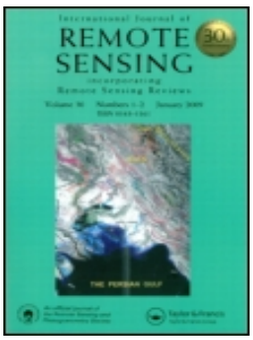

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

A novel self-organizing neuro-fuzzy multilayered classifier for land cover classification of a VHR image

N. E. Mitrakis , C. A. Topaloglou , T. K. Alexandridis , J. B. Theocharis & G. C. Zalidis

To cite this article: N. E. Mitrakis , C. A. Topaloglou , T. K. Alexandridis , J. B. Theocharis & G. C. Zalidis (2008) A novel self-organizing neuro-fuzzy multilayered classifier for land cover classification of a VHR image, International Journal of Remote Sensing, 29:14, 4061-4087, DOI: 10.1080/01431160801891846

To link to this article: https://doi.org/10.1080/01431160801891846

Published online: 14 Jun 2008.

International Journal of Remote Sensing
Vol. 29, No. 14, 20 July 2008, 4061–4087

Taylor & Francis
Taylor & Francis Group

A novel self-organizing neuro-fuzzy multilayered classifier for land cover classification of a VHR image

N. E. MITRAKIS†, C. A. TOPALOGLOU‡, T. K. ALEXANDRIDIS‡, J. B. THEOCHARIS*† and G. C. ZALIDIS§

†Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

‡Laboratory of Remote Sensing and GIS, Faculty of Agronomy, Aristotle University of Thessaloniki, Thessaloniki, Greece

§Laboratory of Applied Soil Science, Faculty of Agronomy, Aristotle University of Thessaloniki, Thessaloniki, Greece

(*Received 7 February 2007; in final form 4 January 2008*)

A novel self-organizing neuro-fuzzy multilayered classifier (SONeFMUC) is introduced in this paper, with feature selection capabilities, for the classification of an IKONOS image. The structure of the proposed network is developed in a sequential fashion using the group method of data handling (GMDH) algorithm. The node models, regarded as generic classifiers, are represented by fuzzy rule-based systems, combined with a fusion scheme. A data splitting mechanism is incorporated to discriminate between correctly classified and ambiguous pixels. The classifier was tested on the wetland of international importance of Lake Koronia, Greece, and the surrounding agricultural area. To achieve higher classification accuracy, the image was decomposed into two zones: the wetland and the agricultural zones. Apart from the initial bands, additional input features were considered: textural features, intensity–hue–saturation (IHS) and tasseled cap transformation. To assess the quality of the suggested model, the SONeFMUC was compared with a maximum likelihood classifier (MLC). The experimental results show that the SONeFMUC exhibited superior performance to the MLC, providing less confusion of the dominant classes in both zones. In the wetland zone, an overall accuracy of 89.5% was attained.

1. Introduction

Land cover classification of remotely sensed images has attracted considerable research interest over the past decades. Along with the numerous applications in the field, several problems have been reported that reduce the accuracy and reliability of the resulting thematic maps. The presence of mixed pixels, the resolution of the acquired images, the reliability of training data, the number of classes and the high degree of spectral overlapping between the classes are key reasons for achieving low classification accuracies.

To tackle the problems encountered in land cover image classification, the research community has turned primarily to two major areas of interest. The first involves the enhancement of features used by the classification algorithm, as applied to the satellite image. Although original bands of satellite sensors remain the basic

*Corresponding author. Email: theochar@eng.auth.gr

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160801891846

source of information in multispectral image classification, advanced features such as topographic information (Richards *et al.* 1982), tasseled cap features (Oetter *et al.* 2001), textural analysis (Haralick and Shapiro 1992) and wavelet decomposition (Dekker 2003) have been developed to reduce the overlapping of the classes in the original feature space. However, the use of new features results in complex models because of the large dimension of the feature space. To arrive at the appropriate feature set, several techniques make use of preprocessing methods, such as principal components analysis (PCA; Li and Yeh 1998) and the time-consuming method of trial-and-error (Shackelford and Davis 2003).

Second, more sophisticated classifiers are found to be a major factor in improving the classification results. The first type of classifiers used mainly statistical parameters (Thomas *et al.* 1987), considered as 'hard' classifiers, such as the maximum likelihood classifier (MLC). However, the assumption of a normal distribution of the data is a major drawback of such classifiers.

The remarkable achievements in the development of fuzzy classifiers, regarded as 'soft' classifiers, provide a fruitful approach (Wang 1990, Bardossy and Samaniego 2002). These classification techniques were found to be more appropriate in tackling the mixed pixels problem (Tso and Mather 2001) because they take into consideration the ambiguities concerning the correct class to which a pixel belongs. However, Foody (1999) proposed that, to resolve the mixed pixel problem, a continuum of classification fuzziness should be defined, so that not only the classifier but also the training and testing stages should be fuzzy.

Another promising type of classifier is derived from the theory of neural networks (Foody 1995, Kavzoglu and Mather 2003). Neural networks are capable of dealing with complex problems, with a high degree of overlapping between the classes, by conducting a non-linear transformation of the original feature space through the layers of the network. Along this direction, a great variety of neural network architectures and training algorithms have been reported in the literature, with productive results (Atkinson and Tatnall 1997, Keramitsoglou *et al.* 2005). Nevertheless, most of the classifiers of this type require the definition of a large number of parameters, such as the number of hidden layers, the number of nodes, the range of the initial weights and the number of training data, which are difficult to decide on. To cope with this problem, Kavzoglu and Mather (2003) proposed a set of guidelines by conducting an extensive search of suggestions reported in the literature. In recent years a considerable amount of research has focused on the development of classifiers combining fuzzy logic and neural networks, thus resulting in superior classification methods (Carpenter *et al.* 1997, Lin *et al.* 2000) for solving remote sensing problems. An alternative approach is to combine different classifier types (Kuncheva *et al.* 2001), by exploiting the best attributes of every classifier. In particular, in land cover classification, Giacinto and Roli (1997) proposed a method of combining the classification results of statistical and neural network classifiers using a modified *k*-nearest-neighbour rule as a metaclassifier. Briem *et al.* (2002) studied the application of boosting, bagging and consensus theory to derive multiple classifiers that outperformed the single classifiers on multisource remote sensing data.

The potential, however, of using a cascaded architecture or a sequential combination of individual classifiers has not been exploited enough, as most of these architectures are based on a hierarchical combination of classifiers (Giacinto and Roli 1997, Kumar *et al.* 1997, Briem *et al.* 2002). Wilkinson *et al.* (1995)

proposed a voting/rejection approach, where an MLC was used in parallel with a neural network in the first stage. In the second stage, an independent neural network was trained to classify the ambiguous pixels of the first stage. Benediktsson and Kanellopoulos (1999) modified the voting/rejection of Wilkinson *et al.* (1995) by using a multisource classifier based on consensus theory instead of an MLC, and used the decision boundary feature extraction (DBFE) method as a tool for reducing the large dimensionality of the input space for hyperspectral data. Although the results showed that the logarithmic opinion pool (LOGP) classifier achieved the best performance, the authors stressed the promising attributes and results of the voting/rejection approach. However, the neural network in the second stage of the algorithm has to be supplied with a large number of training data to achieve a satisfactory classification performance. Hence, the efficiency of the method is highly dependent on the degree of agreement of the classifiers in the first stage of the classification algorithm (Petrakos *et al.* 2001). Moreover, the use of a feature extraction algorithm as a preprocessing tool increased the computational cost.

The aim of the current study was to develop a novel self-organizing neuro-fuzzy multilayered classifier (SONeFMUC) with feature selection attributes, gradually evolved in a self-organizing manner, using the group method of data handling (GMDH) algorithm (Ivakhnenko 1968). Specific objectives include (i) the application of the SONeFMUC to a very high resolution (VHR) IKONOS image using the initial bands and advanced features for land cover classification of a protected wetland and its surrounding agricultural area, and (ii) the validation of the SONeFMUC by comparison with an MLC.

2. Materials and methods

2.1 *Study area*

Lake Koronia is located in a tectonic depression in northern Greece (40′ 41′′ N, 23′ 09′′ E). Its watershed covers an area of 780 km^2, and it formerly drained eastwards into Lake Volvi, then into the sea. The climate of the region is transitional between Mediterranean and temperate. The mean annual precipitation is 455 mm, with a seasonal peak in December and a minimum in August (Mitraki *et al.* 2004). The lake–wetland ecosystem is surrounded by an intensively cultivated agricultural area (figure 1). The dominant agricultural crops are maize, alfalfa and cereals. There is no exploitation of surface water, and the only source of fresh water for irrigation, industrial and urban use is through groundwater resources. Irrigated agriculture is an important economic activity in the area, but the recent development of numerous pump wells has resulted in depletion of the aquifer, and a subsequent decrease in the lake's water level. The industrial sector has also increased in the past decade, discharging untreated effluents in the lake from fabric dyeing, food and dairy processing activities (Zalidis *et al.* 2004).

Because of the above-mentioned pressures, Lake Koronia became progressively more eutrophic, especially after the early 1990s, and is currently hypertrophic (Mitraki *et al.* 2004). Along with the drastic alteration in the water level, which reached a decrease of 80%, the natural ecosystem has suffered severe degradation. There has been a significant loss of volume and habitat heterogeneity in the lake and wetland. The emergent macrophyte community (dominated by *Phragmites australis*) has shifted lakewards and expanded on a recently exposed lake bed (Alexandridis

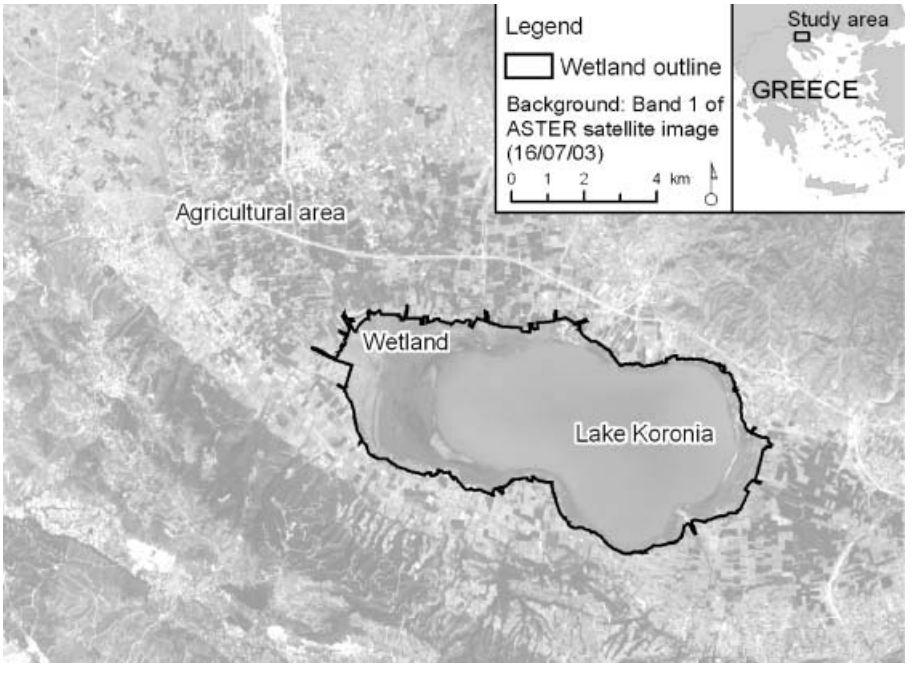

Figure 1. Location map and main land cover of the study area.

et al. 2007). In recognition of its ecological importance, and to prevent further degradation, the lake–wetland system of Lakes Koronia and Volvi is protected by a number of legal and binding actions: it is a Wetland of International Importance according to the Ramsar Convention (site code 57, area 163.88 km^2), a Special Protected Area designated by the implementation of European Directive 79/409/EEC (site code GR1220009, area 156.71 km^2), and a Site of Community Importance following the implementation of the European Habitat Directive 92/43/EEC (site code GR1220001, area 269.47 km^2). The relevant national and local authorities have responded with the identification and mapping of habitats (Hellenic Ministry of Environment 2001), the compilation of the Master Plan for the restoration of Lake Koronia, and the Revised Restoration Plan for Lake Koronia (Zalidis *et al.* 2004). However, continuous monitoring of natural and agricultural environment is required according to the proposed management actions, and to fulfil the obligations to the international and European legislation.

