Approaching the Ocean Color Problem Using Fuzzy Rules

Marco Cococcioni, Giovanni Corsini, *Member, IEEE*, Beatrice Lazzerini, *Member, IEEE*, and Francesco Marcelloni

***Abstract*—In this paper, we propose a fuzzy logic-based approach which exploits remotely sensed multispectral measurements of the reflected sunlight to estimate the concentration of optically active constituents of the sea water. The relation between the concentrations of interest and the subsurface reflectances is modeled by a set of fuzzy rules extracted automatically from the data through a two-step procedure. First, a compact initial rule base is generated by projecting onto the input variables the clusters produced by a fuzzy clustering algorithm. Then, a genetic algorithm is applied to optimize the rules. Appropriate constraints maintain the semantic properties of the initial model during the genetic evolution. Results of the application of the fuzzy model obtained from data simulated with an ocean color model over the channels of the MEdium Resolution Imaging Spectrometer are shown and discussed.**

***Index Terms*—Fuzzy clustering, fuzzy modeling, genetic algorithms, ocean color, remote sensing, TSK-systems.**

I. Introduction

THE OCEAN COLOR analysis aims to determine the concentration of sea water constituents through remotely sensed measurements of the reflected sunlight at specific wavelengths in the visible spectrum. The increasing number of advanced satellite sensors (such as NASA's Sea-viewing Wide Field of view Sensor (SeaWiFS) and the MEdium Resolution Imaging Spectrometer (MERIS) on board the ESA-ENVISAT satellite) makes it possible to exploit large amounts of optical data to support ocean color studies and applications, including monitoring of upwelling phenomena, algae blooms, and marine environmental changes.

Two oceanographic cases can be distinguished [1]: 1) *case I waters*, which correspond to open oceans, where the water optical properties mainly depend on phytoplankton and 2) *case II waters*, which correspond to coastal areas, where the water color also depends on suspended nonchlorophyllous particles and dissolved organic matter (or "yellow substance").

In this paper we are concerned with the color of case II waters, which is much more complex to analyze with classical estimation algorithms than the open ocean water case. The concentrations of the three water constituents (namely, chlorophyll and other pigments contained in phytoplankton, nonchlorophyllous particles, and yellow substance) influence the spectral distribution of the solar light scattered by the water body. According to the widely accepted bio-optical model proposed by Sathyendranath, Prieur, and Morel in [2], the subsurface spectral reflectance R, which is a function of the water constituents, can be described by $R(\lambda) = f(C, X, Y; \lambda)$, where λ is the wavelength, C the chlorophyll concentration, X the scattering coefficient of nonchlorophyllous particles measured at 550 nm, and Y the absorption coefficient of yellow substance measured at 440 nm.

The previous relation represents the sea color direct model. In this paper, we are concerned with the inverse problem, that is we want to estimate the optical parameters from the measured reflectances. As the concentration of each constituent varies independently over a wide range of values, and the relation between the reflectances and the constituents concentration is strongly nonlinear [2], [3], the estimation process is quite complex. Soft computing-based approaches have been recently employed.

Gross *et al.* propose a multilayer perceptron (MLP) to compute phytoplankton pigment concentration (chlorophyll-a plus phaeophytin) from satellite-derived marine reflectances [4]. They apply the MLP to two types of simulated data: The first type is generated by using the bio-optical model and executing simulations in the SeaWiFS sensor spectral bands [5]. The second type is obtained by adding a three-component noise to the first type data. The comparison between MLP and the cubic polynomial that is usually used to perform the identification task shows that MLP guarantees better results for both types of data. Fonlupt uses genetic programming (GP) to estimate phytoplankton concentration in waters of cases I and II [6]. He shows that GP outperforms traditional polynomial fits and rivals artificial neural nets on simulated data in the SeaWiFS sensor spectral bands. Cipollini *et al.* exploit a radial basis function (RBF) neural network to estimate concentrations of phytoplankton, nonchlrophyllous particles and yellow substance in waters of cases I, II, and I+II [3]. They show that RBF outperforms classical algorithms, such as band-ratio, single-band, and multilinear algorithms, typically used in this type of identification problems, on simulated data in the MERIS sensor spectral bands. In particular, in case II waters, where the problem is strongly nonlinear, they prove that the mean square error (mse) can be reduced by one or two orders of magnitude over the error of band-ratio and multilinear algorithms, respectively.

