soil systems

MDPI

Article

Application of Self-Organizing Maps to Explore the Interactions of Microorganisms with Soil Properties in Fruit Crops Under Different Management and Pedo-Climatic Conditions

Francesca Antonucci [1], Simona Violino [1], Loredana Canfora [2], Małgorzata Tartanus [3], Ewa M. Furmanczyk [3], Sara Turci [4], Maria G. Tommasini [4], Nika Cvelbar Weber [5], Jaka Razinger [5], Morgane Ourry [6], Samuel Bickel [7], Thomas A. J. Passey [8], Anne Bohr [9], Heinrich Maisel [10], Massimo Pugliese [11], Francesco Vitali [12], Stefano Mocali [12], Federico Pallottino [1], Simone Figorilli [1], Corrado Costa [1,*] and Eligio Malusà [3,13,*]

1. Centro di Ricerca Ingegneria e Trasformazioni Agroalimentari, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Via Della Pascolare 16, 00015 Monterotondo, Italy; francesca.antonucci@crea.gov.it (F.A.); simona.violino@crea.gov.it (S.V.); federico.pallottino@crea.gov.it (F.P.); simone.figorilli@crea.gov.it (S.F.)
2. Centro di Ricerca Agricoltura e Ambiente, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Via della Navicella 4, 00184 Rome, Italy; loredana.canfora@crea.gov.it
3. The National Institute of Horticultural Research, ul. Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland; malgorzata.tartanus@inhort.pl (M.T.); ewa.furmanczyk@inhort.pl (E.M.F.)
4. RI.NOVA Research and Innovation Soc. Coop., Via Dell'Arrigoni 120, 47522 Cesena, Italy; sturci@rinova.eu (S.T.); mgtommasini@rinova.eu (M.G.T.)
5. Agricultural Institute of Slovenia, Hacquetova ulica 17, SI-1000 Ljubljana, Slovenia; nika.weber@kis.si (N.C.W.); jaka.razinger@kis.si (J.R.)
6. Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark; mo@plen.ku.dk
7. Institute of Environmental Biotechnology, Technische Universitaet Graz, Rechbauerstrasse 12, 8010 Graz, Austria; samuel.bickel@tugraz.at
8. National Institute of Agricultural Botany (NIAB), New Road, East Malling, Kent ME19 6BJ, UK; tom.passey@niab.com
9. Competence Center for Fruit Crops at the Lake of Constance, Schuhmacherhof 6, 88213 Ravensburg, Germany; bohr@kob-bavendorf.de
10. Fordergemeinschaft Okologischer Obstbau EV, Traubenplatz 5, 74189 Weinsberg, Germany; heinrich.maisel@posteo.de
11. Agroinnova, Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco, Italy; massimo.pugliese@unito.it
12. Centro di Ricerca Agricoltura e Ambiente, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Via di Lanciola 12/A, 50125 Firenze, Italy; francesco.vitali@crea.gov.it (F.V.); stefano.mocali@crea.gov.it (S.M.)
13. Centro di Ricerca Viticoltura ed Enologia, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Via P. Micca 35, 14100 Asti, Italy

* Correspondence: corrado.costa@crea.gov.it (C.C.); eligio.malusa@inhort.pl (E.M.)

check for updates

Academic Editor: Jarosław Zawadzki

Received: 10 October 2024
Revised: 16 January 2025
Accepted: 22 January 2025
Published: 26 January 2025

Citation: Antonucci, F.; Violino, S.; Canfora, L.; Tartanus, M.; Furmanczyk, E.M.; Turci, S.; Tommasini, M.G.; Cvelbar Weber, N.; Razinger, J.; Ourry, M.; et al. Application of Self-Organizing Maps to Explore the Interactions of Microorganisms with Soil Properties in Fruit Crops Under Different Management and Pedo-Climatic Conditions. *Soil Syst.* **2025**, *9*, 10. https://doi.org/10.3390/soilsystems9010010

Abstract: Background: Self-organizing maps (SOMs) are a class of neural network algorithms able to visually describe a high-dimensional dataset onto a two-dimensional grid. SOMs were explored to classify soils based on an array of physical, chemical, and biological parameters. Methods: The SOM analysis was performed considering soil physical, chemical, and microbial data gathered from an array of apple orchards and strawberry plantations managed by organic or conventional methods and located in different European climatic zones. Results: The SOM analysis considering the "climatic zone" categorical variables was able to discriminate the samples from the three zones for both crops. The zones were associated with different soil textures and chemical characteristics, and for both crops, the Continental zone was associated with microbial parameters—including biodiversity indices derived from the NGS data analysis. However, the SOM analysis based on the "management method" categorical variables was not able to discriminate the soils between organic and integrated management. Conclusions: This study allowed for the discrimination of soils of medium- and long-term fruit crops based on their pedo-climatic characteristics and associating these characteristics to some indicators of the soil biome, pointing to the possibility of better understanding the interactions among diverse variables, which could support unraveling the intricate web of relationships that define soil quality.

Keywords: apple; neural networks; soil microbiome diversity; strawberry

1. Introduction

Self-organizing maps (SOMs) are a class of neural network algorithms in the unsupervised-learning category that perform a nonlinear projection of a high-dimensional dataset onto a two-dimensional grid [1]. The most popular learning algorithm was developed by Kohonen [2]; hence, SOM architecture is often referred to as Kohonen's model. Since the map preserves the topology in the display, the clustering of the data space, as well as the metric-topological relationships of the data, items are clearly visible, allowing for the exploration of relationships and patterns between parameters as well as structures in a single parameter (e.g., approximating the probability distribution of the parameter) [3,4]. The advantage of having a flexible map instead of a plane specified by principal components analysis allows us to better separate classes [5], making it suitable for clustering, visualization, and extraction of information from multi-dimensional data [6,7]. The dataset used for training the map can be tailored to include or combine parameters relevant to the studied case, for example, combining physical parameters with biological indicators [3–8].

Various studies have applied SOMs to agricultural topics and, interestingly, mainly to soil-related features. For example, this approach was applied to understand the relationship between spatial vegetation patterns and dryland resilience [9] to validate a water stress index [10], or to predict relations between soil chemical parameters [11,12] or physical features [13], individually or together [4]. The application of neural network models to better represent soil complexity is explained considering the complex nature and multifunctionality of soil, its spatial variability, and the timescale of soil processes.

