Multisource Classification of Complex Rural Areas by Statistical and Neural-Network Approaches

L. Bruzzone, C. Conese, F. Maselli, and F. Roli

Abstract

The automatic generation of land-cover inventories by using remote-sensing data is a very difficult task when complex rural areas are involved. The main difficulties are related to the characterization of such spectrally complex and heterogeneous environments and to the choice of an effective classification approach. In this paper, the usefulness of spectral (Landsat-5 Thematic Mapper images), texture (grey-level co-occurrence matrix statistics), and ancillary (terrain elevation, slope, and aspect) data to characterize two complex rural areas in central Italy is quantitatively demonstrated. A statistical and a neural-network classification approach are applied to such a multisource data set, and their classification performances are assessed and compared. The classification performances of the two approaches are quantitatively evaluated in terms of global and conditional Kappa accuracies. The Zeta statistics is used to evaluate the statistical significance of the different classification accuracies obtained by the two approaches by using multisource data.

Introduction

Remote sensing techniques, with properties of being synoptic and objective, easily updated, and being relatively inexpensive, can be very useful in various applications such as environment monitoring, crop inventorying, and geological studies (Swain and Davis, 1978; Curran, 1985). In particular, satellite imagery has been shown to be suited for the automatic generation of land-cover inventories, and several classification methods using such remote sensing data have been proposed (Swain and Davis, 1978; Richards, 1993; Maselli *et al.*, 1992; Serpico and Roli, 1995). Most of these methods yield satisfactory classification accuracies when land-cover classes are spectrally separable. On the contrary, great difficulties are encountered in spectrally complex environments, where several factors (e.g., mixed terrain compositions, terrain irregularities, differences in substrates) affect spectral signatures. In such environments, which are very common in European rural areas, the production of land-cover inventories based on spectral data alone may be unfeasible. The exploitation of other sources of information becomes mandatory. For instance, texture features extracted from remote-sensing images have been proposed (Haralick *et al.*, 1973). Another source of information is represented by ancillary data, in particular, environmental information related to the spatial distribution of the categories of interest (e.g., morphological, geological, climatological data) (Hutchinson, 1982). All these sources of information may be exploited to obtain more accurate land-cover inventories of complex rural areas.

In the literature, several studies on the classification of multisource remote-sensing data have been reported (Lee *et al.*, 1987; Benediktsson *et al.*, 1990; NASA, 1990; IAPR, 1992; Zhuang *et al.*, 1991; Civco and Wang, 1994; Gong, 1994; Wang and Howarth, 1994). Some of them investigated the use of neural-network (NN) classifiers and compared their performances with the ones of classical statistical methods. Benediktsson *et al.* (1990) demonstrated that a multilayer perceptron neural network performed better than statistical methods when applied to Landsat MSS data merged with geographic data, including elevation, slope, and aspect. Zhuang *et al.* (1991) drew similar conclusions for a data set made up of Landsat Thematic Mapper (TM) data merged with land ownership data. In their experimentation, a neural-network classifier provided better classification accuracies than the ones obtained by maximum-likelihood and minimum-distance classifiers.

However, the above studies did not fully investigate to what extent the differences in classification accuracies obtained by statistical and neural-network classification methods were statistically significant. In addition, the works reported in the literature did not consider complex rural areas, nor the exploitation of texture information merged with spectral and ancillary data.

This paper is aimed at investigating the performances of statistical and neural-network classification approaches using spectral, texture, and ancillary data for land-cover inventorying of a complex rural area in central Italy. The different performances of the two classification approaches are evaluated in terms of global and conditional Kappa accuracies. The Zeta statistics is used to evaluate the statistical significance of results.

The two approaches are based on a modified maximum-likelihood (MML) classifier and on a multilayer perceptron (MLP) neural-network classifier, respectively. The spectral

L. Bruzzone is with the Department of Biophysical and Electronic Engineering, University of Genoa, Via Opera Pia 11a, 16145 Genova, Italy, (lore@dibe.unige.it).

C. Conese is with the Centre for the Application of Computer Science in Agriculture (Ce.S.I.A.), Logge Uffizi Corti, 50122 Firenze, Italy.

F. Maselli is with the Institute of Agrometeorology and Environmental Analysis for Agriculture (I.A.T.A.), C.N.R., P. le delle Cascine 18, 50144 Firenze, Italy.

F. Roli is with the Department of Electrical and Electronic Engineering, University of Cagliari, Piazza d'Armi, 09123 Cagliari, Italy.

Photogrammetric Engineering & Remote Sensing, Vol. 63, No. 5, May 1997, pp. 523–533.

0099-1112/97/6305-523$3.00/0

data are Landsat-5 TM scenes, from which texture features were extracted as grey-level co-occurrence matrix statistics (Haralick *et al.*, 1973). The three layers of a digital terrain model (DTM) (i.e., the elevation, slope, and aspect layers) are considered as ancillary data. The use of a feature-selection technique to identify the best subset of multisource data is also illustrated.

Study Area, and Ancillary, Ground, and Spectral Data

The study area corresponds to the basins of the Piomba and Cigno rivers, in the Abruzzo region, in central Italy (Figure 1). The first is a primary basin and covers about 100 km^2; the second is a tributary of the Pescara river and covers an area of about 60 km^2. The major axes of both basins extend from west to east, following most of the Adriatic hydrological network. The terrain is rugged, with elevations ranging from 10 to 740 m and from 80 to 1440 m, respectively. The topographies of the two basins were derived from 15 orthophotomaps (eight for the Piomba basin and seven for the Cigno basin) produced by the Administration of the Abruzzo Region at the scale 1:10,000. These maps were claimed to be of high quality and included contours at a 10-m elevation interval (Regione Abruzzo, 1987).

The basins are mainly covered with agricultural lands, pasture lands, and coniferous and deciduous forests. The main crops are maize, winter wheat, alfalfa, and grape and olive groves, which are generally distributed in the lower areas, whereas pasture lands and forests are distributed in the more rugged, less accessible areas. The ground reference data for the two basins were collected directly by making accurate ground surveys in the spring and summer of 1993. Small plots of approximately 2 to 5 ha were identified during these surveys. Particular attention was paid to the selection of samples representative of the extent and distribution of the land-cover categories in the two basins. Eleven categories, representative of the land cover in the two basins, were identified (Table 1). In addition to the agricultural, forest, and urban classes, a "vegetated badland" class and a "bare badland" class were also considered in order to characterize the environmental degradation phenomenon that occurs in both study areas.

TABLE 1. LAND-COVER CATEGORIES CONSIDERED. FOR EACH CATEGORY, THE PIXELS OF THE TWO BASINS DIGITIZED FROM GROUND REFERENCES WERE DIVIDED INTO A TRAINING SET AND A TEST SET.

