected due to its novelty and the presumed advantages over AHP, and PA because it is an explicit and straightforward weighting method.
4. *Aggregation of indicators*. The OECD and JRC [11] suggest several alternative functional forms that allow indicators to be aggregated, explaining their pros and cons. Depending on the aggregation method used to develop the indices, the results and the conclusions drawn from them may differ from case to case. Thus, the choice of the aggregation method is also subject to criticism relating to the shortcomings of the technique used [40,43,45]. The key issue when selecting a functional method of aggregating indicators is the compensability or marginal rate of substitution among indicators [18]: a) additive linear functions implicitly assume total compensability among indicators, b) multiplicative and geometric functions permit partial compensability, and c) non-compensatory functions assure non-compensability. In order to minimize the subjectivity regarding the method employed to build the composite indicator measuring environmental sustainability, the multicriteria function based on the distance to the ideal point measured by different metrics (i.e., different degrees of compensability) and developed by Díaz-Balteiro and Romero [46] has been chosen for implementation.

Having made the decisions explained above, the composite indicator measuring the environmental sustainability of olive farms (*ENV_SUST*) can be calculated as a function of the normalized values of the 11 AEIs taken into account (I_k), the weights assigned to each of these indicators (w_k) and the compensation parameter (λ), following the expression:

$$ENV_SUST = (1-\lambda)\cdot\left[\underset{k}{Min}(w_k \cdot I_k)\right] + \lambda \cdot \sum_{k=1}^{k=11} w_k I_k. \quad (1)$$

The parameter λ ranges between 0 and 1, thus affecting the degree of compensability among the indicators. Here we consider five values of the compensation parameter ($\lambda = 0$, $\lambda = 0.25$, $\lambda = 0.5$, $\lambda = 0.75$ and $\lambda = 1$), which gives us the three abovementioned possibilities: (a) total compensability ($\lambda = 1$), (b) various degrees of partial compensability ($0 < \lambda < 1$) and (c) zero compensability ($\lambda = 0$).

3.2. Alternative Techniques for Weighting Indicators

As has been previously stated, this paper is focused on the role of weighting techniques when constructing composite indicators, in order to provide further empirical insights about their robustness. For this reason, in this paper we use three exogeneous weighting methods (i.e., based on expert

opinion) to determine the priorities (global weights) (w_k) of the whole set of AEIs used to construct the *ENV_SUST* index: AHP, BWM, and PA.

The AHP was initially developed as a decision-support tool for making complex decisions [47], but it was subsequently adapted to index construction; this technique is also particularly useful for weighting sustainability attributes when constructing composite indicators [13,48]. The implementation of this method involves the following steps: first, the weighting problem is structured as a tree-based hierarchy, where the overall goal of the problem (in our paper, the environmental sustainability of irrigated olive grove) is at the top of the hierarchy. Decision criteria contributing to the main goal are placed at an intermediate level (i.e., biodiversity protection and conservation of natural resources in our case) and decision subcriteria are positioned at the lowest level (the base AEIs in our case). Second, experts individually perform pairwise comparisons at each node of the hierarchy, expressing their preferences as to how much one (sub)criterion should be valued over another, following Saaty's fundamental scale (from 1 −equal importance− to 9 −extreme importance of one (sub)criterion over another). Based on these expert judgments, reciprocal square matrices can be built for each node. Third, the local weights of the sets of criteria (biodiversity protection and conservation of natural resources) and subcriteria (the base AEIs) are calculated using the main eigenvector method proposed by Saaty [49]: $AW = \lambda_{max}W$, where λ_{max} is the maximum eigenvalue of A and W is the vector of local weights. AHP allows some degree of inconsistency in the decision maker's judgments, measured using a consistency ratio that must not exceed predefined values [50]. Fourth, global weights (w_k) of the base AEIs are calculated by multiplying the local weight of each subcriterion (AEI) by the priority of its parent node (its related principle).

The BWM is a novel multi-criteria decision-making technique [19]. Like AHP, this method is suitable for weighting attributes, and although it has only recently been developed, it has already been applied to the construction of composite indicators [51,52]. BWM requires fewer pairwise comparisons than AHP (in the AHP method, the number of comparisons is $n\ (n-1)/2$, while for the BWM, the number of comparisons is $2n-3$), which may lead to more consistent and reliable results. In order to derive the global weights of the base AEIs, BWM entails the following steps: first, as in AHP, the problem is structured as a tree-based hierarchy (overall goal, decision criteria, and decision subcriteria). Second, the best (sub)criterion (i.e., the most important) and the worst (sub)criterion (i.e., the least important) of the set of (sub)criteria are identified by the expert. Third, the preference for the best (sub)criterion over all the other (sub)criteria is determined using a number between 1 and 9, similar to Saaty's fundamental scale. The expert's preferences are then used to generate the Best-to-Others vector: $A_B = (a_{B1}, \ldots, a_{Bk}, \ldots, a_{BK})$, where a_{Bk} shows the preference for the best (sub)criterion B over (sub)criterion k, and $a_{BB} = 1$. Fourth, the preferences of all the (sub)criteria over the worst (sub)criterion are determined using a number between 1 and 9, as in the previous step. This information enables the construction of the Others-to-Worst vector: $A_W = (a_{1W}, \ldots, a_{kW}, \ldots, a_{KW})^T$, where a_{kW} shows the preference for the (sub)criterion k over the worst (sub)criterion W, and $a_{WW} = 1$. Finally, the local weights of decision (sub)criteria and the corresponding indicator of the consistency of responses (ξ^L) are obtained by solving the following linear programming model:

$$\begin{gathered} \min \xi^L, \text{ s.t.} \\ |w_B - a_{Bk}w_k| \leq \xi^L, \text{ for all } k \\ |w_k - a_{kW}w_W| \leq \xi^L, \text{ for all } k \\ \textstyle\sum_k w_k = 1,\ w_k \geq 0, \text{ for all } k. \end{gathered} \tag{2}$$

Using ξ^L, it is possible to calculate a consistency ratio (CR), which must not be higher than 0.25. As in AHP, global weights (w_k) of the base AEIs are calculated by multiplying the local weight of each AEI by the priority of its associated principle.