Soil biological parameters are gaining interest due to the soil biome's critical functions on all soil processes and crop performance [14]. The populations of soil micro-, meso-, and macro-organisms are affected by land use, weather conditions, topography, crop management, etc. [15], which are both spatially and temporarily variable. However, the relationship between soil's physical, chemical, and biological characteristics is not fully understood. It has been shown that some chemical parameters, e.g., pH or organic matter content, are frequently associated with microbial activity [16] but their relationships with microbial biodiversity are still unclear. Patterns of soil microbial populations have been identified at various scales [17,18]; however, representing these relationships and defining classes of soil quality based on complex datasets comprising various kinds of soil parameters is still a challenge [19]. In this respect, SOMs, not relying on the assumption of spatial stationarity [11], can thus capture such complex relationships and patterns in the data, allowing for the representation of both global and local structures that are easier to interpret and suitable for classifying soils with different quality or fertility levels, similarly to what was obtained in other cases [20].

Indeed, the definition of a comprehensive indicator that could include soil's physical, chemical, and biological characteristics would be useful to enable all stakeholders (farmers, advisors, scientists, and policymakers) to make informed evaluations regarding practices, processes, and policies for soil use and management [19]. However, the representation of the index, eventually in relation to the crop, should be easy to interpret and visualize if a practical use should be expected. SOMs could, thus, represent a suitable tool also in this respect.

The objective of this study was to exploit the potential use of SOMs to classify soils based on an array of physical, chemical, and biological parameters. Considering that the practices for soil management differ depending on the cropping system and farm management, medium- (strawberry) and long-term (apple) cropping systems were analyzed, which were managed by conventional and organic methods, gathering data from different European climatic zones.

2. Materials and Methods

2.1. Study Sites

This study utilized data obtained from 21 locations situated in Austria, Denmark, Germany, Italy, Poland, and the UK. The fields selected for the trials planned in the Excalibur project ("https://excaliburh2020.eu/en", accessed on 10 January 2025) represent different climatic conditions and were characterized by diverse field histories and management methods (Supplementary Materials, Table S1). Some of the fields were already hosting the crop (e.g., existing apple orchards—hereafter referred to as the long-term fruit cropping system), while those planned for strawberry plantations were prepared for the establishment of this crop (hereafter referred to as the medium-term fruit cropping system). The fields were managed according to organic farming (ORG) or conventional methods applying integrated pest management (IPM).

2.2. Soil Sample Collection and Analysis

Soil samples were collected from the selected locations at the beginning of the growing season (March–June) in 2021. This time range was due to the differences in climatic conditions, which affected the beginning of the growing season in the different countries. However, the sampling time was suitable to match with the period of dynamic soil microorganism activity. Sampling was performed using Egner's auger (2.5-centimeter diameter) collecting at least 10 subsamples at 0–20 cm depth, which were pooled to form the analytical sample. Four pooled samples (replicates) were collected from each field. The samples were used for physical and chemical analyses, as well as for DNA extraction, to determine bacteria and fungi populations.

Soil texture was determined according to the ISO standard 11277:2020 method [21]. The following soil chemical characteristics were determined: pH; organic matter (OM); total phosphorus (P); total potassium (K); total calcium (Ca); total magnesium (Mg). Mineral elements were determined as previously described [22]. The pH was determined in a water suspension (1:20 vol) with a pH meter. DNA was extracted from 1 g of wet soil following the instructions of the EZNA Soil DNA kit from Omega (Omega Bio-tek, Inc., Norcross, GA, USA). The DNA was subjected to qualitative evaluation by a Nanodrop 1000 spectrophotometer (ThermoFisher, Waltham, MA, USA). The DNA was then filtered and concentrated using 30 kDa Amicon ultra 0.5 mL centrifugal filters (Merk, Darmstadt, Germany) and sequenced by NGS analysis on the Illumina Platform (i.e., MiSeq 2 × 300). Sequencing libraries for bacteria were prepared using V4 16S rRNA gene sequencing from the Earth Microbiome Project ("https://earthmicrobiome.org/protocols-and-standards/16s/", accessed on 12 September 2024) with the Illumina adapters. Sequencing libraries for fungal ITS1 were prepared by amplifying specific locus ITS1f-ITS2 tailed with the Illumina adapters. The bacterial sequences were clustered into Operational Taxonomic Units (OTUs) based on 97% similarity with the SILVA reference database 132 [23]. Fungal taxonomy was assigned using the UNITE database version 8.2, dynamic [24]. Data were evaluated using a standardized bioinformatic pipeline (i.e., QIIME2), which allows for an in-depth analysis of complex microbial communities (diversity, taxonomic composition) and provides first indications for microbial functions.

The total number of bacteria (observed_bact) and fungal (observed_fung) OTU were determined based on the results of NGS analysis. The alpha diversity of the microorganisms' populations (diversity_shannon bact; diversity_shannon fung) was calculated using Shannon index. Pielou index was calculated to describe microbial communities' evenness (evenness_pielou_bact; evenness_pielou_fung).

For each variable considered, the Shapiro–Wilk W test (H_0 = normality) showed a univariate non-normal distribution ($p < 0.05$). For this reason, a non-parametric Kruskal–Wallis test (H_0 = same median) was used to compare, for each variable, the median of the climatic zones. Tests were performed with the software PAST (version 2.17v) [25].

2.3. Self-Organizing Maps (SOMs)

For the SOM analysis, the R package kohonen (3.0.11 version) was used [26]. SOM analysis of the long-term fruit cropping system included 38 samples, while 46 samples were used for the medium-term fruit cropping system. For both cropping systems, 15 variables were selected as follows: silt (%), clay (%), sand (%), pH, OM (g/kg), P (ppm), K (ppm), Ca (ppm), Mg (ppm), observed_bact (number), diversity_shannon_bact (H′), evenness_pielou_bact, observed_fung (number), diversity_shannon_fung (H′), evenness_pielou_fung.

The SOM features two fully connected layers: the input layer and the output layer. Each neuron in the output layer is represented in two ways: by a reference vector, which has as many components as the input variables, and by its position in the grid [27].

The reference vectors collectively formed a codebook, which approximates the space covered by the original dataset (input data or input matrix) and its probability distribution [7]. The map, projecting the codebook onto a grid, enables the identification of clusters of observations with similar characteristics by analyzing all parameters simultaneously.

The SOM analyses of the two datasets (i.e., apple and strawberry) were organized as follows: one vector for each object was created by concatenating the dependent (X; described in Section 2.2) and independent variables (Y), and the SOM was trained by using the distance of an object to a unit as the sum of separate distances for X and Y spaces [11]. A hexagonal SOM grid with 4×4 dimensions was used for the training. The SOM's dimension was determined using the heuristic rule, $\left(5 \cdot \sqrt{N}\right)/2$, aiming to compress the data into a smaller (possibly more manageable) number of patterns for analysis and obtain an acceptable accuracy (of representation of the data vectors) [28]. The number of data layers was 2 and the distance measure used was the sum of squares. The neighbor function used was Gaussian and the number of iterations for the training phase was 100 (random initialization).