2.2 *Dataset used*

An IKONOS bundle image with a 1 m spatial resolution panchromatic image and a 4 m multispectral (three visible and one near-infrared) image covered 134 km^2 of the study area. The image was acquired on 7 August 2005, was clear from clouds and was acquired at a nadir view angle to minimize noise reflectance from topographic effects. A digital elevation model (DEM) of the area was used for orthocorrection of the satellite image. It was produced at 5 m pixel size by contours with a height interval 4 m that were digitized from topographic maps of scale 1 : 5000 using the ANUDEM interpolation method (Hutchinson 1988).

An extensive field survey was conducted at the first days of September 2005 to identify land cover classes that referred mainly to the agricultural and wetland areas, and to collect training and testing samples for the image classification and its accuracy, respectively. Using a Global Positioning System (GPS) receiver embedded in a palm top, 3920 locations were selected at regular intervals along the agricultural road network. The land cover on these locations was identified by visual inspection and subsequently divided into 13 classes. The classification scheme included six crop types, five wetland habitats (following Annex I of Habitats Directive 92/43/EEC) and two ancillary land cover types (following the CORINE Land Cover nomenclature). The sampled points were separated into two sets using the random stratified method: the training set (70%) and the test set (30%).

2.3 *Manipulation of the dataset*

The IKONOS image was orthorectified in the Greek Geodetic Reference System with the use of the DEM and ground control points that were collected from black and white photomaps of 1 m pixel size. The root mean square (RMS) errors were 0.92 m and 2.41 m for panchromatic and multispectral images, respectively. A pan-sharpened image was produced using forward–reverse principal components transforms (Chavez *et al.* 1991), with the panchromatic image replacing the first principal component. This new image was useful for evaluating and minimizing the bias in the samples that were selected in the fieldwork. Atmospheric correction was not applied as it has little effect on classification accuracy when single dates of remotely sensed data are to be classified, as long as the training data from the image to be classified have the same relative scale (corrected, uncorrected) (Kawata *et al.*1990, Song *et al.* 2001).

Advanced features from the multispectral image were calculated using the initial four bands of the image. These features can be categorized into two groups: textural and spectral features. Textural analysis using grey level co-occurrence matrices (GLCMs; Haralick and Shapiro 1992) is common practice in land cover image classification (Lin *et al.* 2000) to decrease the degree of overlap between the various types of classes. The image is raster scanned with sliding windows of $M \times M$ dimensions. A GLCM for each window is calculated, indicating how often different grey levels (i, j) occur with a specific direction ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$) and distance ($d$) between the pixel centres. Assuming that G grey levels occur within the image, a $G \times G$ matrix is computed, with the (i, j)th element of the matrix given as

$$p(i,j) = f_{ij}^{d,\theta} \Big/ \sum_{i}^{N} \sum_{j}^{N} f_{ij}^{d,\theta} \tag{1}$$

where i and j refer to the rows and columns of the matrix, respectively, $f_{ij}^{d,\theta}$ is the frequency of occurrence of grey levels (i, j) separated by a distance d and a direction θ, and N is the total number of pixels in the window for a particular value of d. Originally, 16 measures were proposed that derived from each co-occurrence matrix. Among them, four are considered to be the most important: Contrast, Angular Second Moment (ASM), Correlation and Homogeneity. They are calculated as follows:

$$Contrast = \sum_{i=1}^{G} \sum_{j=1}^{G} |i-j|^2 \, p(i,j) \tag{2}$$

which is a measure of the local variations between a pixel and its neighbours,

$$\textit{Angular}\ \text{Second Monment} = \sum_{i=1}^{G}\sum_{j=1}^{G} p(i,j)^2 \tag{3}$$

which is a measure of uniformity,

$$\textit{Correlation} = \sum_{i=1}^{G}\sum_{j=1}^{G} (i-\mu)(j-\mu)p(i,j)/\sigma^2 \tag{4}$$

which is a measure of how correlated is a pixel to its neighbour, and

$$\textit{Homogeneity} = \sum_{i=1}^{G}\sum_{j=1}^{G} p(i,j)/(1+|i-j|) \tag{5}$$

which measures the closeness of the distribution of elements in the GLCM to its diagonal. The above textural analysis was applied to the four bands of the IKONOS image, providing us with a total of 16 features.

Two different colour spaces were produced from the initial bands and served as input data in the image classification. The first uses intensity (I), hue (H) and saturation (S) as the three positioned parameters (in lieu of R, G and B). This is an advantageous system as it presents colours that are closer match to human colour perception. The intensity represents the total amount of light in a colour, the hue is the property of the colour determined by its wavelength, and the saturation is the purity of the colour (Zhang and Hong 2005). For the calculation of IHS transformation only three bands are needed, so a pseudo-colour RGB composite of IKONOS using channels four, three and two respectively was used.

A second colour space, which is a linear transformation of the four multispectral bands of the IKONOS image, was applied that offers a means to optimize information extracted for vegetation studies, associated with the classes of interest. This transformation is called the tasseled cap (Kauth and Thomas 1976) and has produced three data structure axes that define the vegetation information: Brightness, Greenness and Wetness. These new features are sensor dependent, and until recently they were only available for Landsat 5 and 7 images. Horne (2003), using approximately 200 different scenes of IKONOS images globally, managed to derive tasseled cap coefficients for IKONOS images.

3. Development of the SONeFMUC

3.1 *The SONeFMUC architecture*

The suggested neuro-fuzzy classifier is a multilayered structure, as depicted in figure 2. The network consists of ℓ layers ($\ell=0,\ldots,M$), with the ℓth layer including N_ℓ neurons. The neurons are defined as fuzzy neuron classifiers (FNCs), denoted as $\text{FNC}_j^{(\ell)}$, where $j=0,\ldots,N_\ell$ and $\ell=0,\ldots,M$. The input layer $\ell=0$ includes the m network inputs, x_1, x_2, …, x_m, representing the feature components, while the output layer $\ell=M$ comprises the output node $\text{FNC}_1^{(M)}$, providing the overall network's decision. The neuron models in each layer are regarded here as generic local classifiers, working in a subregion of the feature space and represented by fuzzy rule-based systems. Parent FNCs at each layer are combined to generate a descendant FNC at the next layer, with better classification capabilities. The generic

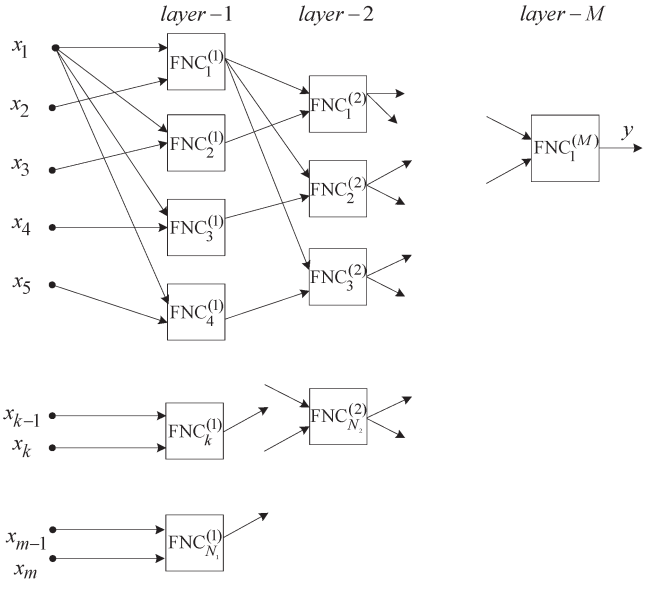

Figure 2. A general SONeFMUC architecture.

node classifiers perform successive feature transformations and decisions, using the distributed fuzzy rule bases and the fuzzy reasoning approach. Hence, the classification task is achieved sequentially by the FNCs arranged along the layered structure of the SONeFMUC.

In a supervised learning scheme, the design of the classifier is based on a set of classified examples, used to establish the association between the pattern attributes and the class labels. Consider that the patterns are distributed over a set of M disjoint classes. Let us assume a training data set comprising N input–output observation pairs: $D_N = \{(x[q], C[q]), q = 1, \ldots, N\}$, where $x[q] = [x_1[q], \ldots, x_m[q]]^T$ denotes the vector of feature components, T denotes the transpose vector, $C[q]$ is the class label for the qth observation, and $C = \{C_1, \ldots, C_M\}$ is the set of classes. For convenience, the features are normalized in the range [0,1], forming the feature space $F = [0,1]^n$.

Unlike conventional classifiers that assume a given pattern to a single class, the suggested SONeFMUC model is a fuzzy classifier performing a map $F \rightarrow [0,1]^M$. Accordingly, they produce a decision output vector $D(x) = [d_1(x), \ldots, d_M(x)]^T$ embracing all classes, where $d_j(x) \in [0,1]$, $j = 1, \ldots, M$ represents the grade of certainty in the assertion that pattern $\boldsymbol{x}$ belongs to class C_j. Based on the classification data, the structure of the SONeFMUC is progressively expanded in layers using a structure learning algorithm.

3.2 *FNC models at the first layer*

The structure of the $\mathrm{FNC}_j^{(1)}$ at layer 1 is shown in figure 3. The generic model consists of two fuzzy rule based-systems: the so-called fuzzy partial description (FPD) and a decision making fuzzy unit (DMFU). Therefore, the FNCs at the first layer are represented as a pair of modules: $\mathrm{FPD}_j^{(1)} = \left\{\mathrm{FPD}_j^{(k)}, \mathrm{DMFU}_j^{(1)}\right\}$, $j = 1, \ldots, N_1$. In the following, we describe the functions of the constituent units.

3.2.1 Fuzzy partial description (FPD). The FPDs are represented by fuzzy rule-based Takagi Sugeno Kang (TSK) systems (Takagi and Sugeno 1985). Instead of

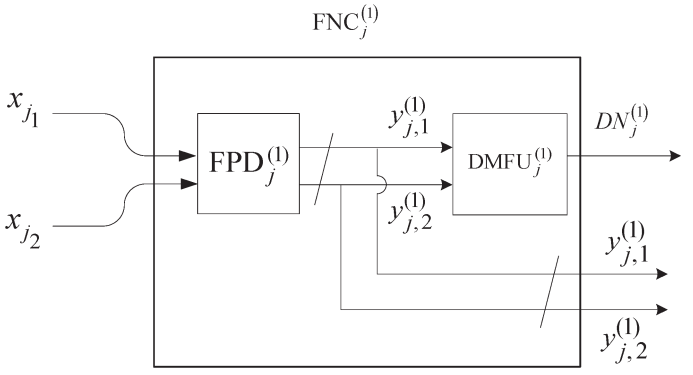

Figure 3. An example of the structure of a node in the first layer of the network. The node incorporates two inputs, two continuous outputs, an FPD unit and a DMFU.

receiving the entire attribute set $\{x_1, x_2,\ldots,x_m\}$, the input vector of $\mathrm{FNC}_j^{(1)}$, denoted by $x_j^{(1)}$, contains a small subset of p features taken from the above set: $x_j^{(1)} = [x_{j1},x_{j2},\ldots,x_{jp}]^T \in \Re^p$. The feature subsets associated with each FPD are derived through recombinations of the original features, a task achieved during the structure learning process. The number of inputs (p) exciting the FPDs at the first layer is a design parameter specified by the user.

