Computers and Electronics in Agriculture 91 (2013) 19–29

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

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

# Yield prediction in apples using Fuzzy Cognitive Map learning approach

E.I. Papageorgiou [a,*], K.D. Aggelopoulou [b], T.A. Gemtos [b], G.D. Nanos [b]

[a] Department of Informatics and Computer Technology, Technological Educational Institute (TEI) of Lamia, 3rd Km Old National Road Lamia-Athens, 35100 Lamia, Greece
[b] University of Thessaly, Department of Agriculture, Crop Production & Rural Environment, Fytoko Street, N. Ionia, Magnesia 38446, Greece

ARTICLE INFO

*Article history:*
Received 15 May 2012
Received in revised form 3 November 2012
Accepted 13 November 2012

*Keywords:*
Fuzzy Cognitive Maps
Modeling
Knowledge representation
Fuzzy sets
Decision making
*Malus*domestica*

ABSTRACT

This work investigates the yield modeling and prediction process in apples (cv. Red Chief) using the dynamic influence graph of Fuzzy Cognitive Maps (FCMs). FCMs are ideal causal cognition tools for modeling and simulating dynamic systems. They gained momentum due to their simplicity, flexibility to model design, adaptability to different situations, and easiness of use. In general, they model the behavior of a complex system, have inference capabilities and can be used to predict new knowledge. In this work, a data driven non-linear FCM learning approach was chosen to categorize yield in apples, where very few decision making techniques were investigated. Through the proposed methodology, FCMs were designed and developed to represent experts' knowledge for yield prediction and crop management. The developed FCM model consists of nodes linked by directed edges, where the nodes represent the main soil factors affecting yield, [such as soil texture (clay and sand content), soil electrical conductivity (EC), potassium (K), phosphorus (P), organic matter (OM), calcium (Ca) and zinc (Zn) contents], and the directed edges show the cause-effect (weighted) relationships between the soil properties and yield. The main purpose of this study was to classify apple yield using an efficient FCM learning algorithm, the non-linear Hebbian learning, and to compare it with the conventional FCM tool and benchmark machine learning algorithms. All algorithms have been implemented in the same data set of 56 cases measured in 2005 in an apple orchard located in central Greece. The analysis showed the superiority of the FCM learning approach in yield prediction.



## 1. Introduction

Soil variability within a farm exists in most soils and regions. This variability interacts with weather, inputs (which practically cannot be applied homogeneously) and the variability of genetic material to produce crop and yield (quantitative and qualitative) variability. These different spatial or temporal variabilities have to be properly managed by the farmers to achieve the best profit with the lowest inputs, thus reducing the adverse environmental effects. Precision agriculture aims at better managing the variability. Several data are collected during each growing period over the years and stored in databases to assist in this management. The problem is that historical data cannot always predict yield variability and final yield. It seems that yield variability is canceling out after 3 years. This was the case in the data reported by Blackmore et al. (2003) for cereals and Fountas et al. (2004) for cotton. It is quite possible to have more stable zones of high and low yielding in perennial crops like orchards. But, in most farms and regions, alternate bearing is encountered in apple orchards and, to a lesser extend, many other fruit crops (Childers et al., 1995). This unpredictable yield and yield variation makes the successful application of variable rate of inputs difficult. It also causes a failure of the development of decision support systems (DSS) for precision agriculture, which is considered one of the reasons of the relatively low adoption.

Precision agriculture applications in fruits and vegetables are rather limited in the literature. Extensive work was reported in citrus (Shumann et al., 2006; Sakai et al., 2007; Maja and Ehsani, 2010; Mann et al., 2011), and vineyards (Bramley et al., 2003), but it is very limited in other crops. In particular, Aggelopoulou et al. (2010, 2011a, 2011b) have worked on precision agriculture applications in apples, Fountas et al. (2011) in olives, and Konopatzki et al. (2009) in pears. Such data would be useful for site-specific yield prediction calculation with proper methods.

Yield prediction in apples, and fruit trees in general, is very important, because it could be used to improve crop management and plan fruit marketing. Furthermore, when yield is predicted site-specifically, the inputs (water, fertilizers, pesticides) and field operations (e.g. fruit thinning) could be applied with variable rates depending on the real needs of the trees in the different areas of the field. As most crop models used did not successfully predict yield and yield spatial variation, different authors have used other parameters to predict them during the growing season. The earlier

* Corresponding author.
*E-mail address:* epapageorgiou@teilam.gr (E.I. Papageorgiou).

0168-1699/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compag.2012.11.008

the prediction, the better it can be used to improve orchard management. In apples, analysis of multispectral images of the fruits (Alchanatis et al., 2007) was used for yield prediction. Digital images of the trees at full bloom (Aggelopoulou et al., 2011b) were used to predict yield variability.