The variables were analyzed considering two groups of categorical variables (dependent): the field management method (two variables, ORG and IPM); and the climatic zone to which the field belonged according to the biogeographical classification defined by the European Environmental Agency [29]. For the apple orchards, these included three categorical variables—Atlantic (A), Continental (C), and Mediterranean (M); for the strawberry plantations, these included three categorical variables—Continental (C), Mediterranean (M), and Alpine (L).

SOM analysis was performed to identify the key characteristics that make up the dominant patterns in the dataset (pattern extraction) and to discover which data items were similar to each other with respect to these characteristics (clustering). The characteristics of the predominant patterns are revealed by the high-dimensional vectors associated with each map node—the location of which is defined by the combination of dimension values that make up its vector. As data elements are mapped to their nearest map nodes after training, clusters of data elements with similar characteristics are formed. Coloring was used to indicate dimensional weights (values) at each node, with a separate component plane showing the values of each dimension of the SOM. The axes and grid nodes corresponded exactly to those of the SOM output map [28–30]. Count plots were obtained to show the number of objects mapped to the individual units. Empty units were depicted in gray. Quality plots were obtained to show the mean distance of objects mapped to a unit to the codebook vector of that unit. The smaller the distance, the better the objects were represented by the codebook vectors.

3. Results

3.1. Univariate Analysis of Long-Term Fruit Cropping System

Significant differences emerged between Continental (C) and Mediterranean (M) climatic zones after the univariate analysis (Table 1) of the following variables: silt, clay, sand, OM, pH, P, K, Mg, observed_bact, observed_fung, and diversity_shannon_fung. When the same analysis was conducted considering the grouping of variable management (integrated vs. organic), significant differences were found only for OM and Mg.

Table 1. Median values of the soil's physical–chemical and microbial parameters in the long-term fruit cropping systems (apple) in the group "climatic zone" for the three categorical variables considered: Atlantic (A), Continental (C), and Mediterranean (M). Different letters indicate a significant difference, $p \leq 0.05$ (Kruskal–Wallis test).

Variables	Climatic Zone		
	A	C	M
Silt	30 [ab]	20 [a]	41 [b]
Clay	8 [ab]	3 [a]	42 [b]
Sand	62 [ab]	77 [b]	17 [a]
OM	17.9 [ab]	23.1 [b]	2.1 [a]
pH	6.3 [ab]	6.6 [a]	8.2 [b]
P	679.5 [ab]	807.5 [b]	19 [a]
K	5089.5 [ab]	2116.5 [b]	284 [a]
Ca	4217.5 [a]	3708 [a]	3915 [a]
Mg	5756 [ab]	2935.5 [b]	402 [a]
observed_bact	569 [ab]	581.5 [b]	553 [a]
diversity_shannon_bact	5.9 [a]	5.9 [a]	5.7 [a]
evenness_pielou_bact	0.9 [a]	0.9 [a]	0.9 [a]
observed_fung	287.5 [ab]	355 [b]	258 [a]
diversity_shannon_fung	4.4 [ab]	4.6 [b]	4.1 [a]
evenness_pielou_fung	0.8 [a]	0.8 [a]	0.8 [a]

3.2. SOM of Long-Term Fruit Cropping System

The SOM for the long-term fruit cropping system considering the "climatic zone" categorical variables returned a pattern in which all three variables were well-discriminated (the left part of Figure 1). The group Atlantic (A) is positioned with one cell on the bottom-left side of the map. The group Continental (C), even if separated from the other two groups, is split into three different positions on the map. This is also visible considering the counts and quality plots in Figure 2. A first sub-group (C1) is positioned with one cell on the bottom-left side of the map; a second sub-group (C2), consisting of five cells, is positioned on the bottom-right side of the map; and a third sub-group (C3) is positioned with one cell on the top-right side of the map. The group Mediterranean (M) is positioned with two cells on the top-left side of the map, well-isolated from the others (Figure 2).

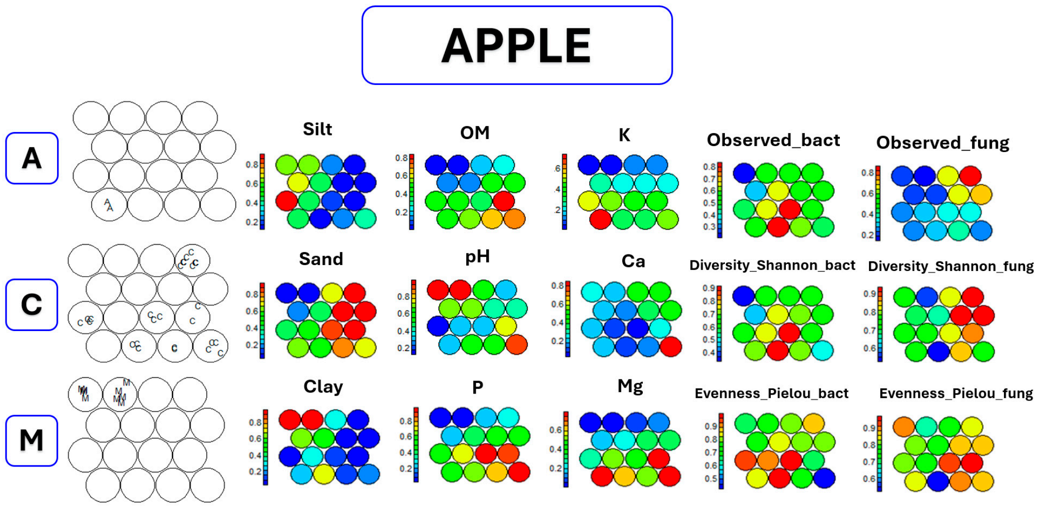

Figure 1. Trained SOM for the long-term fruit cropping system (apple) in the group "climatic zone" for the three categorical variables considered: Atlantic (A), Continental (C), and Mediterranean (M). On the left side, the maps of input records (black and white) split for the three categorical variables considered. On the right side, the heatmaps for each experimental variable show the value distribution among the units in color-code (blue = low value; red = high value).

