Methodological approach for farm typology construction in terms of soil health – the EU case

Dimitre Nikolov[1], Ivan Boevsky[1], Krasimir Kostenarov[1], Ekatherina Tzvetanova[1], Martin Banov[2], Magardich Huliyan[3], Xiaoping Zhang[4], Laura Zavattaro[5], Josef Krasa[6], Tomas Dostal[4], Gunther Carl Liebhard[7], Peter Strauss[7], Zsofia Bakacsi[8], Csilla Hudek[9], Jose A. Gómez[10] and WP4 co-leader Jundi Liu[11].

[1]New Bulgarian University; [2]Agricultural Academy; [3]National Agricultural Advisory Service; [4]Institute of Soil and Water Conservation (ISWC), Chinese Academy of Sciences; [5]Università degli STUdi di Torino, Italy; [6]Czech Technical University, Czech Republic; [7]BOKU Vienna, Austria; [8]Agrartudomanyi Kutatokozpont, Hungary; [9]University of Lancaster, UK; [10]CSIC, Spain; [11]Northwest A&F University, China.

Summary

Soil health is a significant problem in agriculture which demands a tailor-made approach. The study aims to develop a methodological approach for farm typology construction in terms of soil health. TUdi project, under which was made this study, aims to transform unsustainable management of soils in key cropping systems in Europe and China, developing an integrated platform of alternatives to reverse soil degradation. Thus, the focus is on small, medium, and large EU farms, which produce in the three key cropping systems - grassland, cereal-based rotation, and tree crops. It was applied principal component analysis based on which it was constructed four factors, related to soil health. The results from this analysis was used to feed up the cluster analysis together with other significant variables. The developed farm typology consists of four farm types. From practical point of view was introduced a methodology which allow to determine the type of each farm according the TUdi typology.

Key words: *soil health, farm typology, crops, tree crops, cereal-based rotation, grassland*

Методология за конструиране на типология на ферми, свързана със здравето на почвата в страни в ЕС

Резюме

Здравето на почвата е значим проблем в аграрната икономика, които изисква прилагането на индивидуален подход. Целта на настоящата разработка е да се предложи методология за конструиране на типология на ферми, свързана със здравето на почвата. Анализът е реализиран по проект Tudi, който цели трансформирането на неустойчивото управление на почвите в ключови системи в Европа и Китай, разработвайки интегрирана платформа за алтернативи за обръщане на деградацията на почвата. В тази връзкар фокусът е насочен към малки, средни и големи ферми от страни, членки на ЕС, които отглеждат в трите ключови системи на земеделие - зърнени култури, трайни насаждения и пасища. Използван е анализ на главните компоненти за конструирането на четири фактора, свързани със здравето на почвата. Тези фактори заедно с други ключови променливи са анализирани през клъстерен анализ за формирането на четири типа ферми. Предложен е подход за определяне на мястото на всяка една ферма в тази типология.

Ключови думи: *здраве на почвата, типология на ферми, трайни насаждения, пасища, зърнени култури*

1. INTRODUCTION

Soil health problems are a significant issue in the agriculture sector and in general for the well-being of people. The agriculture sector has heterogeneous nature, and often it is challenging to implement new technologies and strategies. In this connection arises the need application of a tailor-made approach. Unfortunately, the heterogeneous farming systems and the difference in their needs sometimes are neglected. The literature review of the different farm typologies shows that the farms are not a monolithic group (Goswami et al., 2014). They should make decisions facing constraints, resource availability, and environmental issues.

Farm typology study recognizes that farmers are not a monolithic group and face differential constraints in their farming decisions depending on the resources available to them and their lifestyle. Ellis (1993) observes that small farmers are always and everywhere typified by internal variations along many lines. Although every farm and farmer is unique in nature, they can be clustered into roughly homogeneous groups. Developing a typology constitutes an essential step in any realistic evaluation of constraints and opportunities that farmers face and helps forwarding appropriate technological solutions, policy interventions (Ganpat and Bekele 2001, Timothy 1994; Vanclay 2005), and comprehensive environmental assessment (Andersen et al. 2009). The heterogeneity of farming systems is created by a host of biophysical (e.g., climate, soil fertility, slope etc.) and socio-economic (e.g., preferences, prices, production objectives etc.) factors (Ojiem et al. 2006).

The selection of factors that define farm typology varies greatly from study to study and may be governed by the purpose of research. For example, farm typologies were used to study appropriate fertilizer application (Tittonell et al. 2006), resource use efficiency (Tittonell et al. 2007), water use efficiency (Senthilkumar et al. 2009), or overall classification of farm types (Bidogeza et al. 2009). Kuivanen et al. (2016) suggest for identification of criteria defining a farm type to be based on the knowledge of local stakeholders, such as extension workers and/or farmers, or derived from the analysis of data collected using farm household surveys which provide a large set of quantitative and qualitative variables to describe the farm household system.