The PA method is a straightforward weighting technique that has proved valuable for determining the priorities of the different attributes of sustainability composite indicators [40,53]. In this method, the expert is asked to directly allocate a fixed number of points (e.g., 10 or 100) among the multiple

criteria (indicators in our case) considered in a decision problem to establish the weight of each criterion [54]. In this paper, we apply this method as follows: first, each expert has to distribute 100 points between the two principles related to the environmental sustainability of irrigated olive grove (biodiversity protection and conservation of natural resources); second, experts allocate another 100 points among the five AEIs linked to biodiversity protection and an additional 100 points among the six AEIs related to the conservation of natural resources. Third, the global weights of the AEIs (w_k) are obtained (in percentage terms) by multiplying the priority of each AEI by the priority of its related principle.

3.3. Data Collection for the Empirical Assessment of Environmental Sustainability

The empirical assessment of the environmental sustainability of the irrigated olives groves in Spain relies on two data gathering sources: a survey of farmers to collect the farm-level technical data needed to calculate the AEIs considered (I_k), and a survey of experts to obtain the weights assigned to each of these indicators (w_k).

3.3.1. Farmer Survey

Due to the dispersion of the olive orchards in Spain, we carried out multi-stage cluster sampling to obtain a representative sample of irrigated olive farms. First, following a random selection of agricultural regions proportional to the total area of olive groves, six agricultural regions in Andalusia were selected (Andalusia accounts for more than 80% of total Spanish olive oil production). Figure 1 shows their location on a map. Second, a number of farms proportional to the area of olive groves in the agricultural region were selected through quota sampling, taking into account the farm size. Third, the selection of olive growers to be interviewed was determined using random route sampling. This procedure yielded a final sample consisting of a total of 480 olive farms. Further details about the sampling procedure can be seen in [23].

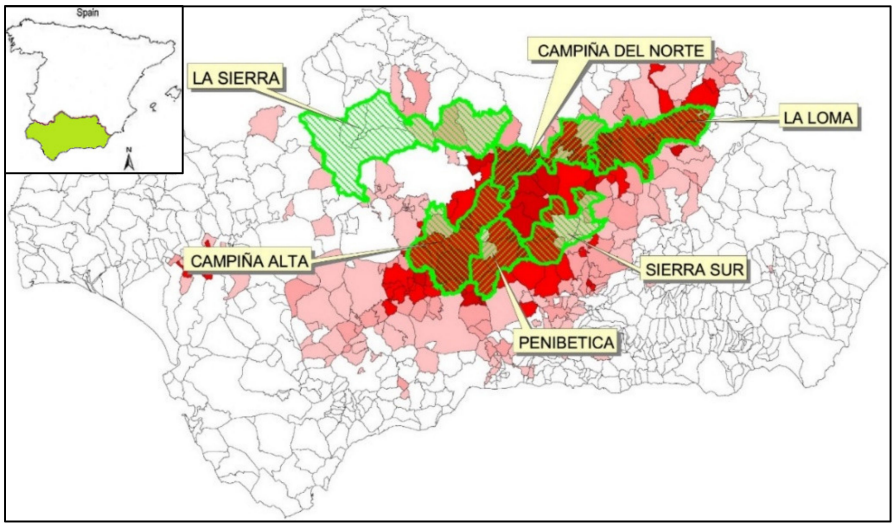

Figure 1. Location of the six agricultural regions included in the survey.

The data collection was carried out via face-to-face interviews with the farmers (lasting around 35 min each) using a structured questionnaire with nine blocks: (1) farm characteristics (location, area of irrigated and rainfed crops, ownership type and farm labor); (2) olive growing characteristics (varieties, plantation age, tree density, type of management: conventional, integrated or organic–and yield); (3) soil and weed management (agricultural practices, weed control—tilled or not tilled—and use of cover crops); (4) olive-pruning (date and desuckering debris management); (5) irrigation system and use of water-soluble fertilizers; (6) fertilization (fertilizers and dosage); (7) crop protection

(chemicals, dosage, and management plans); (8) olive harvest (manual or mechanical); and (9) farmer's socio-economic characteristics (gender, age, professional experience, family size, income share from agriculture, education level, membership of producer organizations, and generational renewal). This information allowed us to calculate the corresponding AEIs (I_k) at the farm level.

The sample of 480 farms is divided into four types of olive orchards: traditional mountain olive groves (rain-fed), traditional low-medium slope olive groves (rain-fed), semi-intensive irrigated olive groves and other types. In this study, the assessment of environmental sustainability focuses on the semi-intensive irrigated type and is based on a subsample of 99 olive farms. The Table 2 summarizes farms' and farmers' characteristics in this subsample:

Table 2. Descriptive statistics from the farmer survey (n = 99).

