Self-Organizing Maps Applied to Soil Conservation in Mediterranean Olive Groves
Jamal Ammouri, Pascale Minet, Malika Boudiaf, Samia Bouzefrane, Meziane Yacoub

To cite this version:

HAL Id: hal-02432848
https://hal.archives-ouvertes.fr/hal-02432848
Submitted on 8 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Self-Organizing Maps Applied to Soil Conservation in Mediterranean Olive Groves

Jamal Ammouri¹, Pascale Minet², Malika Boudiaf³, Samia Bouzefrane¹, Meziane Yacoub¹
¹ CNAM, 292 rue Saint Martin, Paris Cedex 3, France.
² Inria, 2 rue Simone Iff, 75589 Paris Cedex 12, France.
³ Laboratoire Ressources Naturelles UMMTO, Algeria.
Email: jamal.ammouri.auditeur@lecnam.net, pascale.minet@inria.fr, boudiafmalika@yahoo.fr, samia.bouzefrane@cnam.fr, meziane.yacoub@cnam.fr

Abstract

Soil degradation and hot climate explain the poor yield of olive groves in North Algeria. Edaphic and climatic data were collected from olive groves and analyzed by Self-Organizing Maps (SOMs). SOM is a non-supervised neural network that projects high-dimensional data onto a low-dimension topological map, while preserving the neighborhood. In this paper, we show how SOMs enable farmers to determine clusters of olive groves, to characterize them, to study their evolution and to decide what to do to improve the nutritional quality of oil. SOM can be integrated in the Intelligent Farming System to boost conservation agriculture.

Index Terms

Soil conservation, olive groves, conservation agriculture, sustainable farming, self-Organizing Map (SOM), machine learning, intelligent farming system, IoT

I. CONTEXT AND MOTIVATIONS

A. Soil conservation

Context: In Africa, the phenomenon of soil degradation exists because of water deficit, unsuitable farming practices and overexploitation of soil, which does not go together with the edaphic and climatic evolution of the environment. Production yield remains very low with regard to the areas planted. Soil conservation aims at preserving the physical, chemical and...
biological quality of soils. Conservation agriculture boosts production by optimizing the use of agricultural resources and reducing soil degradation through integrated management of soil, water and available biological resources, combined with external inputs. One approach is based on the concept of ecological intensification, which uses soil and ecosystem organisms to increase the availability of nutrients [5] for the plant. One lever of ecological intensification is to use the physical, chemical and biological processes operating at the soil-root interface [6] (i.e. the rhizosphere). The symbiotic associations of plants with rhizosphere microorganisms (mycorrhizal fungi, endophytic and bacteria), very abundant in soils and useful for agriculture, allow a considerable reduction of fertilizers and pesticides [7], while strengthening soil stability. Nutrients will be brought to plants in divided doses according to the plant lifecycle.

Soil biology and conservation agriculture: Earthworms [8], the main players in soil fertility, aerate soil, break down plant residues and organic matter, and transform it into elements easily assimilable by plants. Similarly, soil rich in organic matter captures much more atmospheric CO2, reducing greenhouse gases. However, this recycling is disturbed by the plowing which stimulates the appearance of bacteria carrying diseases for the plants, thus reducing the soil biodiversity. These policies have led to a deterioration of the agro-ecosystems for which attempts are being made to restore their functionality.

B. Irrigation, fertilization and chemigation

Context: The major agricultural issue concerns water resources, which are limited especially in Africa and must be preserved. Despite a vast irrigation program, much remains to be done in terms of controlled irrigation to avoid the waste of water, but also for a suitable use of fertilizers through the chemigation process with NPK fertilizers dissolved in water in a safe and precise dosage.

Irrigation: In addition to the rudimentary solutions used (manual treatment), inadequate or poorly designed irrigation systems can be the source of many problems. Indeed, drip irrigation has shown its effectiveness in terms of water saving compared to the conventional method by sprinkling. Water salinity and soil moisture require rigorous control in the quality of irrigated water.

Chemigation: Due to chemical changes in the soil and other phenomena such as leaching and runoff, most fertilizers and pesticides used in agriculture are largely wasted [5], [7]. This leads to a loss of efficiency and possible risks of contamination of groundwater, aquatic environments and soil.

