European Journal of Operational Research 204 (2010) 355–365

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Interfaces with Other Disciplines

Income prediction in the agrarian sector using product unit neural networks

Carlos R. García-Alonso [a], Mercedes Torres-Jiménez [a,*], César Hervás-Martínez [b]

[a] Department of Management and Quantitative Methods, ETEA, University of Cordoba, Escritor Castilla Aguayo, 14004 Córdoba, Spain
[b] Department of Computing and Numerical Analysis of the University of Córdoba, Spain

ARTICLE INFO

Article history:
Received 14 April 2009
Accepted 29 September 2009
Available online 2 October 2009

Keywords:
Neural networks
OR in agriculture
Product-unit models
Regression

ABSTRACT

European Union financial subsidies in the agrarian sector are directly related to maintaining a sustainable farm income, so its determination using, for example, the farm gross margin is a basic element in agrarian programs for sustainable development. Using this tool, it is possible the identification of the agrarian structures that need financial support and to what extent it is needed. However, the process of farm gross margin determination is complicated and expensive because it is necessary to find the value of all the inputs consumed and outputs produced. Considering the circumstances mentioned, the objectives of this research were to: (1) select a representative and reduced set of easy-to-collect descriptive variables to estimate the gross margin of a group of olive-tree farms in Andalusia; (2) investigate if artificial neural network models (ANN) with two different types of basis functions (sigmoidal and product-units) could effectively predict the gross margin of olive-tree farms; (3) compare the effectiveness of multiple linear, quadratic and robust regression models versus ANN; and (4) validate the best mathematical model obtained for gross margin prediction by analysing realistic farm and farmer scenarios. Results from ANN models, specially the product-unit ones, have provided the most accurate gross margin predictions.

1. Introduction

Since the 1990s, the quest for sustainability in the rural environment has been one of the greatest problems in the European Union (EU) (European Commission, 1999). The EU agrarian sector has critical budget concerns because, although economically it is the least important sector, it does, however, keep a very great portion of the territory stable by avoiding risks of a desertification that could be brought about by a lack of rural population or the abandonment of agrarian activities. The relevance of the agrarian sector in the EU budget requires that financial support be socially justified in terms of environmental maintenance (management of rural areas that includes, for example, avoiding erosion and losses in genetic variability), food quality, efficiency, best practices and so on (Lütz and Bastian, 2002; Amores and Contreras, 2009). Therefore, agrarian income determination (estimated by the farm gross margin) is essential in order to estimate the real financial situation of the agrarian sector, thus enabling the design of the financial support system needed to maintain rural population, that is, to achieve medium-long term sustainability (Pacini et al., 2003).

The problem is that it can be quite complicated to calculate this farm gross margin, basically due to: the existence of mixed farms, extreme variations in the natural environment, sanitary incidences (Sadras et al., 2003), administrative issues, market adjustments and many other reasons. In this complex environment, it is more appropriate to study productive structures or strategic agrarian groups rather than individual cases (farms), even though it is difficult to obtain the data series needed for each type. The strategic group studied here is Andalusian olive-tree farms in dry farming; this group is very important in Andalusia (MAPA, 2003) due to its relevance socially (employment level), economically (Andalusia is the main olive-producing region in Spain, yielding more than 70% of the total production while the other regions produce less than 10% each) and environmentally (IEA, 2000).

The evolution of the gross margin in a strategic agrarian group can be analysed through periodical surveys, by designing a farm-panel or by analysing data from accounting nets (Pacini et al., 2003). All of these methods are expensive and complicated to manage if a representative sample of real farms is required, because hundreds of variables are needed to determine the farm gross margin depending on the number of agrarian activities, the size and its productive structure. It is necessary to determine the value in monetary terms of all the inputs consumed and outputs produced, as well as the value of their structural costs and revenues, financial costs, taxes, insurance costs and so on. The use of all these variables as predictors for estimating the margin would imply a high cost in the acquisition and later computational treatment of the information. Moreover, using all the predictors in the model often results in strong over-fitting and very poor predictions (Meiri and Zahavi, 2006).

* Corresponding author. Tel.: +34 957222192; fax: +34 957222182.
E-mail addresses: cgarcia@etea.com (C.R. García-Alonso), mtorres@etea.com (M. Torres-Jiménez), chervas@uco.es (C. Hervás-Martínez).

0377-2217/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2009.09.033

The first objective of this research was to select a representative and reduced set of easy-to-collect descriptive variables from the larger set of original variables, in order to estimate the gross margin of the farms analysed. According to Liu and Motoda (1998), using this procedure would achieve a better understanding of the final prediction model. Keeping these premises in mind, the prediction models used to estimate gross margin incorporate different techniques for selecting the set of the best predictor variables (efficiency and efficacy) for each case. Of the two possible methodological approaches that have been developed to solve selection problems for descriptive (Sebestyen, 1962), we have applied heuristic techniques because optimal or exact techniques are known to guarantee an optimal solution, though they are only applicable to small-sized sets since they investigate all possible feature subsets under some criteria to look for the best possible configuration (Gatu and Kontoghiorghes, 2006). Moreover, heuristic techniques are able to find good solutions (although unable to guarantee the optimum) in a reasonable amount of time (Kohavi and Sommerfield, 1995). These heuristic techniques include filter methods, the wrapper approach and the non-deterministic search (Liu and Motoda, 2002). The drawback of these heuristic techniques is that the quality of their solutions varies greatly depending on the methods used. As found in other optimization problems, metaheuristic techniques have proved to be efficient alternatives in different domains (Pacheco et al., 2009). Specifically, the application of genetic algorithms for the selection of variables has provided better results than other techniques in many studies (Inza et al., 2001; Wong and Nandi, 2004).

The selection process applied, as a first step, a filter method, using the linear correlations between explicative variables and the gross margin, in order to select those variables that were most closely related to the dependent variable. After that, a process for selecting variables was reapplied in each predicting model, that is: in regression models, a backward step process was applied and for neural network models, we propose a heuristic and non-deterministic method for selecting features from the neural network. This last method includes an evolutive algorithm that modifies the number of connections between the nodes between the intermediate and the input layers through a structural mutation operator. It also uses a selection process that keeps the best solutions – elitism – which facilitates convergence to the global optimum (Rudolph, 1994).

Another objective of this research was to analyse the application of new methodologies like artificial neural networks (ANN) and evolutive algorithms (EA) to solve a real problem in the agrarian sector: the estimation of the gross margin of a selected group of agricultural enterprises. These new techniques have been used in the agrarian sector basically only in production prediction (Torres et al., 2005), cultivation practice planning (Srinivasa et al., 2006) or crop yield prediction (Kaul et al., 2005). We were specifically interested in analyzing the response of these techniques in predicting the gross margin in agrarian farms because there is not much literature available on this subject except for those using standard mathematical programming methods (Recio et al., 2003). For that reason, this research aims to ascertain if ANN models could effectively predict olive-tree farm gross margin using the above mentioned set of easy-to-collect descriptive variables, thus improving the performance of standard regression models.

Standard regression methods are frequently used to resolve prediction problems although there are several drawbacks involved, among them the fact that the functional form needs to be chosen "a priori". Moreover, the variables have to comply with several requirements (like normality, linearity, no multicollinearity, etc.) which rarely occur in reality, as happens in our problem. Consequently, more general basis functions, structured linearly with their corresponding coefficients, are used to design the desired function. Hence, when a function f is approximated to another function g, the response y_i can be written in the form $y_i = g(\mathbf{x}_i) = \sum_{j=1}^{p} \beta_j B_j(\mathbf{x}_i) + \varepsilon_i$, for $i = 1, \ldots, n$; where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ is the vector of coefficients corresponding to basis functions $\mathbf{B} = (B_1, \ldots, B_p)^T$, which are nonlinear transformations of the data vector $\mathbf{x}$, and, finally, ε_i is the error. Thus we extend the above defined class of functions, and each basis function B_i defines a map from input space χ (a range of plausible predictor variables) to real space R. ANN (Bishop, 1995; Haykin, 1999) are an example of basis function models which have received a great deal of attention in the last two decades. The most popular model of ANN is the feed-forward sigmoidal networks, where the activation function of the output unit is its step function, while the activation function of every other computation unit is the sigmoidal function. Research has shown that feed-forward sigmoidal networks can approximate any continuous function, like the gross margin, to an arbitrary degree of accuracy, provided that the hidden layers contain a sufficient number of hidden units. We have also applied another ANN model to estimate the gross margin called the product unit neural network which is based on the product unit function.