	Land-Cover Classes	Number of training pixels	Number of test pixels
1	Winter wheat	984	979
2	Root crop cultivation	145	111
3	Meadow and pasture	642	563
4	Vineyard and orchard	628	521
5	Olive grove with cereals	433	451
6	Bare soil	578	621
7	Riparian forest	154	130
8	Hilly forest	702	752
9	Vegetated badland	162	105
10	Bare badland	199	199
11	Urban area	246	246

As spectral data, we used two Landsat-5 TM quarter scenes (the latitude and the longitude of the center of the quarter scenes were 42°20' and 13°50', respectively). Both

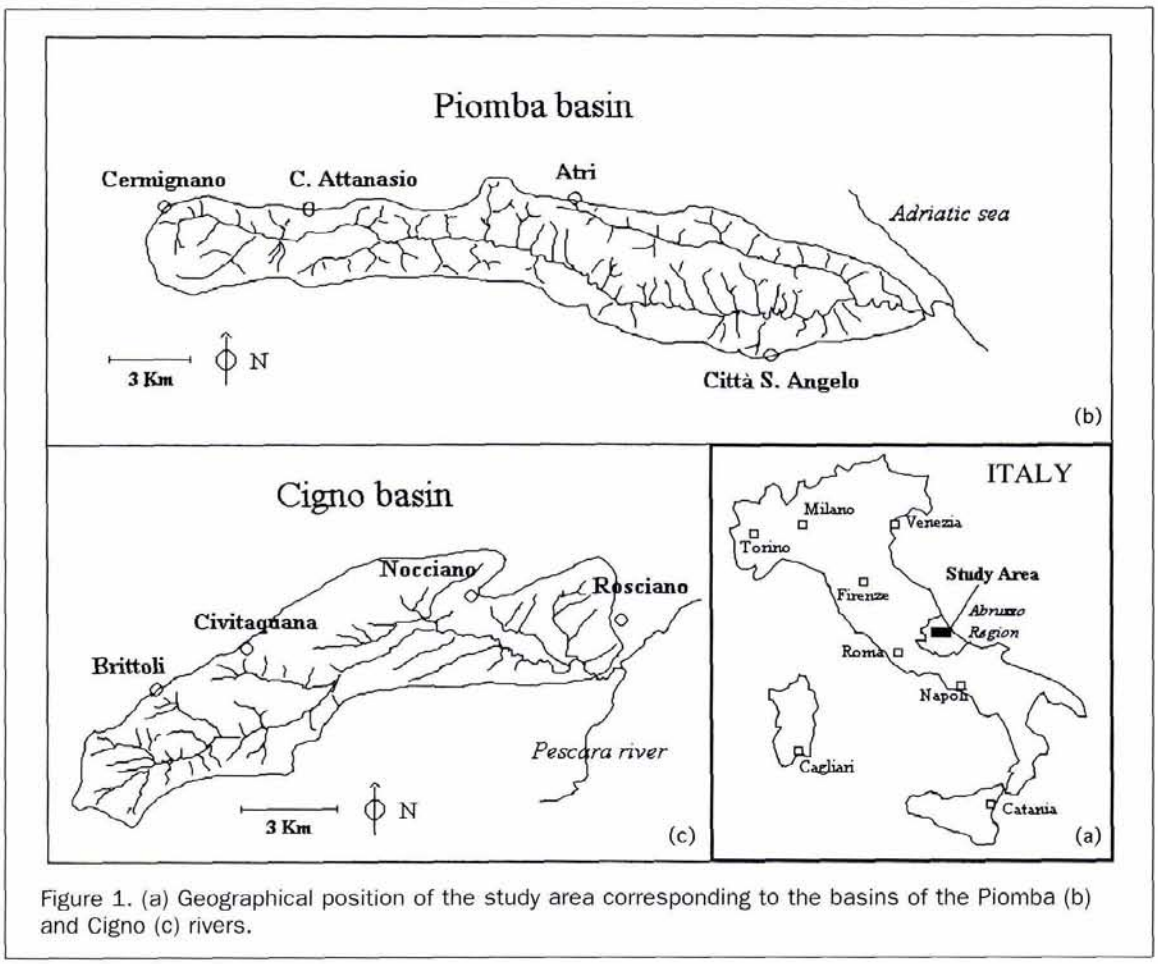

Figure 1. (a) Geographical position of the study area corresponding to the basins of the Piomba (b) and Cigno (c) rivers.

scenes were cloudless and of good quality. The first scene was taken in spring (26 May 1993), when the vegetation reached the peak of its photosynthetic activity, and the second in summer (14 August 1993), during the driest Mediterranean period. As an example, the two multitemporal Band-5 images for both the Cigno and Piomba basins are shown in Figure 2. The multitemporal information in the scenes relating to the same growing season was considered useful to differentiate the cover types in the study area.

Preliminary Data Processing

The preprocessing and information-extraction steps are illustrated in Figure 3. The same preliminary processing was performed for both classification approaches (which were used by the Classification Module shown in Figure 3).

The contour lines of the orthophotomaps were acquired by means of a digitizer table. A *module* (indicated as the DTM Generation Module in Figure 3) took as input such digitized orthophotomaps and provided as output the DTMs of the Piomba and Cigno basins (Geosystems, 1992). A 30- by 30-m pixel size was selected for the DTM data to match the spatial resolution of the TM data. The ground references were digitized using the same pixel size (30 by 30 m). The reference pixels of the two basins were then divided into a training set and a test set by a stratified random sampling technique (Table 1).

The TM scenes were georeferenced to the orthophotomaps. Ground control points in the orthophotomaps were selected for the geometric correction of the Landsat TM data. A bilinear interpolation algorithm was used to resample and georeference the satellite data, with a final pixel size of 30 by 30 m and RMSEs of approximately 1 pixel.

The geometrically corrected and georeferenced TM images were given as input to the Texture Feature Extraction Module, which computed a set of texture features.

At this point of the processing chain, a large set of features (i.e., TM channels, texture features, and ancillary data) was available to characterize the areas to be classified. To reduce the cost of the classification process, we employed the Feature Selection Module (Figure 3) to automatically select the most effective feature subset. The selection concerned spectral and texture features only. DTM data were not considered in the selection process, as the MML approach needed all such data for the computation of the *a priori* probabilities of the land-cover classes.

Texture Analysis

For the present study, we used texture features computed from the grey-level co-occurrence matrix (GLCM) (Haralick *et al.*, 1973). The GLCM constitutes a statistical approach to texture computation that has been successfully tested on remote sensing images for applications such as land-cover mapping of forested and agricultural areas (Peddle and Franklin, 1989; Jacques *et al.*, 1989; Marceau *et al.*, 1990).

The use of the GLCM approach requires the choice of a given number of parameter values for the computation of the grey-level co-occurrence matrix (e.g., the distance and the angle between two resolution cells) (Haralick *et al.*, 1973). Owing to the fine textures of our TM images, we performed the GLCM computation by using one-pixel-distance grey-level spatial-dependence matrices and a 5 by 5 moving window. The texture was assumed to be isotropic and was computed for an angle of zero degrees only. The original 256 grey levels were mapped into 128 levels by an equal-probability quantizing algorithm in order to reduce the cost of the computation of the co-occurrence matrix. Among the features that can be derived from the GLCM, the whole set of 14 texture features suggested by Haralick *et al.* (1973) was computed.