The farm typology evaluation is needed to understand the reason for adoption or rejection of new technology, strategy, and policy. The farms are affected by biophysical (aptitude, soil Ph, soil fertility, etc.) and socio-economic factors (soil health awareness, access to financing, income uncertainty, supply chain security, etc.). The farm typology development in terms of soil health affects the choice of the factors. The farm typology can support the development of the correct tools and policies for a specific group of farms. This will lower the transactional costs and will ensure that it is applied fast enough the right policy to the right group. In the long term, it will make farmers more resilient in the changing environment.

The aim of the study is to develop a methodological approach for farm typology construction in terms of soil health. The study focuses on small, medium, and large farms, which operate in one of the key cropping systems - tree crops, cereal-based rotation, and grasslands in countries in European Union (EU).

This approach bridges soil health problems with socioeconomic, environmental, and technology assessments. Also, links the farming data to environmental. Farm typology determinations is an essential step in any realistic evaluation of constraints and opportunities that farmers face and helps develop appropriate technological solutions, policy interventions, and comprehensive environmental assessment. It can be used to describe the possibilities and implications at larger regional scales of new strategies for promoting soil restoring and best fertilization technologies in agriculture and its inclusion in agricultural and environmental policies. The farm typology in term of soil health was constructed applying two sequential multivariate techniques: principal component analysis (PCA), and cluster analysis (CA).

2. METHODOLOGY FOR SOIL HEALTH FARM TYPOLOGY CONSTRUCTION

2.1. Area of study

It was applied “judgement sampling” approach divided based on economic size and crop systems according to their country representation. As it was mentioned above, it was analyzed small, medium, and large EU farms in the three cropping systems - tree crops, cereal-based rotation, and grasslands.

Farm size is an important structural characteristic of a farm, and many findings provide insight on the relationship between farm size and the economic, social, and environmental resilience of farms. The in-depth literature review shows that farm size can be measured in several different ways, even once the “farm” itself has been defined as an entity. Therefore, universally accepted, consistently used, and commonly agreed definition of “farm size” does not exist. The choice among different criteria and thresholds depends on the purpose for which farms size need to be identified and must necessarily consider limits and characteristics of available data together with the enormous diversity in terms of farms structures. Considering the above-mentioned complexity and difficulties, the applied farm size definition based on standard output is as follows: (i) small farms – 2,000-24,999 euro; (ii) medium farms – 25 000-99,999 euro; (iii) big farms – ≥ 100,000 euro.

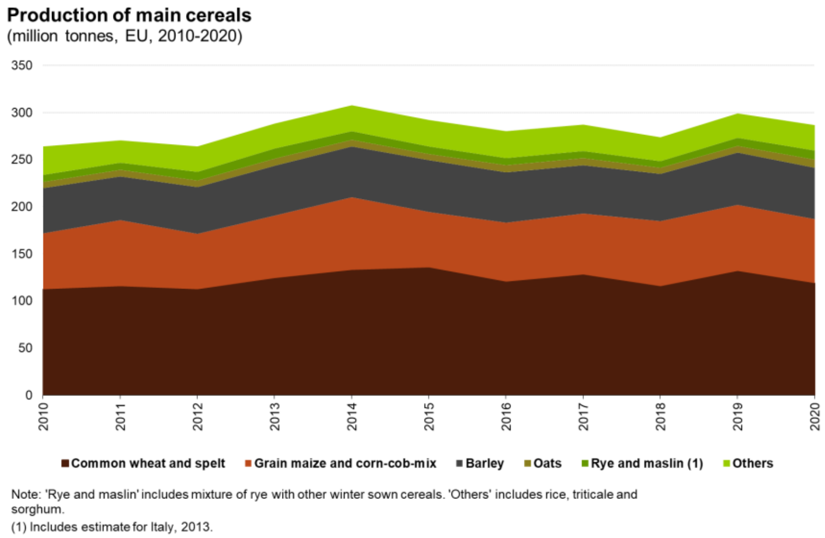
EU agricultural holdings are characterized by rich diversity. EU (Eurostat, 2018c) recognized that in 2016 some agricultural enterprises were specialised in crop production, whether that be where field crop activities are the dominant activity, or where permanent crops (like apples, grapes, and olives) dominate, or indeed horticultural activities. Some farms are specialised in animal production and animal products, whether that be where grazing livestock or granivores (such as pigs and poultry) dominate. Other farms have a mix of crops, mix of livestock, or mix of crops and livestock. About one half (52.5 %) of all farms in 2016 could be categorised as being crop specialist farms; just under one third (31.6 %) of all farms were specialised in field cropping, about one fifth (18.9 %) were specialised in permanent crops, with remainder (1.8 %) being specialist horticultural farms. In this grouping of farms, general field cropping farms that specialised in root crops (such as potatoes and sugar beet), in field vegetables and field crops were the most numerous (accounting for 16.4 % of all EU farms). This was closely followed by specialist cereals, oilseeds, and protein crop farms (15.2 %) of all EU farms. Another one quarter (25.1 %) of the EU's farms were specialist livestock farms, with sheep, goats and other grazing livestock farms (6.2 %) and specialist dairy farms (5.4 %) the most numerous within this group. Mixed farms made up most of the rest (21.1 %), with a small percentage of farms not being classifiable.