The TSK systems considered in this paper provide two outputs. The output vector is denoted as $y_j^{(1)} = \left[y_{j,1}^{(1)},y_{j,2}^{(1)}\right]^T$, where the output variables are normalized in the range $y_{j,k}^{(1)} \in [0,1]$, $k=1,2$. The output space of $\mathrm{FNC}_j^{(1)}$ formed by the components $y_{j,1}^{(1)}, y_{j,2}^{(1)}$ coincides with the class space; that is, the space where the classes are defined. Definition of the classes relates to the class targeting issue. To this end, a target value for both output variables is assigned for each class: $y_d^{(C_j)} = \left[y_{d,1}^{(C_j)},y_{d,2}^{(C_j)}\right]^T$, C_j, $j=1,\ldots,M$. The class ordering is accomplished by following a heuristic scheme determined by the designer. Considering two model outputs for the FPDs provides flexibility for arranging the classes, based on their relative distance within the feature space. Furthermore, an increased number of classes can be considered, allowing SONeFMUC models to handle multiclass problems effectively.

For the sake simplicity, in the remainder of this section we assume that the input and output vector of the FPDs are denoted by $x = [x_1,x_2,\ldots,x_p]^T$ and $y = [y_1,y_2]^T$, respectively. Each premise variable x_i ($i=1,\ldots,p$) is described by two-sided Gaussian fuzzy sets $A_j^{(i)} = \left\{A_{j,L}^{(i)},A_{j,R}^{(i)}\right\}$ ($j=1,\ldots,K_i$) located at a centre value $m_j^{(i)} \in X_i$. The membership functions of the left- and right-hand parts are given by

$$\mu_{j,\mathrm{L}}^{(i)}(x_i) = \exp\left[-\frac{\left(x_i - m_j^{(i)}\right)^2}{\left(\sigma_{j,L}^{(i)}\right)^2}\right], \text{ for } x_i \le m_j^{(i)},$$
$$\mu_{j,\mathrm{R}}^{(i)}(x_i) = \exp\left[-\frac{\left(x_i - m_j^{(i)}\right)^2}{\left(\sigma_{j,\mathrm{R}}^{(i)}\right)^2}\right], \text{ for } x_i > m_j^{(i)} \tag{6}$$

where $\sigma_{j,\mathrm{L}}^{(i)}$ and $\sigma_{j,\mathrm{R}}^{(i)}$ are the membership widths of $A_{j,\mathrm{L}}^{(i)}$ and $A_{j,\mathrm{R}}^{(i)}$, respectively.

Following a grid-type partition approach, we create a total number of $R = \Pi_{i=1}^{r} K_i$ rectangular fuzzy subspaces, $\mathcal{A}^i$, determined by the Cartesian product: $\mathcal{A}^i = A_{i1}^{(1)} \times A_{i2}^{(2)} \times \cdots \times A_{ir}^{(r)}$ $(i = 1, \ldots, R)$. Figure 4 shows an illustrative two-dimensional fuzzy partition after data clustering. The FPDs are described by R TSK-type fuzzy modelling rules of the form:

$$R_m^{(i)} : \text{IF } x_1 \text{ is } A_{i1}^{(1)} \text{ AND} \cdots \text{AND } x_p \text{ is } A_{ip}^{(1)} \text{ THEN } y_1 = g_1^{(i)}(x) \text{ AND } y_2 = g_2^{(i)}(x) \quad (7)$$

Traditionally, the TSK rule functions are represented as linear polynomials of the FPD inputs:

$$g_r^{(i)}(x) = w_{0,r}^{(i)} + w_{1,r}^{(i)} x_1 + \ldots + w_{p,r}^{(i)} x_p,\ r = 1,2 \quad (8)$$

Ignoring the linear terms in equation (8), we are led to rules with crisp consequents, i.e. $g_r^{(i)}(x) = w_{0,r}^{(i)}$.

For each pattern $\boldsymbol{x}$ submitted to the FPD, the firing degree of the rules are calculated as

$$\mu_i(x) = \prod_{j=1}^{R} \mu_{ij}^{(i)}(x) \quad (9)$$

The outputs of the FPDs are then derived by

$$y_r = \sum_{s=1}^{R} \frac{\mu_i(x) g_r^{(i)}(x)}{\sum_{s=1}^{R} \mu_s(x)},\ r = 1,2 \quad (10)$$

Initially, the fuzzy sets $A_j^{(i)}$ are evenly distributed along the x_i axis. To improve the feature transformation of the FPDs, a K-means clustering algorithm (Lee *et al.*

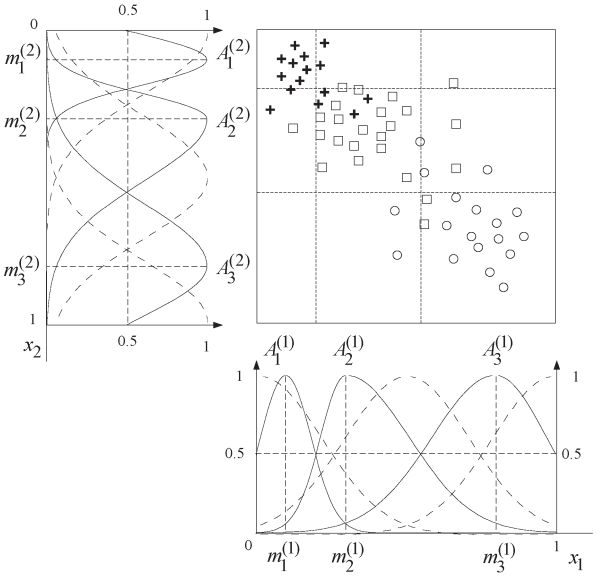

Figure 4. Premise partition after tuning the membership functions using the K-means clustering algorithm.

2001) is applied on the membership centres along each input. The goal is to locate the fuzzy sets in a way that the resulting fuzzy partitions are focused on regions with large data collections. The initial placement of the fuzzy sets (broken lines), as well as their tuning after data clustering (solid lines), is shown in figure 4. The values of $\sigma_{j,\mathrm{L}}^{(i)}$ and $\sigma_{j,R}^{(i)}$ are determined so that consecutive fuzzy sets exhibit a degree of overlapping of 0.5.

Based on the set of rules obtained above, we proceed to a rule base simplification procedure with the scope to reduce the number of rules. Along this direction, each modelling rule $R_m^{(i)}$ is evaluated by computing the percentage concentrations of patterns, $\widetilde{n}_i$, over the entire training set: $\widetilde{n}_i = \frac{n_i}{N} \times 100(\%)$, $i=1,\ldots,R$. n_i stands for the number of patterns that are included in the antecedent part $A_j^{(i)}$ with a degree of firing fulfilling $\mu_i(x) \geq 0.5$. The values of $\widetilde{n}_i$ are then arranged in descending order and compared to a prescribed threshold ξ set up by the user (i.e. $\xi=5\%$). Fuzzy rules exhibiting $\widetilde{n}_i \geq \xi$ are retained while the rest of them are discarded. At the end of this process we obtain a simplified fuzzy model (FPD), including a reduced number of rules $\widetilde{R} < R$. Patterns that are located at fuzzy cells corresponding to removed rules are covered by the neighbouring strong ones. This is made possible by noting that the fuzzy regions are described by Gaussian rather than triangular memberships, of appropriate centres and widths. Accordingly, every pattern takes a stronger or weaker firing, assisting its manipulation by the reduced rule base.

Having determined the premise parameters (means and widths), the outputs of the FPDs are linear with respect to the consequent weights. Therefore, optimal estimates of these parameters can be obtained using the recursive least square estimate (RLSE) method (Goodwin and Sin 1984). Given a specific class targeting, the RLSE method calculates the appropriate values of the consequent weights so that the distance between the FPD outputs and the class targets is minimized. As a result, a supervised learning task is achieved with the following objective: patterns $x_j^{(1)}(q)$ belonging to a particular class C_j should produce an output $y_k^{(1)}(q)$ located in a neighbourhood of the respective class target $y_d^{(C_j)}$, $j=1,\ldots,M$. From this point of view, each FPD realizes a non-linear mapping from the initial feature space to a transformed output space. The FPD outputs can be regarded as transformed versions, $y_j^{(1)} = \mathrm{FPD}_J^{(1)}\left(x_j^{(1)}\right)$, of the input features, which are more separated than the original features. Feature transformation facilitates discrimination of the patterns along the classes, thus leading to more accurate classification results obtained by the following DMFUs.

To show the function undertaken by the feature transformation, an illustrative case is indicated in figure 5 for a four-class problem ($M=4$) with two input features. As can be seen, the class examples in the original feature space (x_1, x_2) are mixed with each other, making discrimination a difficult task. However, the transformed patterns defined by the FPD outputs are more separated because of the fuzzy inference mechanism and the effectiveness of the parameter learning algorithm (RLSE).

3.2.2 Decision making fuzzy unit (DMFU). The DMFU in each FNC follows the associated FPD module, with the aim of determining the degree of support given by the local classifier in each class (figure 3). The continuous outputs of the FPD, $\boldsymbol{y}=[y_1, y_2]^T$, serve as inputs of the DMFU, while its output is an M-dimensional vector, denoted as DN, that provides the soft decision profile of the neuron classifier.

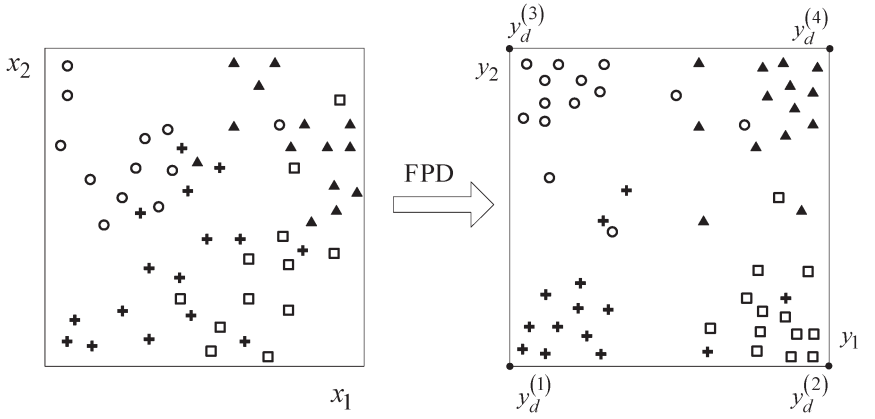

Figure 5. An illustrative graphical representation of the non-linear mapping implemented by the FPD unit for an artificial problem of $M=4$ classes. Class 1 patterns are represented by +, class 2 by □, class 3 by ○ and class 4 by ▲.