In this paper, we propose an integration of fuzzy logic and genetic algorithms (GAs) to solve the inverse problem. Fuzzy logic has been already applied in the field of ocean color remote sensing, but with different aims. In [7], a knowl-

Manuscript received November 25, 2002; revised July 14, 2003. This work was supported by the Italian Ministry of University and Research (MIUR) within the framework of the PRIN project "Classificazione e analisi di immagini iperspettrali in sistemi di telerilevamento." This paper was recommended by Associate Editor D. Goldgof.

The authors are with the Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, 2-56122 Pisa, Italy (e-mail: m.cococcioni@iet.unipi.it; g.corsini@iet.unipi.it; b.lazzerini@iet.unipi.it; f.marcelloni@iet.unipi.it).

Digital Object Identifier 10.1109/TSMCB.2003.822959

edge-guided segmentation and a labeling approach based on an unsupervised fuzzy clustering algorithm in conjunction with image processing techniques are applied to coastal zone color scanner (CZCS) images to determine the boundaries of ocean zones with different water types. In [8], the well-known fuzzy C-means (FCM) algorithm [9] is applied to select and blend satellite ocean color algorithms. Since a universal bio-optical algorithm applicable to all water types is not feasible, different specific algorithms are used for different water types. Thus, it is necessary to identify the correct water type for each satellite image pixel. To this aim, the authors propose first to determine clusters in *in situ* measurements of remote sensing reflectances by the FCM algorithm, and to associate each cluster with a water type. Then, a fuzzy classification method allows assigning satellite image pixels to clusters with which they share spectral characteristics. Thus, for each pixel, the ocean color algorithm, which has been associated with the specific ocean water type, is applied. Since the fuzzy classification method can assign a pixel to a number of plausible classes with different membership values, all the pertinent ocean color algorithms are used. The membership values to each class are used to weight the output of each algorithm, so as to obtain the best estimation of the constituents affecting the upwelling spectral radiance, particularly in the transition zones. Unlike [7], [8], which use fuzzy logic to select and possibly blend satellite ocean color algorithms, we propose a fuzzy system to estimate ocean color for case II waters.

We model the relation between reflectance and optically active constituents by means of a set of fuzzy rules automatically extracted from available data. The rules, which are expressed in Takagi-Sugeno-Kang (TSK) style [10], are identified using a version of the method proposed in [11]. First, we apply an appropriately modified version of the FCM clustering algorithm to extract a compact initial rule-based model from the data. Then, rules are refined using a GA. To preserve the semantic properties of the initial model appropriate constraints on the partition of the input space are forced during the genetic evolution. At the end of the optimization process, the extracted rules can be easily associated with a physical meaning, thus leading transparency to the rules themselves.

We prove the effectiveness of our approach using a set of simulated subsurface reflectances over spectral channels centered around prefixed wavelengths in the visible spectrum of MERIS, the new generation sensor which is on board the ESA-ENVISAT satellite launched in March 2002. More precisely, MERIS provides a set $\boldsymbol{R} = [\overline{R}_1, \ldots, \overline{R}_M]^T$, where $M = 15$ is the number of spectral channels, of estimates of $R(\lambda)$ at prefixed wavelengths λ_j. Actually, $\overline{R}_j, j = 1 \ldots M$ is the average subsurface reflectance over the jth spectral channel centered around the λ_j wavelength. To define the estimation model, we have used a data set (simulated over the channels of MERIS with the method proposed in [2]) consisting of 5000 pairs $(\boldsymbol{R}, y)$, with $\boldsymbol{R} = [\overline{R}_1, \ldots, \overline{R}_8]^T$ and $y = C, X$ or Y. The eight average subsurface reflectances correspond to the eight channels of MERIS in the visible spectrum.