Category	Variable	Mean	St. Dev.
Farm plantation	Olive grove area (ha)	17.2	26.5
	Plantation density (trees/ha)	98.0	23.0
	Plantation age (years)	99.5	101.2
	Average production (kg of olives/ha·year)	6146	1529
Farmers	Age (years)	50.8	10.7
	Time devoted to agriculture (%)	54.7	42.2
	Family labor (person-days/ha·year)	4.9	8.9
	Hired labor (person-days/ha·year)	5.9	8.7
Farm AEIs related to biodiversity protection	Number of olive grove varieties	1.75	0.81
	Index of biological diversity (dimensionless, [0,1])	0.59	0.18
	Pesticide risk (kg live organism/ha·year)	3666	2314
	Percentage of land with other crops (%)	4.77	13.23
	Percentage of non-cultivated land (%)	0.50	2.14
Farm AEIs related to natural resources conservation	Soil erosion (t soil/ha·year)	9.01	6.10
	Soil organic matter (dimensionless, [0,1])	0.71	0.28
	Nitrogen balance (N kg/ha·year)	−7.27	51.93
	Residual herbicide use (kg active matter/ha·year)	839	670
	Irrigation water use (m^3/ha·year)	686	314
	Energy balance (MJ/ha·year)	9990	4387

Source: Own elaboration based on farmer survey.

3.3.2. Expert Survey

In order to weight the contribution of each AEI to the composite indicator measuring the environmental sustainability of olive farms (*ENV_SUST*), a multidisciplinary group of 22 experts was selected following a judgmental sampling method [55]. The expert panel was primarily composed of scientists from universities and research centers (15), but also contained specialists from the Regional Administration (3) and technical services firms (4). Although it is a non-probability sampling technique, the nature of the technical information required and the homogeneity of the group (in terms of their expertise) suggest that the data gathering is reliable and bias-free [56].

The survey was based on one-to-one interviews, and two sessions were conducted with each expert. After an introduction about the objective of the study and the assessment methods, the questionnaires designed for each of those methods (AHP, BWM and PA) were administered. Each interview took approximately 20–30 min. Since the AHP questionnaire was slightly longer than the one for BWM, the short questionnaire conducted for the PA method was included in the BWM session. To avoid order effects (period and carryover effects), a counterbalanced Latin square design was followed [57,58]. Thus, half of the experts were given the AHP questionnaire first, whereas the other half began with the BWM and PA methods (controlling for the period effect). Then, two days later, the experiments were reversed for each group of experts (controlling for the carryover effect).

Finally, it is worth pointing out that from these questionnaires we obtained the individual AEI weights according to each of the three methods used, for each expert in the panel (w_{kj}), with the subscript k denoting the base indicator and the subscript j denoting the expert considered. However, the weights to be included in Equation (1) are the result of the synthesis of the panel's weights. In this regard, we follow Forman and Peniwati [59], who suggest that group decision-making should be performed by aggregating individual weights using the geometric mean for every weighting method (AHP, BWM and PA):

$$w_k = \sqrt[J]{\prod_{j=1}^{j=J} w_{kj}}. \tag{3}$$

4. Results

4.1. Indicator Weighting

The values of the consistency ratios for AHP and BWM, although not reported here due to space constraints, do not exceed the permissible threshold levels and, hence, provide evidence of the high degree of reliability of the experts' responses.

Table 3 shows the summary statistics of the global weights obtained by the three methods implemented (AHP, BWM and PA).

Table 3. Central tendency (mean) and variability (coefficient of variation, CV) of AEI global weights by weighting method.

AEI	AHP		BWM		PA	
	Mean	**CV**	**Mean**	**CV**	**Mean**	**CV**
Biodiversity protection						
NUMVAR	0.021	119.6%	0.025	67.7%	0.029	76.1%
DIVERSIND	0.146	53.3%	0.145	52.8%	0.142	41.8%
PESTRISK	0.104	82.2%	0.078	92.9%	0.085	76.2%
OTHERCROP	0.066	83.2%	0.067	60.7%	0.070	58.0%
NONCULTIV	0.088	83.9%	0.111	76.7%	0.099	57.3%
Natural resources conservation						
EROSION	0.201	51.2%	0.206	43.8%	0.176	46.2%
ORGMAT	0.112	68.6%	0.113	63.1%	0.112	52.0%
NITROGBAL	0.055	63.7%	0.058	51.3%	0.072	38.4%
RESHERB	0.074	105.5%	0.065	71.2%	0.075	65.1%
WATERUSE	0.084	92.1%	0.074	68.5%	0.082	60.8%
ENERGYBAL	0.050	97.9%	0.058	73.4%	0.058	70.4%
Mean		81.9%		65.6%		58.4%

Source: Own elaboration based on expert survey.

There is consensus about the most important AEI for the protection of biodiversity; namely, the index of biological diversity (*DIVERSIND*). This is followed by the pesticide risk (*PESTRISK*) for AHP, and by the percentage of non-cultivated land (*NONCULTIV*) for BWM and PA. Regarding the second principle, the conservation of natural resources, there is also consensus about the most important indicator, soil erosion (*EROSION*), followed by soil organic matter (*ORGMAT*). It is worth noting that the lowest variability of AEI weights is found with the PA method.