Each $y_r \in [0,1]$ $(r=1,2)$ is divided into L_r $(r=1, 2)$ fuzzy sets, namely $\left\{B_1^{(r)},\ldots,B_{Lr}^{(r)}\right\}$. The values of L_r are decided based on the number of classes at hand, so that: $L_1 \times L_2 \geq M$. The fuzzy sets of DMFU are now represented by trapezoidal membership functions, centred at the target values of each class (figure 6). The parameters of the intermediate membership functions are defined as $a_i^{(r)} = (i-1)/(L_r-1)-\rho$, $\beta_i^{(r)} = (i-1)/(L_r-1)+\rho$, $\gamma_i^{(r)} = i/(L_r-1)-\rho$, $\delta_i^{(r)} = i/(L_r-1)+\rho$ for $i=2,\ldots,L_r-1$, $r=1, 2$, where ρ is a small percentage of $1/L_r$ (i.e. $\rho=20\%$) that controls the degree of overlapping between adjacent fuzzy sets. For the left-most and the right-most membership functions we have: $a_1^{(r)} = \beta_1^{(r)} = 0$, $\gamma_1^{(r)} = a_2^{(r)}$, $\delta_1^{(r)} = \beta_2^{(r)}$, $a_{Lr}^{(r)} = \gamma_{Lr}^{(r)}$, $\beta_{Lr}^{(r)} = \delta_{Lr}^{(r)}$, $\gamma_{Lr}^{(r)} = \delta_{Lr}^{(r)} = 1$

Using the grid partition, we obtain $L_1 \times L_2 \geq M$ fuzzy subspaces. An indicative partition of the DMFU for a four-class case ($M=4$) is shown in figure 7. In order to represent $M=4$ classes, we define $L_1=2$ and $L_2=2$ membership functions. For each rectangular subspace we define a classification rule of the form:

$$R_c^{(i)} : \text{IF } y_1 \text{ is } B_{i1}^{(1)} \text{ AND } y_2 \text{ is } B_{i2}^{(2)} \text{ THEN } (y_1[q], y_2[q]) \text{ is } C_j \tag{11}$$

where $C_j \in C$ is the label for class j.

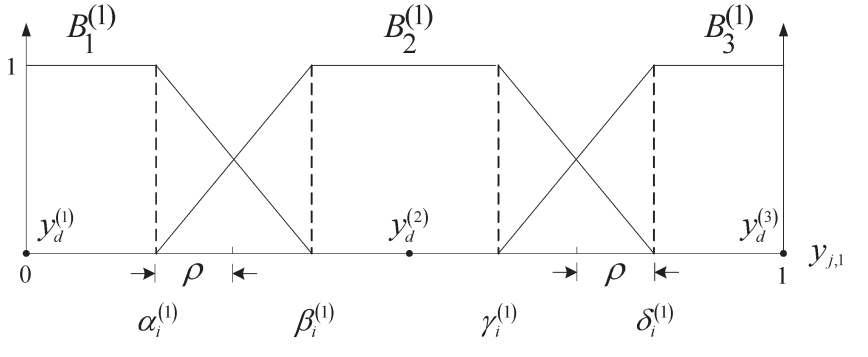

Figure 6. Shape of the membership functions in the premise part of the DMFU with $L_r=3$.

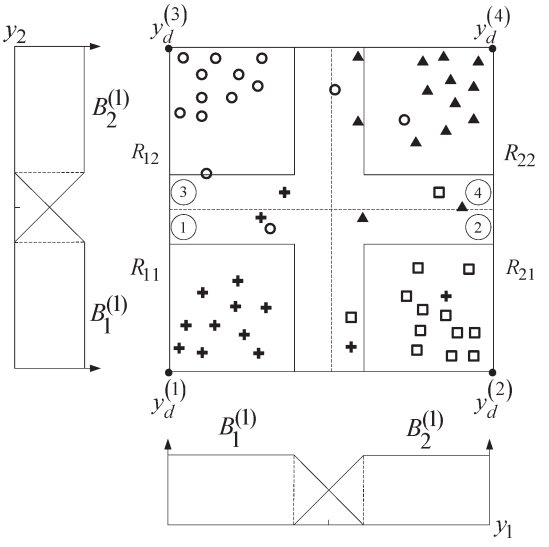

Figure 7. Graphical representation of the class decision regions designed by the DMFU for the artificial problem of four classes.

The fuzzy inference mechanism of the DMFUs is realized in two steps:
Step 1. Calculate the degree of firing of the classes, β_{C_i}, $i=1,\ldots,M$

$$\beta_{C_i} = \mu_{i1}^{(1)}(y_1(q)) \wedge \mu_{i2}^{(2)}(y_2(q)) \tag{12}$$

Step 2. Calculate the normalized firings:

$$\overline{\beta}_{C_i} = \beta_{C_i} \Big/ \sum_{r=1}^{M} \beta_{C_r} \tag{13}$$

The output of the DMFU is a vector, DN, derived as follows:

$$DN = [dn_1, \ldots, dn_m]^T = [\overline{\beta}_{C_1}, \ldots, \overline{\beta}_{C_M}]^T \tag{14}$$

The components $dn_i \in [0,1]$, $i=1,\ldots,M$, represent the degree of support given by the local classifier $\mathrm{FNC}_j^{(1)}$ under the hypothesis that a particular pattern $y[q]$ belongs to class i. The soft outputs of DMFU can be hardened to make a crisp decision upon the class each pattern belongs to, such that $DN(y[q]) = C_x \Rightarrow dn_{c_x} = \max\{dn_v\}$, $v=1,\ldots,M$. For the example considered in figure 5, the class labels can be easily assigned as shown in figure 7. The broken lines split the class decision regions whereas the solid lines represent confident regions, where patterns are classified with a high grade of certainty.

Each $\mathrm{FNC}_j^{(1)}$ provides two output sources: a vector of continuous outputs $y_j = [y_{j,1}, y_{j,2}]^T$ and an M-dimensional vector $DN_j^{(1)}$, the output of the $\mathrm{DMFU}_j^{(1)}$, that includes the soft decision supports for all classes (figure 3). The outputs of the FNCs at layer 1 serve as inputs to the FNCs to be generated in the second layer.

3.3 *FNC models at higher layers*

The general structure of the $\mathrm{FNC}_k^{(\ell)}$ at higher layers ($\ell \geq 2$) is demonstrated in figure 8. Two parent FNCs from the preceding layer are combined to generate a

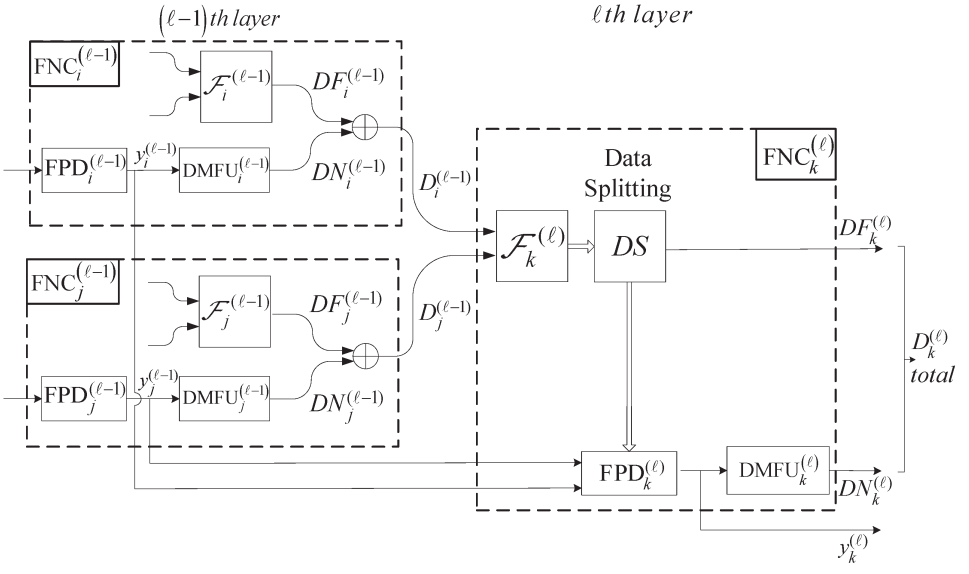

Figure 8. Neuron structure of nodes in higher layers ($\ell \geq 2$).

descendant FNC at that layer. The neuron classifiers are now represented by the triple: $\text{FNC}_k^{(\ell)} = \left\{\text{FPD}_k^{(\ell)}, \text{DMFU}_k^{(\ell)}, \mathcal{F}_k^{(\ell)}\right\}$. To exploit the information acquired by these parent classifiers we make use of a fusion scheme, a common practice in combining classifiers (Kuncheva *et al.* 2001). To this end, a fusion operator ($\mathcal{F}_k^{(\ell)}$) is introduced within each $\text{FNC}_k^{(\ell)}$ to aggregate the outputs of the parent classifiers. The decision output of the fuser is attached to the descendant FNC. Additionally, the fusion algorithm serves as a means to discriminate between those patterns that are currently well classified by the parent classifiers, and those that need further investigation by the FPD unit of the offspring classifier. This procedure gives rise to a data splitting mechanism that offers efficient handling of the data flow and reduction of the computational cost.

The inclusion of the fuser $\mathcal{F}_k^{(\ell)}$ gives rise to a composite decision support vector, denoted as $D_k^{(\ell)}$, at the output of $\text{FNC}_k^{(\ell)}$, which is formulated as described in the sequel. Derivation of $D_k^{(\ell)}$ proceeds along the following steps:

A.1 *Decision fusion*. For a given pattern $\boldsymbol{x}[q]$, the decision outputs of the antecedent classifiers $D_i^{(\ell-1)}$ and $D_j^{(\ell-1)}$ are fused as follows:

$$DF_k^{(\ell)}(x[q]) = \mathcal{F}_k^{(\ell)}\left\{D_i^{(\ell-1)}(x[q]), D_j^{(\ell-1)}(x[q])\right\} \tag{15}$$

where $\mathcal{F}$ denotes a fusion operator. The resulting decision output, $DF_k^{(\ell)}$, includes the certainty grades given by the fuser for all classes

$$DF_k^{(\ell)}(x[q]) = \left[df_{k,1}^{(\ell)}(x[q]), \ldots, df_{k,M}^{(\ell)}(x[q])\right] \tag{16}$$

There are several types of fusion operators reported in the literature. In general, they are divided into two major categories, the class-conscious and the class-indifferent fusion methods, depending on the way they handle the decision information carried out by the combined classifiers. In this paper, four different types of fusion operators are used, namely min, weighted average, fuzzy integral, and decision templates

(Kuncheva *et al.* 2001). For instance, the min operator is a simple and fairly conservative aggregation operator, belonging to the class-conscious methods:

$$df_{k,v}^{(\ell)} = \min\left\{dn_{i,v}^{(\ell)}, dn_{j,v}^{(\ell)}\right\}, \quad v = 1, \ldots, M \tag{17}$$

A.2 *Data splitting (DS)*. The decision profile obtained by the fuser, $DF_k^{(\ell)}$, provides the degree of support that a pattern belongs in each class by combining the outputs of the parent classifiers $\mathrm{FNC}_i^{(\ell-1)}$ and $\mathrm{FNC}_j^{(\ell-1)}$. To ascertain the classification level of the patterns, the values $df_{k,r}^{(2)}$ are compared to a user's defined threshold $df_{k,r}^{(2)} \geq \vartheta, \quad \vartheta \in [0.5,1]$, where ϑ represents the degree of confidence (i.e. $\vartheta = 0.8$) that a pattern belongs to a certain class, $x[q] \in C_r$. The above condition is called here the maximum classification level criterion (MCLC). Based on the MCLC criterion, the entire data set D_N is divided into two disjunctive subsets $J_k^{(\ell)}$ and $V_k^{(\ell)}$, so that $D_N = J_k^{(\ell)} \cup V_k^{(\ell)}$. The subset $J_k^{(\ell)}$ includes those patterns that fulfil the MCLC criterion, that is they are currently well classified with a high grade of certainty. Furthermore, the subset $V_k^{(\ell)}$ contains the rest of the data patterns that are either misclassified by the fuser or correctly classified with a low degree of support.

A.3 *Handling of patterns in* $J_k^{(\ell)}$. In this case, the fuser produces a high grade of certainties for a class, implying that both parent classifiers agree strongly on the same class. Therefore, well-classified patterns contained in $J_k^{(\ell)}$ are handled by the fuser, that is their decision outputs are derived by $DF_k^{(\ell)}(x[q]), \forall\, x[q] \in J_k^{(\ell)}$

A.4 *Handling of patterns in* $V_k^{(\ell)}$. For these patterns, the fuser produces low certainty grades for all classes (below confidence threshold ϑ), which means that either a conflict occurs between the parent FNCs or patterns are correctly classified by the parent classifiers with a weak degree of support. Hence, ambiguous patterns included in $V_k^{(\ell)}$ are fed to $\mathrm{FPD}_k^{(\ell)}$ for further processing. They are subject to an additional feature transformation to improve their discrimination.