Crop production is sensitive to climatic and other natural conditions, and they have significant impact on the quantity and quality of harvests and on crop prices.

- **Cereals**

The harvested production of cereals (including rice) across the EU was 286.5 million tonnes in 2020. This was 12.9 million tonnes less than in 2019, the equivalent of a 4.3 % decline, and 21.4 million tonnes less than the record 307.9 million tonnes recorded in 2014 (Fig. 1). France harvested 57.5 million tonnes of cereals in 2020, one fifth (20.1 %) of the EU’s total harvested production. Germany harvested 43.3 million tonnes (15.1 % of the EU total), Poland a further 35.5 million tonnes of cereals (12.4 % of the EU total) and Spain harvested 26.3 million tonnes (9.2 % of the EU total). (EC, 2021a).

Figure 1. EU production of main cereals 2010 - 2020



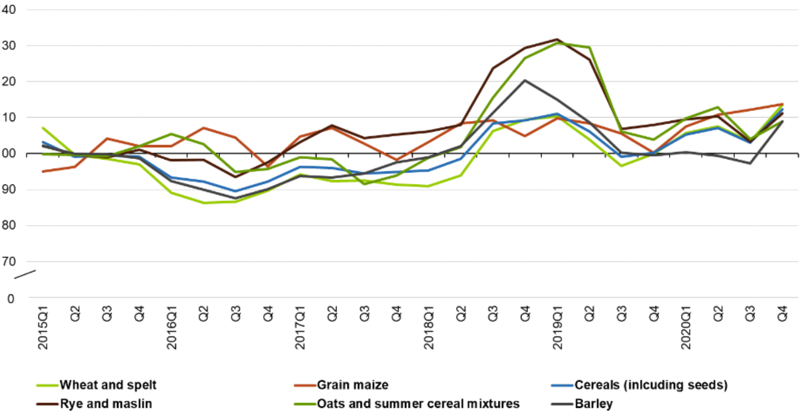
Note: 'Rye and maslin' includes mixture of rye with other winter sown cereals. 'Others' includes rice, triticale and sorghum.
(1) Includes estimate for Italy, 2013.

Source: Eurostat

The overall EU decline in the harvested production of cereals in 2020 was underpinned by steep falls in France (19.2 %, or 13.7 million fewer tonnes) and Romania (-36.3 %, or 11.0 million fewer tonnes). However, there were much higher levels in Poland (up 22.5 %, or 6.5 million tonnes) and Spain (up 32.3 %, or 6.4 million tonnes) (EC, 2021a).

In 2020, the output price of cereals in the EU rose by an average 3.7 % (in nominal terms), in part reflecting the overall lower supply of cereals compared with 2019 (Fig. 2). The provisional average price of wheat and spelt (+5.5 %) and grain maize (+6.3 %) were higher, but there were declines for barley (-3.4 %), oats and summer cereal mixtures (-5.0 %) and rye and maslin (-5.6 %). Over the medium-term, there has been downward pressure on prices as a result of a series of successive and record global harvests. The average price of cereals fell back considerably from the relative highs recorded in 2012 for many Member States. That downward trend began to flatten out in 2016 and for a period between the third quarter of 2018 and the second quarter of 2019 prices rose sharply above the average of 2015 (EC, 2021a).

Figure 2. Cereal price indices development, 2015 - 2020



Development of price indices for cereals
(2015 = 100, EU, 2015-2020)

Source: Eurostat (online data code: apri_pi15_outq)