In principle, all the 14 texture features could have been computed for each of the TM channels of our multitemporal data set, so obtaining a set of 196 texture features (14 texture features by seven TM channels by two scenes). In practice, in order to reduce the resulting computational cost, we computed the 14 texture features for only one of the seven TM channels. In order to automatically select this channel, for each TM band, we computed the mean value of the data class grey-level variance (i.e., the grey-level variance was computed for each land-cover class and the mean value was estimated). The channel with the maximum value of such a "mean variance" was selected. This was regarded as a simple but reasonable strategy to automatically select the TM channel providing the largest amount of texture information, as grey-level variance is a simple measure of texture information. TM channel 5 (i.e., the infrared channel) was selected according to such a strategy.

Feature Selection

Several feature-selection criteria have been proposed in the literature (Fukunaga, 1990; Richards, 1993). For our study, we adopted the feature-selection criterion proposed by Serpico *et al.* (1994), as experimental comparisons with classical criteria proved its greater effectiveness (Serpico *et al.*, 1994). Such a criterion does not transform the original feature space, thus preserving the physical meanings of the features selected. In this way, at the end of the selection, we had information about the most useful original features to discriminate among the land-cover classes considered. The criterion is based on the minimization of the following index:

$$g(f) = \sum_{i=1}^{C} \sum_{j>i}^{C} [p(\omega_i) + p(\omega_j)] Q\left(\frac{d_{ij}(f)}{2}\right) \qquad (1)$$

where f is the vector of the features to be selected, C is the number of classes considered, $p(\omega_i)$ and $p(\omega_j)$ are the *a priori* probabilities of the classes, $Q(x)$ is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} e^{-\frac{\xi^2}{2}} d\xi, \qquad (2)$$

and $d_{ij}(f)$ is given by

$$d_{ij}^2(f) = (\mathbf{m}_i - \mathbf{m}_j)^T \left(\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right)^{-1} (\mathbf{m}_i - \mathbf{m}_j) \qquad (3)$$

where $\mathbf{m}_i$, $\mathbf{m}_j$ and $\boldsymbol{\Sigma}_i$, $\boldsymbol{\Sigma}_j$ are the mean vectors and the covariance matrices for classes ω_i, and ω_j, respectively.

A search algorithm is used to find a subset of features that minimizes the above criterion. In the literature, both optimum search algorithms (e.g., branch and bound methods) and suboptimum ones (e.g., forward selection (FS) and backward selection (BS)) have been proposed (Fukunaga, 1990). In our case, the large set of features considered (42 features: 14 texture features and seven TM channels for each of the two multitemporal scenes) would have made it very time consuming to apply an optimum search algorithm. Therefore, we chose the FS search.

The minimization for our feature selection criterion was obtained for a set of 17 features.

The selected features are given in Table 2. It is worth noting that several texture features were selected; this confirms their importance for the application considered.

Classification Approaches

Two different classification approaches were used. The first was a parametric statistical approach modified by including nonparametric *a priori* probabilities. The second was a nonparametric approach based on a multilayer perceptron neural

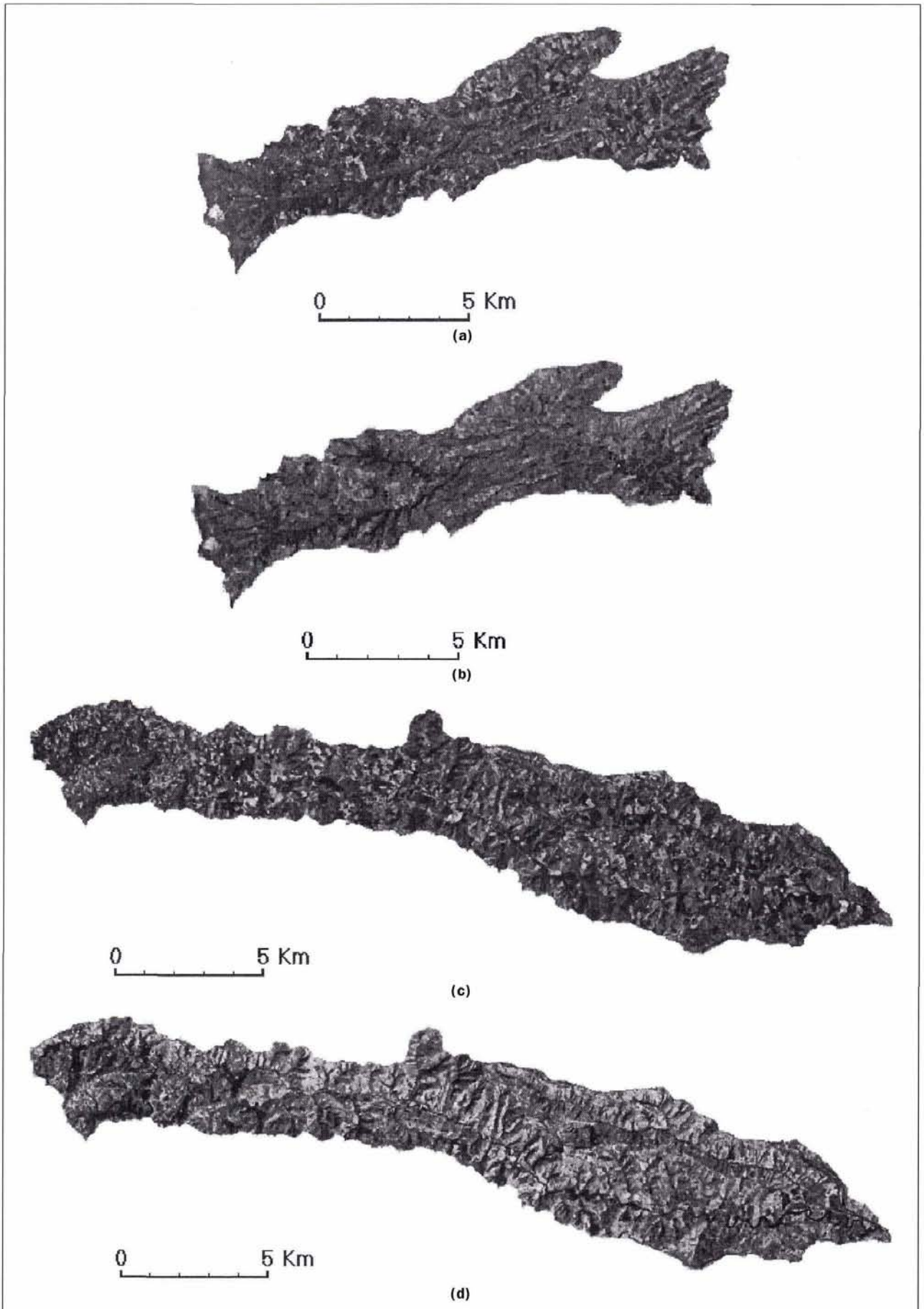

Figure 2. Band 5 of TM for both the Cigno (a and b) and Piomba (c and d) basins. The scenes were acquired in spring (a and c) and summer (b and d).

network. In the following, the main characteristics of both approaches are outlined.