C. IoT, ML and Cloud in smart farming

In smart farming, data are provided by many heterogeneous sources, including meteorological and geological maps, a large variety of sensors and mobile devices (e.g. smart tractor, drone, mobile phone). Static sensors are usually scattered in the crop fields to control the soil moisture and the soil mineral composition, whereas sensors embedded in mobile agricultural machinery can also detect a hydric stress or an attack of pests or diseases. Drones are used to monitor hardly accessible area and are integrated in the IFS platform. Once the information is collected locally to the farm, it is usually locally aggregated/pre-processed at low-cost edge nodes, before being transmitted to the cloud infrastructure. Context-aware data gathering [9] allows the mobile collector to adapt its strategy according to the context of the monitored field.

Process-based crop models are algorithms that dynamically represent (usually at a daily time step) the development of a crop, from sowing to harvest, depending on its environment. Model input variables are climatic data (temperature, rainfall, global solar radiation, wind speed, etc.), soil data (depth accessible to the roots, granulometry, etc.), and the crop (species and variety sown, date of sowing, stocking density, etc.). When the plant model is coupled to a soil model with a water and nitrogen balance, it is possible to know if the quantities required for optimum photosynthesis are available for the plant, and, if not, the quantity that should be provided by the farmer. The decision being taken by the farmers, commands are sent to the actuators to trigger the irrigation, fertigation or chemigation.

The goal is to use Machine-Learning (ML) techniques to design models which are representative of the real world and hence more accurate. These models will be used to improve the accuracy of the decisions taken with regard to smart farming. ML techniques will be applied to help in the following actions: planning how soil and water are used, operation monitoring [10], quality control and product tracking [11], sensing of multiple crop/varieties/seed data at a single time, monitoring and management of IFS.

III. THE SITES STUDIED

This study took place along a North-South axis of 60km in North Algeria as depicted in Fig. 1, where 10 olive groves are considered. All of them have an average of 100 olive trees per hectare of the most widespread olive variety in Kabylia: Chemlal. Traditionally, olive groves are found in nutrient poor marginal lands with no irrigation, no fertilization and on high slopes. Eight olive groves belong to the wilaya of Tizi-Ouzou, with a sub-humid climate, whereas the two other olive groves are in the wilaya of Bouira with a semi-arid climate.

Each olive grove is evaluated according to three aspects:

- geographical with its GPS coordinates, its slope and exposure.
• edaphic: evaluated by 9 edaphic variables:
  – pH indicates whether the soil is neutral \((pH = 7)\), acidic \((pH < 7)\) or basic.
  – CaCO₃ denotes the soil rate of calcium carbonates.
  – OC: the rate of organic carbon is used to compute the organic matter \(MO = 1.72 \times OC\).
  – CEC stands for Cation Exchange Capacity.
  – Nt is the total nitrogen in the soil.
  – C gives the percentage of clay in the soil.
  – \(P_{\text{total}}\) and \(P_{\text{assi}}\) are the amount of phosphorus in the soil and its assimilable part by the plant, respectively.
  – K gives the potassium present in the soil.
• climatic: the monthly rainfall, the monthly minimum and maximum temperatures, the wind direction and speed.
Since the soil variables and climatic variables measured evolve over time, they are measured over two successive years. Each olive grove is abbreviated by its initials followed by 1 or 2 according to the year considered. The olive groves studied are Freha (F), Tirmitine (T), Yakoureene (Y), Tizi Rached (TR), Boghni (B), Guendoul (G), Tizi Rached Yefsa (TRY), Nezla (N), Bouira Sud (BS) and Bouira Sidi Ziane (BSZ). The methodology used for soil characteristics evaluation is the following. Soil characteristics were determined on the composite samples according to standard methods proposed in Jackson (1979). Particle size distribution was measured according to the Robinson pipette method (anic matter oxidation by \(H_2O_2\), shaking in a sodium hexametaphosphate solution). Soil pH was measured in a 1/5 soil distilled water suspension. CaCO₃ was determined using the \(HCl 1M\) volumetric method. Cation exchange capacity was measured according to the ammonium acetate method. Organic C was determined by sulfochromic oxidation. All these variables are measured and gathered into a dataset that is exploited in the following of this paper.