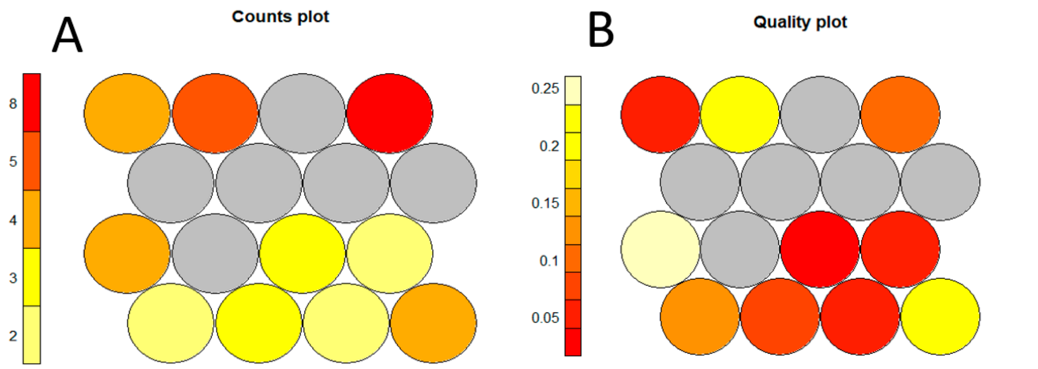

Figure 2. Count (**A**) and quality (**B**) plots for SOM trained for the long-term fruit cropping system (apple) in the group "climatic zone".

Observing the heatmaps for each experimental variable, showing the value distribution among the units in color-code (the right side of Figure 1), it is possible to observe the relationship between variables and groups. In particular, the group Atlantic (A) is characterized by higher values of Mg and K. The group Continental (C1) is characterized by high values of silt and low values of pH and clay. The group Continental (C2) is characterized by high values of sand, OM, P, Mg, Ca, Evenness_Pielou_fung, and all the variables related to bacteria; for the group C2, low values of silt and clay are shown. The group Continental (C3) is characterized by high values of sand, Obs_fung, and Diversity_Shannon_fung and low values of silt and clay. The group Mediterranean (M) is characterized by high values of clay and pH and low values of sand, OM, P, K, Mg, Obs_bact, Diversity_Shannon_bact, Obs_fung, and Diversity_Shannon_fung.

The SOM elaborated considering the two categorical variables related to the "management method", was not able to discriminate between organic and integrated management (Supplementary Materials, Figure S1).

3.3. Univariate Analysis of Medium-Term Fruit Cropping System

Significant differences were found among climatic zones for the following variables: silt, clay, sand, OM, pH, P, K, Mg, and evenness_pielou_bact (Table 2). The same procedure was conducted considering the two grouping variables related to soil management (integrated vs. organic), displaying significant differences concerning the variables silt, sand, pH, K, and Mg (Supplementary Materials, Table S2).

Table 2. Median values of the soil's physical–chemical and microbial parameters for medium-term fruit cropping system (strawberry) performed on the categorical group "climatic zone" for the three categorical variables considered: Continental (C), Mediterranean (M), and Alpine (L). Different letters indicate a significant difference, $p \leq 0.05$ (Kruskal–Wallis test).

Variables	Climatic Zone		
	C	L	M
Silt	20 ^{a}	29 ^{b}	54 ^{b}
Clay	1 ^{a}	2 ^{b}	34 ^{c}
Sand	79 ^{b}	69 ^{a}	12 ^{a}
OM	33.8 ^{b}	2.7 ^{ab}	1.8 ^{a}
pH	7.5 ^{b}	5.7 ^{a}	8.2 ^{c}
P	861.2 ^{b}	9.9 ^{a}	22 ^{a}
K	1079 ^{b}	13.9 ^{ab}	257 ^{a}
Ca	3508 ^{a}	3245 ^{a}	3365 ^{a}
Mg	1866 ^{b}	8751 ^{c}	284 ^{a}
observed_bact	584 ^{a}	641 ^{a}	589 ^{a}
diversity_shannon_bact	5.9 ^{a}	6.0 ^{a}	5.9 ^{a}
evenness_pielou_bact	0.92 ^{ab}	0.93 ^{b}	0.92 ^{a}
observed_fung	293.0 ^{a}	273.5 ^{a}	233.5 ^{a}
diversity_shannon_fung	4.1 ^{a}	4.0 ^{a}	4.1 ^{a}
evenness_pielou_fung	0.7 ^{a}	0.7 ^{a}	0.8 ^{a}

3.4. SOM of Medium-Term Fruit Cropping System

The SOM for the medium-term fruit cropping system (strawberry) and the "climatic zone" categories returned a pattern in which the three zones were discriminated (Figure 3). The group Continental (C), even if separated from the other two groups, is split into two different positions on the map. This is also visible considering the count and quality plots in Figure 4. A first sub-group (C1) is positioned with one cell in the center of the map; a second sub-group (C2), consisting of three cells, is positioned on the bottom-right side of the map. The group Alpine (L), even if separated from the other two groups, is split into two different positions on the map. This is also visible considering the count and quality plots in Figure 4. A first sub-group (L1) is positioned with one cell on the top-left side of the map; a second sub-group (L2), is positioned with one cell on the central-left side of the map. The group Mediterranean (M) is positioned with two cells on the top-right side of the map, well-isolated from the others (Figure 4).

Observing the heatmaps for each experimental variable showing the value distribution among the units in color-code (the right side of Figure 3), it is possible to observe the relationship between variables and groups. The group Continental (C1) is characterized by high values of P, Ca, Evenness_Pielou_fung, and all the variables related with bacteria. The group Continental (C2) is characterized by high values of sand, K, Obs_fung, and Diversity_Shannon_fung and low values of silt, clay, Ca, and Evenness_Pielou_fung. The group Alpine (L1) is characterized by high values of silt and OM and low values of sand and Ca. The group Alpine (L2) is characterized by high values of Mg and low values of clay, pH, P, K, and Ca. The group Mediterranean (M) is characterized by high values of clay and pH and low values of sand, OM, P, Ca, Mg, Obs_fung, Diversity_Shannon_fung, and all the variables related with bacteria.

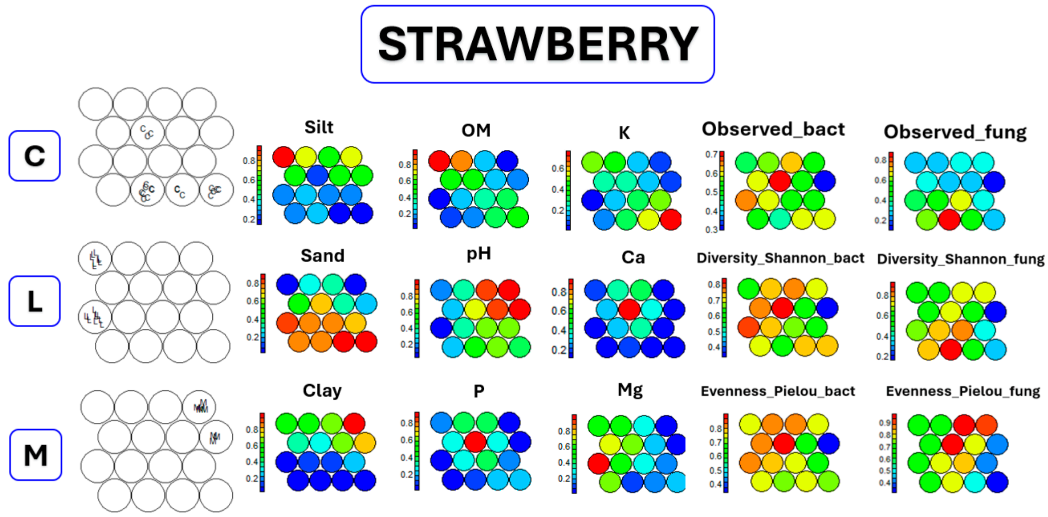

Figure 3. Trained SOM for medium-term fruit cropping system (strawberry) performed on the categorical group "climatic zone" for the three categorical variables considered: Continental (C), Alpine (L), and Mediterranean (M). On the left side, the maps of input records (black and white) split for the three categorical variables considered. On the right side, the heatmaps for each experimental variable show the value distribution among the units in color-code (blue = low value, red = high value).