Product units, introduced by Durbin and Rumelhart (1989), are artificial neurons, unlike more widely used ones because they multiply their inputs instead of adding them. Furthermore, their weights operate as exponents and not as factors. Product units were to allow neural networks to learn multiplicative interactions to an arbitrary degree. For this to occur, there had to be both strong relationships among predictive variables of the model and a large enough space.

Product unit neural networks (PUNN) present an alternative to compute the net input signal with the advantages of increased information capacity and the ability to form higher-order combinations of inputs. Consequently, the network architecture can be reduced because, as has been demonstrated experimentally, PUNN needs a more reduced number of base functions or connections to obtain the same number or fewer errors than other artificial neural networks, including sigmoidal or radial function models (Martínez et al., 2006; Gutiérrez et al., 2009). In order to train PUNN, a rich collection of learning algorithms has been designed and analysed, ranging from local ones, like gradient descent, to more global ones, such as simulated annealing and genetic algorithms (Maniezzo, 1994; Yao and Liu, 1997). On the other hand, while other studies have applied different optimization methods to train PUNN (Janson and Frenzel, 1993; Ismail and Engelbrecht, 1999, 2000, 2002; García-Alonso et al., 2009; Torres et al., 2009), this work proposes an evolutive algorithm for both the architectural design and the estimation of the real-parameters of the PUNN model. This evolutive algorithm was also used for training an artificial neural network model based on sigmoidal unit basis functions (MLPEA), thus comparing the two net models (product units and sigmoidal basis functions) under the same conditions. A traditional ANN model was also applied based on sigmoidal basis functions trained with a back propagation algorithm. The results obtained with all the different ANN models applied were compared to those obtained with regression techniques (linear, quadratic and robust models).

Therefore this paper is the first attempt to predict the gross margin of a farm as a guide to both managerial decisions and policy design through the use of artificial intelligence methodologies. It is structured as follows: Section 2 describes the different prediction methods applied; Section 3 briefly explains both the gross margin calculation and experimental design to estimate the gross margin for Andalusian olive-tree farms in dry farming; the best model obtained in each generalized linear regression model is briefly described statistically in Section 4; the most relevant findings obtained using the PUNN model are described in Section 5 where two representative scenarios are analysed and, finally, some illustrative conclusions are drawn in Section 6.

2. Methodology

When there is no a-priori knowledge of the exact relationship existing between the Y response variable and the set of predictive non-random or deterministic p variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^T$, it is a common practice, due to its simplicity, to design a relationship in which the variable explained is related to explicative variables through a determinist function $f(\boldsymbol{x})$ where some random error component ε is included.

The simplest form of function $f(\boldsymbol{x})$ is linear, but frequently this model is not flexible enough to accurately model the dependent variable in complex environments. That is why alternative models are analysed where coefficients are nonlinear functions. These generalized linear models are structured by a linear combination of nonlinear basis functions as follows:

$$y = f(\mathbf{x}, \beta_j, w) = \sum_{j=0}^{p} \beta_j B_j(\mathbf{x}, w_j) \quad \boldsymbol{x} \in D \subset R^k, \tag{1}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the vector of coefficients, $\mathbf{B} = [B_0(\boldsymbol{x}, w_1), \ldots, B_p(\boldsymbol{x}, w_p)]^T$ is the basis function vector and, finally, w_j are the coefficients associated with the basis functions to introduce nonlinearity into the model, being $B_0(\boldsymbol{x}, w_0 = 1) = 1$. In general, these basis functions are nonlinear transformations of the $\boldsymbol{x}$ vector which is why the range of possible models widens considerably. If we formulate the model proposed in (1) considering the set of data $D_E = \{(\boldsymbol{x}_i, y_i), i = 1, \ldots, k\}$, we see that:

$$y_i = \sum_{j=0}^{p} \beta_j B_j(x_i, w_j) + \varepsilon_i \quad i = 1, \ldots, k. \tag{2}$$

This family of functions can be considered a generalization of response surfaces (Myers and Montgomery, 2002). Thus, the relation $y = f(\boldsymbol{x}) + \varepsilon$ materialises through f functions that can be, in general, m order polynomial equations; finally other variability sources, not included in f, are represented by ε. The response surface of the level two model $y = f(\boldsymbol{x}, \boldsymbol{\beta}) + \varepsilon$, ordered according to its Taylor series, shows this structure (2^{nd} order terms):

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{j<l} \sum_{l=2}^{k} \beta_{jl} x_j x_l + \varepsilon. \tag{3}$$

Researchers from different areas of study have suggested using artificial neural networks as an alternative to surface response models (García et al., 2003; Grznar et al., 2007). Multilayer ANN with forward activation (Haykin, 1999) can be considered predictive models associated with multivariate statistical analysis when k inputs, $\boldsymbol{x}$ variables and, usually, one dependent variable y are considered. In this framework, different alternative basis functions and methods have been proposed, such as: radial basis functions, projection pursuit learning, product unit neural networks and general regression networks.

In this paper, both sigmoidal and product unit basis functions were used to analyse which of these approaches best predicts gross margin in dry farming olive-tree farms. In cases of both sigmoidal and product unit typologies, the best models found were checked, not only for their predictive accuracy but also for their capacity to support real decisional situations.

Product-unit neural networks models (PUNN), introduced by Durbin and Rumelhart (1989) and later developed by Ismail and Engelbrecht (1999, 2000, 2002) and Schmitt (2002), are similar to standard sigmoidal neural networks but are based on multiplicative nodes instead of additive ones. The nonlinear basis functions of the model are constituted by the product of the variables initially included in the problem raised to arbitrary powers.

Given the data set D_E, the regression model can be formulated by a linear combination of product unit basis f function: $D \subset R^k \to R$, so that:

$$y = f(\mathbf{x}) = \sum_{j=0}^{p} \beta_j \left(\prod_{i=1}^{k} x_i^{w_{ji}} \right), \tag{4}$$

where $\beta_j \in R$ and $w_{ji} \in R$ for $j = 0, \ldots, p$ and $x_i > 0$ for $i = 1, \ldots, k$. In these models, basis functions are potential ones like:

$$B_j(\mathbf{x}, w_j) = \prod_{i=1}^{k} x_i^{w_{ji}}. \tag{5}$$

In our case, the network has k inputs (independent variables) in the input layer, m nodes in the hidden layer and one node in the output layer (dependent variable: gross margin), Fig. 1. The activation function of the jth node in the hidden layer is given by (5) where w_{ji} is the weight of the connection between the input node i and the hidden node j.

If the exponents in (5) are {0,1}, we obtain a higher-order unit also known by the name of sigma–pi unit. Therefore, the product units are a generalization of the sigma–pi neural nets which have the capability of implementing higher-order functions and therefore can also implement polynomial functions in a specific instance (Gurney, 1992).

In contrast to the sigma–pi units, the exponents in the product-unit are not fixed and may even take real values. In this way we get more flexible models and avoid the huge number of coefficients involved in the polynomial model. To avoid the problem that could result from networks containing product units that receive negative inputs and weights that are not integers, the values for the input variables (x_i) are limited to positive ones (because, as we know, a negative number raised to some non-integer power yields a complex number).

