Ecological Indicators 143 (2022) 109374

Contents lists available at ScienceDirect

Ecological Indicators

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

Comparison of soil quality indexes calculated by network and principal component analysis for carbonated soils under different uses

Juan Pedro Martín-Sanz [a,b,*], Ana de Santiago-Martín [c], Inmaculada Valverde-Asenjo [a], José Ramón Quintana-Nieto [a], Concepción González-Huecas [a], Antonio L. López-Lafuente [a]

[a] *Universidad Complutense de Madrid, Facultad de Farmacia, U.D. Edafología, Dpt. Química en Ciencias Farmacéuticas, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain*

[b] *Universidad de Santiago de Compostela, Facultad de Farmacia, Dpto. de Edafología y Química Agrícola, Praza do Seminario de Estudos Galegos, s/n, 15705 Santiago de Compostela, A Coruña, Spain*

[c] *Instituto IMDEA Agua, Avenida Punto Com, 2, 28805 Alcalá de Henares, Madrid, Spain*

ARTICLE INFO

Keywords:
Soil quality
Land use
Network analysis
Principal component analysis
Mediterranean calcareous soil

ABSTRACT

There is an urgent need to conserve and improve the quality of agricultural soils in the coming decades. Decision tools capable of providing reliable information about soil quality are needed, and soil quality index (SQI) is one of the most used. Principal component analysis (PCA) is the common methodology to calculate it, however in some cases fails to differentiate soil quality properly. Therefore, the aim of this work is to assess a SQI through a different methodology as network analysis (NTA) and compare it with PCA, assuming that soil uses affect soil qualities differently. From soils with different uses (rainfed, olive grove and forest) network analysis and principal component analysis have been used to select a minimum dataset (MDS) to generate SQI from 36 physical, chemical and biological soil variables. Using NTA, geometric mean of the enzyme activities (GMEAN), bulk density (BD) and phosphatase activity (phos) where selected as indicators, while PCA selected total organic carbon (TOC), free Fe oxides (FeF), crystalline Mn oxides (MnX), pH, electrical conductivity (EC) and percentage of coarse sand (CS). Four SQI were calculated from each MDS through linear and non-linear scoring equations and by additive integration and weights. The SQI generated by NTA were more useful than those generated by PCA, as in addition to having fewer indicators they were able to better differentiate the uses in the study. This greater resolution capacity of the NTA would be the consequence of a better selection of indicators using this method than using PCA.

1. Introduction

The increased demand for food caused by the rise in the world population, estimated to be 25–31 % by 2050 (United Nations, 2019), is a major challenge for humankind. This greater demand poses a risk to soil quality, defined by Karlen et al. (1997) as “the capacity of a specific kind of soil to function, within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality, and support human health and habitation”. Understanding the parameters that determine the quality of agricultural soils can improve their management (Herrick, 2000). In this context, the Soil Health and Food Mission Board has recently proposed a mission to the European Commission entitled “Caring for Soil is Caring for Life”, driven by the pressing need to reduce soil degradation in the European Union and improve the programmes for monitoring soil quality in all the member states, taking into account the variability in soil type, land use

Abbreviations: WHC, Water holding capacity; BD, bulk density; CS, coarse sand; FS, fine sand; EC, electrical conductivity; %$CaCO_3$, $CaCO_3$ equivalent; TN, total nitrogen; TOC, total organic carbon; Pav, phosphorus available; MnA, amorphous Mn oxides; MnF, free Mn oxides; MnX, crystalline Mn oxides; FeA, amorphous Fe oxides; FeF, free Fe oxides; FeX, crystalline Fe oxides; Ninor, Inorganic nitrogen; Ccw, cold water soluble carbon; Ncw, cold water soluble nitrogen; Chw, hot water soluble carbon; Nhw, hot water soluble nitrogen; aglu, α-glucosidase activity; aryln, arylamidase activity; aryls, arylsulphatase activity; bglu, β-glucosidase activity; nag, N-acetyl-glucosaminidase activity; phos, phosphatase activity; pac, acid phosphatase activity; pak, alkaline phosphatase activity; ure, urease activity; dh, dehydrogenase activity; GMEAN, geometric mean of enzyme activities.

* Corresponding author at: Universidad Complutense de Madrid, Facultad de Farmacia, U.D. Edafología, Dpt. Química en Ciencias Farmacéuticas, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain.

E-mail address: juanpmar@ucm.es (J.P. Martín-Sanz).

https://doi.org/10.1016/j.ecolind.2022.109374
Received 4 July 2022; Received in revised form 22 August 2022; Accepted 24 August 2022
Available online 30 August 2022

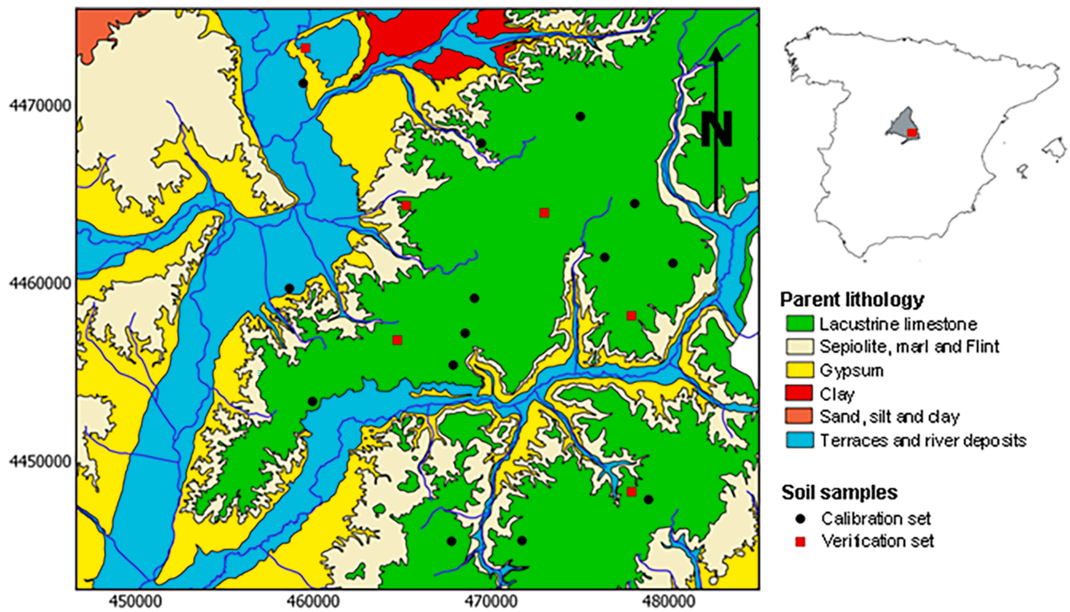

Fig. 1. Location map and distribution of sampling sites.

and climate.

Different types of uses and agricultural management systems may cause alterations in the physical, chemical and biological properties of the soils (Sánchez-Navarro et al., 2015) and negatively affect their quality (Rezapour, 2014). Rainfed agricultural soils are poorer than irrigated soils (Hamidi Nehrani et al., 2020), and ploughed soils have a lower quality than soils with limited ploughing or which are undisturbed (Wander and Bollero, 1999). Marzaioli et al. (2010) observed poorer soil quality in croplands when the spaces between the cultivation rows are left uncovered, and an improvement in quality in croplands with herb cover, while mixed forest soils had the highest quality. A more sustainable agricultural use can therefore be achieved by comparing the quality of agricultural soils in different zones and subject to different management regimes (Karlen et al., 1997), or in the same location but over a period of time (Nortcliff, 2002).

Due to economic and time constraints, it is unfeasible to conduct a systematic study with all the physical, chemical and biological parameters that influence soil quality; in addition, the information provided may be very copious and complex to interpret. It is therefore more practical to select a minimum set of representative quality indicators for the soils to be studied (Bünemann et al., 2018). Ideally, these indicators should (Bünemann et al., 2018; Doran and Zeiss, 2000; Nortcliff, 2002): i) be related with a threat, function or a soil ecosystem service, ii) be simple and cheap to determine, iii) provide reliable measurements with standardised procedures, iv) be sensitive to changes in agricultural management, and v) have limited patterns of spatial–temporal variation.

Soil quality indices (SQI) are used to select and integrate soil quality indicators in a single index, and serve as a management tool to provide land managers with all the most important information to facilitate decision-making in matters of agroecosystem management (Andrews and Carroll, 2001). SQI are calculated from a minimum data set (MDS) formed by indicators selected by means of statistical techniques from an initial set of physical, chemical and biological soil properties.

Principal component analysis (PCA) is the most widely used statistical technique for selecting the indicators in a MDS (Bünemann et al., 2018). However, it has been observed in some cases that this technique is unable to select indicators that reflect the differences in the quality of soils under different management systems (Askari and Holden, 2014; Hamidi Nehrani et al., 2020).

Another statistical technique with the capacity to distinguish the importance of each variable within a group is network analysis (NTA). This type of analysis emerged in the 1930s as a tool in the field of social studies and consists of graphically representing the relations between the people belonging to a group. It is currently used in various areas related with soil and the environment to examine the relations between the study variables and determine their role in the system's functioning (Barberán et al., 2012; Liu et al., 2015; Wang et al., 2019).

We start with the hypothesis that network analysis allows the selection of indicators that represent soil quality and provide information on soil processes, and that the MDS selected with NTA will calculate a SQI with a similar or even greater capacity to differentiate between uses than the MDS obtained with PCA. Assuming that soil use modifies its quality differently, the main aim of this work is to assess a soil quality index (SQI) through network analysis (NTA), and compare it with a widely used method like principal component analysis (PCA). Emphasis is placed on determining the methodology that best classifies and characterises soil quality under three different soil uses: cereal under rainfed management (hereinafter rainfed), olive grove and forest.

2. Material and methods

2.1. Description of the study zone and sampling

The study zone was located to the southeast of the city of Madrid (Spain). The geology of the area is determined by four main geological materials: i) a calcareous formation known as "Calizas del Páramo" of Tertiary-era greyish-white lacustrine limestone alternating with marly limestone, compact marls and sandy reddish marls with boulders that cover most of the study area; ii) gypsum outcrops in the river valley areas; iii) a detrital series consisting mainly of conglomerate, sandstone, sands, clay, marl and levels of flint; and iv) the terraces and river deposits (Fig. 1).