As two parent classifiers are combined at a time, the input vector of the $\mathrm{FPD}_k^{(\ell)}$ is a four-dimensional vector:

$$x_k^{(\ell)} = \left[\left(x_{k,1}^{(\ell)}\right)^T, \left(x_{k,2}^{(\ell)}\right)^T\right]^T = \left[\left(x_i^{(\ell-1)}\right)^T, \left(x_j^{(\ell-1)}\right)^T\right]^T \in \Re^4 \tag{18}$$

Assume that the antecedent classifiers $\mathrm{FNC}_i^{(\ell-1)}$ and $\mathrm{FNC}_j^{(\ell-1)}$ assign a pattern $x_k^{(\ell-1)} \in V_k^{(\ell-1)}$ to the classes C_i* and C_j*, respectively. We check whether patterns $x_i^{(\ell-1)}$ and $x_j^{(\ell-1)}$, which are associated with $x_k^{(\ell)}$, are included in the sets $J_i^{(\ell-1)}$, $J_j^{(\ell-1)}$ or the sets $V_i^{(\ell-1)}$, $V_j^{(\ell-1)}$, respectively. The input parts $x_{k,1}^{(\ell)}, x_{k,2}^{(\ell)} \in \Re^2$ are defined by

$$\begin{aligned}
&\text{IF } x_i^{(\ell-1)} \in V_i^{(\ell-1)} \text{ THEN } x_{k,1}^{(\ell)} = y_i^{(\ell-1)} \\
&\quad \text{ELSE IF } x_i^{(\ell-1)} \in J_i^{(\ell-1)} \text{ THEN } x_{k,1}^{(\ell)} = y_d^{(C_i*)} \\
&\text{IF } x_j^{(\ell-1)} \in V_j^{(\ell-1)} \text{ THEN } x_{k,2}^{(\ell)} = y_j^{(\ell-1)} \\
&\quad \text{ELSE IF } x_j^{(\ell-1)} \in J_j^{(\ell-1)} \text{ THEN } x_{k,2}^{(\ell)} = y_d^{(C_j*)}
\end{aligned}$$

In the output space of $\mathrm{FPD}_k^{(\ell)}$, the classes are ordered following the same class targeting scheme as the one defined for the FNCs at the first layer. Construction of the $\mathrm{FPD}_k^{(\ell)}$ is performed following the steps described in section 3.2. Having obtained the locations of the new transformed features $y_k^{(\ell)}$ ($\mathrm{FPD}_k^{(\ell)}$ outputs) of the patterns in $V_k^{(\ell)}$, decision making is then applied by means of the corresponding $\mathrm{DMFU}_k^{(\ell)}$ module, to obtain the soft decision output vector:

$$DN_k^{(\ell)}(x[q]) = \mathrm{DMFU}_k^{(\ell)}\left\{y_k^{(\ell)}[q]\right\} = \left[dn_{k,1}^{(\ell)}(x), dn_{k,2}^{(\ell)}(x), \ldots, dn_{k,M}^{(\ell)}(x)\right]^T \tag{19}$$

where $dn_{k,j}^{(\ell)} \in [0,1]$ denotes the degree of support for a pattern belonging in class C_j.

A.5 *Overall decision output*. The overall decision profile of $\mathrm{FNC}_k^{(\ell)}$ is formulated as a composite of two parts:

$$D_k^{(\ell)} = DF_k^{(\ell)} \oplus DN_k^{(\ell)} = \left[d_{k,1}^{(\ell)}, d_{k,2}^{(\ell)}, \ldots, d_{k,M}^{(\ell)}\right]^T \tag{20}$$

where the $DF_k^{(\ell)}$ and $DN_k^{(\ell)}$ come from the fuser and the modules pair $\left\{\mathrm{FPD}_k^{(\ell)}, \mathrm{DMFU}_k^{(\ell)}\right\}$, respectively, functioning on the $J_k^{(\ell)}$ and $V_k^{(\ell)}$ data subsets: $D_k^{(\ell)}(x[q]) = DF_k^{(\ell)}(x[q])$, $\forall x[q] \in J_k^{(\ell)}$ and $D_k^{(\ell)}(x[q]) = DN_k^{(\ell)}(x[q])$, $\forall x[q] \in V_k^{(\ell)}$.

When a hard decision is to be made, the patterns are assigned to a class C_r fulfilling the maximum argument principle: $D_k^{(\ell)}(x[q]) = C_r \Rightarrow d_{k,C_r}^{(\ell)} = \max_{j=1,\ldots,M}\left\{d_{k,j}^{(\ell)}\right\}$. The fuzzy-to-crisp transformation operates either on $df_{k,j}^{(\ell)}$ or $dn_{k,j}^{(\ell)}$, depending on whether a pattern belongs to $J_k^{(\ell)}$ or $V_k^{(\ell)}$. It should be noted that especially for the FNCs at layer 2, the decision fusion in equation (15) acts upon the $DN_i^{(\ell-1)}$ and $DN_j^{(\ell-1)}$ provided by the DMFUs of the parent classifiers of layer 1.

The decision mechanism described above exhibits some remarkable merits discussed in the following: (i) first, the fuser exploits the decision supports given by the antecedent classifiers. Patterns that are well classified from their parents have a large support in the fuser's output, resulting in a confident decision for the class they belong to. Provided that a certain confidence level is exceeded, they are excluded from the fuzzy model construction. (ii) Data splitting leads to significant computational savings, a convenient advantage, especially when difficult problems are examined with large data sets and large number of classes. (iii) Exclusion of well-classified data allows the FPDs to focus on those patterns where adequate classification accuracy is not yet achieved. (iv) The antecedent classifiers in each layer are developed following different paths, starting from the original feature space and proceeding through the layered structure of the SONeFMUC. Hence, among the family of feature subspaces investigated, an optimal path is decided on, leading to enhanced classification rates.

3.4 *Structure learning*

The proposed model is generated in a self-organizing manner, by means of the GMDH algorithm (Ivakhnenko 1968), described below. In particular, the structure of the SONeFMUC is not predetermined in advance. Starting from the original system inputs (features), new layers are developed sequentially, until a final topology is obtained, satisfying the performance requirements.

The structure learning of the SONeFMUC proceeds along the following steps:

Step 1. Formulate data sets: The data set D_N is divided into a training data set D_{trn}, a validation data set D_{val} and a testing data set D_{chk}, comprising n_{trn}, n_{val} and n_{chk}, respectively, with $n_{\text{trn}} + n_{\text{val}} + n_{\text{chk}} = N$. The training and validation data sets are used for determining the structure of the individual FNCs, at each layer, while the checking data are used for evaluating the SONeFMUC model obtained at the end of structure learning.

Step 2. Select the model parameters: In this step, we choose the structural parameter values involved in the FPDs, the DMFUs and the fusion mechanism. (i) Select the number p of FPD inputs. For the $\text{FNC}_k^{(1)}$ at layer 1, p takes values in the range $p \in [2,4]$, with the inputs chosen among the set of original features $\{x_1, x_2, \ldots, x_m\}$. All $\text{FNC}_k^{(\ell)}$ at higher layers, $\ell \geq 2$, are developed by combining two parent FNCs of the previous layer ($p=4$). (ii) Select the number of membership functions K_i, $i=1,\ldots,r$, and L_i, $i=1, 2$, used to partition the premise space of $\text{FPD}_k^{(\ell)}$ and $\text{DMFU}_k^{(\ell)}$, respectively. Choice of L_i depends upon the number of classes of the problem at hand. (iii) Select the type of rule consequent functions (polynomial or crisp). (iv) Select the fusion operator $\mathcal{F}_k^{(\ell)}$ among the three class-conscious aggregation rules (min, weighted average, and fuzzy integral) and the class-indifferent alternative (decision templates). (v) Select confidence threshold $\vartheta \in [0.5,0.8]$ that controls the data splitting within each FNC. The greater the value of ϑ, the smaller the number of patterns whose decision is derived by the fuser $\mathcal{F}_k^{(\ell)}\left(DF_{k_\ell}^{(\ell)}\right)$, and the greater the data portion being handled by the pair of modules $\left\{\text{FPD}_k^{(\ell)}, \text{DMFU}_k^{(\ell)}\right\}\left(DN_k^{(\ell)}\right)$. Once specific values of the above parameters are chosen, they are applied uniformly for all FNCs during the structure learning of the SONeFMUC model.

Step 3. Consider the so-called best set, $BS^{(\ell-1)} = \left\{\text{FNC}_i^{(\ell-1)}, i=1,\ldots,W\right\}$, formed at layer $(\ell-1)$. It comprises a number of W individuals that represent the most qualifying FNCs retained at that layer. The outputs of these individuals form the candidate input set used for the construction of the individuals $\text{FNC}_k^{(\ell)}$ at layer ℓ. In particular, the best set $BS^{(0)}$ includes the m original feature components: $BS^{(0)} = \{x_1, x_2, \ldots, x_m\}$. As for the $BS^{(\ell)}$ at the succeeding layers, it is determined as described in step 7.

Step 4. Create the population $P^{(\ell)}$ of candidate FNCs at layer ℓ: The population $P^{(\ell)}$ is formulated by recombining the individuals in $BS^{(\ell-1)}$ obtained at layer $(\ell-1)$. The new individuals to be generated are obtained by combining parent FNCs from $BS^{(\ell-1)}$. Considering all possible combinations ${}^{W}C_2$, we conclude with a total number of $Q^{(\ell)} = \begin{pmatrix} W \\ 2 \end{pmatrix} = W!/(W-2)!2!$ new FNCs that form the population at the current layer: $P^{(\ell)} = \left\{\text{FNC}_k^{(\ell)}, \ell = 1, \ldots, Q^{(\ell)}\right\}$.

Step 5. Construct the $\text{FNC}_k^{(\ell)}$ models of $P^{(\ell)}$: In this step we determine the structure of the $\text{FNC}_k^{(\ell)}$, $\ell = 1, \ldots, Q^{(\ell)}$, by combining its parent modules $\text{FNC}_i^{(\ell-1)}$ and $\text{FNC}_j^{(\ell-1)}$. Combining the parents FNCs means that we make use of both types of outputs being offered, that is the continuous outputs $y_i^{(\ell-1)}$ and $y_j^{(\ell-1)}$ (transformed feature values) and the soft decision vectors $D_i^{(\ell-1)}$ and $D_j^{(\ell-1)}$.

Step 6. Evaluate the FNCs of $P^{(\ell)}$: Each $\text{FNC}_k^{(\ell)}$ is evaluated to assess its approximation and predictive capabilities. For the training data set, calculate the

following error measure:

$$E_{\text{trn},k} = \frac{1}{n_{\text{trn}}} \sum_{j=1}^{2} \sum_{q=1}^{n_{\text{trn}}} \left\{ y_{d,j}^{C}[q] - y_{k,j}^{(\ell)} \right\}^2 + \sum_{q=1}^{n_{\text{trn}}} \left\{ C_d[q] \neq C_j[q] \right\}, \quad k = 1, \ldots, Q^{(\ell)} \tag{21}$$

The first term in equation (29) is the mean squared error function, computing the proximity of the transformed outputs $y_{k,j}^{(\ell)}$ to the respective class targets. The second term determines the total number of misclassifications occurring over the training data set. Assuming that two individuals produce the same number of misclassifications, the one exhibiting the lower mean squared error will be selected, thus providing better placement of the transformed features. The $E_{\text{val},k}$ associated with the validation data set, D_{val}, is calculated in a similar way. The FNCs are evaluated using a weighted average metric: $E_k = (1-a)\, E_{\text{trn},k} + a\, E_{\text{val},k}$, where $a \in [0,1]$ is a weight specified by the user controlling the balance between $E_{\text{trn},k}$ and $E_{\text{val},k}$.