Some advantages of PUNN are their increased information capacity and the ability to form higher-order input combinations. Durbin and Rumelhart (1989) determined empirically that the information capacity of product units (measured by their capacity for learning random Boolean patterns) is approximately $3N$, compared to $2N$ for a network with additive units for a single threshold logic function, where N denotes the number of inputs to the network. Moreover, it is possible to obtain the upper bounds of the Vapnik–Chervonenkis dimension in product-unit neural networks similar to those obtained in sigmoidal neural networks. The Stone–Weierstrass theorem proved that product-unit neural networks are universal approximators (Schmitt, 2002). A disadvantage of this type of nets with respect to standard sigmoidal ones is the greater degree of difficulty for the corresponding training process

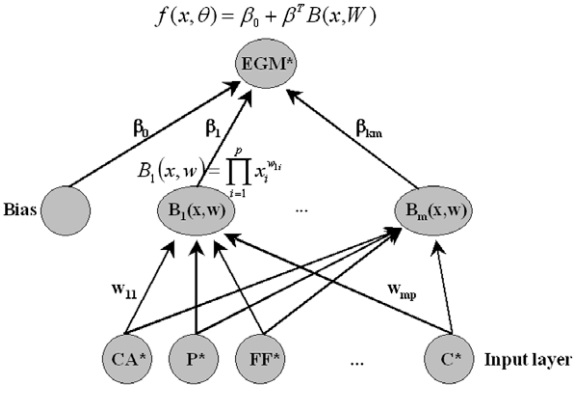

Fig. 1. Product unit neural networks (PUNN) diagram.

since small changes in exponent values can provoke great changes in the error surface. This type of net presents a greater number of local minimums, thus increasing the possibility of getting trapped in them. For this reason, a well known issue is that back-propagation is not efficient in training product units (Janson and Frenzel, 1993). To overcome this problem, we use an evolutionary algorithm as part of the process for the estimation of parameters.

Evolutionary algorithms (EA) are a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm, which uses some mechanisms inspired by biological evolution: reproduction, mutation, recombination, and selection to solve problems that need artificial intelligence methods. Candidate solutions for the optimization problem play the role of individuals in a population and the fitness function determines the environment within which the solutions "live". The evolution of the population then takes place after the repeated application of the above operators. Artificial evolution describes a process involving individual evolutionary algorithms; EA are individual components that participate in an artificial evolution process. They are classified into categories depending on the way in which these individuals are represented, the design of the operators of variation and the selection or reproduction mechanism used. These categories are genetic algorithms (Goldberg et al., 1992), evolutionary strategies (Schewefel, 1993) and evolutive programming (Fogel, 1991).

Evolutionary algorithms often approximate solutions to all types of problems because they ideally do not make any assumption about the underlying fitness framework; this generality is shown by their success in fields as diverse as engineering, art, biology, economics, marketing, genetics, operations research and so on.

Evolutionary computation has been widely used in recent years to evolve neural-network architectures and weights. There have been many applications for parametric learning (Van Rooij et al., 1996) and for both parametric and structural learning (Maniezzo, 1994; Yao and Liu, 1997). On the other hand, different optimization methods have been applied to train PUNN (Janson and Frenzel, 1993; Ismail and Engelbrecht, 1999, 2000, 2002).

In this study the exponents of the variables are estimated and the optimum number of product units is determined in several steps. An evolutionary algorithm was used here, which has points in common with other EA in the bibliography that had previously demonstrated accurate results (Angeline et al., 1994; Yao and Liu, 1997; García et al., 2002, 2003; Torres et al., 2009). The search begins with an initial population, and each repetition updated the population using a population-update algorithm. The population undergoes replication and mutation operations. As Angeline suggested (1994), crossover is not used in this work due to its potential disadvantages in evolving neural networks.

The EA begins by generating N_R nets randomly. First, the total number of hidden nodes is chosen from a uniform distribution in the interval $(0, m]$ where m corresponds to the maximum number of hidden nodes in each one of the population nets. The number of connections between each node in the hidden layer and the nodes in the input layer are also determined by a uniform distribution in the $(0, p]$ interval, p being the number of independent variables. Having defined the topology of the net, weights are assigned to each connection, considering one uniform distribution $[-L, L]$ for those between the input and the hidden layers, and another uniform distribution $[-U, U]$ for those between the hidden and the output layers. After the random generation of the population in N_R nets, the EA takes place until the stopping criterion is fulfilled, as seen in the following sequence:

a. The fitness of every individual in the population is calculated.
b. Individuals are ranked with respect to their fitness.
c. The best individual is copied into the new population.
d. Ten percentage of the best individuals in the population are replicated and substitute the worst 10% of individuals, so the net number N_R remains constant throughout its evolution.
e. Parametric mutation is applied to the best 10% of individuals.
f. Structural mutation is applied to the remaining 90% of individuals.

Parametric mutation affects net weights and consists of a local search algorithm in the space of the weights, a simulated annealing algorithm, where the severity of the mutation depends on the net temperature $T(R)$ defined by:

$$T(R) = 1 - A(R) \quad 0 \leqslant T(R) < 1, \tag{6}$$

where the aptitude $A(R)$ of a net R is calculated as a decreasing function of the root of the mean squared error $E(R)$ from the expression:

$$A(R) = 1/[1 + E(R)] \quad 0 < A(R) \leqslant 1. \tag{7}$$

Parametric mutations consist of adding a Gaussian random variable with mean 0 and standard deviation $\sigma_1 \times T(R)$ to each of the w_{ki} coefficients, while in the rest of the coefficients the Gaussian random variable has a mean 0 and a standard deviation $\sigma_2 \times T(R)$, where $\sigma_1 \ll \sigma_2$.

The structural mutation modifies both the number of hidden nodes and the connections between the nodes in the intermediate layer, and those in the input and output layers, producing a different net topology (Martínez et al., 2006).

3. Farm gross margin calculation and experimental design

Letting n be the number of agrarian activities on a single farm, its gross margin (GM) can be calculated by subtracting total farm agrarian costs (variable production costs (COST_i) and structural costs (STRU) from total revenues (REV_i)).

$$\text{GM} = \sum_{i=1}^{n} \text{REV}_i - \left(\text{STRU} + \sum_{i=1}^{n} \text{COST}_i \right). \tag{8}$$

REV_i can be divided into product sales, diversification revenues, subsidies and other revenues of any agrarian activity. On the other hand, COST_i is structured by input costs, hand labour costs, machinery costs, transformation costs, marketing costs, etc. Finally, the STRU includes energy costs, insurance costs, taxes, infrastructure maintenance, structural hand labour costs, etc. In the end, hundreds of variables are needed, depending on the number of agrarian activities ($i = 1, 2, \ldots, n$), the farm size and its productive structure.

To carry out the experiments, 208 olive-tree farms in dry farming were selected. This number allows a thorough study due to the number of variables requested (over 150 costs, revenues, etc.). This strategic group is very important in Andalusia due to its social (employment level), economic (olive trees are the most important crop in Andalusia) and environmental relevance (European Commission, 1999; IEA, 2000; Fernández et al., 2004).

Before designing the PUNN model and in order to select an appropriate set of variables for predicting gross margin, several premises were established: all the variables selected had to be very easy-to-collect, for example, through an uncomplicated and standard survey process; all the quantitative independent variables had to have a significant relationship (sometimes based on expert knowledge) with the dependent variable (GM); all of them had to deliver quite an accurate description of the socio-economic structure of the farm and the farmer (decision maker) and, finally, some of them had to be directly under the farmer's control (decision maker dependent variables) in order to carry out a sensitivity analysis using the resulting regression models. According to these premises,

we applied a filter method to a first feature selection. From the total original set of descriptive farm variables, those continuous ones that showed significant linear correlation with GM were selected (Table 1): total cultivated area (CA, hectares); *P* (olive-tree production: olives for producing olive oil, measured in tons); FF (family hand labour – in working days – with respect to the total farm hand labour, in percentages) and *T* (the number of farm tractors). The mean altitude (meters) over sea level of the municipality (*H*) was also included despite the fact that there was no significant linear correlation with gross margin because its agronomical relevance in olive-tree production, as will be demonstrated in the scenario analysis presented in Section 5.