IV. Analysis of Soil Characteristics

In this section, we analyze the soil features of olive groves, by means of the nine soil variables measured.

A. Study of correlations

The circle of correlations is depicted in Fig 2. The first axis holds 35.6% of data variability, whereas the second axis holds 29.9% of data variability, leading to a total of 65.5% of data variability on these two axes.

The variables OC, Nt and CEC evolve similarly. They all come from the decomposition of organic matter and indicate the state of the biological balance of soils. In addition, the ratio OC/Nt, called mineralization rate, is often used to evaluate the biological activity of soil. OC, Nt and \(P_{\text{assi}}\) are three important elements for the metabolism of any plant.

Potassium and Clay belong to close axes. Potassium comes from the alteration of clay in soils which have not been fertilized, which is the case of these olive groves.

CaCO₃ and pH evolve similarly. pH evaluates the availability of nutrients. When a soil is calcareous (i.e. CaCO₃), the pH becomes alkaline and the availability of nutrients decreases, phosphorus can be immobilized, as depicted by the antagonist action of \(P_{\text{assi}}\) and CaCO₃ in the correlation circle.

\(P_{\text{assi}}\) is the part of \(P_{\text{total}}\) that can be assimilated by the plant. \(P_{\text{assi}}\) is impacted by the acidity of the soil reflected by its pH and CaCO₃ that blocks the phosphorus.

Once the correlations established, we can reduce the cost of further measurements by evaluating only one variable among all its correlated variables. For instance, we can evaluate only the clay rate instead of evaluating the clay rate and the potassium rate. Similarly, we can evaluate only OC instead of OC and CEC.
B. Use of Self-Organizing Map in soil analysis

The Self-Organizing Map (SOM) [12], [13] is a very popular algorithm, introduced by Teuvo Kohonen in the early 80s as a convenient clustering and visualization tool. It is an unsupervised neural network that projects high-dimensional data onto a low dimension discrete space, usually two-dimensional, called the topological map. The projection preserves the local topology of the initial space so that the data which are close to each other will be mapped to close locations on the map. This corresponds to the fact that similar items excite adjacent neurons in the brain.

SOM is a vector quantization method. It consists of neurons (prototype vectors) organized in a regular low-dimensional grid. Each neuron is a \(d\)-dimensional codebook vector (prototype vector) where \(d\) is equal to the dimension of the input vectors. The neurons are connected to adjacent neurons by a neighborhood relation, which dictates the topology of the map. It is an efficient tool for visualization of multidimensional numerical data. The main advantage of using a SOM is that the results are easy to understand and interpret. The reduction of dimensionality and grid clustering makes it easy to observe similarities in the data.

There are a variety of methods to visualize SOMs. The U-matrix visualizes distances between neighboring map units, and thus shows the cluster structure of the map: high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves. Thus a group of dense map units (neurons) denotes a cluster. Distance zones of neurons must separate several clusters. Any two different clusters are separated by distance zones of neurons.

The data set of dimension 9 (i.e. 9 soil variables studied in our data set) is loaded into Matlab and normalized. There are three built-in map sizes of SOMs: big, medium and small. In our case, we used a big map which allows a finer analysis of the dynamicity of the olive groves. The map size is determined using a heuristic formula, defined in the SOM Toolbox, which depends on the number of samples to be trained and the ratio between the two biggest eigenvalues of the training data. We get a two-dimensional map of 11x9 hexagonal cells, leading to 99 neurons in total, where each neuron is the cell representative. Both the quantization error and the topological error are null: each olive grove is a neuron. Among the 99 cells, 20 are occupied by real olive groves. Figure 3 depicts the U-matrix of our data-set. We observe that the map is organized according to the geographical location of olive groves. Usually, two neurons representing the same olive grove, but for two successive years, belong to close cells. For instance, neurons 1 and 14 representing the olive groves N1 and N2 constitute a cluster which is separated from other groves by a yellow border. However, we observe some singularities. For instance, F1 and F2 are far from each other and separated by a yellow border, similarly Y1 and Y2 are also far from each other.

Component planes are very convenient when one has to visualize a lot of information at once. Each component plane shows the values of one variable in each map unit. That is, each component plane of the SOM consists of the values of the same component in each prototype vector. The component planes are typically visualized by giving each map neuron a color according to the relative value of the respective component in that neuron. By plotting all component planes and comparing them with each other, correlation between variables can be seen.