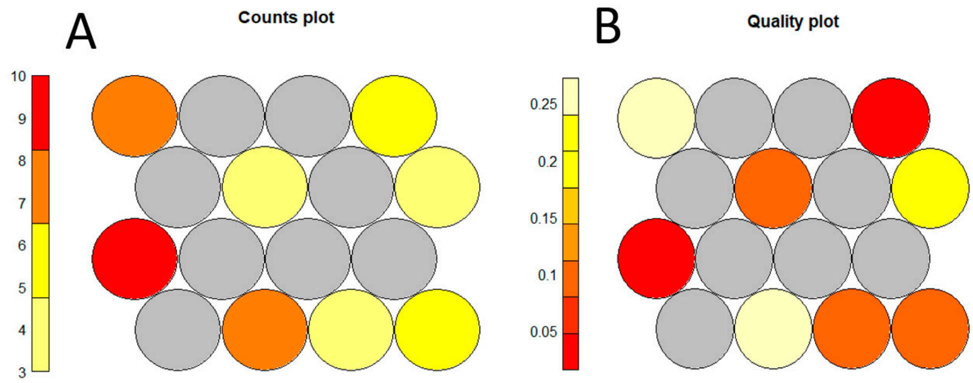

Figure 4. Count (**A**) and quality (**B**) plots for SOM trained for the medium-term fruit cropping system (strawberry), in the group "climatic zone".

The SOM based on the "management method" categorical variables was not able to discriminate between organic and integrated management of this medium-term cropping system (Supplementary Materials, Figure S2).

4. Discussion

The primary function of SOMs is to perform an unsupervised classification. The soil variables considered in this study were well-classified with the SOMs on the basis of climatic zones for both long-term and medium-term cropping systems. This was not clearly emerging when considering the simple analysis of each individual parameter. In particular, the Continental zone was strongly characterized by some physical–chemical parameters, particularly P and Ca, and the totality of biological parameters related to both bacterial and fungal communities, irrespective of the cropping system considered. The other climatic zones were instead characterized by only a few physical or chemical parameters. Clay and pH characterized the Mediterranean zone for both cropping systems, although the fields of each crop were located in different regions with diverse kinds of soils (Table S1). These two parameters are not commonly related, although the mineralogical characteristics of clay can impact on soil pH [31]. In the case of strawberry, OM was associated to soils located in the Alpine zone, while it was included in the Continental zone when soils from apple orchards were considered. The mineralization of soil organic matter is affected by several factors, including climatic conditions (i.e., temperature and moisture), soil physicochemical properties, and the soil's microorganisms [32–34]. However, the impact of the diverse soil management practices between strawberry and apple orchards could also account for this result [35,36]. The analysis of the count and quality plots confirmed that the discrimination of the soils according to the climatic zones was based on a combination of various parameters, highlighting some associations among them that could be expected, being the parameters related (e.g., as in the case of texture variables), as well as between some that could not be assumed a priori (e.g., in the case of the content of some nutrients and microbial features). In some cases, the SOM could have suggested that more than one group was possible to be identified within the same climatic zone. For example, the fields of the Continental zone of the long-term cropping system characterized by high Mg and P content could be differentiated from the others showing high bacterial diversity. Similarly, for the medium-term cropping system, the group of fields showing high silt and OM—which were related to the Alpine zone—could have been distinguished from those showing high sand and Mg; however, this seemed less feasible when considering both the count and quality plots. In the latter case, it shall be considered that sand and silt content are in relation also with the clay content, which was consistently low in both cases. It is intriguing that the Mediterranean zone was characterized by high clay and pH, together with high fungal diversity, although the size of the fungal population was low. This could point to the interaction between the three components of the soil fertility as affected by the climatic conditions.

Soil bacterial and fungal communities are directly driven by climate or soil characteristics [37,38] but also indirectly through processes induced by the plant species [39]. However, it is noteworthy that the SOM analysis carried out was not able to discriminate either cropping system when the variables were categorized according to the management method (organic or conventional). Soil management practices are considered an important factor in shaping the soil fertility, particularly the microbial community [40], e.g., inducing higher microbial richness and diversity [41] and organic matter content [42]. Plant protection, soil tillage, and fertilization are major practices that diverge between organic and conventional crop management, which can modify the soil characteristics in the long-term. However, even though pesticide applications influenced the community abundance of bacteria and fungi [43], the pesticide usage—in either organic farming or integrated pest management of conventional crops—is limited to non-synthetic or to less harmful active substances, respectively. Similarly, tillage practices are also impacting the soil's microbial population and chemical characteristics [44]. The two cropping systems were characterized by a diversified management of soil tillage: less invasive for the long-term apple orchards (where tillage is normally not performed on the row where soil samples were collected) compared to the seasonal cultivation of the medium-term cropping system characteristic of strawberry production. This difference applied also to the fertilization practices, prompting the hypothesis that also different fertilization practices, though impacting the soil's chemical and biological characteristics [45], would not be enough to modify the soil to an extent that allowed for SOM discriminating the two management methods. Indeed, the method of field management was also found to not affect the nematodes' population (size and trophic groups diversity) in samples of the same sites analyzed in this study (unpublished data), which further confirms the difficulty in clearly discriminate soils according to management methods.

Conclusions supporting the results obtained in this study are provided by recent analyses showing that the impact of land use and some soil chemical characteristics on soil microbial community are mainly driven by climatic conditions [18,19]. Nevertheless, in an effort of recognizing patterns in macrofungal communities, SOM was effectively discriminating xerothermal meadows into three plant coenoses, which were characterized by a very different number of fungal species [46]. Mapping environmental data onto SOM patterns that indicate the health of an ecosystem or predict the consequences of environmental changes [47,48] would result in identifying the most important parameters suitable for modeling of the environmental and economic system data and the trade-off analysis useful to propose sustainable environment management [49].