Step 7. Formulate the best set $BS^{(\ell)}$ at layer ℓ: The values of E_k are placed in ascending order. The first individual corresponds to the most qualifying FNC, denoted as $\text{FNC}_*^{(\ell)}$, having the lowest classification error. A collection of W FNCs are retained that exhibit the lower classification errors. These FNCs form the best set $BS^{(\ell)}$ at the current layer, including the highly qualified individuals of the population, $P^{(\ell)}$. The outputs of the FNCs contained in $BS^{(\ell)}$ serve as candidate inputs for the next layer while the remaining ones are discarded.

Step 8. Check for the termination criterion: Assume that the best FNC in $BS^{(\ell)}$, $\text{FNC}_*^{(\ell)}$, exhibits a classification score denoted by $E_*^{(\ell)}$. The best node $\text{FNC}_*^{(\ell)}$ is temporarily regarded as the output of the SONeFMUC model. The model expansion stops when either ℓ reaches a maximum number of layers, $M_{\max}$, or the best performance attained at the current layer exceeds the one obtained at the previous layer: $E_*^{(\ell)} \geq E_*^{(\ell-1)}$. Upon termination of the evolution, proceed to step 10.

Step 9. Determine the inputs to the next layer: Assuming that $E_*^{(\ell)} < E_*^{(\ell-1)}$ and the maximum number of layers is not yet reached, the SONeFMUC model is allowed to expand by including a new layer. The retained individuals $\text{FNC}_k^{(\ell)}$ in $BS^{(\ell)}$ are recombined again to generate the FNCs at layer $(\ell+1)$. Accordingly, their outputs $y_k^{(\ell)}$ and $D_k^{(\ell)}$ are submitted as inputs to the descendant FNCs at the next layer. To this end, go to step 4 for the generation of the new population.

Step 10. Recover the network's architecture: Once the stopping criteria are satisfied for some $M \leqslant M_{\max}$, the node classifier with the best performance, $\text{FNC}_*^{(M)}$, is considered as the ending node of the SONeFMUC, providing the decision outputs of the model. The remaining FNCs at the output layer are discarded. In the following, we perform a reverse flow tracing through the network's structure, moving from the output to the input layer. All nodes at the intermediate layers (the input layer included) having no contribution to the FNC selected at the final layer are removed from the network. As regards the model's inputs, among the original features $\{x_1, \ldots, x_m\}$, a subset of significant features is retained whereas the rest are discarded.

Figure 9 demonstrates the network expansion by the GMDH method. The broken lines indicate the recovered structure of the SONeFMUC. The resulting model is a three-layered network, with a total number of six FNCs. Four features are selected by GMDH, x_1, x_2, x_3 and x_5, from a total of m original features.

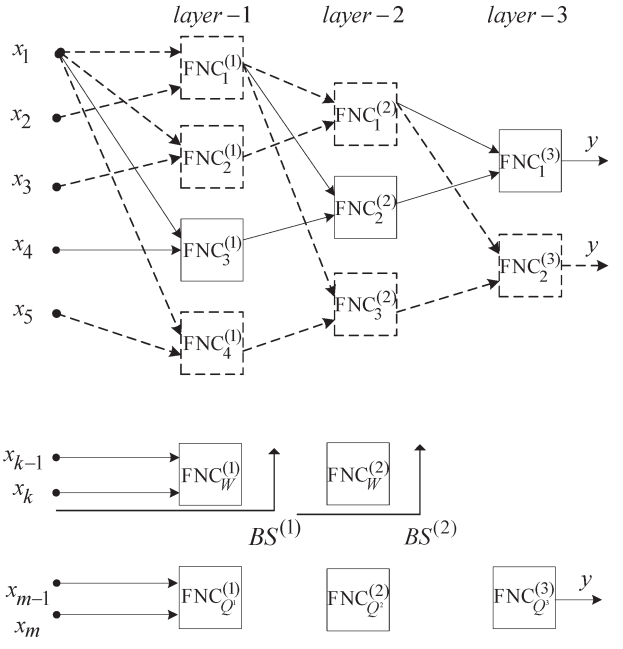

Figure 9. Example of the structure learning process by means of the GMDH algorithm.

The SONeFMUC classifier exhibits some remarkable attributes, distinguishing it from other classification model of the literature: (1) the SONeFMUC is a self-organizing multilayered network whose structure is developed sequentially in a layer-by-layer basis, following a systematic expansion procedure, the GMDH method. Depending on the complexity of the particular classification problem, the network depth is properly controlled so that the performance of the model obtained fulfils the design requirements. (2) The GMDH methodology inherently implements the so-called feature selection task. In the final model recovered at the end of structure learning, only the most important features having a significant contribution to the classification mapping are retained, while the unnecessary ones are discarded. In that respect, the GMDH performs two tasks simultaneously, namely model building and feature selection. (3) The FPD modules of the FNCs perform successive feature transformations through the layers. Starting from the input layer, the original features are repeatedly transformed between the intermediate layer spaces. For higher layers, each new FPD improves the class discrimination, allowing the corresponding DMFUs to draw more accurate classification assignments.

4. Experimental results

4.1 *Application of the SONeFMUC to the study area*

The SONeFMUC network was applied to a multispectral IKONOS image using the set of training samples recognized in the field. Owing to the large number of classes and the spectral overlapping of the feature signatures, we were confronted with misclassification problems, especially in classes that represent vegetation cover. Therefore, based on the pan-sharpened image and after careful photointerpretation, the image was segmented into two zones: the wetland zone, which includes the lake

and its surrounding wetland vegetation, and the agricultural zone. In the wetland zone five classes were recognized: water bodies, phragmites, tamarix, wet meadows, and trees. In the agricultural zone we considered eight classes, six referring to different crop types (maize, alfalfa, cereals, orchards, vegetables, and fallow) and two referring to other land cover types (urban areas and shrubs). The SONeFMUC classifier was applied to both of these zones.

As the structure learning process uses a validation data set to obtain appropriate networks with higher generalization capabilities, the training set was split further in training and validation using 60% and 40% of the original training set, respectively. The MLC classifier used the original training set for the training stage. The testing set was the same for both methods. To initiate the structure learning of the SONeFMUC classifier for each zone, a number of structural parameters have to be decided on: the number of fuzzy sets in each input, the form of the rules and the type of fuser. This task is accomplished by the following thee-step procedure. (1) In the first stage we considered the min fusion operator, and developed through GMDH different network combinations using three or five fuzzy sets along each FPD input and crisp or linear rules. (2) The best network based on the checking data performance was selected as the most appropriate. (3) The remaining three types of fusion (the weighted average, the fuzzy integral, and the decision templates) were used in the network decided in step 2. The network architecture exhibiting the highest classification accuracy on the checking data was selected as the final model.

In all simulations, the GLCMs were calculated using a window size of 7×7, direction $\theta = 0^\circ$ and distance $d = 1$ for the wetland zone, and a window size of 23×23, direction $\theta = 45^\circ$ and distance $d = 1$ for the agricultural zone. The performance of the model obtained was evaluated in terms of four parameters: the confusion error matrix, the overall accuracy, the Khat statistic, and the Z-score with 95% confidence level (Congalton and Green 1999).

4.1.1 Application of the SONeFMUC to the wetland zone. Following the above three-step procedure, a network was selected using three fuzzy sets for the FPD inputs, TSK type rules, and the fuzzy integral as the fusion scheme. The resulting network was a five-layered structure with an overall classification accuracy of 89.5% on the testing data. Table 1 shows the confusion matrix along with the producer's and user's accuracy as a percentage (PA% and UA%, respectively). Additionally, the statistical parameters Khat and Z-score were calculated, and used to assess the quality of the classifier. The SONeFMUC model shows strong agreement between

Table 1. Confusion matrix obtained by application of the SONeFMUC to the testing data set of the wetland zone.

	Wetland map classification						
	Phragmites	Tamarix	Wet meadows	Trees	Water bodies	PA (%)	UA (%)
Phragmites	102	12	7	3	0	90.27	82.26
Tamarix	7	17	2	3	0	54.84	58.62
Wet meadows	4	1	198	0	0	95.65	97.54
Trees	0	1	0	3	0	33.33	75.00
Water bodies	0	0	0	0	21	100.0	100.0
Reference	113	31	207	9	21		

Overall accuracy = 89.5%, Khat = 0.83, Z-score = 33.14

the remotely sensed classification and the reference data, as shown by the high value of the Khat parameter, which is greater than 0.8 (Congalton and Green 1999). As the Z-score of 33.14 was higher than the standard normal critical value $Z_c=1.96$, the classification provided by the SONeFMUC was significantly better than random, at a confidence level of 95%.

As depicted in table 1, the SONeFMUC classified the classes of phragmites, wet meadows and water bodies better than the other classes. The high accuracy in these three classes is attributed to the small spectral overlapping and within-class variance. However, poor classification results were achieved for the tamarix and trees classes. Tamarix was strongly confused with phragmites because their spectral signatures are very similar. The trees class was confused with tamarix and phragmites; these classes represent vegetation cover with similar spectral characteristics.

4.1.2 Application of the SONeFMUC to the agricultural zone. The SONeFMUC model selected was a six-layer network, using five fuzzy sets for each FPD input, crisp type rules, and the min fuser. The model provided an overall accuracy of 74.21% on the testing data set. Table 2 shows the confusion matrix for the agricultural zone, including the corresponding statistical parameters.

The SONeFMUC model exhibited moderate agreement between the remotely sensed classification and the reference data as the Khat value falls in the range 0.4–0.8. As expected, the value of the Z-score is large enough to indicate that the method is significantly better than random selection with a 95% confidence level. The overall classification accuracy was found to be smaller than that attained in the wetland zone, possibly because of the larger number of classes compared to the wetland zone. Nevertheless, the SONeFMUC exhibits an average producer's accuracy of 82.6% on three major classes (maize, alfalfa and cereals), implying that they are classified more accurately than the others, although the dominant classes are overestimated, as shown from the user's accuracy percentages. On the contrary, orchards, vegetables and shrubs were underestimated with a poor producer's accuracy. The explanation for this is that these classes contain different subclasses; for instance, vegetables may include tomatoes, watermelon or eggplants, a situation recognized with fieldwork, thus leading to a large number of misclassifications. A mosaic of land cover map of both zones is illustrated in figure 10(*a*), which depicts the results of the SONeFMUC.

Table 2. Confusion matrix obtained by application of the SONeFMUC to the testing data set of the agricultural zone.