Figure 4 depicts the 9 component planes corresponding to our edaphic variables. The values of components are denormalized so that the values shown on the colorbar are in the original value range. For instance, for CaCO3, we observe that the blue cells like N1, N2, F1, Y1, Y2, B1 and B2 correspond to olive groves where CaCO3 is less than 5. All the other cells are critical...
in green, some are very critical like BS1 and BS2 in yellow. We can also explain the trajectory of Y1Y2, corresponding to Yakourene, where fires are frequent and make vary the physical, chemical and biological characteristics of soils. This explains the increase of 25\% in CEC (the capacity of cationic exchange), which depends on the rate of clay and organic matter. For the trajectory of F1F2 corresponding to Freha where the phosphorus content in the soil is very poor, we observe a decrease of 72\% in P\_assi, and a decrease of 55\% in P\_total. This is due to the absorption of phosphorus in the rhizospheric environment. Freha has a clay soil, we notice an increase of 33.45\% in clay, due to the root activity of the olive tree. We observe a very important decrease of P\_assi and a light increase of CEC and very light increase of OC.

Figure 5 illustrates the Clay rate that is the discriminant variable for the cluster formed by N1 and N2.

V. ANALYSIS OF CLIMATE CHARACTERISTICS

The North Algerian climate is a Mediterranean climate, characterized by a sunny dry season and a wet and cool season with heavy rainfall having large annual variations. The climatic and soil conditions favor the physical, chemical and biological
deterioration of soils. That is why they have to be taken into account into cultivation practices of a sustainable agriculture.

Minimum and maximum monthly temperatures as well as monthly rainfall have been collected from the Algerian National Office of Meteorology over the 2000-2015 years. The average values have been computed month per month for Tizi-Ouzou and Bouira, respectively. In the Tizi-Ouzou region, rainfalls are very strong from November to January with a peak of 158.2 mm in December and a minimum of 1.8 mm in July. The annual rainfall is 826.2 mm. For the Bouira region, the maximum value of 98.4 mm is obtained in December, whereas the minimum value is get in July with 4.4 mm. The annual rainfall is only 560.5 mm.

With regard to monthly temperatures, Tizi-Ouzou records the maximum value of 39.1°C in July and August and a minimum value of 7.1°C in January. For Bouira, the maximum value is reached in July with 37.2°C and a minimum value in February with 4.2°C.

The evapotranspiration is computed according to the Thornthwaite model, taking into account the average monthly temperature and the annual and monthly thermal indices. The annual evapotranspiration is evaluated to 1072.5 mm in Tizi-Ouzou and 998 mm in Bouira, respectively. The maximum monthly evapotranspiration is reached in July and August, two months during which the rainfall are very weak. This causes a hydric stress which coincides with the fruiting and maturation phases of the olives, hence explaining the small yields obtained.

VI. TOWARDS SMART FARMING

In this section, we show how to use the results obtained with SOMs to improve the yield and quality of crops and how to integrate these machine learning techniques into smart farming, which is the next step.

A. How to improve the yield and the quality of crops

Different stresses have been identified:

- the hydric stress during July and August. Irrigation will meet the water requirement of olive trees.
- the fertilizer stress (e.g. NPK stress). By bringing fertilizers during olive growing and maturation, the fruits will be better.
- the disease and pest stress. Many attacks have already been signaled, it is important to react quickly and detect an attack at its beginning, before the whole olive grove is infested.

In addition to these stresses, the aging of olive trees, which can be mitigated by tree pruning, is an additional reason explaining why the yield of olives is small and the quality too. Since the Chemlal olives are used to produce oil, the quality is evaluated by the nutritional value and biochemical properties of the oil produced. To mitigate these stresses while progressing towards a smart and sustainable farming, agroecological techniques should be adopted. For instance, on the one hand associated crops (cereals and legumes, especially fodder crops) enable the facilitation and a better management of hydric stress. On the other hand, covered soil reduces evapotranspiration. To preserve the soil and the environment, irrigation, fertilization and chemigation should be provided only when needed, only where needed and only what is needed but no more.