Some input variables describing the farmer were also included in the analysis because they could influence farm management and therefore affect its financial outcome (gross margin). These variables are related to the farmer's experience and his knowledge of modern cultivation processes: the farmer's age (A: under 40 years old – 1, between 40 and 55 – 2, between 55 and 65 – 3 – and over 65 years of age – 4. Mean: 2.17, standard deviation: 0.89), and the existence or non-existence of agrarian studies on the part of the farmer (AS: yes – 1 – or no – 2. Mean: 1.75, standard deviation: 0.43). Finally, independent variables related to farm activity diversification and possible membership of the farmer in a cooperative were also taken into consideration. Specifically, the I variable denoted if the farmer and/or his family received non-agrarian income from other economic sectors (I: yes – 1 – or no – 2. Mean: 1.29, standard deviation: 0.46), and CP, which showed if the farmer totally or partially handed over his production to a cooperative (CP: yes – 1 – or no – 2. Mean: 1.19, standard deviation: 0.39). Table 1 shows the quantitative variables used to describe the sample selected (208 olive-tree farms in dry farming) and their basic descriptive statistics.

Observations were also identified at the Extremes of the statistical Domain (called OED, considering the variable mean plus and minus three standard deviations until no more OED were detected). The variability in the sample is really relevant because includes both big and small farms and productive and unproductive ones, as well as those with large or small gross margin, all coexisting in the original set (Tables 1 and 2). After analysing these cases, it was decided to include them in the sample after all because there was nothing to prove that they were not representative of the set of activities. For this reason, OED were randomly distributed among training and generalization sets in the partitions designed. In order to take into account extreme observations in the linear regression model, robust regression was also applied because is specially indicated when the errors do not satisfy conditions of normality or when the data contain significant outliers (Yohai and Zamar, 1997).

According to a cross-validation procedure, 10 partitions of the original sample (208 farms) were made in all cases designing the training and generalization sets. Although the cases were selected randomly, they did consider each of the major Andalusian olive-tree regions separately to maintain the territorial representativity of the sample. The training sets included approximately 70% of the 208 farms analysed (147 farms), and the generalization sets comprised the other 30% of the farms (61 farms).

Finally, nine independent variables were considered in the analysis and were initially introduced as predictor variables in all the models applied. Later, in a second phase, a new selection was carried out for each estimation model applied (see Sections 1 and 2 for details), thus searching for the subset of explicative variables that would offer the best predictions for gross margin in each case.

Table 1
Basic descriptive statistics of quantitative variables (surveyed olive-tree farms), 208 observations.

	CA (ha)	*P* (ton)	FF (%)	*H* (m)	*T* (tractors)
Mean	18.35	36.40	43.25	540.38	0.77
Median	8.30	18.00	39.46	549.00	1.00
Standard deviation	28.53	52.50	39.92	202.08	0.97
Variation coefficient (%)	155.48	144.23	92.30	37.39	125.97
Minimum	0.35	0.21	0.00	65.00	0.00
Maximum	240.00	373.93	100.00	1064.00	6.00
Linear correlation with gross margin (*p*-value)	0.354 (0.000)	0.502 (0.000)	−0.168 (0.015)	−0.082 (0.239)	0.247 (0.000)

Table 2
Farm gross margin (10^2 €) – real and estimated (best model PUNN, estimated gross margin) – of surveyed olive-tree farms, 208 observations.

	Real – surveyed	Estimated – best model
Mean	176.40	168.32
Median	49.48	59.02
Standard deviation	569.64	527.05
Minimum	−4649.80	−4747.48
Maximum	3139.45	2973.47

4. Results obtained in estimating gross margin

This section analyses the results obtained in estimating gross margin using all the algorithms selected: linear regression (LR); robust regression (RR); quadratic regression (QR); MLP both with back propagation (MLPBP) and with an evolutive algorithm (MLPEA) and, finally, product-unit neural networks (PUNN).

4.1. Regression models

Standard linear regression was carried out using the selected training sets. Previously a study on the normality of the continuous independent variables had been carried out as well as another on the linearity of the relation. The normality goodness-of-fit test (Kolmogorov–Smirnov) showed that the null hypothesis should be rejected ($\alpha = 0.01$). On the other hand, by applying logarithm transformations, the null hypothesis (normal distributions of the transformed variables) could not be rejected even though the predictive capacity of the resulting models (using transformed variables) decreased. According to these results, the original variables were finally included in linear models. As mentioned above, all the quantitative variables selected, except *H* (height), showed a significant linear correlation (*p*-value 0.05) with the dependent variable gross margin (Table 1). In the first regression model all the variables described in Section 4 (CA, *P*, FF, *T*, *H*, *I*, *A*, AS and CP) were included. Later, progressively, variables whose coefficients surpassed a significance level of 5% were removed since the predictive capacity of the model not only did not deteriorate, but actually improved. Only four predictive variables (CA, *P*, *I* and CP) were finally included in the linear regression models (Table 3 shows the best linear regression model obtained from the 10 partitions, in terms of learning and predicting capacity). In the interpretation of linear regression coefficients, we must keep in mind the scale used in the evaluation of binary variables: *I* and CP. Following the system of numerous binary variable evaluation included in the UCI databases, the scale 1 = yes and 2 = no was used. This scale allows more general models to be obtained than the 0–1 scale, although the value of the coefficients cannot be interpreted as a deviation of the comparison group with respect to the variable criterion. The graph of the residues obtained with linear regression showed the presence of heteroscedasticity in the model and nonlinear patterns, which justifies the application

Table 3
Statistical results obtained from the best models of each methodology.

Method	MSE_T	R_T^2	MSE_G	R_G^2	Best models obtained in 10 partitions training/generalization sets
LR	131875	0.46	77988.5	0.50	$EGM = 393.84 + 4.84P + 3.13CA - 170.10CP - 171.25I$
QR	54775	0.82	122891.1	0.21	$EGM = 0.08(CA)^2 + 20.34(P) - 1.70(CA.I) - 5.02(P.CP) - 3.79(P.R) + 29.83(T.A) - 0.10(P.FF) + 0.15(CA.FF) - 6.43(CA.AS)$
RR	341332.3	0.49	33227.9	0.60	$EGM = -144.69 + 4.25P + 0.14FF - 9.60T + 62.89AE$
MLPEA	23043.6		62249.3		$EGM = -3.02 + 0.01/[1 + \exp -\{1.88AS^*\}] + 3.64/[1 + \exp -\{3.34 + 1.30CA^* + 4.31P^* + 0.65H^* - 0.05CP^*\}] + +0.76/[1 + \exp\{-0.20 + 1.37CA^* + 4.52P^* - 24.69FF^* - 2.88H^* + 0.38A^* - 6.91CP^*\}] - 7.25/[1 + \exp\{-5.70 + 1.22A^* + 0.19AS^* + 9.42T^* - 0.02I^* - 17.93FF^* - 22.14P^* + 5.79CA^*\}] + 0.02/[1 + \exp -\{-225.54AS^* + 655.58T^* - 381.99H^* + 56.81I^* + 356.15FF^* - 1479.45P^* + 346.14CA^*\}]$
PUNN	30013.2		47931.3		$EGM = 1.590 + 0.197(CA^*)^{-0.596}(P^*)^{2.056} + 14.186(CA^*)^{2.010}(P^*)^{-1.253}(H^*)^{1.803}(T^*)^{2.253}(A^*)^{0.312} - 10.086(CA^*)^{2.432}(P^*)^{0.256}(FF^*)^{-2.345}(I^*)^{0.017}(H^*)^{3.432}(T^*)^{2.516}(A^*)^{0.277}(CP^*)^{0.048}$

MSE_T: mean square error (training set). R_T^2: adjusted R^2 coefficient in training set. MSE_G: mean square error (generalization set). R_G^2: adjusted R^2 coefficient in generalization set. LR: linear regression. QR: quadratic regression. RR: robust regression. MLPEA: multilayer perceptron. PUNN: product unit neural networks.

of nonlinear models to estimate the gross margin of the farms analyzed.