The Modified Maximum-Likelihood Approach

The parametric classifier adopted for the present study is based on the Gaussian maximum-likelihood (GML) approach. This method is widely used, especially for practical applications (Swain and Davis, 1978; Curran, 1985), and can be regarded as a reference for comparing the performances of different classification approaches. The classical GML approach was modified by including nonparametric *a priori* probabilities (Maselli *et al.*, 1992). This modification makes the classification process more flexible and more effective in the presence of non-Gaussian spectral distributions, like those related to complex rural areas. The modified *a priori* probabilities were computed as suggested by Skidmore and Turner (1988), that is, they were derived from the grey-level frequency histograms of the land-cover classes considered. In particular, the method devised by Skidmore and Turner (1988) was improved by using ancillary information for the computation of the *a priori* probabilities (Strahler, 1980; Hutchinson, 1982; Maselli *et al.*, 1994). For each pixel, we calculated the product of the frequencies of the spectral and ancillary data on the land-cover classes considered (Maselli *et al.*, 1992; Maselli *et al.*, 1994).

Because the *a priori* probabilities were computed as the product of the frequencies related to different sources of information (i.e., spectral, texture, and ancillary data), a statistical independence was necessary (Maselli *et al.*, 1994). The statistical independence of the ancillary data used in this study can be reasonably assumed. For the spectral and texture features, the independence was achieved by assuming nearly Gaussian distributions and performing a principal component transformation (PCT) on the original features, which had already proved effective for this purpose (Maselli *et al.*, 1992). It is worth noting that the PCT was used only to remove the correlation among the original features, and that no transformed component was disregarded, that is, the PCT was not used to perform feature selection.

TABLE 2. THE SETS OF SPECTRAL AND TEXTURE FEATURES IDENTIFIED BY THE FEATURE-SELECTION ALGORITHM. THE SELECTED FEATURES ARE ORDERED ACCORDING TO THE AMOUNT OF INFORMATION PROVIDED. THE NAMES OF THE TEXTURE FEATURES ARE THE SAME AS USED BY HARALICK *ET AL.* (1973).

Selected Features	Scene Acquired in
TM channel 1	Spring
TM channel 7	Summer
TM channel 5	Spring
TM channel 7	Spring
TM channel 4	Summer
texture feature "Sum of Averages"	Summer
texture feature "Difference of Variances"	Summer
texture feature "Correlation"	Spring
texture feature "Contrast"	Spring
texture feature "Sum of Variances"	Summer
TM channel 4	Spring
texture feature "Variance"	Spring
texture feature "Sum of Averages"	Spring
TM channel 2	Summer
TM channel 3	Summer
TM channel 1	Summer
texture feature "Difference of Entropy"	Summer

The Neural-Network Approach

In this study, we used multilayer feedforward networks (also called "multilayer perceptrons" (MLPs)), trained by means of the error backpropagation learning algorithm (Benediktsson *et al.*, 1990; Hertz *et al.*, 1991; Serpico and Roli, 1995; Foody *et al.*, 1995; Chen *et al.*, 1995).

MLPs are artificial neural-network models resulting from an interconnection of very simple processing elements called network units (or, more simply, neurons). Typically, an MLP network topology consists of multiple layers of neurons with connections only between neurons in neighboring layers. Every connection in the network has a numerical value attached to it that is called "weight." Information is processed starting from one side of the network called "input layer" and moving through successive "hidden layers" to the "out-

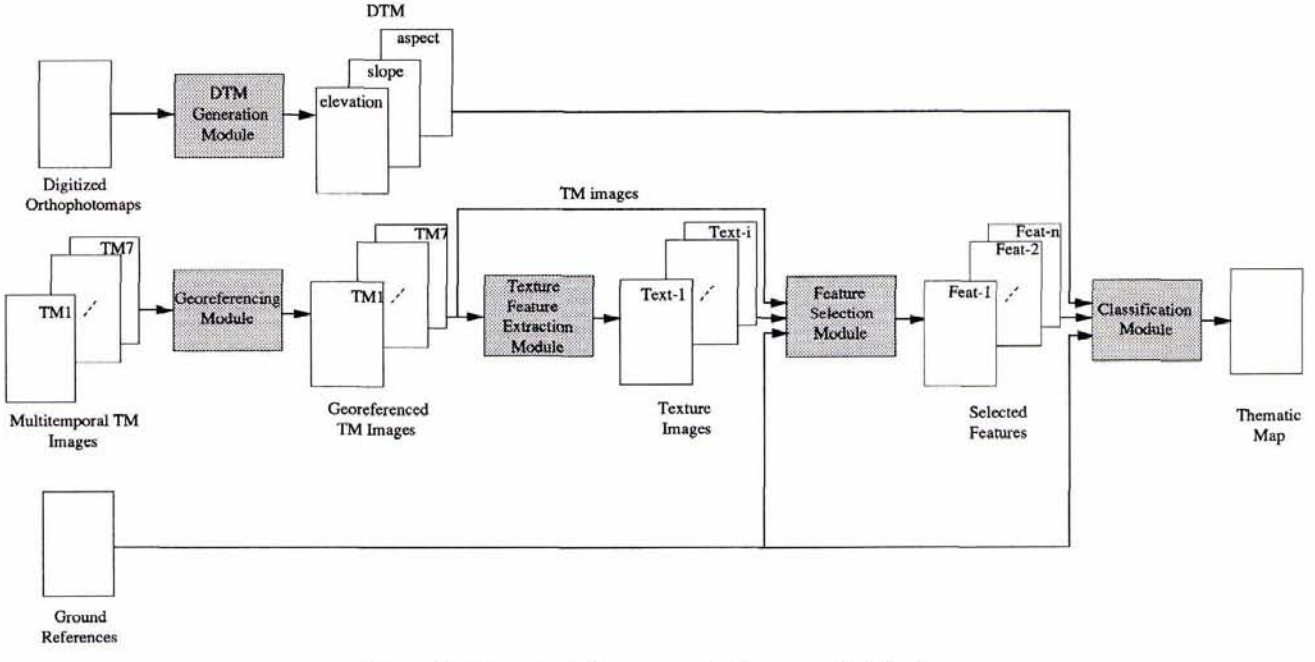

Figure 3. Scheme of the overall data-processing chain.

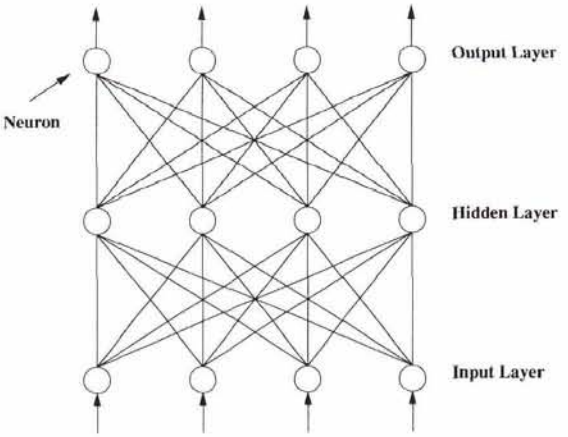

Figure 4. An example of MLP neural-network topology.

put layer." In MLPs, the flow of information is in one direction only (from the input to the output layer); therefore, they are called "feedforward networks." As an example, Figure 4 shows the topology of an MLP neural network with only one hidden layer.