The study area is located at an altitude of 715 m and has a climate between Csa (Mediterranean) and Bsk (semi-arid cold), according to the Köppen-Geiger classification (Climate-data, 2020). Average annual temperature is 13.8 °C (25.7 °C in the warmest month of the year, July, and 4.6 °C in the coldest month, January) and total annual rainfall is 440 mm year-1 (8 mm in the driest month, July, and 65 mm in the month of highest rainfall, October) (data for Arganda del Rey, Madrid, located in the centre of the study area) (Climate-data, 2020). The predominant soils in the area have been characterised as Calcic Luvisols

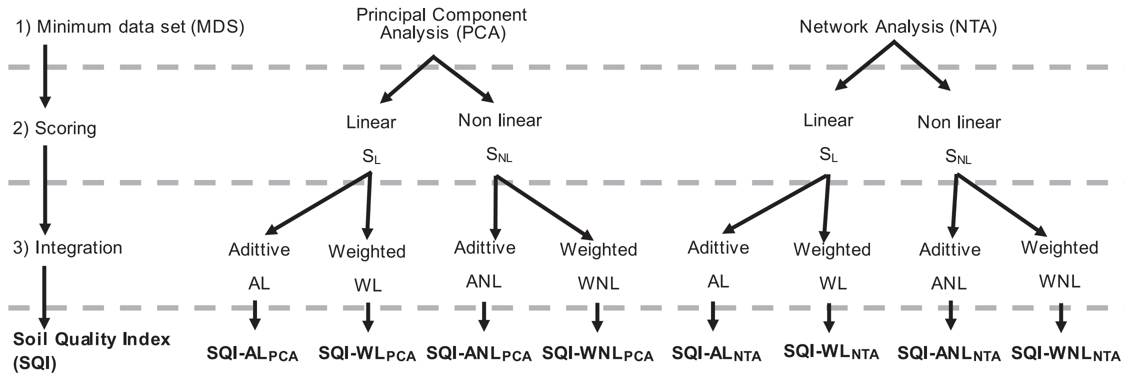

Fig. 2. Diagram of the steps followed to obtain the SQI in this work.

(LVk), Haplic Calcisols (Clh) and Calcaric Regosols (RGc) (Comunidad de Madrid, CSIC, 1990), according to the current FAO nomenclature system (IUSS Working Group WRB, 2015). The main soil use is for agriculture, especially wheat, barley and olive, and the predominant natural plant species are Stipa tenacissima, Quercus ilex, Quercus coccifera and Thymus vulgaris.

The sampling points were selected based on: i) their carbonate origin, as this is the predominant material in the area; ii) their agricultural use – cereal and olive grove – since they are the major crops; iii) the existence of nearby forest soils to serve as a control; and iv) that the three uses (cereal, olive grove and forest) were adjacent to reduce the variability in the soil-forming parameters unrelated to anthropogenization. Based on these criteria, 20 sampling locations were selected. Soil samples were collected in each location in March 2017, in each location three samples were taken, one for each of the three uses, rainfed, olive grove and forest, thus generating a final number of 60 samples (20 samples for each use studied) (Martín-Sanz et al., 2018).

Composite samples were collected at each point, generated from five subsamples according to a crosswise diagram aligned with the points of the compass, taking one subsample at the central point and four at the ends at a distance of 5 m from the centre. The sampling depth was from 0 to 30 cm, as this is the typical ploughing depth. The subsamples were mixed *in situ*; one fraction of the compound sample was refrigerated for subsequent biological analysis, and another unrefrigerated fraction was air-dried for all the other analyses.

2.2. Soil parameters analysed

To ensure the construction of representative soil quality indices with the highest possible capacity to discriminate between uses, the soils were characterized as described below by means of 36 physical, chemical and biological parameters related with soil functionality.

Water holding capacity (WHC) and bulk density (BD) were determined using the Soil Survey Staff procedure (2014) on the unaltered sample taken at each sampling point. Soil particle size was determined using sieved air-dried samples, and total sand (S), coarse sand (CS), fine sand (FS), silt and clay were differentiated according to the ISSS with Robinson's pipette method, after oxidation of the organic matter with H_2O_2 (ISRIC, 2002).

Chemical properties were analysed in <2 mm sieved air-dried samples. pH and electrical conductivity (EC) were determined in soil:water suspensions of 1:2.5 and 1:5 (w:v) respectively (ISRIC, 2002). Total organic carbon content (TOC) was determined by the wet-oxidation method of Walkley and Black (1934), assessing the excess of dichromate with an automatic Metrohm 888 TRITANDO and Metrohm 665 DOSIMAT valuator. Total nitrogen (TN) was determined by combustion with a LECO CHNS-932 analyser. The TOC/TN ratio was calculated. Cold water soluble carbon (Ccw) and hot water soluble carbon (Chw), and cold water soluble nitrogen (Ncw) and hot water soluble nitrogen (Nhw) were extracted according to the method of Ghani et al. (2003) and quantified with a TC/TN ANALYTIKJENA MICRO N/C analyser. Soluble inorganic N (Ninor) content was calculated as the sum of soluble NO_3^-, NO_2^- and NH_4^+ content; anionic forms were extracted from a 1:5 soil:water ratio (w:v) and determined with a Metrohm 761 COMPACT IC ion chromatograph with an automatic Metrohm 838 ADVANCED SAMPLE PROCESSOR carousel; and soluble NH_4^+ content was extracted according to Keeney and Nelson (1982) and quantified by means of UV–visible spectrophotometry with a TECAN NANOQUANT INFINITE M200 PRO multiwell plate reader. Equivalent calcium carbonate content (%$CaCO_3$) was analysed with the acid neutralisation method (ISRIC, 2002). Assimilable phosphorus (Pav) was determined with the method of Olsen and Sommers (1982) and quantified by UV–visible spectrophotometry. Amorphous Fe and Mn oxide content (FeA and MnA) were obtained by extraction with oxalate acid, and free Fe and Mn oxides (FeF and MnF) were obtained by extraction with citrate-dithionite (ISRIC, 2002), and quantified by means of AAS (Analytikjena NovAA 300); crystalline Fe and Mn oxides (FeX and MnX) were obtained by difference with the former.

To identify the biological characteristics of the soils, in the month after sampling, enzyme activities were determined in <2 mm sieved soil samples which were refrigerated until the corresponding analyses were carried out. A total of nine enzyme activities involved in the main biogeochemical soil nutrient cycles were determined according to standard ISO 20130 (ISO, 2018): i) related with the C cycle: alpha-glucosidase (aglu) and beta-glucosidase (bglu); ii) related with the N cycle: arylamidase (aryln), *N*-acetyl-glucosaminidase (nag) and urease (ure); iii) related with the P cycle: phosphatase (phos), acid phosphatase (pac) and alkaline phosphatase (pak); and iv) related with the S cycle: arylsulphatase (aryls). Additionally, endocellular dehydrogenase enzyme activity (dh) was determined as an indicator of the activity of living microbial populations according to Schaefer (1963). The geometric mean of these enzyme activities, GMEAN, was used as a global indicator of enzyme activity due to its sensitivity and because it has less temporal variability than the enzyme activities determined individually (García-Ruiz et al., 2008; Paz-Ferreiro and Fu, 2016).

2.3. Statistical analyses

2.3.1. Determination of the soil quality index (SQI)

The initial set of 20 sampling locations were randomly split into two different sets. One set, denoted the calibration set, was used to obtain the SQI with 70 % of the total samples (14 sampling locations with three soil uses, making a total of 42 soil samples); and another set, denoted the verification set, was used with the 30 % remaining sites (six sampling locations with three soil uses making a total of 18 soil samples) to test the SQI. The SQI were determined following three steps (Fig. 2): 1) the representative indicators in the minimum data set (MDS) were selected from the total set of soil parameters analysed; 2) the MDS indicators were transformed into scores; and 3) the indicators were integrated to form the soil quality index (SQI).

2.3.1.1. Selection of indicators: Minimum dataset (MDS). Two different approaches were used on all 42 samples in the calibration set, regardless of use, to select the variables that form the MDS: i) principal component analysis (PCA), as this is the most widely used methodology (Askari and Holden, 2014), and ii) the more novel network analysis (NTA).

The selection of the MDS with PCA was done according to Andrews et al. (2002) with the SPSS v.23 software, considering the principal components (PC) with an eigenvector $\lambda > 1$ and which explain at least 5 % of the total accumulated variance. The matrix of components was obtained via Varimax rotation, which minimises the number of variables with higher loading values in each PC and makes it easier to interpret the results (Peris et al., 2008). The variable was selected with the highest rotated loading in absolute value in each PC identified, together with the variables that differed from that value by 10 % (Andrews et al., 2002). In order to reduce the redundancy of the variables in the MDS, when several variables in a PC fulfilled the previous conditions, the Pearson's correlations ($P < 0.05$) between them were taken into account. In the case of correlations, the variables with the highest loading values were selected as indicators, and in the absence of correlations between the variables in the same PC, all the variables were selected as indicators.

In network analysis, a network or graph is the graphic representation of the study variables represented as points called nodes, and of the relations existing between these variables represented by means of lines known as edges. These relations can be summarised in the form of a diagram in what is known as adjacency matrix. Pearson's bivariate correlation matrix was used ($P < 0.05$) as an adjacency matrix in the Gephi 0.9.2 software to create the network of soils in this study.

To maintain a parallelism with the methodology used to determine the MDS from PCA, the nodes in the NTA are divided into modules. These modules represent the structures of communities between the variables, a greater number of edges are found between nodes belonging to the same module than between nodes belonging to other modules (Newman, 2006). The separation into modules is generally done by optimising a function known as modularity, which represents the number of edges in a module minus the number of edges expected in an equivalent network with random edges. This maximisation of modularity produces a better separation of the nodes in the network into communities (Newman, 2006). Gephi 0.9.2 uses the Blondel algorithm to extract the communities of the nodes (Blondel et al., 2008); this algorithm is based on the optimisation of the difference in modularity and consists of two iterative steps. The first step considers each node i as a differentiated module, then considers all the neighbouring nodes j for each node i and assesses the gain in modularity that occurs when going from node i to the module of node j. Node i becomes part of the module for which the modularity gain is positive and maximum. This process is repeated sequentially for all the nodes until the distribution into modules cannot be improved any further. The second phase of the algorithm generates a new network in which the nodes are the modules generated in the previous phase, and the weights of the edges are the number of edges in each module and of the edges that interconnect the different modules generated. Once this second phase has been completed, the algorithm is iterative, and is repeated until no more changes can be generated and the modularity is maximum. Blondel et al. (2008) do not include a modularity range beyond which the network has a modular structure; they only refer to the fact that this modularity is within the range of [-1,1]. According to Newman and Girvan (2004), a modularity value of zero would indicate that the separation into communities is random, whereas values close to one would show that the network has a strongly modular structure; in practice the typical modularity values are located between 0.3 and 0.7. In this study it was therefore considered that the network generated from the soil variables in the analysis had a modular structure if the modularity value was over 0.3, otherwise the network would have a single module.