The relationship between biotic and chemical–physical variables within the Continental climatic zone highlighted possible positive and negative correlations among them. For example, for the medium-term fruit crop, it was possible to observe a positive relationship between the bacteria population (all parameters) and the high content of some nutrient elements (P and Ca), which has been observed under different conditions as well [50]. The high content of Ca is generally associated to high pH in soils, and could be also associated to high P content (as calcium phosphate is a common soil mineral). Their association with the bacteria community could depend on the capacity of bacteria species to solubilise P compounds with low solubility containing also calcium. For the long-term fruit crop, texture components resulted differently associated to microbial community indices. The richness and diversity of the soil fungal population is known to be positively correlated to some soil fractions [51]. Various studies have demonstrated that the diversity and composition of soil microbial communities are linked to soil pH and nutrients [17,52–54]. Moreover, a study at the continental level pointed out an effect of soil mineral nutrient levels on the microbiome [37]. Phosphorus, potassium, and calcium contents correlated with both bacterial and fungal communities' diversity, like the results from the SOM analyses presented here. Interestingly, the SOM classified sand content within the Continental zone together with the soil microorganisms' community parameters, a textural characteristic that was also associated with potential bacterial and fungal pathogens [37], common species of agricultural soils such as those included in the present study.

Grouping the input data according to the climatic zone of the study sites allowed for identifying consistent associations between the variables, irrespective of the cropping system. Interestingly, SOM analysis was able to classify different soil types based on variables such as color, texture, and drainage class with almost 92% correctness [55]. The discrimination power of SOM was much higher compared to PCA and hierarchical cluster analysis (HCA) to evaluate sediment quality of 40 sites from three estuaries using physical, chemical, and ecotoxicological variables [56]. SOM was able to differentiate a group of stations characterized by lower contamination and toxicity, thus with lower environmental risk, and not requiring further steps of a multilevel assessment framework, which were instead clustered together with other sites by the HCA or PCA analyses. The use of various biological parameters, including biochemical (phospholipid fatty acids) and genetic (RFLPs) variables of the soil microbial community in an SOM analysis allowed us to discriminate soils from two different regions of Australia [57]. The associations of various chemical and physical variables and the indices characterizing microbial communities that were pointed out in the current study through SOM could improve our understanding of the relationships between the factors defining the soil quality toward an overall classification of soil quality.

Nevertheless, despite their strengths, SOMs are not without limitations. One challenge is the interpretability of the resulting maps, as the visualization may become cluttered or overly complex with very-high-dimensional data. Furthermore, while SOMs can highlight correlations and patterns, they do not provide direct causal insights, necessitating further analysis to draw definitive conclusions. An example of these challenges is reflected in a large study applying SOM on a multidimensional soil dataset surveying the Central Valley of Chile [4]. Even though the pH is considered a critical factor in shaping soil microbial communities [58], and has been found associated in our study to various microbial parameters, it was the only parameter of that study that did not have important associations with other parameters. On the other hand, the study pointed out that the content of organic matter related to soil series characteristics and management practices, which was not possible to confirm in the present study. The availability of a wider dataset in terms of samples and specific metadata, including soil management practices, could provide a higher classification power to the SOM analysis of soils, useful for practical purposes.

5. Conclusions

The advent of sophisticated computational techniques has significantly transformed the landscape of data analysis. Among these, multivariate analysis techniques such as SOMs have emerged as powerful tools for analyzing and interpreting complex datasets by visualizing data of a different nature. The present study allowed us to discriminate soils of medium- and long-term fruit crops based on their pedo-climatic characteristics and associating these characteristics to some indicators of the soil biome. However, it was not possible, for both crops, to discriminate between soils managed with organic farming or integrated (i.e., conventional) methods. Nevertheless, by facilitating the exploration of interactions among such diverse variables, particularly those associated to soil microbiome or other biological features, SOMs could become a method to derive meaningful insights from a complex matrix such as soil, helping to unravel the intricate web of relationships that define its quality.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/soilsystems9010010/s1, Table S1: List of fields and their characteristics used for the SOM analysis. Table S2. Median value of each variable analyzed for the long-term (apple) and medium-term (strawberry) cropping systems in the group "management" for the two categorical variables considered: integrated (I) and organic (O). Different letters indicate significant differences, $p \leq 0.05$ (Kruskal–Wallis test). Figure S1: Mapping plot of the trained SOM performed for the long-term cropping system on the categorical group "management" for the two categorical variables considered: integrated (I), organic (O). Figure S2: Mapping plot of the trained SOM performed for the medium-term cropping system on the categorical group "management" for the two categorical variables considered: integrated (I), organic (O).

Author Contributions: Conceptualization, F.A., S.V., L.C., C.C. and E.M.; methodology F.A., S.V., L.C., C.C. and E.M.; software, S.F., F.A., F.P., C.C., S.V. and L.C.; validation, C.C., L.C. and E.M.; formal analysis, F.A., S.V., L.C., M.T., E.M.F., S.T., M.G.T., N.C.W., M.O., S.B., T.A.J.P., A.B., H.M., F.V., F.P., S.F. and C.C.; investigation, F.A., S.V., L.C., M.T., E.M.F., S.T., M.G.T., N.C.W., M.O., S.B., T.A.J.P., A.B., H.M., F.V., F.P., S.F. and C.C.; resources, F.A., S.V., L.C., M.T., E.M.F., S.T., M.G.T., N.C.W., M.O., S.B., T.A.J.P., A.B., H.M., F.V., M.P., F.P., S.F. and C.C.; data curation, F.A., S.V., L.C., M.T., E.M.F., S.T., M.G.T., N.C.W., M.O., S.B., T.A.J.P., A.B., H.M., F.V., M.P., J.R., F.P., S.F. and C.C.; writing—original draft preparation, F.A., S.V., L.C., E.M. and C.C.; writing—review and editing, F.A., S.V., L.C., E.M., S.M. and C.C.; visualization, F.A., S.V., L.C., E.M. and C.C.; supervision, F.A., S.V., L.C., E.M. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was performed under the EXCALIBUR project, funded by the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 817946).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because they are part of an ongoing study and finalization of the full dataset deriving from the project, for which funding has been received. Requests to access the datasets should be directed to the corresponding authors.

Acknowledgments: The authors would like to acknowledge the critical support of the anonymous reviewers for improving the manuscript.