4.1.1. Inter-rater Reliability

Before undertaking the comparison of the weighting methods, we assess the degree of agreement between the experts' AEI weights, using the intraclass correlation coefficient (ICC). Unlike the traditional

correlation coefficient, based on paired observations, the ICC simultaneously considers the group agreement. Higher ICC values indicate higher inter-rater reliability, with 1 indicating perfect agreement and 0 only indicating random agreement [60]. There are 10 different forms of ICC depending on: (1) the statistical model; one-way or two-way models, according to whether the source of variation comes from objects or subjects (raters), respectively; (2) whether raters are considered as random or fixed effects (two-way random-effects model or two-way mixed-effects model, respectively); and (3) the type of agreement: absolute agreement (for the same object, similar scores among raters) or consistency (for the same object, similar ranking among raters) [61,62].

Since the experts are not randomly selected and we are interested in assessing whether or not the AEI weights are equal within each weighting method, we estimate the absolute agreement among experts using a two-way mixed-effects model. The resulting ICCs for AHP, BWM and PA are presented in Table 4.

Table 4. Intraclass correlation coefficients and 95% confidence intervals for the three methods.

Weighting Method	ICC(3,k) Coefficient	Lower Bound	Upper Bound
AHP	0.89	0.82	0.96
BWM	0.93	0.87	0.97
PA	0.92	0.85	0.97

Source: Own elaboration based on expert survey.

To interpret ICC, Cicchetti [63] gives some guidelines: <0.40, 0.40–0.59, 0.60–0.74, and >0.74 for poor, fair, good and excellent reliability, respectively. Additionally, Koo and Li [64] give slightly different intervals: <0.5, 0.5–0.75, 0.75–0.9 and >0.9 for poor, moderate, good and excellent reliability, respectively. In our case, the degree of agreement can be regarded as good to excellent in all three methods. Thus, despite the sample size limitation, all three methods produce consistent assessments of AEI weights. Notwithstanding these outcomes, as commented above, PA produces the least variability in experts' assessments in terms of the coefficient of variation (see Table 3).

4.1.2. Multivariate Comparison of Weights from the Three Methods

In order to compare the AEI weights from the three methods, a within-subjects multivariate analysis of variance (MANOVA) design was implemented. The MANOVA not only reduces the chance of Type I error but can also account for the correlation among the dependent variables [65], and therefore has more power to detect differences among groups [66,67]. The experimental design met the assumptions relating to the measurement of the dependent variables at interval scale, the independence of observations, and adequate sample size (more observations than the number of dependent variables; in this case, 22 observations vs. 11 variables).

Regarding the additional assumptions of the MANOVA, we conclude: (1) the visual check of scatterplots suggests that the condition of linearity (no curvilinear pattern between all pairs of indicators) is met; (2) the conditions of homogeneity of variance-covariance matrices (Box'M statistic = 124.0, p-value = 0.686) and homogeneity of variances (minimum Levene statistic = 0.895, p-value = 0.414) are fulfilled; (3) there is no multicollinearity among the AEIs (maximum r = 0.580 < 0.90 [67]); (4) none of the three methods satisfy the multivariate normality assumption, however, since we have equal group sizes, the MANOVA is robust given the absence of multivariate outliers [68,69]; and (5) no multivariate outliers were identified using Mahalanobis distance (minimum probability equals 0.066). According to this evidence, the MANOVA can be applied to check for significant differences of means among weighting methods. The results of the MANOVA are shown in Table 5.

As Table 5 shows, three out of four multivariate criteria test statistics suggest there is no statistically significant difference in means. Furthermore, as Kuhfeld [70] points out, in the event of a discrepancy between Roy's Largest Root and the other three test statistics, the effect should be considered to not

be significant. In summary, the results suggest that the AEI weights do not depend on the weighting method (AHP, BWM, and PA).

Table 5. Multivariate comparison of AEI weights by weighting method.

Within-Subjects Effect		Value	F	Hypothesis d.f.	Error d.f.	*p*-Value
Statistic	Pillai's Trace	0.541	1.441	18.000	70.000	0.140
	Wilks' Lambda	0.522	1.449	18.000	68.000	0.138
	Hotelling's Trace	0.794	1.455	18.000	66.000	0.136
	Roy's Largest Root	0.589	2.289	9.000	35.000	0.039

Source: Own elaboration based on expert survey.

It is subject to discussion whether the consistency of the three methods depends on the use of the same group of experts. Although further research would be needed in this regard, a random subsampling of the 22 experts was carried out (assigning 8 to AHP, 7 to BWM, and 7 to PA) to compare the AEI weights, yielding the same conclusion based on the MANOVA test (*p*-values: 0.190, 0.256, 0.338 and 0.158, respectively).

4.2. Assessing the Environmental Performance of Irrigated Olive Farms

Table 6 shows the main descriptive statistics of the 15 distributions of the composite indicator *ENV_SUST* (3 weighting methods × 5 values of lambda) obtained for the 99 olive farms sampled. As can be clearly observed, the index values calculated vary more due to the compensation parameter λ than due to the weighting method. In fact, while there is no statistically significant difference among the means and the variances of the *ENV_SUST* distributions obtained for every single λ, it can be proved that the average values significantly decrease as the compensation parameter decreases.

Table 6. Descriptive statistics of the composite indicator (*ENV_SUST*) by compensation parameter (λ) and weighting method.