In order to select the indicators within each module, we considered the importance of the nodes in each module within the network as a whole. This was done by determining the eigenvector centrality of each node. This centrality is defined mathematically as the eigenvector of the adjacency matrix for the greatest eigenvalue in this matrix. This interpretation of the centrality index is since a node is important not only because of the number of edges it has, but also because of the importance of the nodes with which it is related (Girvan and Newman, 2002). In the case of NTA, the MDS indicators were selected in each module considering the modules that had a maximum eigencentrality value of at least 0.75. This limit was established to select the most representative variables and avoid considering possible modules that have very little importance in the set. This condition is similar to the procedure followed with PCA, in which PCs that explain at least 5 % of the variance were selected.

In each module that met the previous condition, all the variables with up to 10 % less than the maximum eigencentrality value of the module were selected, following the same procedure as in the PCA method, where the variables over a maximum of 10 % from the absolute maximum of the loadings of each PC were selected. If several variables met this condition, the variables with a higher eigencentrality value were selected as MDS indicators, as in the PCA. Also, as in the PCA, the Pearson's correlations between the preselected variables in each module were taken into account. In the case of correlation between the preselected variables in a module, the variable with the highest eigenvector centrality was chosen as the indicator of its module. If there were two or more variables with the same eigencentrality, the indicators were the variables with the highest absolute weighted value; this is a centrality measure that considers both the number of relationships of a node and their weight.

2.3.1.2. Transformation of the MDS indicators. Two different transformations – linear and non-linear – were carried out on each indicator to reduce the effect of scale between the different MDS indicators and to normalise all their values to the range [0,1].

The linear transformation (S_L) (Andrews et al., 2002) of the MDS indicators considered whether a higher value of the indicator was beneficial for soil quality ("more is better"), in which case Equation (1) was used; and if a lower value of the indicator was beneficial for soil quality ("less is better"), Equation (2) was used. For indicators with an optimum range ("mid-point"), the transformation was built using Equations (1) and (2) as appropriate and equalising the values of the indicator in the optimum range to a value of 1.

$$S_L = x - x_{Min}/x_{Max} - x_{Min} \tag{1}$$

$$S_L = 1 - (x - x_{Min})/(x_{Max} - x_{Min}) \tag{2}$$

In both equations, x is the value of the indicator, x_{Max} is the maximum value of the indicator in all the samples analysed, and x_{Min} is the minimum value of the indicator in all the samples analysed.

Sigmoidal curves were used in the non-linear transformation (S_{NL}) (Bastida et al., 2006; Hussain et al., 1999) following Equation (3):

$$S_{NL} = a/\left(1 + (x/x_0)^b\right) \tag{3}$$

where a is the maximum value obtained by the sigmoidal curve (in this study $a = 1$), x is the value of the indicator, x_0 is the mean value of this indicator in all the samples analysed, and b is a coefficient equal to −2.5 for the "more is better" indicators, and to 2.5 for the "less is better"

Table 1
Mean and standard deviation values of the soil physical, chemical and biological properties studied. Different letter following mean values indicate statistically significant differences (P < 0.05).

Soil variables	Units	Rainfed		Olive grove		Forest	
		Mean	SD	Mean	SD	Mean	SD
WHC	%	28.8 b	4.2	30.6 b	6.21	46.5 a	19.8
BD	g/cm3	1.68 a	0.13	1.67 a	0.11	1.47 b	0.18
Sand	%	25.6 a	10.4	28.9 a	8.82	30.1 a	10.1
CS	%	13.8 a	9.97	15.0 a	7.77	15.9 a	9.8
Silt	%	47.4 a	9.83	43.2 a	11.3	43.7 a	11.1
Clay	%	27.0 a	6.82	27.9 a	8.74	26.2 a	6.97
FS	%	11.8 a	4.45	13.9 a	4.80	14.24 a	4.99
pH	–	8.08 a	0.23	8.26 a	0.27	7.9 b	0.32
EC	10^{-3} dS/m	177.9 a	53.1	165.5 a	62.6	184.9 a	55.4
%$CaCO_3$	%	13.5 a	15.9	19.3 a	18.3	22.7 a	24.5
TN	%	0.18 b	0.04	0.17 b	0.06	0.39 a	0.17
TOC	%	1.09 b	0.40	0.96 b	0.61	3.52 a	1.65
TOC/TN	–	6.26 b	1.96	5.26 b	1.88	9.19 a	2.29
Pav	mg/kg	55.4 a	48.6	22.0 b	18.8	25.0 b	18.3
MnA	mg/kg	87.3 a	52.6	78.4 a	62.3	109.3 a	100.7
MnF	mg/kg	130.2 a	64.2	133.1 a	77.1	150.8 a	101.6
MnX	mg/kg	42.9 a	21.2	54.7 a	35.0	41.5 a	28.0
FeA	mg/kg	0.43 a	0.25	0.38 a	0.21	0.43 a	0.28
FeF	mg/kg	6.38 a	2.69	5.07 a	2.33	5.80 a	3.71
FeX	mg/kg	5.96 a	2.60	4.68 a	2.20	5.37 a	3.53
Ninor	mg/kg	28.6 a	39.1	28.6 a	41.8	24.8 a	10.2
Ccw	mg/kg	14.6 b	4.75	13.9 b	5.68	26.1 a	12.5
Ncw	mg/kg	6.14 b	8.74	3.88 b	4.43	10.4 a	10.1
Chw	mg/kg	62.5 b	21.4	51.5 b	18.2	106.6 a	33.9
Nhw	mg/kg	7.70 b	2.68	8.27 b	4.45	16.3 a	10.3
aglu	U/g	0.038 b	0.012	0.029 b	0.011	0.053 a	0.018
aryln	U/g	0.027 b	0.010	0.021 b	0.010	0.036 a	0.010
aryls	U/g	0.013 b	0.006	0.009 b	0.007	0.041 a	0.027
bglu	U/g	0.147 b	0.051	0.128 b	0.063	0.294 a	0.142
nag	U/g	0.013 b	0.009	0.007 b	0.004	0.029 a	0.022
phos	U/g	0.624 b	0.606	0.324 b	0.411	1.133 a	0.728
pac	U/g	0.113 b	0.057	0.094 b	0.100	0.527 a	0.404
pak	U/g	0.368 b	0.250	0.256 b	0.208	0.891 a	0.481
ure	U/g	0.010 a	0.005	0.009 a	0.005	0.012 a	0.005
dh	U/g	0.097 b	0.070	0.062 b	0.039	0.178 a	0.087
GMEAN	U/g	0.055 b	0.022	0.037 b	0.018	0.113 a	0.051

indicators. The indicators with optimum values ("mid-point") were transformed using a piecewise function: a score value of 1 for all the indicator values (x) within the optimum range, and the score transformations "more is better" or "less is better" were used when the values of the indicator differed from the optimum.

2.3.1.3. Integration of the MDS indicators into the SQI. The linear and non-linear scores of the indicators were integrated into a single index in two ways: additive and weighted (SQI-W), following Equations (4) and (5) respectively:

$$SQI-A = \sum_{i=1}^{n} S_i/n \tag{4}$$

$$SQI-W = \sum_{i=1}^{n} W_i S_i \tag{5}$$

where S_i is the transformed linear (L) or non-linear (NL) score for each indicator, n is the number of MDS indicators, and W_i is the weight of each indicator.

In the case of MDS selected by PCA, W_i is a value between 0 and 1 that corresponds to the weight of each indicator calculated according to Equation (6).

$$W_i = \frac{\%VarPC_i}{\%VarTotal} / \sum_{i=1}^{n} \frac{\%VarPC_i}{\%VarTotal} \tag{6}$$

where %VarPCi is the percentage of variance explained by the PC for indicator i, %VarTotal is the percentage of variance explained by all the PCs in the MDS, and n is the maximum number of PCs selected. The sum term is needed when there are two or more indicators for the same PC.

As in the case of the MDS obtained with PCA, the W_i in the MDS selected using NTA was calculated in two stages according to Equation (7): i) the ratio was obtained between the eigencentrality value of the module to which the indicator belongs and the sum of all the eigencentrality values of all the network variables, and ii) the final weight W_i was calculated as the relation between the ratio calculated at (i) and the sum of all the ratios obtained for all the indicators.

$$W_i(NTA) = \frac{EC(M_i)}{EC_{Total}} / \sum_{i=1}^{n} \frac{EC(M_i)}{EC_{Total}} \tag{7}$$

where $EC(M_i)$ is the sum of the eigencentrality value of all the variables in module M_i, EC_{Total} is the sum of the eigencentrality value of all the variables in the network, and n is the maximum number of M selected. The sum term is needed when there are two or more indicators for the same M_i.

2.3.1.4. SQI obtained and their comparison between the different uses. The result of this whole process was a total of eight SQI, differentiated based on the type of transformation and integration used for each of the two MDS identified (one by PCA and another by NTA) (Fig. 2). The methodology, PCA or NTA, that produced SQI with a better ability to

Table 2
Loading coefficients of the variables analysed for the principal components (PC) that comply with the condition of $\lambda > 1$. In bold, for each PC, the variables that comply with the condition of belonging to the range of loadings between the absolute maximum value and 10 %. The variables in italic and the underlined loading values identify the indicators selected taking into account Pearson's bivariate correlations (P < 0.05).

Soil variables	Principal component (PC)					
	PC1	PC2	PC3	PC4	PC5	PC6
WHC	0.50	−0.69	0.01	−0.20	0.22	−0.01
BD	−0.29	0.70	−0.33	−0.40	−0.16	−0.04
Sand	0.03	−0.39	0.31	−0.05	0.07	**0.84**
CS	0.04	−0.25	−0.29	−0.13	−0.16	***0.85***
Silt	0.07	0.28	−0.50	−0.06	−0.26	−0.70
Clay	−0.08	0.09	0.45	0.07	0.22	−0.14
FS	0.12	−0.36	**0.84**	0.05	0.22	0.08
pH	−0.23	−0.14	0.00	***−0.76***	−0.11	0.12
EC	0.42	0.09	0.04	0.05	***0.79***	−0.13
%CaCO$_3$	0.15	−0.73	0.53	−0.22	0.15	−0.04
TN	0.51	−0.40	0.07	0.50	0.20	0.43
TOC	***0.91***	−0.11	0.07	0.10	0.09	0.15
TOC/TN	**0.84**	0.19	0.10	−0.08	−0.02	−0.03
Pav	0.07	−0.13	0.15	0.22	−0.03	−0.13
MnA	−0.08	0.79	−0.23	−0.09	0.28	−0.36
MnF	−0.10	0.81	0.16	0.02	0.16	−0.33
MnX	0.06	0.13	***0.90***	−0.02	−0.19	0.16
FeA	0.14	0.61	0.22	−0.25	−0.33	−0.12
FeF	−0.07	***0.93***	0.03	−0.02	−0.09	−0.19
FeX	−0.07	**0.92**	0.01	0.00	−0.07	−0.18
Ninor	0.40	0.06	−0.48	−0.14	−0.16	0.10
Ccw	0.22	−0.04	0.12	0.65	−0.13	0.14
Ncw	0.47	−0.26	0.25	0.18	0.58	0.22
Chw	0.61	−0.33	−0.28	0.36	0.30	0.15
Nhw	0.01	0.01	0.07	0.01	0.05	0.00
aglu	0.79	−0.05	0.20	0.30	0.19	0.05
aryln	0.55	0.05	0.05	0.38	0.24	−0.01
aryls	**0.86**	−0.29	0.09	0.22	−0.17	−0.02
bglu	0.77	−0.11	0.30	0.26	0.08	−0.05
nag	0.48	−0.12	−0.06	0.64	−0.07	−0.14
phos	0.61	−0.25	0.15	0.25	0.51	−0.13
pac	0.63	−0.09	−0.34	0.09	0.42	0.10
pak	0.80	−0.30	−0.05	0.11	0.35	0.09
ure	0.03	0.17	−0.19	0.41	−0.65	−0.21
dh	0.72	0.02	−0.24	−0.10	−0.04	−0.03
GMEAN	**0.90**	−0.17	−0.11	0.31	0.14	−0.05
Eigenvalue	11.79	5.90	3.95	2.77	2.21	2.00
Variance (%)	32.76	16.40	10.96	7.68	6.15	5.57
Cumulative variance (%)	32.76	49.16	60.12	67.8	73.95	79.52

differentiate between soil uses was selected as the best methodology for assessing soil quality in this study.