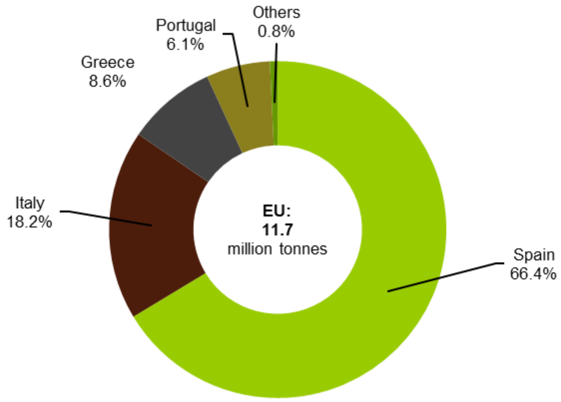
Source: Eurostat

- **Trees**

The EU produces a wide range of fruit, berries, and nuts. An estimated 36.8 million tonnes were harvested in 2020, of which 14.3 million tonnes were pome fruit (apples and pears), 11.4 million tonnes were citrus fruit (such as oranges, satsumas, and lemons), 6.5 million were stone fruit (such as peaches, nectarines, apricots, cherries and plums), 2.7 million tonnes were sub-tropical and tropical fruit (such as figs, kiwis, avocadoes and bananas), 1.3 million tonnes were nuts and 0.7 million tonnes were berries. Spain and Italy are the main EU producers of fruit, but for some specific fruit other Member States were key producers. Commercial apple production to take place in all Member States. Broadly speaking, three in every ten apples produced in the EU (30.0 %) were harvested in Poland in 2020. The other principal apple-producing Member States were Italy (20.8 % of the EU total) and France (13.7 %). By contrast, orange production and peach production are much more restricted by climatic conditions; over 90 % of all oranges and peaches produced in the EU came from Spain, Italy, and Greece (EC, 2021a).

The EU is the largest producer of olive oil in the world, accounting for around two thirds of global production. Most of the world’s production comes from southern Europe, northern Africa and the Near East, as 95 % of the olive trees in the world are cultivated in the Mediterranean region. The total harvested production of olives for olive oil in the EU was 11.7 million tonnes in 2020 (Fig. 3). This was 2.0 million tonnes more than the production level in 2019 but still 1.2 million tonnes less than in 2018. The overall rise in 2020 was due to a higher harvested production in Spain, which accounted for about 66 % of all EU production in 2020. The production of olives for olive oil in Spain was 7.8 million tonnes in 2020, some 2.1 million tonnes more than in 2019. There was little change in the harvested production in Italy (+0.4 % at 2.1 million tonnes) but a moderate rise in Greece (+5.1 % to 1.0 million tonnes, albeit far below the 1.8 million tonnes produced in 2012). By contrast, there was a one fifth reduction (-22.0 %) in the production level of olives in Portugal in 2020 (EC, 2021a).

Figure 3. Production of olives for olive oil, % of EU total, 2020



Source: Eurostat

The EU is big player on the world's wine market; between 2014 and 2018 it accounted for 65 % of global production, 60 % of consumption and 70 % of exports, with 45 % of the wine-growing areas in the world. The total harvested production of grapes for wine in the EU was an estimated 24.1 million tonnes in 2020. This was 1.8 million tonnes more than in 2019, although still down on the 25.7 million tonnes in 2018. Each of the three largest wine grape-producing Member States recorded higher production levels: Italy (+4.4 %), Spain (+20.2 %) and France (+7.2 %) (EC, 2021a).

Considering the period 1990-2018 provided by CLC, it can be highlighted that northern countries and south countries of Europe has increased the proportion of grassland in their lands while the central countries of Europe has reduced it. However, in the last periods between 2009 to 2018 it is shown that most of the East and Northern countries increased their proportion of grassland while southern and western have reduced it. Both databases clearly show that central European countries reduced the proportion of grasslands.

- **Grasslands**

Permanent grasslands are so far from the most important type of grasslands in Europe with a higher representation than temporary grassland. LUCAS shows that there was a clear reduction of permanent grasslands in the western part of Europe and an increase in some eastern and northern countries of the EU, where the percentage of permanent grasslands is low. On the contrary, it has been found a generalized increase of the temporary grassland all over Europe. Grazed areas have been maintained all over Europe for the 2009 to 2018 period. Silvopasture is a practice with a low representativeness in Europe that has been maintained in the last years. Livestock presence is specialized to different European areas with those big animals like horses and bovines more associated to northern and central countries and those small animals mostly living in the South part of Europe.

2.2.Data collection

The needed information for farm typology determination was collected based on a questionnaire for available farm/experiment data survey. The following analysis is using a sample of 416 observations from 6 countries – Austria, Bulgaria, Czech Republic, Hungary, Italy, and Spain. The survey was performed using different methods such as face-to-face (Bulgaria), online (Hungary), and mixed (other countries) due to the COVID-19 outbreak. It was applied “judgement sampling” approach divided based on economic size and crop systems according to their country representation.

The questionnaire has three parts:

- General information - The descriptive information covers some of the socioeconomic characteristics of each farm such as age, gender, and education of the manager. Also, it collects some of the main characteristics of the farm such as agricultural type, agricultural system, structure of the cropping system, total ha managed by the farmer, does the farmer has a livestock, or irrigation.
- Information on farm soil health awareness – It is focused on the level of identification of soil health problems, the soil quality information and its application in the farm management, and the current implement tools which support soil heath management. It evaluated by 12 questions analyzing: if the farmers recognize the soil health as an issue; which soil problems are identified in the farm; do the farmers have enough information concerning soil quality parameters; do they use soil quality information to decide the soil management; do they use any central/national databases of soil analytical parameters and digital maps; do the respondents analyze and record the qualitative soil data; do the farmers use precision farming; from where they take the information and how it is used in farm management; do the farmers apply nutrient management plan.
- Socio-economic and environmental information - These questions are focused on the economic, environmental, and social variables which are needed to determine the typology. The economic variables include economic size and information related to the income and cost to calculate the gross margin of the farm. Even if there is no unified definition among the countries, in the study it was applied the classification described in table 1. The environmental variables investigate the predominant soil system in the farm, humus horizon, pH, biological activity, soil health awareness. The social questions are focused on how the farmers determine their environment in terms of demand, restoration practices development, and policies. The survey includes evaluation of: access to financing for soil restoration practices; the level of specific training and equipment for soil restoration practices; the level of unified terminology regarding soil quality; the level of society’s and consumers’ interest and demand for environmentally