Conflicts of Interest: Sara Turci and Maria G. Tommasini were employed by the company RI.NOVA Research and Innovation Soc. Coop. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Information Sciences; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2001; pp. 105–175.
2. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [CrossRef]
3. Vesanto, J.; Alhoniemi, E. Clustering of the Self-organizing Map. *IEEE Trans. Neural Netw.* **2000**, *11*, 586–600. [CrossRef] [PubMed]
4. Rivera, D.; Sandoval, M.; Godoy, A. Exploring soil databases: A self-organizing map approach. *Soil Use Manag.* **2015**, *31*, 121–131. [CrossRef]
5. Hulle, M.M.V. *Faithful Representations and Topographic Maps: From Distortion-to Information-Based Self-Organization*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2000; ISBN 0471345075.
6. Penn, B. Using self-organizing maps to visualize highdimensional data. *Comput. Geosci.* **2005**, *31*, 531–544. [CrossRef]
7. Herbst, M.; Casper, M.C. Towards model evaluation and identification using Self-organizing Maps. *HESS* **2008**, *12*, 657–667. [CrossRef]
8. Astel, A.; Tsakovski, S.; Barbieri, P.; Simeonov, S. Comparison of self-organizing maps classification approach with cluster and principal components analisis for large environmental data sets. *Water Res.* **2007**, *41*, 4566–4578. [CrossRef]
9. Kéfi, S.; Génin, A.; Garcia-Mayor, A.; Guirado, E.; Cabral, J.S.; Berdugo, M.; Guerber, J.; Solè, R.; Maestre, F.T. Self-organization as a mechanism of resilience in dryland ecosystems. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2305153121. [CrossRef]
10. Kumar, S.; Islam, A.R.M.T.; Hasanuzzaman, M.; Salam, R.; Khan, R.; Islam, M.S. Preliminary assessment of heavy metals in surface water and sediment in Nakuvadra-Rakiraki River, Fiji using indexical and chemometric approaches. *JEM* **2021**, *298*, 113517. [CrossRef]
11. Iyer, R.T.; Krishnan, M.T. Spatial prediction of soil micronutrients using supervised self-organizing maps. *J. Agric. Food Res.* **2024**, *15*, 101033. [CrossRef]
12. Žibret, G.; Šajn, R. Hunting for geochemical associations of elements: Factor analysis and self-organizing maps. *Math. Geosci.* **2010**, *42*, 681–703. [CrossRef]
13. Ramdani, A.I.; Chandra, V.; Finkbeiner, T.; Vahrenkamp, V. Multi-scale geophysical characterization of microporous carbonate reservoirs utilizing machine learning techniques: An analog case study from an upper Jubaila formation outcrop, Saudi Arabia. *Mar. Petrol. Geol.* **2023**, *152*, 106234. [CrossRef]
14. Stone, D.; Ritz, K.; Griffiths, B.G.; Orgiazzi, A.; Creamer, R.E. Selection of biological indicators appropriate for European soil monitoring. *Appl. Soil Ecol.* **2016**, *97*, 12–22. [CrossRef]
15. Bangroo, S.A.; Najar, G.R.; Achin, E.; Truong, P.N. Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the North Kashmir forest Himalayas. *Catena* **2020**, *193*, 104632. [CrossRef]
16. Fließbach, A.; Oberholzer, H.R.; Gunst, L.; Mäder, P. Soil organic matter and biological soil quality indicators after 21 years of organic and conventional farming. *Agric. Ecosyst. Environ.* **2007**, *118*, 273–284. [CrossRef]
17. Cao, H.; Chen, R.; Wang, L.; Jiang, L.; Yang, F.; Zheng, S.; Wang, G.; Lin, X. Soil pH, total phosphorus, climate and distance are the major factors influencing microbial activity at a regional spatial scale. *Sci. Rep.* **2016**, *6*, 25815. [CrossRef] [PubMed]
18. Labouyrie, M.; Ballabio, C.; Romero, F.; Panagos, P.; Jones, A.; Tedersoo, L.; van der Heijden, M.G.A.; Orgiazzi, A. Interaction effects of pH and land cover on soil microbial diversity are climate-dependent. *Environ. Microbiol.* **2024**, *26*, e16572. [CrossRef] [PubMed]
19. 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.; et al. Soil quality—A critical review. *Soil Biol. Biochem.* **2018**, *120*, 105–125. [CrossRef]
20. Kaski, S.; Kohonen, T. Exploring data analysis by self-organising map: Structures of welfare and poverty in the world. In *Neural Networks in Financial Engineering, Procedings of the Third International Conference on Neural Networks in the Capital Markets, London, UK, 11–13 October 1995*; Refenes, A.-P.N., Abi-Mostafa, Y., Moody, J., Weigend, A., Eds.; World Scientific Singapore: Singapore, 1996; pp. 498–507.
21. *Standard ISO 11277:2020*; Soil Quality — Determination of Particle Size Distribution in Mineral Soil Material — Method by Sieving and Sedimentation. 3rd Edition. International Organization for Standardization: Geneva, Switzerland, 2020.