2.4. Statistical analysis

Prior to the PCA analysis, all the variables were tested for normality and homogeneity of variance through the Kolmogorov-Smirnov test, Levene's test and the visual evaluation of histograms. The one-way analysis of variance (ANOVA) was used to establish any significant difference between soil parameters due to the various soil uses, and to identify if any of the SQI distinguished between the soil uses in the study. Statistical analyses were done using SPSS v.23.0 software.

3. Results

3.1. Effect of soil use on soil properties

From the analysis of the variance carried out on the soil variables (Table 1), it can be seen that the use affects a large part of them. Forest soils were characterised by higher rates of TOC and TN and greater concentrations of their soluble forms Ccw, Chw, Ncw and Nhw, implying that this use also has the greatest enzyme activity (with the exception of ure which did not vary between uses) and WHC, in addition to minor values of pH and BD. Rainfed soils had the highest concentrations of Pav, which was the only variable that showed significant differences between the agricultural uses.

3.2. Selection of the minimum data set (MDS) and calculation of the weights (Wi)

3.2.1. MDS generated from a principal component analysis, MDS_{PCA}

Eight PCs were obtained in the PCA with eigenvalues of over 1. However, this number was reduced to the six first axes, as they explained at least 5 % of the variance (Table 2). The variables with the highest loading in the PC1 were total organic carbon (TOC), geometric mean of enzyme activities (GMEAN), arylsulfatase activity (aryls) and the TOC/TN ratio (Table 2). The variable TOC had a higher loading score in this PC and was selected as an indicator of PC1 in the MDS, as all the other preselected variables were related with it (Table 3). The variables with the highest loading in PC2 were free Fe oxides (FeF) and crystalline Fe oxides (FeX) (Table 2). As these two variables showed intercorrelations (P < 0.05) (Table 3), the variable FeF was selected as an indicator for PC2 in the MDS due to its higher loading. In PC3 the variables crystalline Mn oxides (MnX) and fine sand (FS) showed the highest loadings on this axis (Table 2). MnX was selected as an indicator as it was correlated with the other preselected variable in PC3 (Table 3) and had the highest loading. For PC4 and PC5 the indicators were pH and electrical conductivity (EC) respectively. Finally, in PC6 coarse sand (CS) and sand had the highest loading, and CS was selected as indicator due to its correlation with sand and its higher loading in the PCA. In summary, the weights of the indicators derived from the MDS_{PCA} calculated according to Equation (6) are shown in Table 3 and were subsequently used to calculate the SQI.

3.2.2. MDS generated from network analysis, MDS_{NTA}.

Three modules – M1, M2 and M3 – were identified in the network analysis according to Pearson's bivariate correlations, based on the algorithm in the Gephi 0.9.2 software (Fig. 3).

In M1 (Table 4), the variables GMEAN, TOC, b-glucosidase (b-glu), arylsulphatase (aryls), acid phosphatase (pac) and *N*-acethyl-glucosaminidase (nag) had the maximum eigencentrality value (1). Following the same procedure used to obtain the MDS from PCA, all the variables at a distance of 10 % from this maximum eigencentrality score were considered in M1, that is, those within the range [1, 0.9], in addition to the previous variables: total nitrogen (TN), hot-water extractable carbon (Chw), a-glucosidase (a-glu), alkaline phosphatase (pak), TOC/TN ratio and arylamidase (aryln) were selected. It was necessary to verify whether there were any correlations among these variables to reduce redundancies (Table 5). All these variables were intercorrelated, so the M1 indicator was selected taking into account the highest eigenvector centrality value (1); and from the variables that met this condition, the weighted degree was chosen to select the M1 indicator, being GMEAN the variable chosen due to its higher weighted degree.

In M2, bulk density (BD) had the highest eigencentrality value (0.914027), with no other variables in the range [0.914027, 0.822624], so BD was selected as indicator. In M3, phosphatase activity (phos) had the highest eigencentrality value (0.913891). Only the eigenvector centrality of nitrogen soluble in cold water (Ncw) belongs to the preselection range for this module [0.913891, 0.8225]. As both variables wereintercorrelated (Table 5), phos was selected as the indicator for this module due to its higher eigenvector centrality.

The weights of the indicators derived from the MDS_{NTA} calculated according to Equation (7) are shown in Table 4 and were subsequently used to calculate the SQI.

Table 3
Pearson's bivariate correlations (*P < 0.05, **P < 0.01) and number of samples between the variables in bold in Table 2.

Soil variables		PC1				PC2		PC3		PC4	PC5	PC6
		TOC	aryls	GMEAN	TOC/TN	FeF	FeX	MnX	FS	pH	EC	Sand
PC1	aryls	**0.895****										
		40										
	GMEAN	**0.920****	**0.908****									
		42	40									
	TOC/TN	**0.794****	**0.601****	**0.682****								
		42	40	42								
PC2	FeF	0.054	0.187	0.175	0.078							
		41	39	41	41							
	FeX	0.032	0.183	0.159	0.057	**0.997****						
		41	39	41	41	41						
PC3	MnX	0.260	0.176	0.194	0.243	0.260	0.237					
		38	36	38	38	38	38					
	FS	0.234	0.128	0.157	0.093	−0.110	−0.087	**0.480****				
		42	40	42	42	41	41	38				
PC4	pH	**−0.486****	**−0.520****	**−0.570****	−0.289	−0.199	−0.187	−0.128	0.107			
		40	39	40	40	39	39	36	40			
PC5	EC	0.225	0.028	0.214	0.291	−0.016	−0.010	−0.127	0.204	−0.110		
		41	39	41	41	40	40	37	41	39		
PC6	Sand	−0.121	−0.221	−0.218	−0.040	**−0.599****	**−0.590****	0.055	0.261	0.160	0.081	
		39	37	39	39	38	38	35	39	37	38	
	CS	−0.173	−0.175	−0.289	−0.117	**−0.591****	**−0.584****	−0.231	−0.141	0.295	−0.113	**0.752****
		40	38	40	40	39	39	36	40	39	39	38

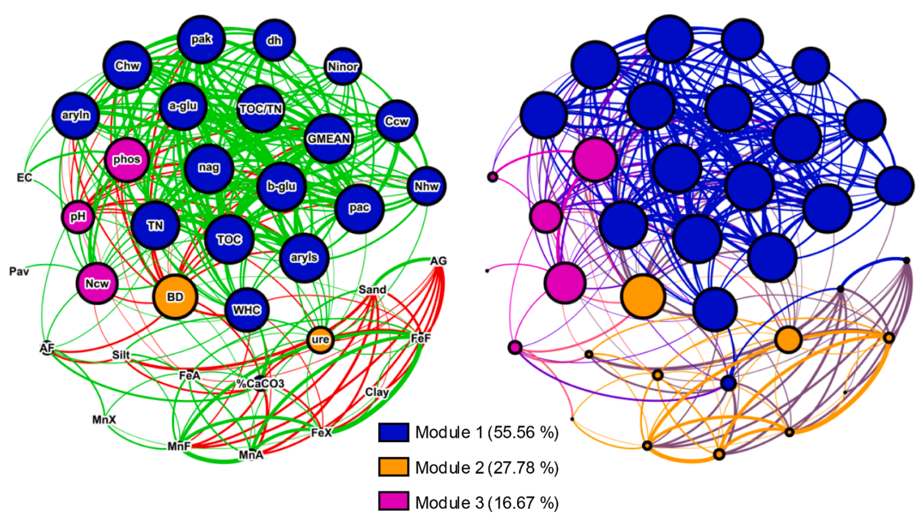

Fig. 3. Graph of the network resulting from considering Pearson's bivariate correlations (P < 0.05), Fruchterman Reingold layout. The modules identified by the Gephi software, M1, M2 and M3, are shown in different colours, green lines represent positive correlations and red lines represent negative correlations.

3.3. Calculation of the SQI

3.3.1. SQI_{PCA}

To calculate the scores (S_i) of the MDS indicators selected by PCA, the indicators TOC, FeF, MnX and EC were considered for the soils in this study as the "more is better" type, due to their positive role in the correct functioning of the soils (Brady and Weil, 2014). In contrast, the indicator pH and CS were considered to be "less is better", as pH has a negative influence on the availability of the main micronutrients in the pH values where these soils are located (range of 7.3–8.3) (Kabata-Pendias and Pendias, 2001), and coarse particle size negatively influences soil nutrient retention capacity (Brady and Weil, 2014).

The additive SQI (SQI-A) based on the MDS determined by PCA was finally expressed as follows:

$$SQI - A_{PCA} = (S_{TOC} + S_{FeF} + S_{\mathrm{MnX}} + S_{pH} + S_{EC} + S_{CS})/6 \tag{8}$$

where S denotes the linear or non-linear score of the indicator shown. The $SQI\text{-}AL_{PCA}$ and $SQI\text{-}ANL_{PCA}$ were obtained for each soil by substituting the linear and non-linear scores in Equation (8).

The SQI integrated by weights (SQI-W) from the MDS determined by PCA with the weights corresponding to each indicator (Table 6) was:

$$SQI - W_{PCA} = 0.412S_{TOC} + 0.206S_{FeF} + 0.138S_{MnX} + 0.097S_{pH} + 0.077S_{EC} + 0.07S_{CS} \tag{9}$$

Once again, S denotes the linear or non-linear score of the indicator shown. The $SQI\text{-}WL_{PCA}$ and $SQI\text{-}WNL_{PCA}$ were obtained for each soil by substituting the corresponding scores in Equation (9).