As expected, the mean squared error is considerable in both the training and generalization sets (MSE_T and MSE_G). The adjusted R^2 coefficient in the training sets was between 40% and 50% in most of the 10 partitions analysed. In generalization sets, the percentage of gross margin variability was also within 40–50% (50% in the best model, Table 3) in the resulting models.

The second generalized linear model was based on quadratic regression (response surface). Again, non-significant variables were removed (significance level 5%). In addition to the variables included in the best linear model, four independent variables appeared through interactions: the farmer's age (A), the existence of agrarian studies (AS), the number of tractors (T), and the percentage of family hand labour (FF). These quadratic models showed better statistical results than linear ones in training sets (lower MSE_T and higher adjusted R^2 coefficients, within 70–82% of designed partitions), although the MSE_Gs obtained using generalization sets, being under 22%, were higher than those obtained in linear models and the adjusted R^2 coefficients. It is evident that the application of the quadratic model produces an over-fitting process and denotes this model's low predictive capacity compared even to the standard linear one (Table 3 shows the best quadratic model found in the 10 partitions). Precisely because of this, the possibility of using surface models of generalized response, like PUNN models, was considered as an alternative to quadratic or polynomial models of any other m order.

Due to the existence of extreme observations, the MM regression, a robust regression technique, was also used (Yohai et al., 1991; Yohai and Zamar, 1997). The Wald test was applied to determine whether a regression coefficient was zero. On comparing the results obtained with the other regression models applied (linear and quadratic), we can see that although the best model of robust regression offered the most reduced value in MSE_G (Table 3), its mean results were worse than those obtained by other regression techniques. Robust regression results obtained in the 10 partitions of the sample (Table 4) showed greater standard deviations; probably the position of extreme observations in each sample design (in training or generalization sets) was the cause of this variability.

Table 4
MSE_G descriptive statistics in 10 generalization partitions.

Model	Mean	SD	Best	Worst
PUNN	71624.70	15125.89	47931.30	93381.19
MLPEA	88263.37	23004.23	62249.30	102314.90
MLPBP	165357.78	71901.72	77734.40	229680.56
LR	170661.22	68577.21	77988.54	276342.85
QR	171384.93	31736.37	122891.10	222663.50
RR	149471.58	92182.03	33227.99	252349.84

SD: standard deviation.

4.2. Results for MLP and PUNN models

The input variables were scaled in the interval [0.1,0.9] and the output variables in the interval [1,2] for PUNN models, while in MLP models, it was at [0.1,0.9] for input and [0.1,0.9] for output variables. The former case (also depending on the potential function shape) gave output values far from zero and, the latter the output range avoids the saturation of output values due to the shape of the sigmoidal functions. The new scaled variables are named CA^*, P^*, FF^*, etc., for the input variables, and estimated gross margin (EGM) for the output variable.

The first model that was analyzed was a MLPBP. A weight elimination process was applied (Williams, 1994) to decide how many neurons should be included in the hidden layer. It was decided to start out with a network of considerable size and gradually eliminate nodes and unnecessary weights (pruning) until a network size providing satisfactory results could be reached.

In order to test the predictive capability of both MLPEA and PUNN in a high-dimensional space, an EA was used to determine the architecture and to train the weights of the basis function models. Using the structural mutation operator, our EA can indirectly prune input variables. By deleting connections, this operator can remove any input variable.

The EA was run 30 times, for 200 generations, with the following parameters: the exponents w_{ji} were begun in the interval $[-5,5]$ and the coefficients β_j in $[-5,5]$. The size of the population was $N_R = 2000$. The number of nodes that could be added or removed in structural mutation was within the interval [1,2]. The number of connections that could be added or removed in a structural mutation also fell within the interval [1,6].

For the MLP neural networks, the best model is obtained using the EA (MLPEA, five nodes in the hidden layer with 34 coefficients; $MSE_G = 62249.3$) and includes all the input variables. Applying MLPBP, the best model obtained was not as good ($MSE_G = 77734.4$). For PUNN, the best model is obtained using three nodes in the hidden layer with 19 coefficients and has a $MSE_G = 47931.3$ (Table 3).

Results obtained in both training and generalization sets using MLPEA were better than those obtained in previous analyses. If we compare the best MLPEA and PUNN models (Table 3), results in the training sets were very similar although somewhat higher in the PUNN model (23043.6 quadratic units, cu, for MLPEA compared to 30013.2 cu for PUNN); in generalization sets, PUNN mod-

els were also found to be better (62249.3 cu, for MLPEA as compared to 47931.3 cu for PUNN), taking into account that the interactions among variables are very relevant (see Section 5).

Furthermore, taking into account the average results obtained in the 10 generalization groups (Table 4), PUNN models were also better than MLPEA, with fewer quadratic errors, and the variability estimated by the MSE was higher when analysing MLPEA models than when considering PUNN ones. The MLPBP model reached better mean results than linear and quadratic regression models but worse ones than evolutionary MLPEA and PUNN models. RR reached the lowest inferior average quadratic error in its best model but its results also presented the greatest variability (with a variation coefficient over 60%). On average, the RR was better than the quadratic and linear regression models, and even the MLPBP model, but proved to be worse than the MLPEA and PUNN models.

A non-parametric Friedman test for k related samples was carried out to analyse the significance of the differences between MSE_G means. The results showed the existence of significant differences (p-value = 0.000) and so PUNN models were those finally selected to estimate gross margin since they could be considered to be an appropriate alternative to analyze our complex system.

Table 2 shows the differences between both real and estimated gross margin, considering the best PUNN model. Both the range and the variation coefficient decrease slightly in the estimated gross margin series, but results do not differ significantly in the mean (t-test, p-value = 0.402). From a decisional point of view, it is very important to verify that the real variability of the sample remains practically constant in the estimated gross margin range, so the PUNN model recognizes the characteristics of the environment analysed.

The most relevant coefficient in estimated gross margin (PUNN) determination is the constant, which can be considered an estimated gross margin mean estimator. An exception to this profile can be examined in both extremes of the gross margin range (Fig. 2) where the $2\text{nd}(B_2) - \text{CA}^*, P^*, H^*, T^*$ and A^* – and the $3\text{rd}(B_3) - \text{CA}^*, P^*, H^*, T^*, A^*, I^*$ and CP^*-coefficients are dominant. In the rest of the gross margin range $1\text{st}(B_1) - \text{CA}^*$ and P^* – and $2^{nd}(B_2)$ coefficients are dominant but sometimes the third one (B_3) is still important.

5. Relevant findings for decision makers

Once the accuracy of the PUNN algorithm was demonstrated and in order to test the interpretability and the utility of the best model obtained (PUNN, Table 3), two realistic farm and farmer profiles (scenarios) were developed, setting up different variable sets and varying the values of the rest in order to obtain their respective estimated gross margin curve families. These scenarios were developed based on cluster analysis (bottom-up hierarchical analysis and k-means) according to previous analyses of the strategic behaviour of Andalusian olive-tree farms (Fernández et al., 2004).

From a decision maker's point of view, it is more important to predict an accurate gross margin range by using an easy-to-handle methodology that gives interpretable estimated gross margins, than to obtain a perfect mathematical estimated gross margin adjustment to reality. To sum up, is the best PUNN model capable of classifying real farms appropriately according to their gross margin? Or, is the best PUNN model capable of predicting the gross margin correctly using only descriptive variables? If so, a new and interesting tool could be systematically used to predict real gross margin, minimizing periodical adjustments. Taking these questions into account, our best PUNN model can be used:

(i) To determine the relevance of individual descriptive variables in the EGM.
(ii) To estimate farm membership within a gross margin interval, rather than an exact gross margin, in order to determine its relative financial situation. Frequency analysis (quintiles, using real gross margin scores) has been used to check PUNN best model capacity to locate the farm's estimated gross margin accurately.