friendly products; the level of farmers' awareness and knowledge level of environmental issues; the level of political will to support delivery of environmental goods and services by farmers; the level of farmers' uncertainty of income; the level of secure supply chain and certainty of demand for farm products; the level of implementation of technology (experience, attitude, access). In addition, some marketing questions are asked related to information sources and willingness to participate in future TUdi materials.

2.3.Principle component analysis

Principle component analysis is a type of factor analysis (Pearson, 1901). It determines the minimum number of variables that are enough to describe a specific problem.

PCA is a linear transformation of data into a new coordinate system based on variables' correlation or covariation. As a result, it is constructed a factor that incorporates several significant variables. The factor notated by F_1 could have the following formula:

$$F_1 : q_1 * y_1 + q_2 * y_2 + q_i * y_i, \qquad (1)$$

where q_j are the coefficient of the linear combination with i = 1 to m (number of variables included in the factor), and y_1, y_2,…,y_m are the included variables.

PCA will group the variables into different factors which explain some specific features of the farm typology. These factors will be used as variables to feed a cluster analysis for determination of each farm type. The farm type will contain different ratios of some of the factors. PCA gives which are features of the clustered groups.

PCA application pass through following steps - initial evaluation and factor determination. The analysis is based on correlation because the variables are in different scales. It will give more information examining the relations between variables. The output of PCA is standardized using the following approach: subtract the variable's mean value from each figure and divide the result by the standard deviation of the transformed vector.

Initial evaluation

The first step is to determine the number of factors that could be built. This analysis is based on the following information – analysis of the descriptive statistics, Kaiser – Meyer – Olkin test for the sample adequacy (KMO coefficient), Bartlett's sphericity test, and correlation analysis. Bartlett's sphericity test tests the hypothesis that there is no significant correlation between at least two variables.

The sample should meet the following criteria to be able to proceed with PCA:

- the KMO coefficient should be above 0.7;
- Bartlett's sphericity test should have a significance level below the accepted level of significance (0.05);

● the correlation between two variables should be between 0.9 and 0.3. The decision for removing variables that don't meet the defined criteria will be based on the conclusions of all data analyses.

The number of factors will be determined based on the percentage of the explained variance. The accepted threshold is above 50% explained variance of the group of variables.

Factor determination

The analysis aims to identify those variables which are correlated and to group them into factors. However, it is needed to build uncorrelated factors between each other because the aim is to differentiate independent groups of farms. Thus, the rotation method that will be applied is the orthogonal - Varimax. The factors are assumed to be independent, and the factor loading is the correlation between the factor and variables (Field 2009).

It was constructed four factors based on the principal component analysis:

- *The Social environment* factor is based on the 5 variables which describe the environment around the farm respondent related to the interests and demand of environmentally friendly products as well as technology development, secure supply chain and political support.
- The second factor is called *Soil health problems*. It includes the following identified and related (correlated) soil health problems in the respondents' farms: soil structure (aggregate stability); land /soil waterlogging; surface compaction; subsurface compaction; soil erosion – sheet erosion; soil erosion – depositional areas.
- The third factor, called *Soil knowledge*, is based on the knowledge sources for soil health analysis and its usage in farm management.
- The fourth factor is *Soil restoration* constructed based on access to financing for soil restoration practices; the level of specific training and equipment for soil restoration practices; the level of unified terminology regarding soil quality.

The factor components are determined based on the rotated component matrix (Table 1). The name of the factors is based on the variables that have the higher contribution.