3.3.2. SQI_{NTA}

The score of the MDS indicators GMEAN and phos generated from the network analysis was calculated according to "more is better" transformations, because they positively influenced soil quality (Adetunji et al., 2017; Paz-Ferreiro and Fu, 2016). In contrast, the "less is better" consideration was used for the indicator BD due to its negative relationship with soil organic matter (Li et al., 2019), and its increased

Table 4
Soil variables belonging to the modules identified with the Gephi software and their respective eigencentrality values. The variables with eigencentrality in the range between the maximum eigencentrality value minus 10% are shown in bold, and from among these variables the variables selected finally for the MDS_{NTA} are in italics.

Modules					
M1			M2		
Soil variables	Eigencentrality	Weighted degree	Soil variables	Eigencentrality	Weighted degree
GMEAN	***1***	***12.54***	***BD***	***0.9140***	***−9.00***
TOC	**1**	**12.32**	ure	0.5526	3.63
bglu	**1**	**11.53**	MnA	0.2217	1.82
aryls	**1**	**11.15**	FeF	0.2161	1.93
pac	**1**	**10.60**	FeA	0.2101	1.16
nag	**1**	**9.65**	MnF	0.1785	1.97
TN	**0.9853**	**10.20**	FeX	0.1697	1.55
Chw	**0.9809**	**10.51**	Silt	0.1518	−1.20
aglu	**0.9707**	**10.82**	Clay	0.0446	0.28
pak	**0.9707**	**10.52**	MnX	0.0438	1.68
TOC/TN	**0.9662**	**9.62**			
aryln	**0.9537**	**8.24**			
WHC	0.8959	5.60	M3		
Ccw	0.8483	7.75	Soil variables	Eigencentrality	Weighted degree
dh	0.8477	9.07	***phos***	***0.9139***	***8.57***
Nhw	0.7833	7.28	**Ncw**	**0.8451**	**6.42**
Ninor	0.7557	6.65	pH	0.6682	−5.74
%$CaCO_3$	0.3087	−1.83	AF	0.2743	1.46
Sand	0.1208	−2.71	EC	0.1961	1.74
AG	0.1157	−3.33	Pav	0.0451	0.42

due to the use of agricultural machinery (Brady and Weil, 2014).

The additive SQI (SQI-A) based on the MDS determined by network analysis finally presented the expression:

$$SQI - A_{NTA} = (S_{GMEAN} + S_{BD} + S_{phos})/3 \quad (10)$$

where S denotes the linear or non-linear score of the indicator shown. The $SQI\text{-}AL_{NTA}$ and $SQI\text{-}ANL_{NTA}$ were obtained for each soil by substituting the corresponding scores in Equation (10).

The SQI integrated by weights (SQI-W) from the MDS determined by network analysis with the weights corresponding to each indicator (Table 6) was:

$$SQI - W_{NTA} = 0.745S_{GMEAN} + 0.133S_{phos} + 0.122S_{BD} \quad (11)$$

Once again, S denotes the linear or non-linear score of the indicator shown. The $SQI\text{-}WL_{NTA}$ and $SQI\text{-}WNL_{NTA}$ were obtained for each soil by substituting the corresponding scores in Equation (11).

3.4. Comparison of the SQI calculated by PCA and NTA

3.4.1. SQI applied to the soils in the calibration set

The SQI calculated from the PCA and using the additive integration $SQI\text{-}AL_{PCA}$ and $SQI\text{-}ANL_{PCA}$ showed significant differences between the forest and olive grove use ($P < 0.05$) (Fig. 4), without clearly differentiating the uses studied. In the case of integration by weights, the indices $SQI\text{-}WL_{PCA}$ and $SQI\text{-}WNL_{PCA}$ differentiated forest and cultivation uses ($P < 0.05$).

The soil quality indices calculated from the MDS generated by network analysis (NTA) with the linear transformations ($SQI\text{-}ANL_{NTA}$ and $SQI\text{-}WL_{NTA}$), and the soil quality index obtained by non-linear transformations and integrated by weights ($SQI\text{-}WNL_{NTA}$), showed significant differences between the three uses ($P < 0.05$), with higher SQI values in the forest use, followed by the rainfed and finally the olive grove use (Fig. 4). In contrast, the linear index integrated by addition ($SQI\text{-}AL_{NTA}$) only had differences between agricultural and forest uses, with the latter showing the highest SQI scores (Fig. 4). Therefore, 75 % of the SQI obtained by NTA were capable of differentiating between the three soil uses studied, whereas no SQI-PCA showed these differences.

3.4.2. SQI applied to the soils in the validation set

In the case of SQI_{PCA}, only the $SQI\text{-}WNL_{PCA}$ showed differences between soil uses, and was sensitive to the differences between forest and agricultural uses ($P < 0.05$), while the other PCA-related SQI were not sensitive to use ($P < 0.05$) (Fig. 5). Among the SQI_{NTA}, $SQI\text{-}AL_{NTA}$, $SQI\text{-}WL_{NTA}$ and $SQI\text{-}WNL_{NTA}$ were sensitive to the differences between forest and agricultural uses ($P < 0.05$) (Fig. 5). Therefore, 75 % of SQI-NTA could differentiate between agricultural (rainfed and olive grove) and forest uses, whereas this percentage fell to 25 % for SQI-PCA.

4. Discussion

The aim of this study was to verify the possibility of using network analysis to select the indicators that form soil quality indexes. To determine the suitability of this type of analysis for this task, the classification of variables, indicators and SQI obtained by NTA were compared with those obtained by PCA, as this is the most widely used methodology in the determination of SQI.

The PCA identified six indicators, in order of higher to less importance considering the percentage of explained variance: TOC, FeF, MnX, pH, EC and CS. Total organic carbon has a capital role in soil quality due to its influence in a broad biological, physical and chemical soil properties (Hamidi Nehrani et al., 2020), being one of the most used indicators (Bünemann et al., 2018; Zornoza et al., 2015). The free Fe oxides were inversely correlated with %$CaCO_3$ (0.559; $P < 0.01$), this indicator would represent variations in the decarbonation-rubefaction processes between the studied soils, soils subject to a higher decarbonation would show higher concentrations of Fe oxides and vice versa (Loeppert, 1986). The rubefaction process would also be responsible for the positive relationship between the clay content and the FeF (0.419; $P < 0.01$) and MnX (0.323; $P < 0.05$), since these oxides can precipitate forming coatings on the clays (Sipos et al., 2019). The importance of MnX as indicators of soil quality would be due to the greater chemical reactivity and the greater nutrient retention capacity that they have with respect to Fe oxides (Chao and Theobald, 1976), in addition, Mn oxides influence stabilization-destabilization processes of soil organic matter (Li et al., 2021) being a critical component in the decomposition of plant remains, especially those rich in lignin (Keiluweit et al., 2015; Li et al.,

Table 5
Pearson's bivariate correlations (*P < 0.05, **P < 0.01) between the variables in bold in Table 4.

Soil variables		M1												M2	M3
		GMEAN	TOC	bglu	aryls	pac	nag	TN	Chw	aglu	pak	TOC/TN	aryln	BD	phos
M1	TOC	0.920** 42													
	bglu	0.908** 40	0.891** 40												
	aryls	0.908** 40	0.895** 40	0.861** 39											
	pac	0.887** 42	0.838** 42	0.781** 40	0.821** 40										
	nag	0.841** 40	0.768** 40	0.740** 39	0.791** 40	0.653** 40									
	TN	0.816** 38	0.860** 38	0.651** 36	0.788** 37	0.677** 38	0.667** 37								
	Chw	0.739** 42	0.730** 42	0.755** 40	0.635** 40	0.633** 42	0.551** 40	0.605** 38							
	aglu	0.852** 42	0.821** 42	0.796** 40	0.801** 40	0.685** 42	0.721** 40	0.723** 38	0.726** 42						
	pak	0.853** 42	0.818** 42	0.732** 40	0.659** 40	0.756** 42	0.579** 40	0.658** 38	0.694** 42	0.804** 42					
	TOC/TN	0.682** 42	0.794** 42	0.664** 40	0.601** 40	0.534** 42	0.503** 40	0.506** 38	0.531** 42	0.695** 42	0.693** 42				
	aryln	0.646** 39	0.559** 39	0.688** 38	0.600** 38	0.482** 39	0.483** 38	0.516** 36	0.710** 39	0.708** 39	0.481** 39	0.408** 39			
M2	BD	−0.544** 41	−0.513** 41	−0.530** 39	−0.537** 40	−0.478** 41	−0.338* 40	−0.650** 38	−0.513** 41	−0.582** 41	−0.575** 41	−0.315* 41	−0.495** 39		
M3	phos	0.661** 36	0.497** 36	0.557** 36	0.382* 36	0.380* 36	0.465** 36	0.488** 35	0.500** 36	0.651** 36	0.670** 36	0.377* 36	0.649** 35	−0.531** 36	
	Ncw	0.439** 42	0.491** 42	0.448** 40	0.400* 40	0.388* 42	0.319* 40	0.398* 38	0.473** 42	0.579** 42	0.592** 42	0.315* 42	0.453** 39	−0.485** 41	0.697** 36

Table 6
Summary of the indicators selected for the MDS with PCA and the weights used to calculate the SQI-W_{PCA} and summary of the indicators selected for the MDS with NTA, and the weights used to calculate the SQI-W_{NTA}.

Indicators	PC related	PC variance (%)	Weight (W_i)	Indicators	Module related	Module eigencentrality	Weight (W_i)
TOC	1	32.76	0.412	GMEAN	1	16.5	0.745
FeF	2	16.4	0.206	BD	2	2.70	0.122
MnX	3	10.96	0.138	phos	3	2.94	0.133
pH	4	7.68	0.097				
EC	5	6.15	0.077				
CS	6	5.57	0.070				

2021). Soil pH is one of the most widely used indicators in soil quality studies (Andrés-Abellán et al., 2019; Bünemann et al., 2018; Gómez et al., 2009; Paz-Kagan et al., 2016; Rahmanipour et al., 2014). The natural soils in this study showed more acidic pH values than the agricultural soils ($P < 0.05$), possibly due to their greater vegetation cover and/or the mixture of superficial horizons with deeper ones in the soils dedicated to cultivation due to agricultural practices (Rahmanipour et al., 2014). In the study soils, EC would represent soluble forms of nutrients as reflected by its positive correlations with Ncw (0.429; $P < 0.01$) and Chw (0.321; $P < 0.05$). The influence of electrical conductivity on soil quality has been referred to by other authors (Jahany and Rezapour, 2020; Paz-Kagan et al., 2016; Rezapour et al., 2015), identifying a decrease in soil quality with lower electrical conductivity values. Zhang and Hou (2012) and Rezapour (2014) relate the presence of coarse sand with lower quality soils by affecting the ability of soils to retain nutrients and accelerating erosion processes. In this study, the percentages of coarse sand were negatively correlated with the percentages of clays (−0.550; $P < 0.01$) and silts (−0.351; $P < 0.05$), with the concentrations of oxides: MnA (−0.573; $P < 0.01$), MnF (−0.609; $P < 0.01$), FeA (−0.362; $P < 0.05$); FeF (−0.591; $P < 0.01$), FeX (−0.584; $P < 0.01$); and with urease enzymatic activity (−0.481; $P < 0.01$), so this indicator would provide the SQI with information about soil texture.