Each neuron is a simple processor which usually has many different input connections, coming from other neurons in the MLP network, and only one output which is usually sent to many other neurons. A neuron i computes a net input (net_i) from the outputs o_j of other neurons and from the weights of the connections. Typically, such a net input is a weighted sum, and a numerical value, called "bias" (b_i), is added to the net input. In MLPs, a function S, called "activation function," is applied to the net input. Formally, denoting by w_{ij} the weight of the connection from neuron i to neuron j, the computation performed by each neuron, except the input ones, can be expressed as follows (Figure 5):

$$net_i = \sum_j w_{ij} o_j + b_i \tag{4}$$

$$o_i = S(net_i) \tag{5}$$

In our experiments, we used a sigmoid function as the activation function: i.e.,

$$S(net_i) = (1 + e^{-net_i})^{-1} \tag{6}$$

as is usually done for MLPs (Hertz *et al.*, 1991).

One of the most commonly used training schemes for MLPs is the error backpropagation (EBP) learning algorithm. It is a learning algorithm by examples. In the most commonly used version, called "training by patterns," a set of training samples representative of each of the data classes to be recognized is selected and repeatedly given as input to the network (each training iteration is called "epoch"). The EBP algorithm adjusts the values of the network connections so that the activations of the output neurons matches more closely the desired output values. This allows a correct classification of each training sample. Typically, there is an output neuron for each data class, and an input sample is classified as belonging to a given class if the related output neuron has the highest activation among all the output neurons. Therefore, for each input sample, the EBP algorithm aims to maximize the activation value of the neuron related to the correct class and to minimize the activation values of all the other output neurons. This is usually performed by minimizing iteratively the following summed squared error (SSE) function:

$$\text{SSE} = \sum_p \sum_k (o_{pk} - t_{pk})^2 \tag{7}$$

where p indexes the training samples and k indexes the output neurons of the network. In many cases, the mean SSE, called MSE, is used; it is computed by averaging the SSE function over the training samples. For each training sample, the terms t_{pk}, called "targets," represent the desired activation values for the output neurons (typically, just one target is set to one and the others to zero). The EBP algorithm minimizes the above error function by a gradient descent technique which changes the weight values according to the following rule:

$$\Delta w_{ij} = -\varepsilon \frac{\partial \text{SSE}}{\partial w_{ij}} \tag{8}$$

where Δw_{ij} is the weight change and ε is the so-called "learning rate."

Detailed discussions of the equations implementing such a minimization technique and of the complete derivation of the weight updating rule can be found in Hertz *et al.* (1991).

Experimental Results

Experimentation Planning

Experiments were performed in order to assess both the importance of texture and ancillary data to characterize complex rural areas and to evaluate the degrees of effectiveness of both classification approaches in exploiting such multisource information. Therefore, classification accuracies were evaluated by characterizing the data set in Table 1 using the following sets of features:

- spectral features (i.e., the nine TM channels in Table 2);
- spectral and texture features (i.e., the features in Table 2);
- spectral and ancillary features (i.e., the nine TM channels in Table 2 and elevation, slope, and aspect features); and
- spectral, texture, and ancillary features (i.e., the 17 features in Table 2 and the elevation, slope, and aspect features).

The application of the MML approach was preceded by a PCT on the spectral and texture data in order to produce uncorrelated features. It is worth noting that the ancillary data were used only to modify the *a priori* probabilities of the land-cover classes (Maselli *et al.*, 1994), as the statistical distribution of such data was very far from normal. Consequently, a maximum number of 17 features were used as input to the MML classifier.

For the NN approach, several MLP architectures were considered to select the best for each of the above sets of features. In all cases, the number of input units was equal to the number of input features. All the networks had eleven

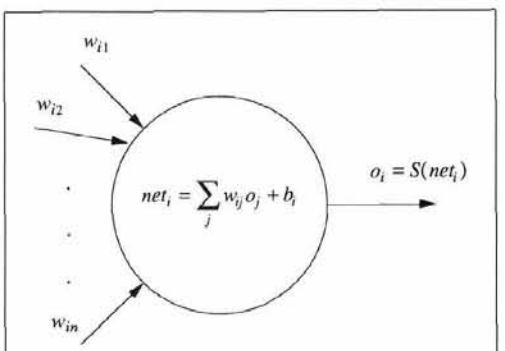

Figure 5. Neuron computation.

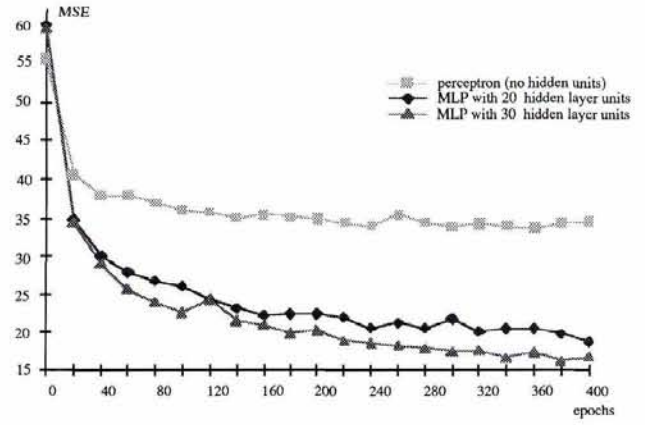

Figure 6. Mean-square-error (MSE) decay as a function of the number of training epochs. Three curves are shown that are related to three MLP networks with no hidden units, 20 hidden units, and 30 hidden units, respectively. All the networks had 20 input units and 11 output units.

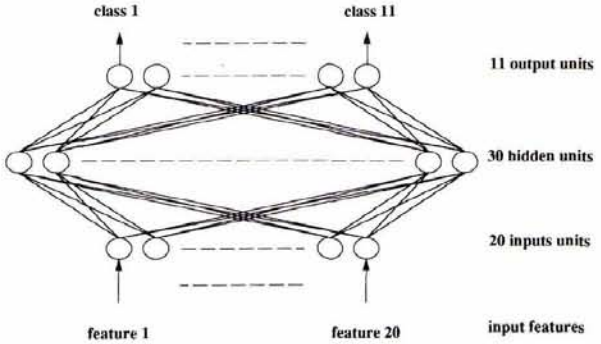

Figure 7. The MLP neural network selected to classify the data set characterized by spectral, texture, and ancillary features (i.e., the 17 features in Table 2 and three ancillary features). An architecture with 20 input units (one for each input feature), 30 hidden units, and 11 output units (one for each land-cover class) was adopted.

output units, one unit for each data class. All the architectures were such that the number of weights was much smaller than the number of pixels in the training set, in accordance with the simplified rule suggested in Baum and Haussler (1989). All the neural networks were trained by the EBP learning algorithm according to a "training by pattern" criterion and using different "learning rates" (i.e., 0.1, 0.05, and 0.01) (Hertz *et al.*, 1991). As a convergence criterion, an MSE smaller than 0.05 was required. Training was stopped when convergence was reached, or after 400 "epochs." Some examples of MSE decay are given in Figure 6: they are related to the networks used to classify the data set composed of the 20 multisource features. For classification purposes, each test pixel was assigned to the class corresponding to the output unit with the highest activation. Concerning the data set characterized by the whole feature set (i.e., spectral, texture, and ancillary features), the best classification performances for the considered test set (Table 1) were achieved by using a three-layer architecture with 30 hidden units and at a learning rate equal to 0.01 (Figure 7).