Table 1. Rotated Component Matrix

Variable	**Description**	**Factors** **1 – Social environment**	 **2-Soil health problems**	 **3 – Soil knowledge**	 **4 – Soil restoration**

Variable	Description	Factors			
		1 – Social environment	2-Soil health problems	3 – Soil knowledge	4 – Soil restoration
II11_1	Soil structure (aggregate stability)	-.097	**.617**	-.161	-.053
II11_3	Land /soil waterlogging	.217	**.518**	.006	-.378
II11_5	Surface compaction	.100	**.698**	-.044	.063
II11_6	Subsurface compaction	.013	**.696**	-.045	.002
II11_7	Soil erosion – sheet erosion	-.044	**.729**	-.076	.082
II11_9	Soil erosion – depositional areas	-.079	**.701**	-.147	.037
II20_3	From literature	.095	-.056	**.733**	.020
II20_4	From leaflets	.164	-.111	**.790**	.038
II20_5	From YouTube and other social networks	.052	-.144	**.762**	-.006
II20_7	From ministries	-.050	-.115	**.664**	.201
III30_1	Access to financing for soil restoration practices	.406	.088	.113	**.674**
III30_2	The level of specific training and equipment for soil restoration practices	.409	.059	.158	**.709**
III30_3	The level of unified terminology regarding soil quality	.276	-.029	.046	**.749**
III30_4	The level of society's and consumers' interest and demand for environmentally friendly products	**.723**	-.081	.082	.055
III30_5	The level of farmers' awareness and knowledge level of environmental issues	**.592**	.009	-.010	.266
III30_6	The level of political will to support delivery of environmental goods and services by farmers	**.656**	-.030	.085	.223
III30_8	The level of secure supply chain and certainty of demand for farm products	**.717**	-.038	.031	.103
III30_9	The level of implementation of technology (experience, attitude, access)	**.653**	.119	.085	.154

Sources: Authors' calculations

2.4.Cluster analysis

Cluster analysis is a group of statistical procedures which aims to discover a structure within a complex set of data. The different elements (the different farms/respondents) are combined into clusters. The farms within the cluster have some degree of similarity among themselves (homogeneity) according to the different variables. Different clusters are relatively distinct from each other (heterogeneity). The variables are the economic and social indicators for every farm collected by a survey. Depending on the studied problem, the cluster analysis can build a classification which may become a basis of classification of new observations (Anderberg, 1973).

- *Data requirements*

The initial set of variables have been delivered by a multistage process of literature study, brainstorming trough the experts and stakeholders. The cluster analysis (CA) is fed up with the factors constructed using PCA and other variables collected via questionnaire such as farm size and the agricultural system. As both methods (PCA and CA) are in fact a form of data reduction methods they can be used together. PCA lowers the number of variables/factors then the cluster analysis determines the clusters based on few factors with low correlation (Ding and He, 2004).

- *Number of clusters*

The choice of the clusters' number is one of the most important decisions, conducting CA. There are different ways to determine the number of clusters. It is applied a two-step approach using hierarchical and K-means methods. Agglomerative hierarchical clustering methods solve a problem iteratively - the method starts from one cluster, which includes all respondents/farms, and ends at the point in which each respondent/farm is a cluster.

On the first step, the Euclidean distance and Ward method (which analyzes the variance of clusters) are applied to derive the optimal number of clusters within the iterations in hierarchical clustering.

On the next step, it was applied K-means cluster analysis to find best solution for the number of clusters that is derived from hierarchical clustering. One of its main characteristics is that the method can give different solutions using different number of clusters. Thus, it was combines the both methods – hierarchical and K-means. The number of clusters was derived using hierarchical cluster analysis while K-means cluster analysis determines the clusters based on socio-economic and soil quality characteristics of the farms.

- *The algorithm for K-means cluster analysis*

K-means cluster analysis tries to find the optimal clusters of farms by minimizing the sum of the squared error (of the factors derived from PCA) over all K clusters. Because the squired error lowers by increasing the number of clusters it can be minimized only for a given, predetermined number of clusters (Jain, 2010). Application of the K-means algorithm follows the steps:

1. Selection of the initial partition with K clusters.
2. Generation of a new partition by assigning each pattern to its closest cluster center.
3. Computation of new cluster centers.

The procedure has to be repeated until the final clusters are found.

The K-means cluster analysis requires three decisions to be made: number of clusters (which we will determine by the hierarchical clustering), cluster initialization (which SPSS chooses randomly), and distance metric – we are using Euclidean metric. As a result, K-means finds spherical clusters in data which suits our expectation about the data that is generated.

- *Interpretation of results*

Final clusters give the opportunity to interpret what are the typical characteristics for a particular cluster. The characteristics are all data which is gathered by the survey, including economic size, income, expenditures, efficiency, soil quality, crop type and social variables. The interpretation and description of the clusters is done by summary descriptive statistics like the mean and variance (King, 2015). There might be applied standard tests of significance on the differences between group means like ANOVA. As a result, it will be determined what is typical socio-economic characteristic for each cluster related with the soil health. That information can be used in subsequent analysis.

The cluster analysis was made with six variables – the factors determined by the PCA, size of the farm, and farm cropping system. As a result, it was constructed four clusters:

Cluster 1 name is *Intensive Large Farms* – average amount of the farms in cluster 1 are large, and average amount of crop types are cereal-based rotation. They use the land intensively, which leads to medium soil health problems. They estimate positively the soil restoration practices but the degree is low. These farmers feel medium negative social environment about their problems. They feel lack of knowledge (low negative).