	Method	Mean	Min.	Max.	St. Dev.	CV
$\lambda = 1$	AHP	0.549	0.140	0.783	0.118	21.5%
	BWM	0.533	0.151	0.782	0.116	21.9%
	PA	0.536	0.106	0.777	0.120	22.3%
$\lambda = 0.75$	AHP	0.414	0.105	0.587	0.090	21.8%
	BWM	0.402	0.113	0.587	0.089	22.1%
	PA	0.404	0.079	0.582	0.091	22.6%
$\lambda = 0.5$	AHP	0.278	0.070	0.512	0.066	23.6%
	BWM	0.270	0.076	0.496	0.065	23.9%
	PA	0.272	0.053	0.510	0.066	24.4%
$\lambda = 0.25$	AHP	0.143	0.035	0.483	0.050	35.0%
	BWM	0.139	0.038	0.475	0.050	35.7%
	PA	0.139	0.026	0.482	0.050	36.1%
$\lambda = 0$	AHP	0.007	0.000	0.454	0.052	700.9%
	BWM	0.007	0.000	0.454	0.052	696.6%
	PA	0.007	0.000	0.454	0.052	718.5%

Source: Own elaboration based on expert and farmer surveys.

Although the discussion about the most suitable value of the compensation parameter to measure farms' environmental performance is beyond the scope of this paper, it is worth pointing out that composite indicators based on complete compensability (i.e., $\lambda = 1$) have been criticized because trade-offs between base indicators could be considered incompatible with the concept of

sustainability [13,40]. It is thus reasonable to opt for indexes that allow partial compensability (i.e., *ENV_SUST* for $0 < \lambda < 1$). In any case, the selection of the most suitable value of λ is an issue that remains open for discussion in future studies [18].

The assessment of the rankings of farms produced by the composite indicator *ENV_SUST* using the three weighting methods, for each of the five values considered for the compensation parameter ($\lambda = 0$, $\lambda = 0.25$, $\lambda = 0.5$, $\lambda = 0.75$ and $\lambda = 1$), is carried out using the Kendall's coefficient of concordance, or Kendall's W [71], mathematically:

$$W = \frac{12S^2}{m^2(k^3 - k)} - \frac{3(k+1)}{k-1}, \tag{4}$$

where $m = 3$ (AHP, BWM, and PA), $k = 99$ (number of farms), r_{ij} = ranking of farm i by method j, $R_i = \sum_{j=1}^{m} r_{ij}$ and $S^2 = \sum_{i=1}^{k} R_i^2$. To test the null hypothesis of no agreement among the methods, that is $W = 0$, the statistic to be used is $m(k-1)W \sim \chi^2_{k-1}$.

As the results in Table 7 show, the ranking of the irrigated olive farms based on their environmental performance (i.e., values of composite indicator *ENV_SUST*), for any lambda considered, does not depend on the weighting method used (AHP, BWM, or PA). Furthermore, when considering the 15 rankings simultaneously, the overall Kendall's *W* indicates a strong level of concordance (Kendall *W's* = 0.705). This indicates that, regardless of the weighting method or the compensation parameter, all *ENV_SUST* measurements provide similar rankings of the sampled irrigated olive farms.

Table 7. Kendall's coefficient of concordance of farms' sustainability ranking with the three weighting methods.

	Compensation Parameter					
	λ = 1	**λ = 0.75**	**λ = 0.5**	**λ = 0.25**	**λ = 0**	**Overall**
Kendall's *W*	1.000	0.987	0.988	0.988	1.000	0.705
p-value	0.000	0.000	0.000	0.000	0.000	0.000

Source: Own elaboration based on experts' and farmers' survey.

5. Conclusions

Measuring environmental sustainability in agriculture through the construction of composite indicators is a widespread practice, although it is a tough task. An especially challenging aspect is the choice of the most appropriate methods to normalize, weight and aggregate the large set of base agri-environmental indicators usually considered. In particular, the results of environmental composite indicators are heavily dependent on how the base indicators are weighted (i.e., if the indicator weights can accurately synthetize the relative importance of each AEI included in the index built). For this reason, our main aim in this paper was to analyze the robustness of three alternative weighting methods (AHP, BWM and PA). To that end, we consulted the opinions of a panel of experts and compared the results obtained for an environmental index implemented in the real-world case study of irrigated olive farms in Spain.

In light of the results, we can identify three main findings: first, there is a high level of consistency in experts' assessments of AEI weights derived from the three weighting methods; second, there are no statistically significant differences in the means of the AEI weights estimated with the three methods; and third, the values of the composite indicator built (*ENV_SUST*) using the three alternative weighting techniques produce similar rankings of the irrigated olive farms in terms of their environmental performance. Further evidence regarding the consistency of the weighting methods is needed to confirm whether this finding is generalizable. In any case, it can be hypothesized that similarly consistent results can be expected whenever the composite indicator construction (i.e., weighting) is focused on the assessment of technical concepts, where expert opinion is rooted in empirical knowledge.

Overall, these findings provide useful empirical insights into the robustness of the two multicriteria methods, AHP and BWM, and the more straightforward PA, as weighting techniques to be used when constructing composite environmental indicators. However, although the three methods are valid, feasible tools to determine the weights of the individual base indicators, it is worth noting that the PA could be cumbersome if there is a large number of indicators to be included in a single index (e.g., more than six) [44].

Beyond the methodological focus, it is also worth pointing out that the *ENV_SUST* composite indicator implemented in this study is sufficiently stable and methodologically sound to be used in the design of targeted policy interventions aimed at measuring the environmental sustainability of farms. In this context, there is still room for research on the practical implementation of this environmental index to analyze the heterogeneous environmental performance among farms (i.e., determining reference and threshold values) and track changes in agricultural practices (i.e., irrigation or fertilization). This practical information could be useful for policy-makers in the design of the results-based agri-environmental programs, to set the level of payments to be granted to each particular farm or any other policy instruments with a similar purpose (i.e., fiscal or qualitative rewards).