The performances of both classification approaches were assessed by using error matrices. The overall and conditional Kappa coefficients of agreement were derived from such matrices as described by Congalton *et al.* (1983), Hudson and Ramm (1987), and Rosenfield and Fitzpatrick-Lins (1986). The Kappa coefficient was regarded as an effective measure to assess the degrees of usefulness of the different types of features used and the degrees of effectiveness of the two classification approaches in exploiting them (Rosenfield and Fitzpatrick-Lins, 1986). The Zeta statistics was used to assess quantitatively if the differences in the Kappa values obtained in our experiments were statistically significant (Congalton *et al.*, 1983). In particular, this statistical test was carried out to compute the significance of the differences in classification accuracy when texture and ancillary data were used. The Zeta statistics was also used to assess the significance of the differences in classification accuracy between the NN and MML approaches.

Experimental Results

The classification accuracies obtained for the test set containing only the spectral data were unsatisfactory (Table 3a), with a Kappa coefficient less than 0.6 for both classification approaches. Several land-cover classes were poorly classified. The MML approach suffered from the high spectral complexities of the rural areas considered, so it yielded larger differences among data-class accuracies than the ones provided by the NN approach.

Considerable improvements were obtained by adding ancillary or texture data to the spectral-feature set. As an example, the value of the Kappa coefficient for the NN approach increased from 0.590 to 0.681 when ancillary features were added (Table 3a). For both classification approaches, each of these two types of data increased the values of the Kappa coefficients by approximately 0.1, as compared with the values obtained by using spectral data only (Table 3a). A large in-

TABLE 3. VALUES OF THE KAPPA COEFFICIENTS AND OF THE ZETA STATISTICS RELATED TO THE APPLICATIONS OF THE TWO CLASSIFICATION APPROACHES USING DIFFERENT FEATURE SETS TO CHARACTERIZE THE SELECTED DATA SET. (A) VALUES OF THE GLOBAL KAPPA COEFFICIENT AS A FUNCTION OF THE CLASSIFICATION APPROACH AND OF THE FEATURE SET USED. (B) VALUES OF THE ZETA STATISTICS TO ASSESS THE SIGNIFICANCE OF THE DIFFERENCES IN KAPPA ACCURACY WHEN USING DIFFERENT FEATURE SETS (THE SPECTRAL FEATURE SET IS UTILIZED AS A REFERENCE SET). (C) VALUES OF THE ZETA STATISTICS TO ASSESS THE SIGNIFICANCE OF THE DIFFERENCES IN KAPPA ACCURACY BETWEEN THE NN AND THE MML APPROACHES USING DIFFERENT FEATURE SETS. ALL DIFFERENCES ARE SIGNIFICANT AT THE 99 PERCENT CONFIDENCE LEVEL.

Kappa Coefficient	Spectral Data	Spectral and Ancillary Data	Spectral and Texture Data	Spectral and Ancillary and Texture Data
MML	0.529	0.608	0.616	0.653
NNs	0.590	0.681	0.699	0.765

(a)

Zeta Statistics	Spectral and Ancillary Data	Spectral and Texture Data	Spectral and Ancillary and Texture Data
MML using Spectral Data	6.91	7.70	11.09
NNs using Spectral Data	8.33	10.07	16.70

(b)

Zeta Statistics	Spectral Data	Spectral and Ancillary Data	Spectral and Texture Data	Spectral and Ancillary and Texture Data
NNs vs. MML	5.28	6.75	7.71	10.91

TABLE 4. ERROR MATRICES FOR THE MML (a) AND THE NN (b) CLASSIFICATION APPROACHES WHEN APPLIED TO THE DATA SET WITH ALL FEATURES.

(a)

Class	1	2	3	4	5	6	7	8	9	10	11	Row total	Commission error (%)
1	790	6	134	37	53	38	16	27	14	8	5	1128	30.0
2	5	62	3	4	3	1	1	0	0	0	0	79	21.5
3	29	7	304	5	34	44	0	30	3	6	12	474	35.9
4	47	21	18	424	46	10	1	2	1	4	23	597	29.0
5	72	14	46	32	296	21	7	13	5	8	14	528	43.9
6	4	0	42	1	2	459	6	76	4	1	0	595	22.9
7	1	0	0	4	1	1	51	39	1	0	0	98	48.0
8	3	1	7	2	7	19	41	548	4	2	0	634	13.9
9	7	0	1	2	5	17	6	9	53	21	9	130	59.2
10	16	0	0	5	2	8	1	8	18	138	38	234	41.0
11	5	0	8	5	2	3	0	0	2	11	145	181	19.9
Column total	979	111	563	521	451	621	130	752	105	199	246	4678	
Omission error (%)	19.3	44.1	46.0	18.6	34.4	26.1	60.8	27.1	49.5	31.7	41.1		

(b)

Class	1	2	3	4	5	6	7	8	9	10	11	Row total	Commission error (%)
1	840	5	76	22	42	40	9	13	8	11	4	1070	21.5
2	4	69	2	7	1	3	0	0	0	0	0	86	19.8
3	44	4	407	7	23	44	1	16	1	6	1	554	26.5
4	22	16	22	433	40	25	3	4	2	0	9	576	24.8
5	40	11	28	32	311	22	10	9	4	5	2	474	34.4
6	1	0	18	0	3	457	1	10	1	1	0	492	7.1
7	5	0	1	4	7	11	102	23	20	0	0	173	41.0
8	6	1	4	1	4	4	3	673	1	0	0	697	3.4
9	6	0	0	0	2	4	0	0	53	5	5	75	29.3
10	3	4	2	1	5	8	1	3	14	155	5	201	22.9
11	8	1	3	14	13	3	0	1	1	16	220	280	21.4
Column total	979	111	563	521	451	621	130	752	105	199	246	4678	
Omission error (%)	14.2	37.8	27.7	16.9	31.0	26.4	21.5	10.5	49.5	22.1	10.6		

crease in the Kappa coefficient was obtained for the NN approach when texture features were added to the spectral feature set (0.699 vs. 0.590).

The independence of the information provided by the texture and ancillary features was indicated by the increase in the Kappa accuracy obtained when such features were simultaneously added to the spectral feature set (Table 3a). For both classification approaches, the classification accuracies improved over the ones obtained by using either ancillary and texture features separately. The increase in accuracy was smaller for the MML approach, for which the Kappa value was 0.653. The NN approach benefited more from the use of the whole multisource data set and reached a Kappa value of 0.765 (Table 3a).

The statistical significance of the differences in classification accuracy due to the texture and ancillary features (Table 3a) was assessed by the Zeta statistics test (Table 3b). All the differences in classification accuracy given in Table 3a were highly significant (Table 3b). This proved both the usefulness of texture and ancillary data in characterizing complex rural areas and the capabilities of both approaches to exploit multisource information. It should be noted that the increases in classification accuracy due to the simultaneous use of ancillary and texture features were highly significant for both classification approaches (11.09 and 16.70 for the MML and NN approaches, respectively).

The Zeta statistics was also used to assess the statistical significance of the greater efficiency of the NN approach in terms of classification accuracy (Table 3c). For all the types of features considered, the greater efficiency of this approach was highly significant (Table 3c). The Zeta values increased when the texture and ancillary features were considered. In particular, the difference in classification accuracy between the MML and NN approaches when using the whole feature set (Kappa=0.653 vs. Kappa=0.765, as given in Table 3a) resulted in a high Zeta value (Zeta = 10.91). This confirmed that the NN approach better handles multisource information.