	Agriculture map classification								PA	UA
	Alfalfa	Cereals	Maize	Orchards	Vegetables	Fallow	Shrubs	Urban	(%)	(%)
Alfalfa	149	8	23	3	12	3	2	0	74.13	74.50
Cereals	18	190	2	5	4	23	14	2	90.48	73.64
Maize	21	1	147	8	11	0	1	0	83.05	77.78
Orchards	6	1	2	21	2	1	1	0	50.00	61.76
Vegetables	5	1	3	2	15	0	0	0	32.61	57.69
Fallow	2	8	0	1	2	33	2	2	55.00	66.00
Shrubs	0	1	0	2	0	0	7	0	25.93	70.00
Urban	0	0	0	0	0	0	0	28	87.50	100.0
Reference	201	210	177	42	46	60	27	32		

Overall accuracy=74.21%, Khat=0.67, Z-score=34.84

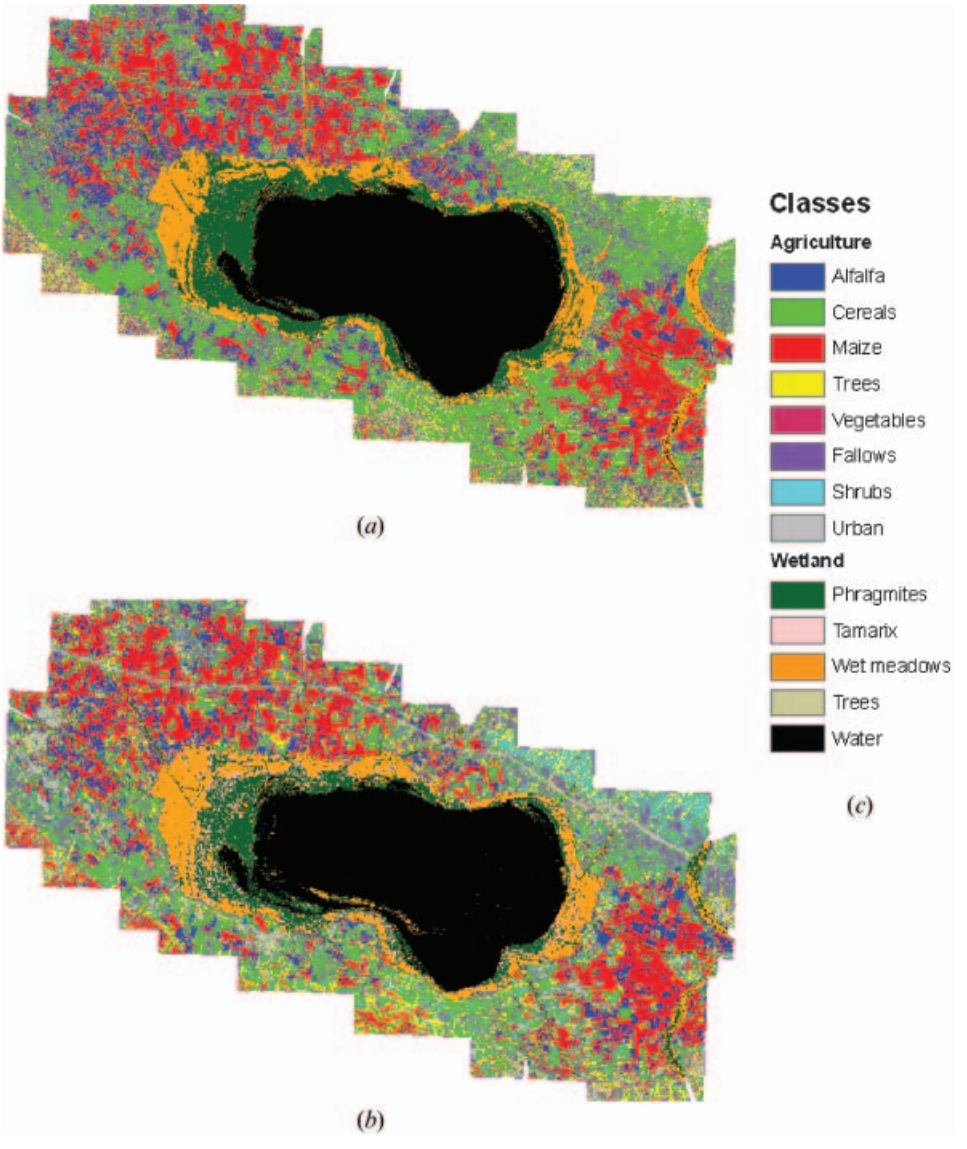

Figure 10. Mosaic of IKONOS land cover classification of wetland and agricultural zone using (*a*) the SONeFMUC and (*b*) the MLC. The legend is presented in (*c*).

4.2 *Comparison with the MLC*

To validate the classification results of the SONeFMUC algorithm, the IKONOS image was classified using the MLC, a traditional spectral probabilistic 'hard' classifier. As both the SONeFMUC and the MLC were applied on the same testing set, a McNemar test with continuity correction based on a χ^2 distribution was used to compare the performance of the two algorithms, as suggested by Foody (2004).

4.2.1 MLC classification of the wetland zone. Table 3 presents the confusion matrix for the MLC on the testing data. A high overall performance of 84.51% was obtained, which was 5% lower than that of the SONeFMUC. The Khat statistic was

Table 3. Confusion matrix obtained by application of the MLC to the testing data set of the wetland zone.

	Wetland map classification						
	Phragmites	Tamarix	Wet meadows	Trees	Water bodies	PA (%)	UA (%)
Phragmites	87	10	9	7	0	76.99	76.99
Tamarix	19	15	1	0	0	48.39	42.86
Wet meadows	7	6	197	0	0	95.17	93.81
Trees	0	0	0	2	0	22.22	100.0
Water bodies	0	0	0	0	21	100.0	100.0
Reference	113	31	207	9	21		

Overall accuracy=84.51%, Khat=0.74, *Z*-score=25.48

found to be 0.74, suggesting a moderate agreement between the remotely sensed classification and the reference data, as opposed to the SONeFMUC, where the agreement was very strong. Additionally, the *Z*-score was calculated as 25.48, which is significantly greater than Z_c, showing that the classification was better than random.

The MLC exhibits a good performance on classifying wet meadows and water. However, compared to the SONeFMUC, the phragmites classification gave poor results; most omission errors were found with tamarix, although phragmites cover a large area of the wetland. A low level of accuracy was also achieved in trees and tamarix as in the SONeFMUC algorithm, which shows the inefficiency of either of these two algorithms in classifying these classes.

4.2.2 MLC classification of the agricultural zone. The MLC was also applied in the agricultural zone, with the error confusion matrix shown in table 4. A moderate overall accuracy of 71.7% was obtained, slightly lower than that of SONeFMUC (approximately 74%). The Khat value was 0.65 and *Z*-score 33.69, indicating a fair agreement but suggesting that the classification result is not random.

Based on the statistical parameters, the MLC algorithm performed similarly to the SONeFMUC model in the agricultural zone. However, useful information about the performance of the two classification methods could be derived by further

Table 4. Confusion matrix obtained by the application of MLC to the testing data set of the agricultural zone.

	Agriculture map classification								PA	UA
	Alfalfa	Cereals	Maize	Orchards	Vegetables	Fallow	Shrubs	Urban	(%)	(%)
Alfalfa	119	7	5	0	4	0	0	0	59.20	88.15
Cereals	25	162	3	6	4	19	4	0	77.14	72.65
Maize	27	0	144	2	5	0	0	0	81.36	80.90
Orchards	12	7	15	34	6	3	1	0	80.95	43.59
Vegetables	18	3	10	0	22	0	0	0	47.83	41.51
Fallow	0	23	0	0	5	38	3	0	63.33	55.07
Shrubs	0	3	0	0	0	0	19	0	70.37	86.36
Urban	0	5	0	0	0	0	0	32	100.0	86.49
Reference	201	210	177	42	46	60	27	32		

Overall accuracy=71.7%, Khat=0.65, *Z*-score=33.69

analysing the error confusion matrix. For example, in the first three classes (maize, cereals and alfalfa), the dominant classes in the zone, the mean producer's accuracy of the MLC is lower (72.6%) than the SONeFMUC's (82.6%), indicating an underestimation by the MLC in these classes. However, fallow is classified better in the MLC than in the SONeFMUC, although this class is confused in both classifiers with cereals. This is because they are spectrally similar as the satellite image was acquired in August when cereals are harvested and fallows represent low vegetation cover. A mosaic of land cover maps of both zones is illustrated in figure 10(*b*), showing the results obtained by the MLC.

4.3 *Discussion of the results*

The classification algorithms SONeFMUC and MLC were applied to the wetland and agricultural zones of the satellite image, providing differing results. In the wetland zone, the SONeFMUC was found more suitable because the overall accuracy was higher and Khat showed a higher reliability of the results. McNemar's test on the two methods also demonstrated that their performance was significantly different in the wetland zone, $\chi^2_{1,0.95} > 3.84$ (see table 5).

The results followed a similar trend in the agricultural zone, where the SONeFMUC offered a higher overall accuracy and higher Khat. However, there was insufficient statistical evidence that SONeFMUC performed significantly better. This may be due to the larger number of classes in the agricultural zone compared to the wetland zone (eight and five, respectively), and the similar spectral characteristics of the crops. However, as regards the dominant classes, the SONeFMUC was found to be more suitable than the MLC, which had overestimated them.

Useful information was derived using a visual assessment of land cover maps produced by both classifiers. Figure 11 illustrates a region of the thematic map in the agricultural zone, showing the superior performance of the suggested approach. The SONeFMUC yields fewer pixel misclassifications, thus producing more homogeneous fields. Moreover, the MLC exhibits a large confusion between alfalfa and maize (blue and red colours, respectively) in many fields, a situation not confronted in the SONeFMUC.

A similar visual assessment was carried out in the wetland zone, as illustrated in figure 12. The overestimation of tamarix compared with phragmites by the MLC can be observed, a result also appearing in the confusion matrix shown in table 3. The SONeFMUC classified the stretch phragmites uniformly, which is the correct land cover type. Additionally, at the lake's shore where the depth of the water is low, a large number of pixels were classified errroneously by the MLC as phragmites and

Table 5. Comparison of the classification performance of the SONeFMUC with the MLC in the wetland and agricultural zones, using statistical parameters and McNemar's test.

	Wetland		Agricultural	
	SONeFMUC	MLC	SONeFMUC	MLC
Overall accuracy (%)	89.5	84.51	74.21	71.7
Khat	0.83	0.74	0.67	0.65
Z-score	33.14*	25.48*	34.84*	33.69*
χ^2	7.84*		0.82*	

*95% confidence level.

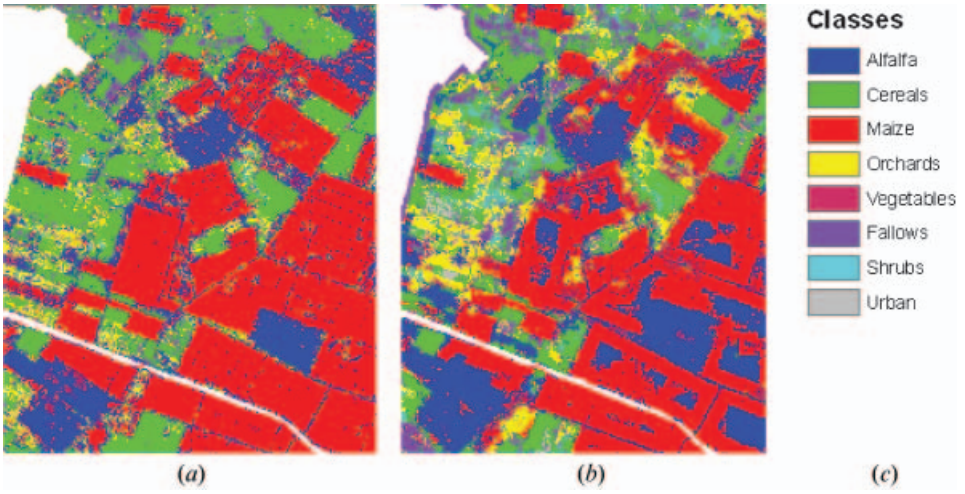

Figure 11. A subset of the land cover map produced (*a*) the SONeFMUC and (*b*) the MLC in the agricultural zone east of the wetland. The legend is presented in (*c*).

wet meadows; these pixels, however, belong to water, as correctly labelled by the SONeFMUC.