Author Contributions: Conceptualization, J.A.G.-L.; methodology, J.A.G.-L. and M.A.; formal analysis, M.A., J.A.G.-L., and M.D.G.-B.; writing—original draft preparation, J.A.G.-L., M.A., and M.D.G.-B.; writing—review and editing, J.A.G.-L., M.A., and M.D.G.-B.; funding acquisition, J.A.G.-L. and M.D.G.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Andalusian Department of Economy and Knowledge and the European Regional Development Fund (ERDF) through the research project FINAGUA (UCO-1264548). These funding institutions had no involvement in the conduction of the research nor the preparation of the paper.

Acknowledgments: The authors are gratefully to anonymous reviewers for their comments and suggestions, which have substantially improved the quality of the paper. We are also grateful for the collaboration of the 22 academics and professionals who took part in the panel of experts weighting the AEIs (Francisco Bruno, Juan Cano, Juan Castro, Sergio Colombo, Juan Domínguez, Manuel Estévez, José A. Gómez, Emilio González, Rubén Granado, Carlos Gutiérrez, Javier Hidalgo, Lorenzo León, Salvador Nadal, Ángel Navas, Rafaela M. Ordóñez, Luis Padilla, José M. Penco, Macario Rodríguez, Raúl de la Rosa and Anastasio J. Villanueva; the two remaining experts prefer to remain anonymous). The authors also thank the participation of the 480 farmers sampled for providing actual information about the agricultural management of their olive groves.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hansen, J.W. Is agricultural sustainability a useful concept? *Agric. Syst.* **1996**, *50*, 117–143. [CrossRef]
2. Pretty, J. Agricultural sustainability: Concepts, principles and evidence. *Philos. Trans. R. Soc. B* **2008**, *363*, 447–465. [CrossRef]
3. Bockstaller, C.; Guichard, L.; Makowski, D.; Aveline, A.; Girardin, P.; Plantureux, S. Agri-environmental indicators to assess cropping and farming systems: A review. In *Sustainable Agriculture*; Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 725–738. [CrossRef]
4. Lin, B.; Chen, X.; Yao, H.; Chen, Y.; Liu, M.; Gao, L.; James, A. Analyses of landuse change impacts on catchment runoff using different time indicators based on SWAT model. *Ecol. Indic.* **2015**, *58*, 55–63. [CrossRef]
5. Maes, J.; Liquete, C.; Teller, A.; Erhard, M.; Paracchini, M.L.; Barredo, J.I.; Grizzetti, B.; Cardoso, A.; Somma, F.; Petersen, J.-E.; et al. An indicator framework for assessing ecosystem services in support of the EU Biodiversity Strategy to 2020. *Ecosyst. Serv.* **2016**, *17*, 14–23. [CrossRef]
6. Mastrorilli, M.; Rana, G.; Verdiani, G.; Tedeschi, G.; Fumai, A.; Russo, G. Economic evaluation of hydrological ecosystem services in Mediterranean river basins applied to a case study in southern Italy. *Water* **2018**, *10*, 241. [CrossRef]
7. Yli-Viikari, A.; Hietala-Koivu, R.; Huusela-Veistola, E.; Hyvönen, T.; Perälä, P.; Turtola, E. Evaluating agri-environmental indicators (AEIs)—Use and limitations of international indicators at national level. *Ecol. Indic.* **2007**, *7*, 150–163. [CrossRef]