The NTA identified three indicators, GMEAN, BD and phos, and reduced the number of indicators compared to those selected by PCA by 50 %. GMEAN is an index capable of condensing the information on enzyme activities into a single value (García-Ruiz et al., 2008; Paz-Ferreiro et al., 2012; Wang et al., 2012). This capacity for condensation was seen in this study in the positive and significant correlations ($p < 0.001$) between GMEAN and all the enzyme activities determined. BD is a key variable in soil functioning (Gajda et al., 2016) that is widely used as a soil quality indicator (Bünemann et al., 2018), and is related with other physical, chemical and biological variables (Al-Shammary et al., 2018). It is particularly worth noting the relation it tends to present with organic matter (Heuscher et al., 2005; Topa et al., 2021). Indeed, in the soils in this study, BD was negatively correlated with TOC (-0.513; $P < 0.001$), TN (-0.650; $P < 0.001$), Cws (-0.502; $P < 0.001$), Nws (-0.485; $P < 0.01$), Chw (0.513; $P < 0.001$) and Nhw (-0.337; $P < 0.05$). These relations explain the greater BD of agricultural soils compared to forest soils, as they are affected by losses in organic matter content due to oxidation caused by ploughing and due to processes of compaction by agricultural machinery (Havaee et al., 2014; Khormali et al., 2009). Enzyme activities can act as soil quality indicators, and can be affected either by pollutants or other anthropic factors (Rao et al., 2014); they also fulfil most of the criteria required in a good indicator, as they are representative of a soil function, operative, integrative, easy to measure and sensitive to soil handling and structure (Adetunji et al., 2017). Among all enzyme activities, phosphatase activity has been identified as one of the most sensitive to soil use and to the soil organic matter content. It is characteristic of this activity to be suppressed by the use of inorganic P fertilizers (Caravaca et al., 2002; Janes-Bassett et al., 2022; Zhang et al., 2018; Zornoza et al., 2007), therefore, it is an especially useful parameter for differentiating between land uses. Phosphatase activity has been established as indicator of soil quality in several studies (Puglisi et al., 2006; Zornoza et al., 2008, Zornoza et al., 2007), being a variable usually selected to create the SQI (Andrés-Abellán et al., 2019; Zhou et al., 2020). This study highlights the capacity of enzyme activities, except for urease activity, as indicators of use, as they were higher in the forest use and lower in agricultural soils due to their lower levels of organic matter. Phosphatase activity showed the highest activity in this study, and the widest variation between the three uses studied; it was also the enzyme activity that revealed the greatest difference between rainfed and olive grove uses, probably due to the different management in these soils in terms of the use of fertilizers. In view of this, the indicators selected by NTA in this study were representative of soil processes involved in soil quality, specifically related with the organic matter dynamic within the different soil uses studied.

As shown in the results, the ordination of the variables into modules in the NTA was similar to that obtained in the PCA, and indicated that the most important variables in these soils are organic matter and enzyme activity, followed by inorganic colloids in the soil. This relation between NTA and PCA is not a result of chance, as both analyses are based on the use of the same correlation matrix. Therefore, the NTA and the algorithm used to determine the modules in this study have demonstrated their capacity to identify edaphological characteristics of the variables.

Four SQI were calculated based on the indicators selected by NTA and used both for soils in the calibration set and for soils in the validation set; their capacity to differentiate uses was then compared with that of the SQI obtained by PCA (Figs. 4 and 5). The greatest capacity was observed in the SQI_{NTA}, which were the only SQI capable of differentiating between the three uses for the calibration set; they also had a greater capacity than the SQI_{PCA} for differentiating between agricultural (rainfed and olive grove) and forest uses for the verification set, possibly because the indicators selected by NTA better reflected the differences between uses. All the indicators selected by NTA showed significant differences between agricultural and forest uses, whereas of the indicators selected by PCA (TOC, FeF, MnX, pH, EC and CS), only 33.3 % (TOC and pH) revealed differences between these uses. Values of SQI_{NTA} and SQI_{PCA} obtained in this study indicated that forest soils have higher SQI values than permanent croplands which agrees with the results of other authors (Marzaioli et al., 2010; Rezapour, 2014).

Based on these considerations, in this study NTA was used to generate a series of SQI that represent the quality of the soils studied, with a lower number of indicators than the SQI_{PCA} and a better differentiation capacity than the SQI_{PCA} when applied to both the soils in the calibration set and the soils in the verification set. To judge from the results obtained, we believe that network analysis may be useful for both the study of soil quality and in other areas of soil science and offers an analytical tool that should be explored in future research.

5. Conclusions

In this study, assuming that different land uses generate soils with different qualities, the ability to select soil quality indicators through network analysis (NTA) that form a minimum data set (MDS) with which

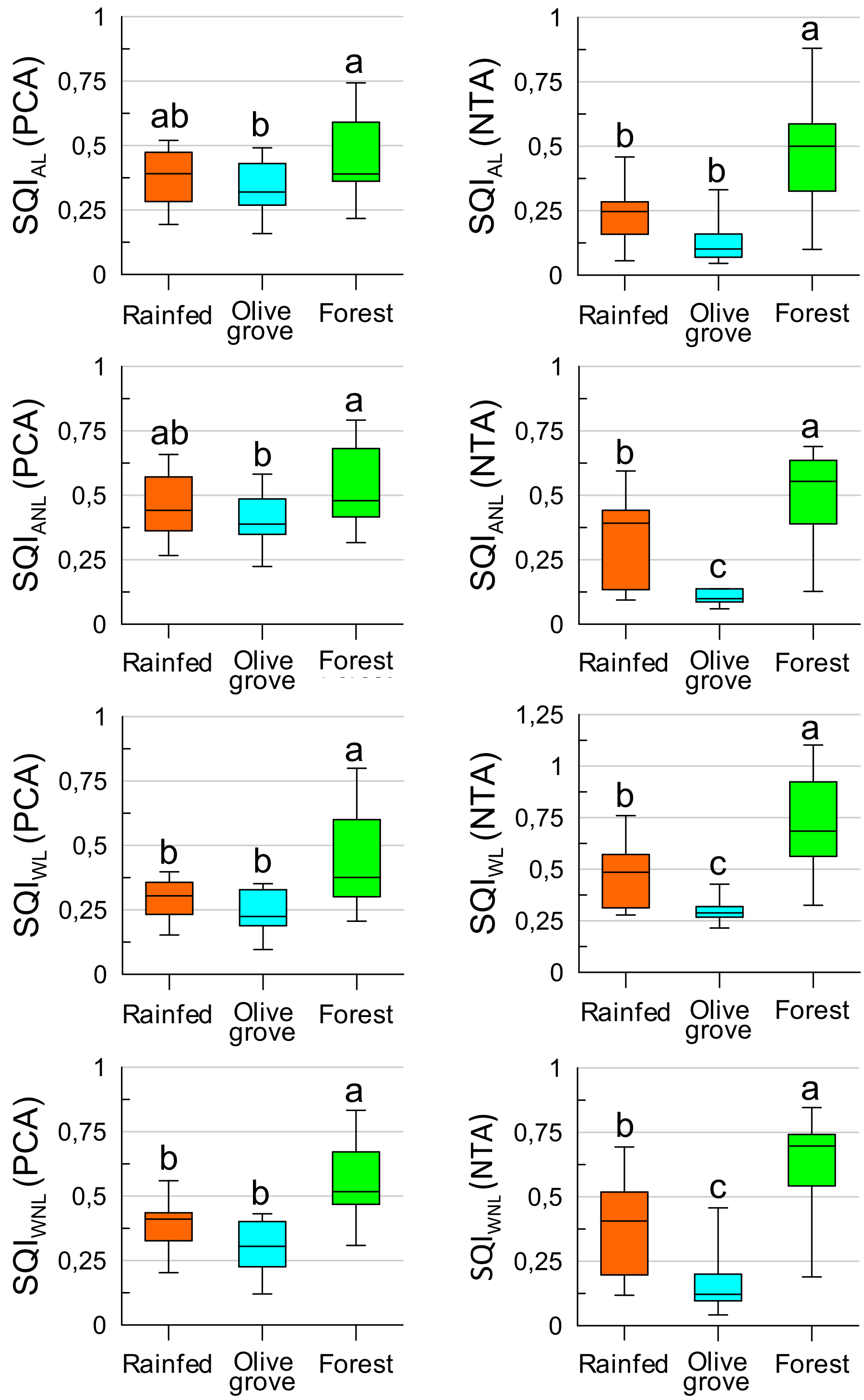

Fig. 4. Soil quality index (SQI) scores for rainfed, olive grove and forest soil uses applied to calibration set and calculated from the minimum data set (MDS) selected by principal component analysis (PCA) and network analysis (NTA), with linear (L) or non-linear scores (NL) and with additive integration (A) or integration by weights (W). Letters indicate significant differences ($P < 0.05$) between uses by one-way ANOVA.