The performances of the two classification approaches when using the whole feature set (i.e., spectral, texture, and ancillary features) are also described by the error matrices in Tables 4a and 4b. According to the Kappa values in Table 3a, the NN approach yielded higher classification accuracies than the MML approach for most of the land-cover classes.

Plates 1a and 1b present the classification results obtained by the two approaches by using the whole feature set to characterize the basin of the Cigno river. The two thematic maps seem to exhibit a good degree of similarity, even though some differences affect a few land-cover classes. For example, the hilly forest class has similar extents in the two thematic maps but is spatially distributed in different ways. The same is true for the bare-soil class. The riparian forest class was almost completely omitted by the MML approach. These differences can be ascribed to the non-optimal performances of the two approaches for these classes (Figure 8).

Even though the overall classification accuracy obtained by using the whole feature set can be considered acceptable for practical applications, it should be pointed out that misclassification errors were unequally distributed among the land-cover classes (Figure 8). In particular, classes 2, 3, 7, 9, and 11 were poorly classified by the MML approach; the worst results yielded by the NN approach concerned classes 5 and 7. This unequal distribution may be due to the difficulties inherent in the data set used, such as the spectral complexities of the cover classes, the presence of mixed pixels, and possible inaccuracies in the collection of the reference samples and in the resampling of the TM images.

Nevertheless, the contextual information provided by the texture features proved very useful, as well as the information contained in the ancillary data. All this information was better exploited by the NN approach. The main reason for the

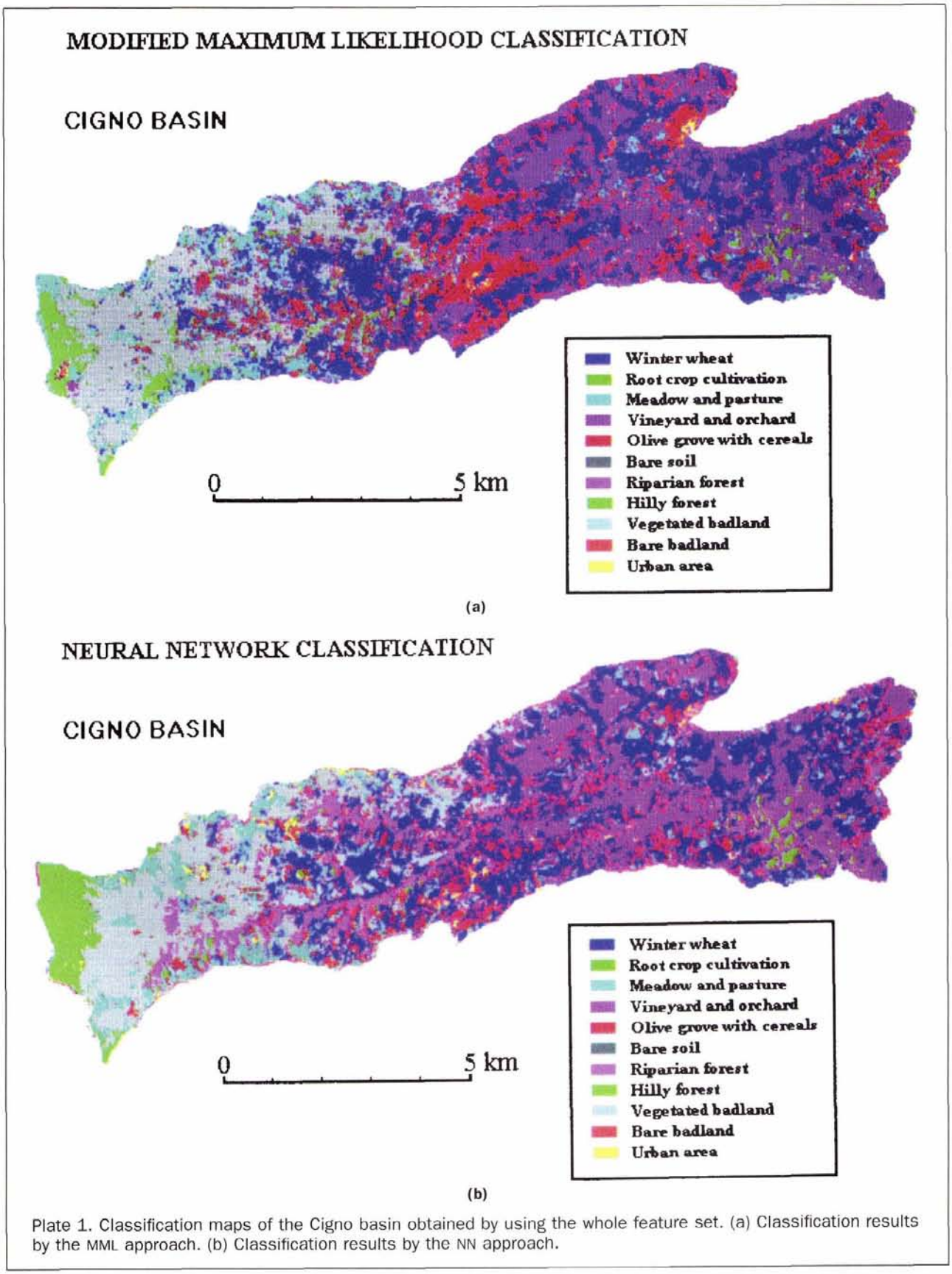

Plate 1. Classification maps of the Cigno basin obtained by using the whole feature set. (a) Classification results by the MML approach. (b) Classification results by the NN approach.

different behaviors of the two approaches is the difficulty with representing texture and ancillary information in parametric form. It is worth noting that the MML approach did not use ancillary data as input features but used them only to modify the *a priori* probabilities of the land-cover classes considered. However, additional experiments not reported in

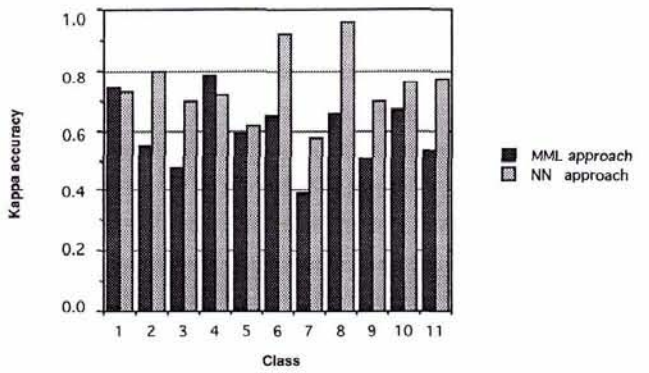

Figure 8. Values of the conditional Kappa coefficients obtained by the two classification approaches by using the whole feature set for each of the 11 land-cover classes.

this paper showed that improvement in classification accuracy could not be obtained by using ancillary features as input data to the MML classifier. This evidence is reasonable, as the statistical distribution of ancillary data is very far from normal.