8. Piorr, H.-P. Environmental policy, agri-environmental indicators and landscape indicators. *Agric. Ecosyst. Environ.* **2003**, *98*, 17–33. [CrossRef]
9. Sabiha, N.E.; Salim, R.; Rahman, S.; Rola-Rubzen, M.F. Measuring environmental sustainability in agriculture: A composite environmental impact index approach. *J. Environ. Manag.* **2016**, *166*, 84–93. [CrossRef]
10. Seidel, C.; Heckelei, T.; Lakner, S. Conventionalization of organic farms in Germany: An empirical investigation based on a composite indicator approach. *Sustainability* **2019**, *11*, 2934. [CrossRef]
11. OECD (Organization for Economic Co-operation and Development); JRC (Joint Research Centre). *Handbook on Constructing Composite Indicators. Methodology and User Guide*; OECD Publishing: Paris, France, 2008. [CrossRef]
12. Dobbie, M.J.; Dail, D. Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecol. Indic.* **2013**, *29*, 270–277. [CrossRef]
13. Gan, X.; Fernandez, I.C.; Guo, J.; Wilson, M.; Zhao, Y.; Zhou, B.; Wu, J. When to use what: Methods for weighting and aggregating sustainability indicators. *Ecol. Indic.* **2017**, *81*, 491–502. [CrossRef]
14. Becker, W.; Paruolo, P.; Saisana, M.; Saltelli, A. Weights and importance in composite indicators: Mind the gap. In *Handbook of Uncertainty Quantification*; Ghanem, R., Higdon, D., Owhadi, H., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 1187–1216. [CrossRef]
15. El Gibari, S.; Gómez, T.; Ruiz, F. Building composite indicators using multicriteria methods: A review. *J. Bus. Econ.* **2018**, *89*, 1–24. [CrossRef]
16. Krajnc, D.; Glavič, P. How to compare companies on relevant dimensions of sustainability. *Ecol. Econ.* **2005**, *55*, 551–563. [CrossRef]
17. Gómez-Limón, J.A.; Riesgo, L. Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. *J. Environ. Manag.* **2009**, *90*, 3345–3362. [CrossRef] [PubMed]
18. Gómez-Limón, J.A.; Sanchez-Fernandez, G. Empirical evaluation of agricultural sustainability using composite indicators. *Ecol. Econ.* **2010**, *69*, 1062–1075. [CrossRef]
19. Rezaei, J. Best-worst multi-criteria decision-making method. *Omega* **2015**, *53*, 49–57. [CrossRef]
20. Saisana, M.; Saltelli, A.; Tarantola, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2005**, *168*, 307–323. [CrossRef]
21. Baquero, O.F.; Gallego-Ayala, J.; Gine-Garriga, R.; de Palencia, A.J.F.; Perez-Foguet, A. The influence of the human rights to water and sanitation normative content in measuring the level of service. *Soc. Indic. Res.* **2017**, *133*, 763–786. [CrossRef]
22. Gómez-Limón, J.A.; Riesgo, L. Sustainability assessment of olive groves in Andalusia: A methodological proposal. *New Medit* **2012**, *11*, 39–49.
23. Gómez-Limón, J.A.; Arriaza, M. *Evaluación de la Sostenibilidad de las Explotaciones de Olivar en Andalucía*; Analistas Económicos de Andalucía: Málaga, Spain, 2011.
24. Gómez Calero, J.A. (Ed.) *Sostenibilidad de la Producción de Olivar en Andalucía*; Consejería de Agricultura y Pesca, Junta de Andalucía: Sevilla, Spain, 2009.
25. Taguas, E.V.; Gómez, J.A. Vulnerability of olive orchards under the current CAP (Common Agricultural Policy) regulations on soil erosion: A study case in Southern Spain. *Land Use Policy* **2015**, *42*, 683–694. [CrossRef]
26. Rodríguez-Entrena, M.; Arriaza, M.; Gómez-Limón, J.A. Determining economic and social factors in the adoption of cover crops under mower control in olive groves. *Agroecol. Sustain. Food Syst.* **2014**, *38*, 69–91. [CrossRef]
27. Rey, P.J. Preserving frugivorous birds in agro-ecosystems: Lessons from Spanish olive orchards. *J. Appl. Ecol.* **2011**, *48*, 228–237. [CrossRef]
28. Carpio, A.J.; Oteros, J.; Tortosa, F.S.; Guerrero-Casado, J. Land use and biodiversity patterns of the herpetofauna: The role of olive groves. *Acta Oecol.* **2016**, *70*, 103–111. [CrossRef]
29. Hermosin, M.C.; Calderon, M.J.; Real, M.; Cornejo, J. Impact of herbicides used in olive groves on waters of the Guadalquivir river basin (southern Spain). *Agric. Ecosyst. Environ.* **2013**, *164*, 229–243. [CrossRef]
30. Gómez-Limón, J.A.; Arriaza, M.; Villanueva, A.J. Typifying irrigated areas to support policy design and implementation: The case of the Guadalquivir river basin. *Irrig. Drain.* **2013**, *62*, 322–329. [CrossRef]
31. Expósito, A.; Berbel, J. Drivers of irrigation water productivity and basin closure process: Analysis of the Guadalquivir River Basin (Spain). *Water Resour. Manag.* **2019**, *33*, 1439–1450. [CrossRef]