5.1. Scenario 1: standard olive-tree farms in dry farming located at a medium-low altitude (river valley), $\text{EGM} = f(\text{CA}^*, P^*, \text{FF}^*)$

In this first scenario, a standard olive-tree farm and farmer are considered as follows: the agrarian enterprise is located at medium-low altitude over sea level $(H^* = 0.335, 358.46\,\text{m})$; it has

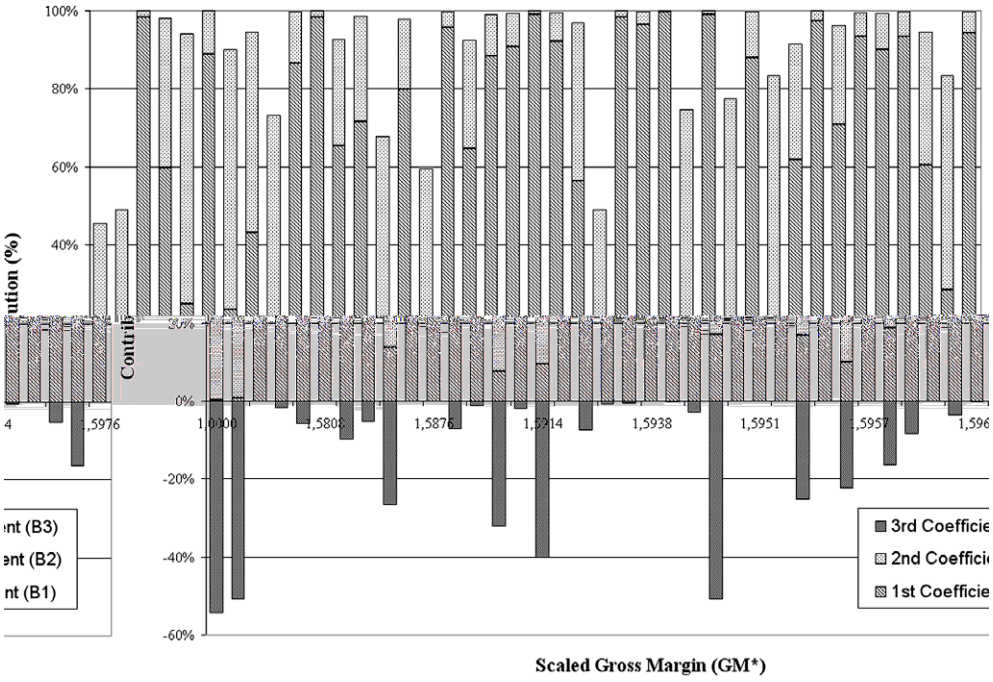

Fig. 2. Relative contributions (%) of different best model coefficients (PUNN) excluding the constant, an example considering only the lower scale estimated gross margin* values [1, 1.5975].

one tractor, ($T^* = 0.233$) which is very common in the Andalusian agrarian environment; the farmer is middle-aged to elderly ($A^* = 0.633$, from 55 to 65 years old), also very normal in the target rural environment; the farm receives non-agrarian income ($I^* = 0.1$); and, finally, the farmer is a member of a cooperative ($CP^* = 0.1$). Considering this agrarian profile, the independent variables considered were: cultivated area (CA^*), olive tree production (P^*) and the percentage of family hand labour (FF^*). In order to avoid undesirable results, a yield constraint (the maximum technical yield for standard olive-tree cultivation was 12 tons of olives per hectare) was also included because a small farm cannot have very large production.

As expected, when FF^* was fixed at 0.1 (non-family hand labour employed on the farm), in small-medium farms ($CA^* \leqslant 0.5$) the greater the production, the greater the EGM. Nevertheless, and surprisingly, in medium-big farms ($CA^* \geqslant 0.6$ approximately) with low production ($P^* \leqslant 0.4$), EGM reaches a maximum (Fig. 3A). This behaviour is a result of a strong interaction between EGM and P^* and CA^* and is reasonable because the greater a farm is, the greater its possibility of obtaining revenues from other agrarian activities (mixed structure). Obviously, the estimated gross margin is not completely due to olive tree production (very small in this case); other crops and livestock as well as other complimentary activities that provide non-agrarian revenues can make a contribution.

As soon as the production grows (except in the very big farms mentioned in the previous paragraph), the shape of the estimated gross margin surfaces changes. All of them have a minimum that identifies a critical CA^*. Smaller sized farms with medium and high production have a very stable financial situation. Medium sized farms cannot maintain the same EGM level as those obtained by the smaller ones because of higher structural and cultivation costs. In bigger farms, due to scale economies and a greater capacity for diversification, the estimated gross margin increases a bit at the end of the CA^* range (Fig. 3A). The EGM has a minimum that depends on P^* and is located in the smaller olive-tree farms.

Family hand labour has a very important and negative incidence on EGM in medium-sized olive-tree farms (Fig. 3B). Increasing family participation (valued at market price) in olive-tree farm cultivation lowers EGM, which could jeopardise financial stability. The decision maker does not usually calculate the worth of his family's and his own work in the agrarian enterprise so, for him, the EGM should show a more profitable situation. The inexistence of family hand labour on the farm ($FF^* = 0.1$) shows a more business-oriented productive structure. Smaller sized farms have a chance to survive due to favourable EGM expectations, which also occurs when the family has other income sources.

5.2. Scenario 2: standard olive-tree farms in dry farming where family hand labour is a relevant production factor (FF = 50%) and where the income source and farmer cooperative affiliation are analysed, $EGM = f(CA^*, P^*, I^*, CP^*)$

Most Andalusian olive-tree farms receive non-agrarian incomes (I) coming from other economic sectors, and totally or partially hand over their production to a cooperative (CP). In order to calibrate the impact of both variables in EGM evolution, a second farmer and farm profile was designed as follows: family hand labour ($FF^* = 0.5, 50\%$); average altitude ($H^* = 0.6, 689.375$ m); one tractor ($T^* = 0.233$); and, finally, the farmer is middle-aged to elderly ($A^* = 0.633$, 55–65 years old). Now the independent variables were: cultivated area ($CA^*, [0.1, 0.9] <> [0.35$ ha, 240 ha$]$), production [$P^* = 0.6$ (233.785 ton) and $P^* = 0.8$ (327.215 ton)], non-agrarian income [$I^* = 0.1$ (has non-agrarian income?: Yes) and $I^* = 0.9$ (has non-agrarian income?: No)] and cooperative membership [$CP^* = 0.1$ (is a cooperative member?: Yes) and $CP^* = 0.9$ (is a cooperative member?: No)]. Both family curves show (Fig. 4) that cooperative membership is a positive factor in EGM which is more important than the reception of non-agrarian revenues, also positive. These positive effects are more relevant in medium-sized to bigger farms because both variables (I^* and CP^*) appear only in the 3rd(B_3) coefficient of the PUNN model. The EGM in smaller to medium-sized farms (from $CA^* = 0.1$ to $CA^* = 0.5$), considering the defined scenario, is not sensitive to I^* and CP^* variation, but the farm needs non-agrarian revenues to balance its financial situation because of farm size and also because cooperative membership solves both its transformation and commercialisation problems.