To improve the yield and the quality of the olive oil produced, a smart farming application running on a smartphone or any mobile device, will be of considerable help to olive growers. It will be able to:

- Display a dashboard of various indicators (e.g. soil moisture, humidity, NPK content, chlorophyl level) computed from data collected in the field by means of various sensors. Static sensors are deployed in the crops, whereas mobile sensors are
embedded in agricultural machinery and drones. Multispectral sensors of agricultural drones are able to build a cartography of soils (e.g., geological map) and crops (evaluating the hydric stress or the maturity level). Data can also be generated by other sources like Geographical Information Systems (GIS).

- Alert the farmer if some values exceed the thresholds (e.g., a hydric stress, a NPK stress, a pest/disease stress);
- Help the farmer to take a decision (e.g., irrigation, fertilization, chemigation);
- Generate the commands to actuators opening or closing the valves of irrigation/fertilization/chemigation pipes according to the decision taken.

B. An architecture example

The smart farming application requires the availability of data evaluating the needs of olive trees with regard to what is already provided by the soil or the climate. For this purpose, an architecture based on the interconnection of wireless sensors and actuators to an edge node serving as wireless gateway to the cloud needs to be deployed. The edge node is usually located in the farm. An example of a wireless meshed sensor network based on the Time Slotted Channel Hopping technology (TSCH) deployed in a peach-orchard is described in [14]. Other technologies are also possible like ZigBee, also based on the IEEE 802.15.4 standard, or LoRaWAN for long range communications. TSCH and ZigBee provide the advantages of an upstream and downstream communication with a more reliable communication, whereas LoRaWAN consumes less energy. The communication from the gateway edge node to the private/public Cloud can be done either via the cellular network (e.g., GPRS, 3G, 4G, and 5G) if the farm coverage is good [15], via WiFi otherwise. The Cloud is connected to the IP backbone.

![Fig. 6: An example of architecture for smart farming.](image)

After a pre-processing by the edge to clean and remove outliers from data gathered from sensors, these data are stored in a warehouse, database, or a GIS, where they can be analyzed and retrieved using machine-learning techniques such as SOM to help the farmer to take accurate decisions and smartly manage his olive groves. To avoid expensive battery replacement of sensor and actuator nodes deployed, solar panels are used to provide energy. To detect pest or disease attacks, a drone (UAV) or a mobile robot embedded in an agricultural machinery is used to draw a map of the olive growth. If inspecting (by air for a drone, in the field for a mobile device embedded in an agricultural machine) the whole olive grove is considered as too much expensive, the trajectory followed by the drone/mobile device is optimized according to the context. Initially, the drone/mobile device has to visit a random and uniform selection of olive trees. If no pest/disease attack is detected, the olive grove monitoring is done. Otherwise, as soon as an olive tree is detected attacked, its neighbors are also visited for damage assessment. The trajectory is dynamically updated according to the attacks detected.

C. A new method of grove fertilization

An important result of the soil analysis reported in this paper is: the roots of the olive tree use some strategies (e.g., photoassimilates) to adapt to nutrient-deficient environments to create an environment favorable to its development and its sustainability. This root system generates exudates of rhizodeposits to improve mineral nutrition of the olive tree. These results provide olive growers with a new method of grove fertilization based on the rhizosphere management for sustainable olive growing.

VII. CONCLUSION

In an Intelligent Farming System for a sustainable agriculture, wireless sensors are deployed to measure parameters enabling the computation of the amount of water needed by the olive trees in real-time. Data acquired from the olive groves are processed by smart algorithms including Machine Learning ones to compute the evapotranspiration model of olive trees and detect when to trigger irrigation. In such a case, commands are delivered to actuators to start and stop irrigation. In this paper, we have more particularly shown how to use self-organizing maps (SOMs) to identify olive grove clusters with similar features, characterize
each cluster and show the evolution over time of each olive grove. With SOM, it becomes possible to warn the farmer when some specific action needs to be done in the case of hydric stress, NPK stress, pest/disease attack. We are convinced that SOMs are very useful to sustainable agriculture by an early detection of anomalies and a fast reaction.

As a further work, we would like to extend this work in order to better understand the water balance, determine the rate in trace elements, particularly boron, identify the soil microflora and study the temporal dynamics and bioavailability of nutrients.

REFERENCES