Besides, a large number of approaches like crop models, algorithms and statistical tools have been proposed and used for yield prediction in precision agriculture using historical data. Correlation and multiple linear regression (MLR) have been commonly used to predict yield and identify important factors influencing yield (Kravchenko and Bullock, 2000; Park et al., 2005; Gutiérrez et al., 2008; Huang et al., 2010), but the results are not so encouraging due to the existence of polynomial and interaction terms, which were not considered (Kitchen et al., 2003). In the case of MLR analysis, the description of linear relationships between crop parameters and site variables is limited, and the results may be misleading when these relationships are not linear (Kitchen et al., 2003; Schultz et al., 2000; Miao et al., 2006). Another common approach is combining multivariate techniques, like principal component analysis (PCA) and factor analysis (FA), with multiple regressions (Jiang and Thelen, 2004; Huang et al., 2010; Fortin et al., 2010). These methods were used to minimize the problem caused by interacting variables, facilitate the interpretation of complex relationships, reduce the dimensionality of the dataset or select a subset of appropriate variables from a large data set (Huang et al., 2010). Subsequent attempts have been made by applying artificial intelligence principles and soft computing techniques in precision agriculture for spatial analysis and crop management (Drummond et al., 2003; Savin et al., 2007; Jutras et al., 2009; Huang et al., 2010).

Artificial neural networks (ANNs), as no-linear statistical techniques, have also been applied to investigate yield response to soil variables (Jutras et al., 2009; Fortin et al., 2010; Park et al., 2005; Effendi et al., 2010). Specifically, ANN analysis has been applied in precision agriculture for spatial analysis and crop management (Kitchen et al., 2003; Drummond et al., 2003; Liu et al., 2001; Irmak et al., 2006). In the case of ANNs, the observed dataset of the selected variables is fitted to describe the problem by adjusting the weights of linkages connecting input and output variables and can be regarded as multivariate non-linear analytical tools. The ANNs can be combined with other artificial intelligence techniques or other statistical methods to benefit from the advantages of ANN modeling, and to also avoid some of their limitations such as the need for large amounts of data for training. Huang et al. (2010) summarized the soft computing techniques and their applications in agricultural and biological engineering.

An alternative soft computing technique, the fuzzy cognitive mapping (FCM), could be used for yield prediction and is used in our case study for apple yield prediction. FCM is a method for analyzing and depicting human perception of a given system with the development of a conceptual model, which is not limited by exact values and measurements (Kosko, 1986). The advantageous modeling features of FCMs, such as simplicity, adaptability and capability of approximating abstractive structures, encourage us to use them for complex problems. They gained momentum due to their dynamic characteristics and learning capabilities (Salmeron, 2009). The learning approaches for FCMs are concentrated on learning an adjacency matrix, based either on expert intervention and/or on the available historical data. According to the available type of knowledge, the learning techniques can be categorized into three groups: Hebbian-based, population-based and hybrid combining the main aspects of Hebbian-based and evolution-based type learning algorithms. The most used learning approaches in the literature involve the non-linear Hebbian learning (NHL) algorithm and the genetic algorithm learning (Papageorgiou, 2012), which are the most efficient in FCM training.

FCMs have been widely used in many different scientific fields such as engineering, business and management, environment, medicine and telecommunications (Papageorgiou, 2012). In agriculture, the FCM methodology with its learning capabilities was applied in cotton yield prediction (Papageorgiou et al., 2009, 2011), producing also a modeling tool for helping farmers make decisions in precision agriculture (Papageorgiou et al., 2011). The aim of the present study was to construct the FCM model for classifying yield in apples based on experts' knowledge and then to use efficient learning approach to train this model with field data and therefore exploit yield predictions. The NHL method was used for yield classification and its inference capabilities were compared with the FCM tool without learning and with the most used and known machine learning algorithms. It was shown that the NHL method for FCMs gives better prediction accuracies than those obtained with the use of conventional FCMs and machine learning techniques.

## 2. Material and methods

### 2.1. The data

The present study was carried out in a commercial apple orchard located in Agia area, central Greece (22°35′33″E, 39°40′28″N) in a 5 ha field in 2005. The main cultivar was Red Chief grafted on MM106 with Golden Delicious as pollinator. The trees were planted at 3.5 m × 2 m, trained as free palmette and intensively cultivated including regular irrigation and fertilization, winter and summer pruning and precise hand thinning 2 weeks after petal fall. This work is a part of an experiment that lasted 3 years (2005–2007) including soil, yield, and quality mapping. Yield was consistent over the two of the 3 years of the study (2005 and 2006). In the third year (2007) yield was lower than the two previous years because a number of trees died out (Aggelopoulou et al., in press). There were no sights of alternate bearing in the bearing trees due to proper weather conditions during fruit set, the intensive cultivation practices applied and the low vegetative vigor of apple cv. Red Chief. In this paper the data of year 2005 are presented.

For yield mapping, apples were collected manually in September 2005 and placed in plastic bins along the tree rows at commercial harvest time (Fig. 1). Yield per 10 trees was weighted and the geographical position in the center of the 10 trees was recorded using a hand-held computer with GPS (Trimble pathfinder).

In December 2005, 20 soil samples were taken before winter crop fertilization to a sampling depth of 0–30 cm. The sampling positions were geo-referenced using a hand-held computer with GPS. The samples were air-dried, passed through a 2 mm sieve and analyzed for the following properties: soil texture (% sand, %

**Fig. 1.** Orchard under study with the harvesting bins placed along the rows. Apples from groups of 10 adjacent trees were weighed to create yield maps.