5.3. Gross margin prediction and mixed farms

By dividing both the original gross margin and estimated gross margin samples into quintiles (Table 5), the farms surveyed can be classified into five intervals: farms with a very small gross margin

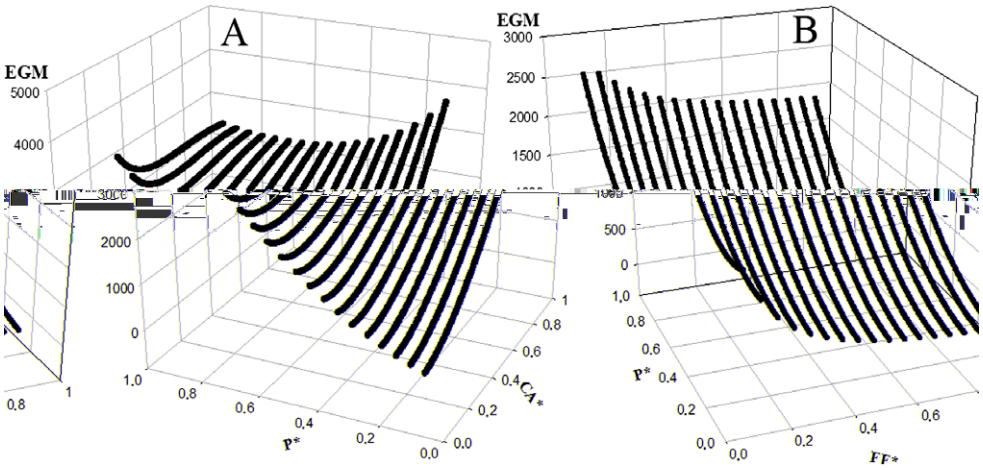

Fig. 3. Estimated gross margin (10^2 €) evolution taking into consideration cultivated area (CA^*) [0.1,0.9], production (P^*) [0.1,0.9] and family hand labour (days worked, %) over total hand labour in the farm (FF^*) [0.1,0.9] (to facilitate curve comprehension, the scaling process of estimated gross margin – EGM – has been reversed). (A) $FF^* = 0.1$ (non-family hand labour employed on the farm). (B) $CA^* = 0.5$ (medium-size farms). $H^* = 0.335(358.46$ m$)$, $T^* = 0.233$ (1 tractor), $A^* = 0.633$ (3 <> 55 to 65 years old), $I^* = 0.1$ (does the farmer have non-agrarian income?: Yes) and $CP^* = 0.1$ (is he a cooperative member?: Yes).

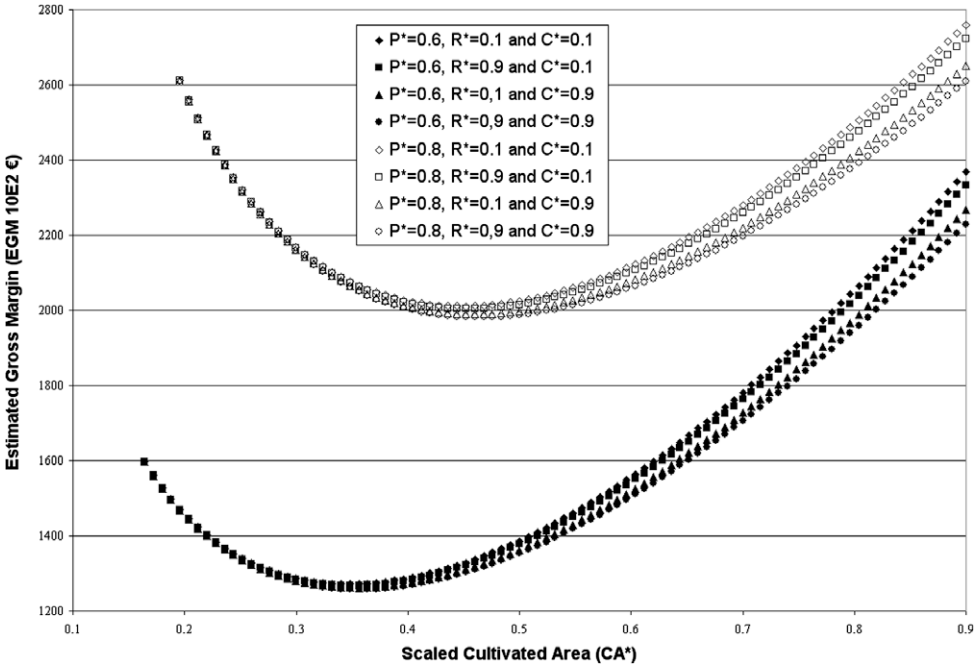

Fig. 4. Estimated gross margin (10^2 €) evolution considering cultivated area (CA^*) [0.1,0.9], production (P^*) – $P^* = 0.6$ and $P^* = 0.8$, does the farm receive non-agrarian income? (Yes – $I^* = 0.1$- or No – $I^* = 0.9$) and does the farmer give its production – all or partially – to a cooperative? (Yes – $CP^* = 0.1$ – or No – $CP^* = 0.9$). The rest of the variables were fixed as follows: $FF^* = 0.5, H^* = 0.6, T^* = 0.233$ and $A^* = 0.633$ (to facilitate curve comprehension, the scaling process of estimated gross margin has been reversed). CA^*: scaled cultivated area [0.1,0.9] <> 0.35 ha,240 ha]. P^*: scaled production. $P^* = 0.6$ (233.785 ton) and $P^* = 0.8$ (327.215 ton). $I^* = 0.1$ (has the farmer non-agrarian revenues?: Yes) and $I^* = 0.9$ (has he non-agrarian revenues?: No). $CP^* = 0.1$ (is he a cooperative member?: Yes) and $CP^* = 0.9$ (is he a cooperative member?: No). $FF^* = 0.5(50\%), H^* = 0.6(689.375\ \text{m}), T^* = 0.233$ (1 tractor), $A^* = 0.633$ (3 <> 55–65 years old).

Table 5
Probability of both real and estimated gross margin – the PUNN best model-being in the same quintile and having the same probability of being shifted to another quintile.

Quintile defined by gross margin	#Obs.	Probability of being in the same quintile	Probability of being shifted to an adjacent quintile	Probability of being shifted two or more quintiles	# of crops: mean and (standard deviation)[a]
q_1	41	0.4878	0.1951	0.3171	1.69 (1.11)
q_2	42	0.3810	0.5476 (0.3810 to $q.1$)	0.0714	1.67 (1.15)
q_3	42	0.4286	0.4524 (0.2857 to $q.2$)	0.1190	1.60 (0.89)
q_4	42	0.4048	0.4524 (0.2619 to $q.3$)	0.1429	2.00 (2.00)
q_5	41	0.6585	0.2195	0.1220	1.60 (1.34)

#Obs.: number of observations. #: number. $q.$: quintile, gross margin 10^2 € (q_1: first quintile [−4649.8,7.036], q_2: second quintile [7.036,29.728], q_3 third quintile [29.728,79.572], q_4 fourth quintile [79.572,261.581] and q_5 fifth quintile [261.581,3139.45]).
[a] Considering only the observations (farms) shifted two or more quintiles.

(less than 7.04×10^2 €), farms with a small gross margin (from 7.04×10^2 € to 2.993×10^3 €), intermediate farms (gross margin from 2.993×10^3 € to 7.957×10^3 €), farms with a large gross margin (from 7.957×10^3 € to 2.616×10^4 €) and farms with a very large gross margin (over 2.616×10^4 €). These intervals shift a little bit if the estimated gross margin is considered, but this displacement is very small. Dividing both the gross margin and estimated gross margin ranges according to these quintile limits, the probability that an observation – a farm – will shift more than one interval (two or more quintiles) is low. Only farms with a very small gross margin (first quintile, Q_1) show a relatively high probability of being misclassified (0.32). In the rest of the intervals, this probability is always under 0.15, therefore the best model's (PUNN) ability to predict gross margin as well as to classify the olive-tree farms in terms of estimated gross margin is very good (Pearson's χ^2 p-value: 0.000). On average, the number of crops in misclassified farms is greater but not significantly different (one way ANOVA) than the original 1.35 crops per farm, so it is not possible to state that the gross margins of mixed farms are more difficult to predict.