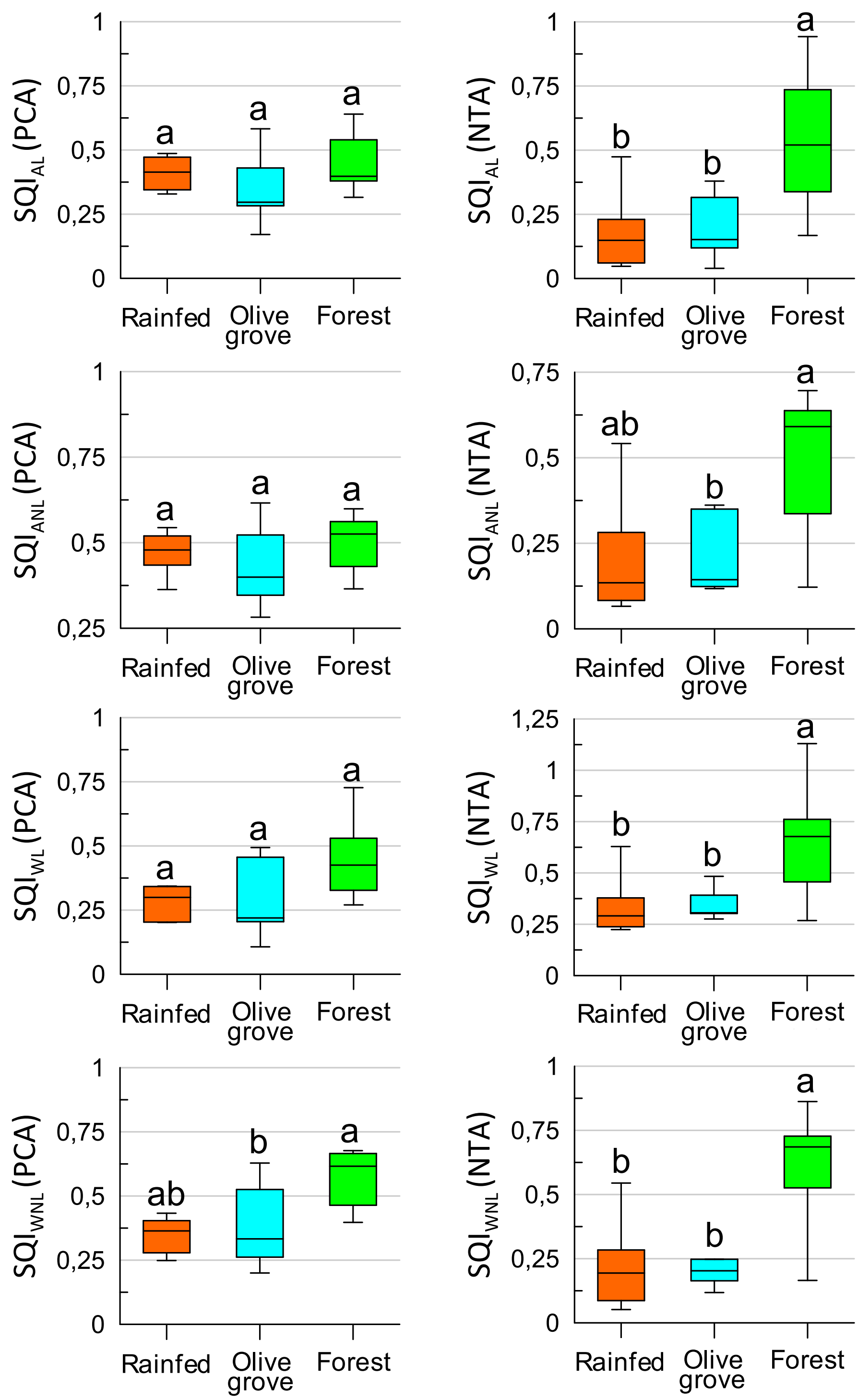

Fig. 5. Soil quality index (SQI) scores for rainfed, olive grove and forest uses applied to verification set and calculated from the minimum data set (MDS) selected by principal component analysis (PCA) and network analysis (NTA), with linear (L) or non-linear scores (NL) and with additive integration (A) or integration by weights (W). Letters indicate significant differences ($P < 0.05$) between uses by one-way ANOVA.

to calculate soil quality indices (SQI) capable of differentiating soil quality has been evaluated. To achieve this, carbonated soils with three different land uses (rainfed, olive grove and forest) were analysed and the results obtained by NTA were compared with those obtained by principal component analysis (PCA), which is the method commonly used to calculate the SQI.

NTA has allowed obtaining a MDS with fewer indicators than the one obtained by PCA. The indicators selected by NTA better reflected the differences between land uses. This caused the SQI obtained by NTA to be more useful than those obtained by PCA by better differentiating between the land uses studied in the carbonated soils studied. The study showed differences in the quality of agricultural soils (rainfed and olive groves) compared to the surrounding natural soils. We believe that the NTA may be useful in future soil quality studies and in other types of edaphological studies.

CRediT authorship contribution statement

Juan Pedro Martín-Sanz: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Ana de Santiago-Martín:** Conceptualization, Investigation, Writing – review & editing, Supervision. **Inmaculada Valverde-Asenjo:** Conceptualization, Investigation, Writing – review & editing, Supervision. **José Ramón Quintana-Nieto:** Methodology, Formal analysis, Resources, Supervision. **Concepción González-Huecas:** Project administration, Resources. **Antonio L. López-Lafuente:** Project administration, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the Madrid Autonomous Region through the CARESOIL PROGRAMME (Ref.S2018/EMT-4317), by the research projects SANTANDER/COMPLUTENSE (Ref. PR75/18-21618) and by the European Union-NextGenerationEU through the Ministry of Universities and the call CT31/21 of the UCM. We especially wish to thank Ms. Pru Brooke-Turner for the revision of the English version of the manuscript and the comments and suggestions of the anonymous referees.

References

Adetunji, A.T., Lewu, F.B., Mulidzi, R., Ncube, B., 2017. The biological activities of β-glucosidase, phosphatase and urease as soil quality indicators: a review. J. Soil Sci. Plant Nutr. 17, 794–807. https://doi.org/10.4067/S0718-95162017000300018.

Al-Shammary, A.A.G., Kouzani, A.Z., Kaynak, A., Khoo, S.Y., Norton, M., Gates, W., 2018. Soil bulk density estimation methods: a review. Pedosphere 28, 581–596. https://doi.org/10.1016/S1002-0160(18)60034-7.

Andrés-Abellán, M., Wic-Baena, C., López-Serrano, F.R., García-Morote, F.A., Martínez-García, E., Picazo, M.I., Rubio, E., Moreno-Ortego, J.L., Bastida-López, F., García-Izquierdo, C., 2019. A soil-quality index for soil from Mediterranean forests. Eur. J. Soil Sci. 70, 1001–1011. https://doi.org/10.1111/ejss.12798.

Andrews, S.S., Carroll, C., 2001. Designing a soil quality assessment tool for sustainable. Ecol. Appl. 11, 1573–1585.

Andrews, S.S., Karlen, D.L., Mitchell, J.P., 2002. A comparison of soil quality indexing methods for vegetable production systems in Northern California. Agric. Ecosyst. Environ. 90, 25–45. https://doi.org/10.1016/S0167-8809(01)00174-8.

Askari, M.S., Holden, N.M., 2014. Indices for quantitative evaluation of soil quality under grassland management. Geoderma 230–231, 131–142. https://doi.org/10.1016/j.geoderma.2014.04.019.

Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 6, 343–351. https://doi.org/10.1038/ismej.2011.119.

Bastida, F., Luis Moreno, J., Hernández, T., García, C., 2006. Microbiological degradation index of soils in a semiarid climate. Soil Biol. Biochem. 38, 3463–3473. https://doi.org/10.1016/j.soilbio.2006.06.001.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008 (10), P10008.

Brady, N.C., Weil, R.R., 2014. The Nature and Properties of Soils, 14th ed. Pearson.

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – a critical review. Soil Biol. Biochem. 120, 105–125.

Caravaca, F., Masciandaro, G., Ceccanti, B., 2002. Land use in relation to soil chemical and biochemical properties in a semiarid Mediterranean environment. Soil Tillage Res. 68, 23–30. https://doi.org/10.1016/S0167-1987(02)00080-6.

Chao, T.T., Theobald, P.K., 1976. The significance of secondary iron and manganese oxides in geochemical exploration. Econ. Geol. 71, 1560–1569. https://doi.org/10.2113/gsecongeo.71.8.1560.

Comunidad de Madrid, CSIC, 1990. Mapa de Asociaciones de Suelos de la Comunidad de Madrid. Consejería de Agricultura y Cooperación Comunidad de Madrid.

Doran, J.W., Zeiss, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. Appl. Soil Ecol. 15, 3–11. https://doi.org/10.1016/S0929-Get.

Gajda, A.M., Czyz, E.A., Dexter, A.R., 2016. Effects of long-term use of different farming systems on some physical, chemical and microbiological parameters of soil quality. Int. Agrophysics 30, 165–172. https://doi.org/10.1515/intag-2015-0081.

García-Ruiz, R., Ochoa, V., Hinojosa, M.B., Carreira, J.A., 2008. Suitability of enzyme activities for the monitoring of soil quality improvement in organic agricultural systems. Soil Biol. Biochem. 40, 2137–2145. https://doi.org/10.1016/j.soilbio.2008.03.023.

Ghani, A., Dexter, M., Perrott, K.W., 2003. Hot-water extractable carbon in soils: A sensitive measurement for determining impacts of fertilisation, grazing and cultivation. Soil Biol. Biochem. 35, 1231–1243. https://doi.org/10.1016/S0038-0717(03)00186-X.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. PNAS 99, 7821–7826. https://doi.org/10.1073/pnas.122653799.

Gómez, J.A., Álvarez, S., Soriano, M.A., 2009. Development of a soil degradation assessment tool for organic olive groves in southern Spain. Catena 79, 9–17. https://doi.org/10.1016/j.catena.2009.05.002.

Hamidi Nehrani, S., Askari, M.S., Saadat, S., Delavar, M.A., Taheri, M., Holden, N.M., 2020. Quantification of soil quality under semi-arid agriculture in the northwest of Iran. Ecol. Indic. 108, 105770 https://doi.org/10.1016/j.ecolind.2019.105770.

Havaee, S., Ayoubi, S., Mosaddeghi, M.R., Keller, T., 2014. Impacts of land use on soil organic matter and degree of compactness in calcareous soils of central Iran. Soil Use Manag. 30, 2–9. https://doi.org/10.1111/sum.12092.

Herrick, J.E., 2000. Soil quality: An indicator of sustainable land management? Appl. Soil Ecol. 15, 75–83. https://doi.org/10.1016/S0929-1393(00)00073-1.

Heuscher, S.A., Brandt, C.C., Jardine, P.M., 2005. Using soil physical and chemical properties to estimate bulk density. Soil Sci. Soc. Am. J. 69, 51–56. https://doi.org/10.2136/sssaj2005.0051a.

Hussain, I., Olson, K.R., Wander, M.M., Karlen, D.L., 1999. Adaptation of soil quality indices and application to three tillage systems in southern Illinois. Soil Tillage Res. 50, 237–249. https://doi.org/10.1016/S0167-1987(99)00012-4.

ISO, 2018. ISO 20130 Soil quality-Measurement of enzyme activity patterns in soil samples using colorimetric substrates in micro-wells plates.

ISRIC, 2002. Procedures for Soil Analysis, 3rd ed. International Soil Reference and Information Center, Wageningen.

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. FAO.

Jahany, M., Rezapour, S., 2020. Assessment of the quality indices of soils irrigated with treated wastewater in a calcareous semi-arid environment. Ecol. Indic. 109, 105800 https://doi.org/10.1016/j.ecolind.2019.105800.