22. Sas-Paszt, L.; Sumorok, B.; Derkowska, E.; Trzciński, P.; Lisek, A.; Grzyb, Z.S.; Sitarek, M.; Przybył, M.; Frąc, M. Effect of microbiologically enriched fertilizers on the vegetative growth of strawberry plants under field conditions in the first year of plantation. *J. Res. Appl. Agric. Eng* **2019**, *64*, 29–37.
23. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mah, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [CrossRef] [PubMed]
24. Abarenkov, K.; Zirk, A.; Piirmann, T.; Pöhönen, R.; Ivanov, F.; Nilsson, R.H.; Kõljalg, U. *UNITE QIIME Release for Fungi 2*; Version 04.02.2020; UNITE Community: London, UK, 2020. [CrossRef]
25. Hammer, Ø.; Harper, D.A. Past: Paleontological statistics software package for educaton and data anlysis. *Palaeontol. Electron.* **2001**, *4*, 1.
26. Wehrens, R.; Wehrens, M.R. Package 'kohonen', Version 3.0.11 R Topics Documented; 2015; 25. Available online: https://www.rdocumentation.org/packages/kohonen/versions/3.0.11 (accessed on 9 October 2024).
27. Kalteh, A.M.; Hjorth, P.; Berndtsson, R. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environ. Model. Softw.* **2008**, *23*, 835–845. [CrossRef]
28. Clark, S.; Sisson, S.A.; Sharma, A. Tools for enhancing the application of self-organizing maps in water resources research and engineering. *Adv. Water Res.* **2020**, *143*, 103676. [CrossRef]
29. European Environmental Agency (EEA). Biogeographical Regions in Europe. Available online: https://www.eea.europa.eu/data-and-maps/figures/biogeographical-regions-in-europe-2 (accessed on 20 November 2024).
30. Park, Y.; Céréghino, R.; Compin, A.; Lek, S. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. *Ecol. Modell.* **2003**, *160*, 265–280. [CrossRef]
31. Yu, T.R. *Chemistry of Variable Charge Soils*; Oxford University Press: Oxford, UK, 1997. [CrossRef]
32. Han, L.; Sun, K.; Jin, J.; Xing, B. Some concepts of soil organic carbon characteristics and mineral interaction from a review of literature. *Soil Biol. Biochem.* **2016**, *94*, 107–121. [CrossRef]
33. Malý, S.; Fiala, P.; Reininger, D.; Obdržálková, E. The relationships among microbial parameters and the rate of organic matter mineralization in forest soils, as influenced by forest type. *Pedobiologia* **2014**, *57*, 235–244. [CrossRef]
34. Sierra, C.A.; Trumbore, S.E.; Davidson, E.A.; Vicca, S.; Janssens, I. Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. *J. Adv. Model. Earth Syst.* **2015**, *7*, 335–356. [CrossRef]
35. Ceccon, C.; Panzacchi, P.; Scandellari, F.; Prandi, L.; Ventura, M.; Russo, B.; Millard, P.; Tagliavini, M. Spatial and temporal effects of soil temperature and moisture and the relation to fine root density on root and soil respiration in a mature apple orchard. *Plant Soil* **2011**, *342*, 195–206. [CrossRef]
36. Meyer, M.; Diehl, D.; Schaumann, G.E.; Munoz, K. Analysis of biogeochemical processes in plastic-covered soil during establishment period in strawberry cultivation. *SN Appl. Sci.* **2020**, *2*, 1749. [CrossRef]
37. Barnett, S.E.; Youngblut, N.D.; Buckley, D.H. Soil characteristics and land-use drive bacterial community assembly patterns. *FEMS Microb. Ecol.* **2020**, *96*, fiz194. [CrossRef] [PubMed]
38. Labouyrie, M.; Ballabio, C.; Romero, F.; Panagos, P.; Jones, A.; Schmid, M.W.; Mikryukov, V.; Dulya, O.; Tedersoo, L.; Bahram, L.; et al. Patterns in soil microbial diversity across Europe. *Nat. Comm.* **2023**, *14*, 3311. [CrossRef]
39. Lauber, C.L.; Strickland, M.S.; Bradford, M.A.; Fierer, N. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* **2008**, *40*, 2407–2415. [CrossRef]
40. Tsiafouli, M.A.; Thébault, E.; Sgardelis, S.P.; De Ruiter, P.C.; Van Der Putten, W.H.; Birkhofer, K.; Hemerik, L.; de Vries, F.T.; Bardgett, R.D.; Brady, M.V.; et al. Intensive agriculture reduces soil biodiversity across Europe. *Glob. Change Biol.* **2015**, *21*, 973–985. [CrossRef] [PubMed]
41. Venter, Z.S.; Jacobs, K.; Hawkins, H.J. The impact of crop rotation on soil microbial diversity: A meta-analysis. *Pedobiologia* **2016**, *59*, 215–223. [CrossRef]
42. Lal, R. Challenges and opportunities in soil organic matter research. *Eur. J. Soil Sci* **2009**, *60*, 158–169. [CrossRef]
43. Geiger, F.; Bengtsson, J.; Berendse, F.; Weisser, W.W.; Emmerson, M.; Morales, M.B.; Ceryngier, P.; Liira, J.; Tscharntke, T.; Winqvist, C.; et al. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. *BAAE* **2010**, *11*, 97–105. [CrossRef]
44. Zuber, S.M.; Villamil, M.B. Meta-analysis approach to assess effect of tillage on microbial biomass and enzyme activities. *Soil Biol. Biochem.* **2016**, *97*, 176–187. [CrossRef]
45. Dincă, L.C.; Grenni, P.; Onet, C.; Onet, A. Fertilization and soil microbial community: A review. *Appl. Sci.* **2022**, *12*, 1198. [CrossRef]
46. Adamczyk, J.J. Applications of self-organizing map for patterning macrofungal diversity of xerothermic swards. *Ecol. Res.* **2011**, *26*, 547–554. [CrossRef]
47. Chon, T.S. Self-organizing maps applied to ecological sciences. *Ecol. Inform.* **2011**, *6*, 50–61. [CrossRef]
48. Li, T.; Sun, G.; Yang, C.; Liang, K.; Ma, S.; Huang, L. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. *Sci. Total Environ.* **2018**, *628*, 1446–1459. [CrossRef] [PubMed]

49. Shanmuganathan, S.; Sallis, P.; Buckeridge, J. Self-organising map methods in integrated modelling of environmental and economic systems. *Environ. Model. Softw.* **2006**, *21*, 1247–1256. [CrossRef]
50. Wakelin, S.; Mander, C.; Gerard, E.; Jansa, J.; Erb, A.; Young, S.; Condron, L.; O'Callaghan, M. Response of soil microbial communities to contrasted histories of phosphorus fertilisation in pastures. *Appl. Soil Ecol.* **2012**, *61*, 40–48. [CrossRef]
51. Xia, Q.; Rufty, T.; Shi, W. Soil microbial diversity and composition: Links to soil texture and associated properties. *Soil Biol. Biochem.* **2020**, *149*, 107953. [CrossRef]
52. Constancias, F.; Terrat, S.; Saby, N.P.A.; Horrigue, W.; Villerd, J.; Guillemin, J.P.; Biju-Duval, L.; Nowak, V.; Dequiedt, S.; Ranjard, L.; et al. Mapping and determinism of soil microbial community distribution across an agricultural landscape. *MicrobiologyOpen* **2015**, *4*, 505–517. [CrossRef] [PubMed]
53. Fierer, N.; Jackson, R.B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 626–631. [CrossRef]
54. Tedersoo, L.; Anslan, S.; Bahram, M.; Drenkhan, R.; Pritsch, K.; Buegger, F.; Padari, A.; Hagh-Doust, N.; Mikryukov, V.; Gohar, D.; et al. Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and plant species effects in northern Europe. *Front. Microbiol.* **2020**, *4*, 1953. [CrossRef]
55. Mutalib, S.; Jamian, S.F.; Abdul-Rahman, S.; Mohamed, A. Soil classification: An application of self organising map and k-means. In Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications (IEEE), Cairo, Egypt, 29 November–1 December 2010; pp. 439–444. [CrossRef]
56. Alvarez-Guerra, M.; González-Piñuela, C.; Andrés, A.; Galán, B.; Viguri, J.R. Assessment of self-organizing map artificial neural networks for the classification of sediment quality. *Environ. Int.* **2008**, *34*, 782–790. [CrossRef]
57. Mele, P.M.; Crowley, D.E. Application of self-organizing maps for assessing soil biological quality. *Agric. Ecosyst. Environ.* **2008**, *126*, 139–152. [CrossRef]
58. Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **2009**, *75*, 5111–5120. [CrossRef] [PubMed]