Cluster 2 name is *Grassland Small Farms.* They have soils with high health quality. The level of soil restoration is low negative because they do not need strong soil restoration practices that leads to low negative knowledge about soil restoration problems. Social environment is medium negative for them.

Cluster 3 name is *Cereal Diversified Farms*. It consists of farms from all sizes, mainly (in high degree) cereal-based rotation. They have high positive social environment. They estimate soil health problems as low positive. Their knowledge and soil restoration practices are at low positive level**.**

Cluster 4 name is *Tree Small Farms*. Their soil health problems are medium. They have average access to knowledge, but do not apply soil restoration practices (medium negative). Their social environment is neither positive nor negative.

The four clusters can be used to determine the type of a particular farm or to classify each farm into which cluster it falls. To achieve this, the farm must be evaluated on each of the factors involved in the construction of the clusters. First, the farm is classified by size and crop type. After that, an assessment is made on soil problems, soil restoration, social environment, and soil knowledge. The evaluation is done on a scale from 1 to 6, as an explanation of the individual evaluations are given in table 2.

Table 3 gives the values of the indicators that the farms must have in order to be classified in the corresponding cluster.

For example, a large farm that grows cereal-based rotation crops, has soil problems that the farmer estimates as 3, soil restoration estimation - 4, social environment - 2 and soil knowledge with 3 is classified in cluster 1 as a large intensive farm.

Table 2. The four clusters and their main characteristics

Cluster 1: Intensive Large Farms	Cluster 2: Grassland Small Farms	Cluster 3: Cereal Diversified Farms	Cluster 4: Tree Small Farms
Large farms	Small farms	Medium size farms	Small farms
Cereal-based rotation	Grassland systems	Cereal-based rotation	Tree crops system
Soil problems – average (1, 2, 3)*	No soil problems (4, 5, 6)*	Very little or no soil problems (3, 4 ,5, 6)*	Soil problems – average (1, 2, 3, 4)*
Soil restoration – There are SR (3, 4, 5, 6)*	Soil restoration – low level (1, 2, 3, 4)*	Soil restoration - There are SR (4, 5, 6)*	Soil restoration – insignificant (1, 2, 3, 4)*
Social environment – average negative (1, 2, 3, 4)*	Social environment – average negative (1, 2, 3, 4)*	Social environment – strongly positive (3, 4, 5, 6)*	Social environment – neutral (2, 3, 4)*
Soil knowledge - low negative (1, 2, 3, 4)*	Soil knowledge – low negative (1, 2, 3 or 4)*	Soil knowledge – low positive (3, 4, 5, 6)*	Soil knowledge – large positive (3, 4, 5, 6)*

*The number in parenthesis corresponds to the estimation scale in table 2

Table 3. Scale for estimation. F3 to F5 corresponds to the factors used for K-means cluster analysis

Scale Value	Definition of scale for F3 Estimate soil problems of the pilot farm	Definition of scale for F4. Estimate soil restoration in the farm	Definition of scale for F5 Estimate the social environment	Definition of scale for F5 Estimate the soil knowledge
1	The farm is strongly exposed to soil health problems	Missing or very insignificant restoration practices	Very negative social environment	Missing or very insignificant soil knowledge
2	The farm is exposed on an average degree to soil health problems	Insignificant restoration practices	Negative social environment	Insignificant soil knowledge
3	The farm has low level/small soil health problems	low level (but existing) of soil restoration practices	Slightly negative social environment	low level of soil knowledge
4	The soil health is positive but still some problems can arise with soil health	There are soil restoration practices	Slightly positive social environment	There is some soil knowledge
5	The soil health is positive and very small soil health problems can arise	There are restoration practices, but more can be done for restoration of the	Positive social environment	Solid soil knowledge but more can be done

Scale Value	Definition of scale for F3 Estimate soil problems of the pilot farm	Definition of scale for F4. Estimate soil restoration in the farm	Definition of scale for F5 Estimate the social environment	Definition of scale for F5 Estimate the soil knowledge
		soils		
6	The soil is in perfect health	There is very high level of soil restoration practices	Very positive social environment	Significant soil knowledge

3. CONCLUSION

Developing a typology constitutes an essential step in any realistic evaluation of constraints and opportunities that farmers face and helps forwarding appropriate technological solutions, policy interventions and comprehensive environmental assessment. The aim of the study is to develop a methodological approach for farm typology construction in terms of soil health. The study focuses on small, medium, and large farms in the key cropping systems - tree crops, cereal-based rotation, and grasslands in countries in European Union.

Based on principle component analysis are determined the six factors, which define the farmers in terms of soil health: Economic size, Cropping systems, Soil health problems, Social environment, Soil knowledge, Soil restoration. The K-means cluster analysis is used to make the typology of the farms. The four types of the farms are:

Intensive Large Farms cluster which includes large farms, with average soil problems, soil restoration, average negative social environment, and low negative soil knowledge.