32. Rodríguez-Entrena, M.; Arriaza, M. Adoption of conservation agriculture in olive groves: Evidences from southern Spain. *Land Use Policy* **2013**, *34*, 294–300. [CrossRef]
33. Carmona-Torres, C.; Parra-López, C.; Sayadi, S.; Chirosa-Ríos, M. A public/private benefits framework for the design of polices oriented to sustainability in agriculture: An application to olive growing. *Land Use Policy* **2016**, *58*, 54–69. [CrossRef]
34. Rodríguez-Entrena, M.; Colombo, S.; Arriaza, M. The landscape of olive groves as a driver of the rural economy. *Land Use Policy* **2017**, *65*, 164–175. [CrossRef]
35. Ness, B.; Urbel-Piirsalu, E.; Anderberg, S.; Olsson, L. Categorising tools for sustainability assessment. *Ecol. Econ.* **2007**, *60*, 498–508. [CrossRef]
36. Bockstaller, C.; Guichard, L.; Keichinger, O.; Girardin, P.; Galan, M.-B.; Gaillard, G. Comparison of methods to assess the sustainability of agricultural systems: A review. In *Sustainable Agriculture*; Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 769–784. [CrossRef]
37. Van Cauwenbergh, N.; Biala, K.; Bielders, C.; Brouckaert, V.; Franchois, L.; Garcia Cidad, V.; Hermy, M.; Mathijs, E.; Muys, B.; Reijnders, J.; et al. SAFE. A hierarchical framework for assessing the sustainability of agricultural systems. *Agric. Ecosyst. Environ.* **2007**, *120*, 229–242. [CrossRef]
38. Bell, S.; Morse, S. *Sustainability Indicators. Measuring the Incommensurable?* Earthscan: London, UK, 2008. [CrossRef]
39. Pannell, D.J.; Glenn, N.A. A framework for the economic evaluation and selection of sustainability indicators in agriculture. *Ecol. Econ.* **2000**, *33*, 135–149. [CrossRef]
40. Talukder, B.; Hipel, K.W.; vanLoon, G.W. Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques. *Resources* **2017**, *6*, 66. [CrossRef]
41. Pollesch, N.L.; Dale, V.H. Normalization in sustainability assessment: Methods and implications. *Ecol. Econ.* **2016**, *130*, 195–208. [CrossRef]
42. Munda, G. Measuring sustainability: A multi-criterion framework. *Environ. Dev. Sustain.* **2005**, *7*, 117–134. [CrossRef]
43. Böhringer, C.; Jochem, P.E.P. Measuring the immeasurable—A survey of sustainability indices. *Ecol. Econ.* **2007**, *63*, 1–8. [CrossRef]
44. Zardari, N.H.; Ahmed, K.; Shirazi, S.M.; Yusop, Z.B. *Weighting Methods and their Effects on Multi-Criteria Decision Making Model Outcomes in Water Resources Management*; Springer: Cham, Switzerland, 2015.
45. Rowley, H.V.; Peters, G.M.; Lundie, S.; Moore, S.J. Aggregating sustainability indicators: Beyond the weighted sum. *J. Environ. Manag.* **2012**, *111*, 24–33. [CrossRef]
46. Díaz-Balteiro, L.; Romero, C. In search of a natural systems sustainability index. *Ecol. Econ.* **2004**, *49*, 401–405. [CrossRef]
47. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw-Hill: New York, NY, USA, 1980.
48. Molinos-Senante, M.; Gómez, T.; Garrido-Baserba, M.; Caballero, R.; Sala-Garrido, R. Assessing the sustainability of small wastewater treatment systems: A composite indicator approach. *Sci. Total Environ.* **2014**, *497*, 607–617. [CrossRef]
49. Saaty, T.L. Decision-making with the AHP: Why is the principal eigenvector necessary. *Eur. J. Oper. Res.* **2003**, *145*, 85–91. [CrossRef]
50. Dodd, F.J.; Donegan, H.A.; McMaster, T.B.M. A statistical approach to consistency in AHP. *Math. Comput. Model.* **1993**, *18*, 19–22. [CrossRef]
51. Wang, T.; Fu, Y. Constructing composite indicators with individual judgements and best–worst method: An illustration of value measure. *Soc. Indic. Res.* **2019**. [CrossRef]
52. Omrani, H.; Alizadeh, A.; Amini, M. A new approach based on BWM and MULTIMOORA methods for calculating semi-human development index: An application for provinces of Iran. *Socio-Econ. Plan. Sci.* **2019**, 100689. [CrossRef]
53. van Calker, K.J.; Berentsen, P.B.M.; Romero, C.; Giesen, G.W.J.; Huirne, R.B.M. Development and application of a multi-attribute sustainability function for Dutch dairy farming systems. *Ecol. Econ.* **2006**, *57*, 640–658. [CrossRef]
54. Pöyhönen, M.; Hämäläinen, R.P. On the convergence of multiattribute weighting methods. *Eur. J. Oper. Res.* **2001**, *129*, 569–585. [CrossRef]

55. Purvis, G.; Louwagie, G.; Northey, G.; Mortimer, S.; Park, J.; Mauchline, A.; Finn, J.; Primdahl, J.; Vejre, H.; Vesterager, J.P.; et al. Conceptual development of a harmonised method for tracking change and evaluating policy in the agri-environment: The Agri-environmental Footprint Index. *Environ. Sci. Policy* **2009**, *12*, 321–337. [CrossRef]
56. Jager, J.; Putnick, D.L.; Bornstein, M.H. More than just convenient: The scientific merits of homogeneous convenience samples. *Monogr. Soc. Res. Child. Dev.* **2017**, *82*, 13–30. [CrossRef]
57. Richardson, J.T.E. The use of Latin-square designs in educational and psychological research. *Educ. Res. Rev.* **2018**, *24*, 84–97. [CrossRef]
58. Bradley, J.V. Complete counterbalancing of immediate sequential effects in a Latin square design. *J. Am. Stat. Assoc.* **1958**, *53*, 525–528. [CrossRef]
59. Forman, E.H.; Peniwati, K. Aggregating individual judgments and priorities with the analytic hierarchy process. *Eur. J. Oper. Res.* **1998**, *108*, 165–169. [CrossRef]
60. Hallgren, K.A. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor. Quant. Methods Psychol.* **2012**, *8*, 23–34. [CrossRef] [PubMed]
61. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [CrossRef] [PubMed]
62. McGraw, K.O.; Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1996**, *1*, 30–46. [CrossRef]
63. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **1994**, *6*, 284–290. [CrossRef]
64. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]
65. Huberty, C.J.; Morris, J.D. Multivariate analysis versus multiple univariate analyses. *Psychol. Bull.* **1989**, *105*, 302–308. [CrossRef]
66. Field, A.; Miles, J.; Field, Z. *Discovering Statistics Using R.*; Sage Publications: Los Angeles, CA, USA, 2012.
67. Tabachnick, B.G.; Fidell, L.S.; Ullman, J.B. *Using Multivariate Statistics*; Pearson: Boston, MA, USA, 2013.
68. Mardia, K.V. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika* **1971**, *58*, 105–121. [CrossRef]
69. Vallejo, G.; Ato, M. Robust tests for multivariate factorial designs under heteroscedasticity. *Behav. Res. Methods* **2012**, *44*, 471–489. [CrossRef]
70. Kuhfeld, W.F. A note on Roy's largest root. *Psychometrika* **1986**, *51*, 479–481. [CrossRef]
71. Teles, J. Concordance coefficients to measure the agreement among several sets of ranks. *J. Appl. Statist.* **2012**, *39*, 1749–1764. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).