Janes-Bassett, V., Blackwell, M.S.A., Blair, G., Davies, J., Haygarth, P.M., Mezeli, M.M., Stewart, G., 2022. A meta-analysis of phosphatase activity in agricultural settings in response to phosphorus deficiency. Soil Biol. Biochem. 165, 108537 https://doi.org/10.1016/j.soilbio.2021.108537.

Kabata-Pendias, A., Pendias, H., 2001. Trace Elements in Soils and Plants, 3rd ed, Trace Elements in Soils and Plants, Third Edition. CRC, Boca Ratón. https://doi.org/10.1201/9781420039900.

Karlen, D.L., Mausbach, M.J., Doran, J.W., Cline, R.G., Harris, R.F., Schuman, G.E., 1997. Soil quality: a concept, definition, and framework for evaluation (a guest editorial). Soil Sci. Soc. Am. J. 61 (1), 4–10.

Keeney, D.R., Nelson, D.W., 1982. Nitrogen-Inorganic forms. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2- Chemical and Microbiological Properties. American Society of Agronomy & Soil Science Society of America, Madison, Winsconsin, pp. 643–698.

Keiluweit, M., Nico, P., Harmon, M.E., Mao, J., Pett-Ridge, J., Kleber, M., 2015. Long-term litter decomposition controlled by manganese redox cycling. Proc. Natl. Acad. Sci. U. S. A. 112, E5253–E5260. https://doi.org/10.1073/pnas.1508945112.

Khormali, F., Ajami, M., Ayoubi, S., Srinivasarao, C., Wani, S.P., 2009. Role of deforestation and hillslope position on soil quality attributes of loess-derived soils in Golestan province, Iran. Agric. Ecosyst. Environ. 134, 178–189. https://doi.org/10.1016/j.agee.2009.06.017.

Li, S., Li, Q., Wang, C., Li, B., Gao, X., Li, Y., Wu, D., 2019. Spatial variability of soil bulk density and its controlling factors in an agricultural intensive area of Chengdu Plain, Southwest China. J. Integr. Agric. 18, 290–300. https://doi.org/10.1016/S2095-3119(18)61930-6.

Li, H., Santos, F., Butler, K., Herndon, E., 2021. A critical review on the multiple roles of manganese in stabilizing and destabilizing soil organic matter. Environ. Sci. Technol. 55, 12136–12152. https://doi.org/10.1021/acs.est.1c00299.

Liu, L., Wang, Z., Ju, F., Zhang, T., 2015. Co-occurrence correlations of heavy metals in sediments revealed using network analysis. Chemosphere 119, 1305–1313. https://doi.org/10.1016/j.chemosphere.2014.01.068.

Loeppert, R.H., 1986. Reactions of iron and carbonates in calcareous soils. J. Plant Nutr. 9, 195–214. https://doi.org/10.1080/01904168609363437.

Martín-Sanz, J.P., Valverde-Asenjo, I., de Santiago-Martín, A., Quintana-Nieto, J.R., González-Huecas, C., López-Lafuente, A.L., Diéguez-Antón, A., 2018. Enzyme activity indicates soil functionality affectation with low levels of trace elements. Environ. Pollut. 243, 1861–1866. https://doi.org/10.1016/j.envpol.2018.10.029.

Marzaioli, R., D'Ascoli, R., De Pascale, R.A., Rutigliano, F.A., 2010. Soil quality in a Mediterranean area of Southern Italy as related to different land use types. Appl. Soil Ecol. 44, 205–212. https://doi.org/10.1016/j.apsoil.2009.12.007.

Newman, M.E.J., 2006. Modularity and community structure in networks. PNAS 103 (23), 8577–8582.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E - Stat Nonlinear, Soft Matter Phys. 69, 1–16. https://doi.org/10.1103/PhysRevE.69.026113.

Nortcliff, S., 2002. Standardisation of soil quality attributes. Agric. Ecosyst. Environ. 88, 161–168. https://doi.org/10.1016/S0167-8809(01)00253-5.

Olsen, S.R., Sommers, L.E., 1982. Phosphorus. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), Methods of Soil Analysis, Part 2 – Chemical and Microbiological Properties Microbiological and BioChemical Properties. American Society of Agronomy & Soil Science Society of America, Madison, Winsconsin, pp. 403–430.

Paz-Ferreiro, J., Fu, S., 2016. Biological indices for soil quality evaluation: perspectives and limitations. L. Degrad. Dev. 27, 14–25. https://doi.org/10.1002/ldr.2262.

Paz-Ferreiro, J., Gascó, G., Gutiérrez, B., Méndez, A., 2012. Soil biochemical activities and the geometric mean of enzyme activities after application of sewage sludge and sewage sludge biochar to soil. Biol. Fertil. Soils 48, 511–517. https://doi.org/10.1007/s00374-011-0644-3.

Paz-Kagan, T., Ohana-Levi, N., Herrmann, I., Zaady, E., Henkin, Z., Karnieli, A., 2016. Grazing intensity effects on soil quality: A spatial analysis of a Mediterranean grassland. Catena 146, 100–110. https://doi.org/10.1016/j.catena.2016.04.020.

Peris, M., Recatalá, L., Micó, C., Sánchez, R., Sánchez, J., 2008. Increasing the knowledge of heavy metal contents and sources in agricultural soils of the European Mediterranean region. Water. Air. Soil Pollut. 192, 25–37. https://doi.org/10.1007/s11270-008-9631-1.

Puglisi, E., Del Re, A.A.M., Rao, M.A., Gianfreda, L., 2006. Development and validation of numerical indexes integrating enzyme activities of soils. Soil Biol. Biochem. 38, 1673–1681. https://doi.org/10.1016/j.soilbio.2005.11.021.

Rahmanipour, F., Marzaioli, R., Bahrami, H.A., Fereidouni, Z., Bandarabadi, S.R., 2014. Assessment of soil quality indices in agricultural lands of Qazvin Province. Iran. Ecol. Indic. 40, 19–26. https://doi.org/10.1016/j.ecolind.2013.12.003.

Rao, M.A., Scelza, R., Acevedo, F., Diez, M.C., Gianfreda, L., 2014. Enzymes as useful tools for environmental purposes. Chemosphere 107, 145–162. https://doi.org/10.1016/j.chemosphere.2013.12.059.

Rezapour, S., 2014. Response of some soil attributes to different land use types in calcareous soils with Mediterranean type climate in north-west of Iran. Environ. Earth Sci. 71, 2199–2210. https://doi.org/10.1007/s12665-013-2625-3.

Rezapour, S., Najari, S., Ghaemian, N., 2015. The impacts of long-term intensive agriculture on the Vertisol properties in a calcareous region. Environ. Monit. Assess. 187 https://doi.org/10.1007/s10661-015-4453-2.

Sánchez-Navarro, A., Gil-Vázquez, J.M., Delgado-Iniesta, M.J., Marín-Sanleandro, P., Blanco-Bernardeau, A., Ortiz-Silla, R., 2015. Establishing an index and identification of limiting parameters for characterizing soil quality in Mediterranean ecosystems. Catena 131, 35–45. https://doi.org/10.1016/j.catena.2015.02.023.

Schaefer, R., 1963. Dehydrogenase activity as a measurement of the global biological activity of soil. Ann. l'Institut Pasteur 105, 326–331.

Sipos, P., Tóth, A., Kis, V.K., Balázs, R., Kovács, I., Németh, T., 2019. Partition of Cd, Cu, Pb and Zn among mineral particles during their sorption in soils. J. Soils Sediments 19, 1775–1787. https://doi.org/10.1007/s11368-018-2184-z.

Topa, D., Cara, I.G., Jităreanu, G., 2021. Long term impact of different tillage systems on carbon pools and stocks, soil bulk density, aggregation and nutrients: A field meta-analysis. Catena 199, 105102.

United Nations, 2019. World population prospects 2019-Data Booklet.

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter and proposed modification of chromic acid tritation. Soil Sci. 37, 317–327.

Wander, M.M., Bollero, G.A., 1999. Soil quality assessment of tillage impacts in Illinois. Soil Sci. Soc. Am. J. 63, 961–971. https://doi.org/10.2136/sssaj1999.634961x.

Wang, L., Han, X., Liang, T., Guo, Q., Li, J., Dai, L., Ding, S., 2019. Discrimination of rare earth element geochemistry and co-occurrence in sediment from Poyang Lake, the largest freshwater lake in China. Chemosphere 217, 851–857. https://doi.org/10.1016/j.chemosphere.2018.11.060.

Wang, Q., Liu, J., Wang, Y., Guan, J., Liu, Q., Lv, D., an, 2012. Land use effects on soil quality along a native wetland to cropland chronosequence. Eur. J. Soil Biol. 53, 114–120. https://doi.org/10.1016/j.ejsobi.2012.09.008.

Zhang, J., Hou, P., 2012. Soil quality change during reversal of desertification in agro-pastoral transition zone of northern China. Adv. Mater. Res. 599, 859–864. https://doi.org/10.4028/www.scientific.net/AMR.599.859.

Zhang, X., Yang, Y., Zhang, C., Niu, S., Yang, H., Yu, G., Wang, H., Blagodatskaya, E., Kuzyakov, Y., Tian, D., Tang, Y., Liu, S., Sun, X., Luo, Y., 2018. Contrasting responses of phosphatase kinetic parameters to nitrogen and phosphorus additions in forest soils. Funct. Ecol. 32 (1), 106–116.

Zhou, Y., Ma, H., Xie, Y., Jia, X., Su, T., Li, J., Shen, Y., 2020. Assessment of soil quality indexes for different land use types in typical steppe in the loess hilly area. China. Ecol. Indic. 118, 106743 https://doi.org/10.1016/j.ecolind.2020.106743.

Zornoza, R., Mataix-Solera, J., Guerrero, C., Arcenegui, V., Mayoral, A.M., Morales, J., Mataix-Beneyto, J., 2007. Soil properties under natural forest in the Alicante Province of Spain. Geoderma 142, 334–341. https://doi.org/10.1016/j.geoderma.2007.09.002.

Zornoza, R., Mataix-Solera, J., Guerrero, C., Arcenegui, V., Mataix-Beneyto, J., Gómez, I., 2008. Validating the effectiveness and sensitivity of two soil quality indices based on natural forest soils under Mediterranean conditions. Soil Biol. Biochem. 40, 2079–2087. https://doi.org/10.1016/j.soilbio.2008.01.014.

Zornoza, R., Acosta, J.A., Bastida, F., Domínguez, S.G., Toledo, D.M., Faz, A., 2015. Identification of sensitive indicators to assess the interrelationship between soil quality, management practices and human health. Soil 1, 173–185. https://doi.org/10.5194/soil-1-173-2015.