Grassland Small Farms cluster which consists small farms with no soil problems, low soil restoration level, average negative social environmental, and low negative soil knowledge.

Cereal Diversified Farms cluster which incorporates cereal-based medium size farms with very little or no soil problems, soil restoration, strongly positive social environmental, and low positive soil knowledge.

Tree Small Farms cluster contains tree small farms with average soil problems, insignificant soil restoration, neutral social environment, and large positive soil knowledge.

It has been introduced a methodology how to determine the type of the farm according the TUdi typology. For that purpose, the farm has to be estimated for all six factors, using a six-degree scale.

Developing a typology of the farms in terms of soil health is important matter because it can help to be generated information for the overall condition of soils and farms. The derivation of a plausible typology can assist the individual farmer in understanding the depth of the problem, deeper understanding the condition of the soils on his own farm compared to other farms. Finally, it can help in making a technological decision regarding soil health. On the other hand, the classification is important for academics to deepen the study of problem soils, as well as in the search for solutions for their rehabilitation. Finally, the public authorities can use the typology to make strategic decisions, derive soil policies, and build workable solutions at the state or local level. From this point of view, the present study is a basis for further research to expand and deepen the knowledge of soil health and to propose workable solutions in this direction.

Acknowledgements

The study was developed with the financial help of the European Union’s Horizon 2020 Research and Innovation action under the project “Transforming Unsustainable management of soils in key agricultural systems in EU and China. Developing an integrated platform of alternatives to reverse soil degradation” - TUdi with grant agreement No 101000224.

LITERATURE

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2, 559–572.

Field, A. (2009) Discovering Statistics Using Spss For Windows, 3td ed. Sage, London

Anderberg, M. (1973) Cluster Analysis for Applications, Elsevier Inc., ISBN: 978-0-12-057650-0.

Ding, Ch. and He, X. (2004). Cluster Structure of K-means Clustering via Principal Component Analysis. Proc 8th Pacific-Asia conf on advances in knowledge discov data mining (PAKDD. 414-418. 10.1007/978-3-540-24775-3_50.

Jain, A. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/J.PATREC.2009.09.011.

King, R. (2015) Cluster Analysis and Data Mining : An Introduction. Dulles, Virginia: Mercury Learning & Information.

Bidogeza JC, Berentsen PBM, de Graaff J, Lansink AGJMO (2009) A typology of farm households for the Umutara Province in Rwanda. Food Sec 1:321–335.

Ellis F, (1993). Peasant Economics: Farm Households and Agrarian Development, 2nd edn. Cambridge University Press, Cambridge.

Eurostat, (2018c). Farms and Farmland in the European Union – Statistics. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Farms_and_farmland_in_the_European_Union_-_statistics. The_evolution_of_farms_and_farmland_from_2005_to_2016.

EC, (2021a). Agricultural production – crops, Eurostat, November 2021.

Ganpat W, Bekele I (2001) Looking For The Trees In The Forest: Farm Typology As A Useful Tool In Defining Targets For Extension. In: Lindner JR (ed) Proceedings of the 17th Annual Conference of the Association for International Agricultural and Extension Education. Baton Rouge, Louisiana.

Goswami, R., Chatterjee, S. & Prasad, B. (2014). Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: Study from coastal West Bengal, India. Agric. Food Econ. 2, 1–24.

Kuivanen KS, Michalscheck M, Descheemaeker K, Adjei-Nsiah S, Mellon-Bedi S, Groot JCJ, et al. (2016). A comparison of statistical and participatory clustering of smallholder farming systems±A case study in Northern Ghana. Journal of Rural Studies.; 45: p. 184-198.

Ojiem J, Ridder N, Vanlauwe B, Giller KE (2006) Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems. Int J Agric Sust 4:79–93Soule MJ (2001) Soil management and the farm typology: do small family farms manage soil and nutrient resources differently than large family farms? Agric Resour Econ Rev 30:179–188.

Senthilkumar K, Bindraban PS, de Boer W, de Ridder N, Thiyagarajan TM, Giller KE (2009) Characterising rice-based farming systems to identify opportunities for adopting water efficient cultivation methods in Tamil Nadu, India. Agric Water Manage 96:1851–1860.

Tittonell P, Leffelaar PA, Vanlauwe B, van Wijk MT, Giller KE (2006) Exploring diversity of crop and soil management within smallholder African farms: a dynamic model for simulation of N balances and use efficiencies at field scale. Agric Syst 91:71–101.

Tittonell P, Vanlauwe B, de Ridder N, Giller KE (2007) Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: soil fertility gradients or management intensity gradients. Agric Syst 94: 376–390.

Timothy WO (1994) Identifying target groups for livestock improvement research: the classification of sedentary livestock producers in western Niger. Agric Syst 46:227–237.

Vanclay JK (2005) Using a typology of tree-growers to guide forestry extension. Ann Trop Res 27:97–103.