By using the structure learning algorithm (GMDH), the proposed method accomplished the feature selection task. The network generated for the wetland zone used only eight out of 26 original features as inputs. The subset of significant features selected by GMDH includes: bands 1–3, the ASM co-occurrence feature from the third band, greenness and brightness from the tasseled cap features, and intensity and hue from the IHS features. In the agricultural zone, the resulting network used only 12 out of the 26 initial features. These features consist of the four bands, correlation from the first band, homogeneity from the third band, ASM from the fourth band, the three tasseled cap features, and intensity and hue from the IHS features. On the contrary, the MLC does not provide this capability; therefore,

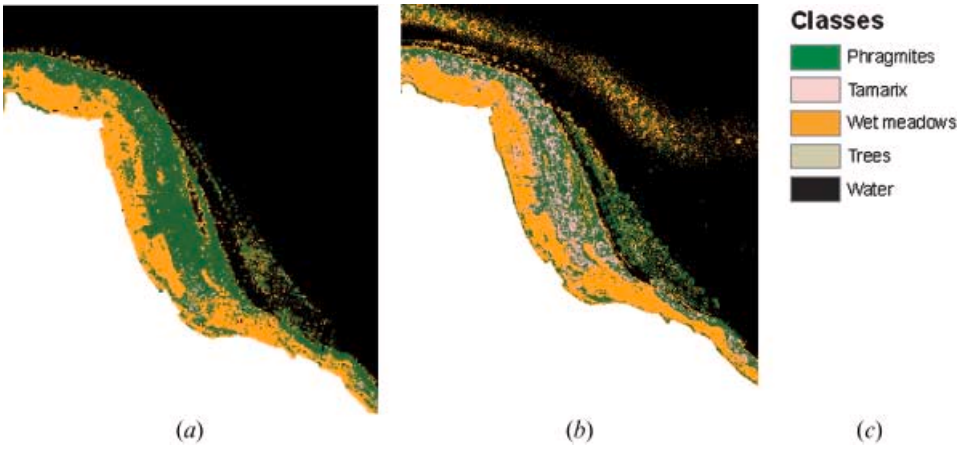

Figure 12. A subset of the land cover map produced with (*a*) the SONeFMUC and (*b*) the MLC in the wetland zone south of the lake. The legend is presented in (*c*).

feature selection is based on the designer's experience or on a time-consuming trial-and-error process.

5. Conclusions

In this paper, the SONeFMUC was proposed and applied to land cover classification of a VHR image in a protected area of high ecological interest. To improve classification accuracy, the image was divided into two zones: the wetland zone, where land cover corresponded to five habitat classes, and the agriculture zone with eight crop classes. The performance of the SONeFMUC was compared to that of the MLC.

A high classification accuracy of 89.5% was obtained by the SONeFMUC in the wetland zone (Khat=0.83). The suggested classifier was able to discriminate between the dominant habitat classes: phragmites, wet meadows and water bodies. In the agricultural zone, the performance of the SONeFMUC was lower, showing an overall accuracy of 74.21% (Khat=0.67). The basic reason for this was the large spectral overlapping between the crop classes, together with the larger number of classes as compared to the wetland zone. The dominant classes of the agricultural zone (alfalfa, maize and cereals) were distinguished to a satisfactory degree, as the average producer's accuracy was 82.6%. Finally, it should be noted that the proposed method was able to select the appropriate input features for each zone, leading to a higher classification accuracy.

Performance comparisons with the MLC verified the efficiency of the proposed SONeFMUC. In both zones, the SONeFMUC achieved a higher overall accuracy and higher Khat value than the MLC. In particular, in the wetland zone the overall accuracy of the MLC was 5% lower and the Khat revealed a poorer classification quality. In addition, McNemar's test between the two methods proved that their performance was significantly different. In the agricultural zone, the quality statistics of the MLC was lower but comparable to the SONeFMUC. Nevertheless, the MLC underestimated the dominant crops, contrary to the SONeFMUC, making the use of the suggested method more appropriate. Apart from the initial bands, more informative features are extracted from the multispectral image, namely textural and spectral features. Future work will explore the capabilities of our network using wavelet transformation as an additional input.

Acknowledgements

This study was funded by 'Pythagoras II', a research grant awarded by the Managing Authority of the Operational Programme 'Education and Initial Vocational Training' of Greece, which is partially funded by the European Social Fund, European Commission.

References

Alexandridis, T., Takavakoglou, V., Zalidis, G. and Crisman, T.L., 2007, Remote sensing and GIS techniques for selecting a self-sustainable scenario for Lake Koronia. *Environmental Management*, **39**, pp. 278–290.

Atkinson, P.M. and Tatnall, A.R.L., 1997, Introduction: neural networks in remote sensing. *International Journal of Remote Sensing*, **18**, pp. 699–709.

Bardossy, A. and Samaniego, L., 2002, Fuzzy rule-based classification of remotely sensed imagery. *IEEE Transactions on Geosience and Remote Sensing*, **40**, pp. 362–374.

Benediktsson, J.A. and Kanellopoulos, I., 1999, Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geosience and Remote Sensing*, **37**, pp. 1367–1377.

Briem, J.G., Benediktsson, J.A. and Sveinsson, R.J., 2002, Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geosience and Remote Sensing*, **40**, pp. 2291–2299.

Carpenter, G.A., Gjaja, M.N., Gopal, S. and Woodcock, C.E., 1997, ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. *IEEE Transactions on Geosience and Remote Sensing*, **35**, pp. 308–325.

Chavez, P.S., Jr., Sides, S.C. and Anderson, J.A., 1991, Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogrammetric Engineering and Remote Sensing*, **57**, pp. 295–303.

Congalton, R.G. and Green, K., 1999, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Boca Raton: Lewis Publishers).

Dekker, R.J., 2003, Texture analysis and classification of ERS SAR images for map updating of urban areas in the Netherlands. *IEEE Transactions on Geosience and Remote Sensing*, **41**, pp. 1950–1958.

Foody, G.M., 1995, Land cover classification using an artificial neural network with ancillary information. *International Journal of Geographical Information Systems*, **9**, pp. 527–542.

Foody, G.M., 1999, The continuum of classification fuzziness in thematic mapping. *Photogrammetric Engineering and Remote Sensing*, **65**, pp. 443–451.

Foody, G.M., 2004, Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 627–633.

Giacinto, G. and Roli, F., 1997, Ensembles of neural networks for soft classification of remote sensing images. In *Proceedings of the European Symposium on Intelligent Techniques, European Network for Fuzzy Logic and Uncertainty Modelling in Information Technology*, 19–21 March 1997, Bari, Italy, pp. 166–170.

Goodwin, G.C. and Sin, K.S., 1984, *Adaptive Filtering and Control* New Jersey: Prentice-Hall).

Haralick, R.M. and Shapiro, L.G., 1992, *Robot and Computer Vision*, vol. 1, pp. 457–462 (Boston: Addison-Wesley).

Hellenic Ministry of Environment 2001, *Identification and Description of Habitat Types in Areas of Interest for Nature Conservation*, Final Report (Athens, Greece: Hellenic Ministry of Environment, Physical Planning and Public Works: Environmental Planning Division).

Horne, J.H., 2003, A tasseled cap transformation for IKONOS images. In *Proceedings of the ASPRS Annual Conference*, 5–9 May 2003, Anchorage, Alaska (CD-Rom).

Hutchinson, M.F., 1988, Calculation of hydrologically sound digital elevation models. In *Third International Symposium on Spatial Data Handling*, 17–19 August, Sydney (Ohio: International Geographical Union), pp. 117–133.

Ivakhnenko, A.G., 1968, The group method of data handling; a rival of the method of stochastic approximation. *Soviet Automatic Control*, **1–3**, pp. 43–45.

Kauth, R.J. and Thomas, G.S., 1976, The tasseled cap – a graphical description of the spectral-temporal development of agricultural crops as seen by Landsat. In *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, 29 June–1 July 1976, West Lafayette, Indiana (New York: Institute of Electrical and Electronics Engineering), pp. 4B41–4B51.

Kavzoglu, T. and Mather, P.M., 2003, The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, **24**, pp. 4907–4938.

Kawata, Y., Ontani, A., Kusaka, T. and Ueno, S., 1990, Classification accuracy for the MOS-1 MESSR data before and after the atmospheric correction. *IEEE Transactions on Geoscience Remote Sensing*, **28**, pp. 755–760.

KERAMITSOGLOU, I., SARIMVEIS, H., KIRANOUDIS, C.T. and SIFAKIS, N., 2005, Radial basis function neural networks classification using very high spatial resolution satellite imagery: an application to the habitat area of Lake Kerkini (Greece). *International Journal of Remote Sensing*, **26**, pp. 1861–1880.

KUMAR, A.S., BASU, S.K. and MAJUMDAR, K.L., 1997, Robust classification of multispectral data using multiple neural networks and fuzzy integral. *IEEE Transactions on Geosience and Remote Sensing*, **35**, pp. 787–790.

KUNCHEVA, L.I., BEZDEK, J.C. and DUIN, R.P.W., 2001, Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, **34**, pp. 299–314.

LEE, H.M., CHEN, C.M., CHEN, J.M. and JOU, Y.L., 2001, An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Transactions on Systems, Man and Cybernetics*, **31**, pp. 426–432.

LI, X. and YEH, A.G.O., 1998, Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, **19**, pp. 1479–1500.

LIN, C.-T., LEE, Y.-C. and PU, H.-C., 2000, Satellite sensor image classification using cascade architecture of neural fuzzy network. *IEEE Transactions on Geosience and Remote Sensing*, **38**, pp. 1033–1043.

MITRAKI, C., CRISMAN, T.L. and ZALIDIS, G., 2004, Lake Koronia: shift from autotrophy to heterotrophy with cultural eutrophication and progressive water-level reduction. *Limnologica*, **34**, pp. 110–116.

OETTER, D.R., COHEN, W.B., BERTERRETCHE, M.K., MAIERSPERGER, T.K. and KENNEDY, R.E., 2001, Land cover mapping in an agricultural setting using multiseasonal Thematic Mapper data. *Remote Sensing of Environment*, **76**, pp. 139–155.

PETRAKOS, M., BENEDIKTSSON, J.A. and KANELLOPOULOS, I., 2001, The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion. *IEEE Transactions on Geosience and Remote Sensing*, **39**, pp. 2539–2546.

RICHARDS, J.A., LANDGREBE, D.A. and SWAIN, P.H., 1982, A means for utilizing ancillary information in multispectral classification. *Remote Sensing of Environment*, **12**, pp. 463–477.

SHACKELFORD, A.K. and DAVIS, C.H., 2003, A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on Geosience and Remote Sensing*, **41**, pp. 1920–1932.

SONG, C., WOODCOCK, C.E., SOTO, K.C., LENNEY, M.P. and MACOMBER, S.A., 2001, Classification and change detection using Landsat TM data: when and how to correct atmospheric effects. *Remote Sensing of Environment*, **75**, pp. 230–244.

TAKAGI, T. and SUGENO, M., 1985, Fuzzy identifications and its application to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, **15**, pp. 116–132.

THOMAS, I.L., BENNING, V.M. and CHING, N.P., 1987, *Classification of Remotely Sensed Images* (Bristol: Adam Hilger).

TSO, B. and MATHER, P.M., 2001, *Classification Methods for Remotely Sensed Data* (London: Taylor & Francis).

WANG, F., 1990, Fuzzy supervised classification of remote sensing images. *IEEE Transactions on Geosience and Remote Sensing*, **28**, pp. 194–201.

WILKINSON, G.G., FIERENS, F. and KANELLOPOULOS, I., 1995, Integration of neural and statistical approaches in spatial data classification. *Geographical Systems*, **32**, pp. 1–20.

ZALIDIS, G.C., TAKAVAKOGLOU, V. and ALEXANDRIDIS, T., 2004, *Revised Restoration Plan of Lake Koronia*, [in Greek, English summary] (Aristotle University of Thessaloniki, Thessaloniki: Department of Agronomy, Laboratory of Applied Soil Science).

ZHANG, Y. and HONG, G., 2005, An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. *Information Fusion*, **6**, pp. 225–234.