Discussion and Conclusions

In this paper, two different approaches to the classification of complex rural areas by use of multisource data have been described. The main purpose of our investigation was to quantitatively assess, also from the viewpoint of statistical significance, both the usefulness of multisource data for the characterization of these areas and the capabilities of the two approaches to exploit such data in an effective way. Even though the results of our study cannot be considered final, nevertheless, some interesting conclusions can be drawn.

The reported results point out the importance of ancillary and texture data to characterize complex rural areas. Satisfactory and statistically significant accuracies were achieved only when both texture and ancillary information were used. The spectral complexities and the heterogeneities of the two rural areas resulted in spectral information that was insufficient for an accurate discrimination among the land-cover classes. In particular, the importance of the ancillary data was related to the strong influence of the environmental factors on the distribution of the land-cover types. The structural differences in the land-cover types explain the discrimination potential of the texture features. Similar situations can be assumed to occur in many other rural areas, for which the use of multisource data is then particularly appropriate.

The two classification approaches described in this paper have proved to be suited for multisource classification of complex rural areas. The NN approach provided better overall and conditional accuracies than did the MML approach, and this difference in accuracy increased when texture and ancillary data were used. The high Zeta values reported proved the statistical significance of the effectiveness of the NN approach in exploiting multisource data. This indicates that the nonparametric approach used can provide more satisfactory accuracies than the parametric MML in practical applications. On the other hand, the NN approach requires a complex and expensive design phase (e.g., concerning the choice of a suitable architecture) and a much longer training time. Moreover, the high sensitivity of the NN approach to unrepresentative training samples (Bischof *et al.*, 1992) should be taken into account, especially in practical applications.

Acknowledgments

This research was partially conducted within the framework of the cooperative research project MEDALUS II (Mediterranean Desertification and Land Use), funded by the EEC (Environment Programme, Contract EV5V-CT92-0166), and partially within the framework of the Italian project "Sviluppo di metodi integrati di classificazione agroecologica tramite dati di telerilevamento per la gestione delle risorse naturali," funded by the Italian Space Agency. Both supports are gratefully acknowledged.

The authors wish to thank Dr. P. Pellegretti for his helpful suggestions and comments, Dr. A. Rodolfi for her cooperation in processing ground data, and the anonymous reviewers for their constructive criticisms.

References

Baum, E.B., and D. Haussler, 1989. What size network gives valid generalization, *Neural Computation*, 1:151–160.

Benediktsson, A., P.H. Swain, and O.K. Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):550–552.

Bischof, H., W. Schneider, and A.J. Pinz, 1992. Multispectral classification of Landsat images using neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–490.

Chen, K.S., Y.C. Tzeng, C.F. Chen, and W.L. Kao, 1995. Land-cover classification of multispectral imagery using a dynamic learning neural network, *Photogrammetric Engineering & Remote Sensing* 61(4):403–408.

Civco, D.L., and Y. Wang, 1994. Classification of multispectral, multitemporal, multisource spatial data using artificial neural networks, *Proc. 1994 Annual ASPRS/ACSM Convention*, Reno, Nevada, 1:123–133.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, *Photogrammetric Engineering & Remote Sensing*, 49(12):1671–1678.

Curran, P.J., 1985. *Principles of Remote Sensing*, Longman, New York.

Foody, G.M., M.B. McCulloch, and W.B. Yates, 1995. Classification of remotely sensed data by an artificial neural network: Issues related to training data characteristics, *Photogrammetric Engineering & Remote Sensing*, 61(4):391–401.

Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, Second Edition, Academic Press Inc., New York.

Geosystems, 1992. *Software Geosys*, Geosystems S.r.l., Florence, Italy.

Gong, P., 1994. Integrated analysis of spatial data from multiple sources: An overview, *Canadian Journal of Remote Sensing*, 20(4):349–359.

Haralick, R.M., K. Shanmugan, and I. Dinstein, 1973. Textural features for image classification, *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621.

Hertz, J., A. Krogh, and R.G. Palmer, 1991. *Introduction to the Theory of Neural Computation*, Addison Wesley Pub. Co., The Advance Book Program.

Hudson, W.D., and C.W. Ramm, 1987. Correct formulation of the Kappa coefficient of agreement, *Photogrammetric Engineering & Remote Sensing*, 53(4):421–422.

Hutchinson, C.F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement, *Photogrammetric Engineering & Remote Sensing*, 48(1):123–130.

IAPR TC7, 1992. *Proc. Int. Workshop on Multisource Data Integration in Remote Sensing for Land Inventory Applications*, Wageningen Agricultural Univ. Pub., Delft, The Netherlands.

Jacques, P., M. Massart, and J. Wilmet, 1989. Approache multitemporelle et texturale des donnes SPOT multispectrales pour le suivi agricole en Afrique centrale, *Proc. IGARSS 89*, Vancouver, B.C., 2:480–483.

Lee, T., J.A. Richards, and P.H. Swain, 1987. Probabilistic and evi-

dential approaches for multisource data analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 25(3):283–293.

Marceau, D.J., P.J. Howarth, J.M. Dubois, and D.J. Gratton, 1990. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):513–519.

Maselli, F., C. Conese, L. Petkov, and R. Resti, 1992. Inclusion of prior probabilities derived from a nonparametric process into the maximum likelihood classifier, *Photogrammetric Engineering & Remote Sensing*, 58(2):201–207.

Maselli, F., C. Conese, T. De Filippis, and M. Romani, 1994. Integration of ancillary data into a maximum likelihood classifier with nonparametric priors, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(2):2–11.

NASA, 1990. *Proc. Int. Workshop on Multisource Data Integration in Remote Sensing*, NASA Conference Publication 3099, Maryland.

Peddle, D.R., and S.E. Franklin, 1989. High resolution satellite image texture for moderate relief terrain analysis, *Proc. IGARSS 89*, Vancouver, B.C., 2:653–654.

Regione Abruzzo, 1987. *Ortofotocarte, Scale 1:10,000*, SELEA Ed., Florence, Italy.

Richards, J.A., 1993. *Remote Sensing Digital Image Analysis*, Second Edition, Springer-Verlag, New York.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.

Serpico, S.B., P. Pellegretti, and L. Bruzzone, 1994. Feature selection for remote-sensing data classification, *Proc. of the European Symposium on Satellite Remote Sensing (EUROPTO94)*, Rome, Italy, 2315:569–577.

Serpico, S.B., and F. Roli, 1995. Classification of multisensor remote-sensing images by structured neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):562–578.

Skidmore, A.K., and B.J. Turner, 1988. Forest mapping accuracies are improved using a supervised non parametric classifier with SPOT data, *Photogrammetric Engineering & Remote Sensing*, 54: 1415–1421.

Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data, *Remote Sensing of Environment*, 10:135–163.

Swain, P.H., and S.M. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York.

Wang, M., and P.J. Howarth, 1994. Multi-source spatial data integration: problems and some solutions, *Canadian Journal of Remote Sensing*, 20(4):360–367.

Zhuang, X., B.A. Engel, M.F. Baumgardner, and P.H. Swain, 1991. Improving classification of crop residues using digital land ownership data and Landsat TM imagery, *Photogrammetric Engineering & Remote Sensing*, 57(11):1487–1492.

(Received 7 February 1995; revised and accepted 18 April 1996; revised 24 July 1996)