6. Conclusions

The results (estimated gross margin) obtained from the best model (PUNN, Table 3) can be considered to be very reasonable from a technical and economic point of view. PUNN can be considered an alternative methodology along with MLP. In agrarian production (olive-tree farms in dry farming), it is very complicated to determine the exact income of rural families or farms except by using very expensive and unwieldy methodologies. Based on neural networks, the model obtained (PUNN), using easy-to-collect descriptive and structural variables, can determine an accurate estimated gross margin of the farms and classify an agrarian enterprise according to its gross margin. This classification is essential in sustainability studies. Obviously, the model obtained has to be adjusted periodically because of environmental evolution but this process will still be cheaper and easier to manage than, for example, a massive survey. From a mathematical point of view, the best PUNN function selected can model unreal situations that need to be constrained. It is very rare, for example, to obtain elevated production levels in high mountain farms where olive trees cannot

easily survive. Likewise, it is also very unlikely to find high production levels in small farms (except in high density plantations in irrigated farming). Therefore, the formula obtained (best model) must be interpreted carefully and filtered by designing appropriate technical constraints.

In all of the farmer and farm profiles analysed, the estimated gross margin evolution – curve families – could be explained from a technical point of view. Except for AS (agrarian studies on the part of the farmer), which was removed from the final model, all independent descriptive and structural variables were included in the best PUNN model obtained. Each one's individual relevance is different and must be analysed in detail according to realistic scenarios based on expert knowledge.

The PUNN models always included CA (cultivated area) and *P* (olive-tree production); CP (cooperative membership) and *I* (non-agrarian incomes) were also considered very relevant and were included, respectively, in nine and in eight of the 10 partitions carried out. In spite of the variability of the sample that provides different algebraic model structures, there is a basic nucleus of independent variables (CA, *P*, CP, *I* and, maybe, *A*) that explain farm gross margin in a very complex environment.

The best PUNN model obtained can be used to predict gross margin and classify agrarian enterprises, and can also be a mathematical function to be maximised in a multi-criteria nonlinear model incorporating appropriate technical constraints.

PUNN models give the decision maker a relevant ranking of independent variables and show the interactions among them. In addition, PUNNs allow the most appropriate models to be selected when a combination of specific independent variables must be analyzed. Therefore, they offer different views of the same reality (represented by different combinations of independent variables) which can be consulted by decision makers who must depend on real circumstances.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Spanish Inter-Ministerial Commission of Science and Technology (MICYT) (TIN 2008-06681-C), FEDER funds and the P08-TIC-3745 Project of the "Junta de Andalucía" (Spain).

References

Amores, A.F., Contreras, I., 2009. New approach for the assignment of new European agricultural subsidies using scores from data envelopment analysis. Application to olive-growing farms in Andalusia (Spain). European Journal of Operational Research 193, 718–729.

Angeline, P.J., Saunders, G.M., Pollack, J.B., 1994. An evolutionary algorithm that constructs recurrent neural networks. IEEE Transactions on Neural Networks 5 (1), 54–65.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, UK.

Durbin, R., Rumelhart, D., 1989. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. Neural Computation 1, 133–142.

European Commission (EU), 1999. Agriculture in the European Union – Statistical and Economic Information. <http://www.europa.eu.int/comm/agriculture/agrista/>.

Fernández, L., Carbonero, M., García-Alonso, C.R., Martín, J.M., Pérez, P.P., 2004. Strategic groups in andalusian olive tree farms. Acta Horticulturae 655, 209–217.

Fogel, D.B., 1991. Systems Identification Simulated Evolution. A Machine Learning Approach. Gim Press.

García-Alonso, C.R., Guardiola, J., Hervás, C., 2009. Logistic evolutionary product-unit neural networks: Innovation capacity of poor Guatemalan households. European Journal of Operational Research 195 (2), 543–551.

García, N., Hervás, C., Muñoz, J., 2002. Multiobjective cooperative coevolution of artificial neural networks. Neural Networks 15 (10), 1255–1274.

García, N., Hervás, C., Muñoz, J., 2003. Covnet: A cooperative coevolutionary model for evolving artificial neural networks. IEEE Transactions on Neural Networks 14 (3), 575–596.

Gatu, C., Kontoghiorghes, E., 2006. Branch-and-bound algorithms for computing the best-subset regression models. Journal of Computational and Graphical Statistics 15 (1), 139–156.

Goldberg, D.E., Deb, K., Clark, J.H., 1992. Genetic algorithms, noise, and the sizing of populations. Complex Systems 6, 333–362.

Grznar, J., Prasad, S., Tata, J., 2007. Neural networks and organizational systems: Modeling non-linear relationships. European Journal of Operational Research 181 (2), 939–955.

Gurney, K.N., 1992. Training nets of hardware realizable sigma–pi units. Neural Networks 5 (2), 289–303.

Gutiérrez, P.A., Hervás, C., Carbonero, M., Fernández, J.C., 2009. Combined projection and kernel basis functions for classification in evolutionary neural networks. Neurocomputing 72, 2731–2742.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Macmillan, New York.

IEA (Instituto de Estadística de Andalucía), 2000. Anuario de Estadísticas Agrarias y Pesqueras de Andalucía. <http://www.juntadeandalucia.es/agriculturaypesca/publico/>.

Inza, I., Merino, M., Larrañaga, P., Quiroga, J., Sierra, B., Girala, M., 2001. Feature subset selection by genetic algorithms and estimation of distribution algorithms – A case study in the survival of cirrhotic patients treated with tips. Artificial Intelligence in Medicine 23 (2), 187–205.

Ismail, A., Engelbrecht, A.P., 1993. Training products units in feedforward neural networks using particle swarm optimisation. In: Proceedings of the International Conference on Artificial Intelligence (ICAI99), Durban, South Africa, pp. 36–40.

Ismail, A., Engelbrecht, A.P.,2000. Global optimization algorithms for training product unit neural networks. In: IJCNN'2000 International Joint Conference on Neural Networks, vol. 1, IEEE Computer Society, Los Alamitos, CA.

Ismail, A., Engelbrecht, A.P., 2002. Pruning product unit neural networks. In: IEEE International Joint Conference on Neural Networks, Honolulu, Hawai.

Janson, D.J., Frenzel, J.F., 1993. Training product unit neural networks with genetic algorithms. IEEE Expert 8 (5), 26–33.

Kaul, M., Hill, R.L., Walthall, C., 2005. Artificial neural networks for corn and soybean yield prediction. Agricultural Systems 85, 1–18.

Kohavi, R., Sommerfield, D., 1995. Wrappers for feature subset selection. Artificial Intelligence 97 (1-2), 273–324.

Liu, H., Motoda, H., 1998. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, Boston.

Liu, H., Motoda, H., 2002. On issues of instance selection. Data Mining and Knowledge Discovery 6, 115–130.

Lütz, M., Bastian, O., 2002. Implementation of landscape planning and nature conservation in the agricultural landscape –A case study from saxony. Agriculture Ecosystems and Environment 92, 159–170.

Maniezzo, V., 1994. Genetic evolution of the topology and weight distribution of neural networks. IEEE Transactions on Neural Networks 5, 39–53.

MAPA (Ministerio de Agricultura, Pesca y Alimentación), 2003. Anuario de Estadística Agroalimentaria, Ministerio de Agricultura, Pesca y alimentación, Madrid.

Martínez, A.C., Martínez, F., Hervás, C., García, N., 2006. Evolutionary product unit

Van Rooij, A.J.F., Jain, L.C., Johnson, R.P., 1996. Neural networks training using genetic algorithms. Machine Perception and Artificial Intelligence, vol. 26. World Scientific, Singapore.

Williams, P.M., 1994. Bayesian regularization and pruning using a Laplace prior. Neural Computation 7, 117–143.

Wong, M.L.D., Nandi, A.K., 2004. Automatic digital modulation recognition using artificial neural network and genetic algorithm. Signal Processing 84 (2), 351–365.

Yao, X., Liu, Y., 1997. A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks 8 (3), 694–713.

Yohai, V.J., Zamar, R.H., 1997. Optimal locally robust *M*-estimates of regression. Journal of Statistical Planning and Inference 57, 73–92.

Yohai, V.J., Stahel, W.A., Zamar, R.H., 1991. A procedure for robust estimation and inference in linear regression. In: Stahel, W.A., Weisberg, S.W. (Eds.), Directions in Robust Statistics and Diagnostics, Part II. Springer-Verlag, New Work.