We demonstrate that the results of the application of the model to the estimation of the concentrations of chlorophyll, dissolved organic matter and suspended nonchlorophyllous particles have revealed good generalization properties and estimation accuracy. Further, to verify the goodness of our approach, we compare our system with the RBF and MLP neural networks proposed in [3] and [12] and applied to the same simulated data. We show that although the three different approaches have similar performance, our system can provide a meaningful description of the relation between the reflectances and the optically active parameter values.

II. The Fuzzy System

The prediction of the three optically active parameters C, X, and Y from the average spectral subsurface reflectances is a typical problem of identifying an unknown nonlinear system $y = f_y(\boldsymbol{R})$, with $y = C, X$ or Y, based on a set of N available input–output data $(\boldsymbol{R}_k, y_k)$, with $k = 1 \ldots N$. This problem has been extensively studied in the literature. The identification process is often performed by constructing local linear models around selected operating points and then combining these models to attain the system model. One of the most used models is the Takagi-Sugeno-Kang (TSK) model, which joins good identification performance to a reasonable interpretability of the model.

A. TSK Model

A TSK model consists of a set of fuzzy rules in the form [10]

$$\mathrm{r_i}: \textbf{if } \bar{R}_1 \textbf{ is } A_{i,1} \textbf{ and } \ldots \textbf{ and } \bar{R}_M \textbf{ is } A_{i,M}$$
$$\textbf{then } o_i = p_{i,1}\bar{R}_1 + \ldots + p_{i,M}\bar{R}_M + p_{i,M+1}, \quad i = 1 \ldots P$$

where $\bar{R}_1, \ldots, \bar{R}_M$ are, in our case, the M average subsurface reflectances, $o_i, i = 1 \ldots P$ is the consequent of the ith rule, $p_{i,1}, \ldots, p_{i,M+1}$ are real numbers, and $A_{i,1}, \ldots, A_{i,M}$ are fuzzy sets defined on the domains of $\bar{R}_1, \ldots, \bar{R}_M$, respectively. The model output y is computed by aggregating the conclusions inferred from the individual rules as follows:

$$y = \frac{\sum_{i=1}^{P} \beta_i o_i}{\sum_{i=1}^{P} \beta_i} \tag{1}$$

where $\beta_i = \prod_{j=1}^{M} A_{i,j}(\bar{R}_j)$ is the degree of activation of the ith rule. The antecedent of each rule determines a region of the input space by means of a conjunction of fuzzy clauses that contain the input variables. The consequence is a mathematical function which approximates the behavior of the system to be identified in the region fixed by the antecedent.

The TSK model generation task includes two subtasks: the *structure identification* and the *parameter identification*. The structure identification determines the number of rules and the variables involved in the rule antecedents. The parameter identification estimates the parameters which define the membership functions of the fuzzy sets in the antecedents and the parameters which identify the consequent functions. The parameter identification is obtained in two steps: First, data-driven techniques are used to identify the fuzzy sets which compose the antecedent of the rules. Then, the parameters of the mathematical functions in the consequent part are estimated. In the TSK models found in the literature, these mathematical functions are typically linear functions (the resulting TSK model is called first-order TSK

fuzzy model). Thus, the estimation process can be carried out by standard linear least-squares methods.

The structure identification is generally performed by exploiting fuzzy clustering algorithms. In particular, the number of rules is equal to the number of clusters which compose the partition assessed to be the best with respect to an appropriate validity index. Given the training set of data $Z = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N]$, generated by the direct model, where $\boldsymbol{z}_k = (\boldsymbol{R}_k, y_k) \in \Re^{M+1}$, $\boldsymbol{R}_k$ is the kth vector of average subsurface spectral reflectances and y_k is one of the three parameters C, Y and X, a fuzzy clustering algorithm is applied to compute the fuzzy partition matrix U whose ikth element $u_{i,k} \in [0,1]$ expresses the membership degree of point $\boldsymbol{z}_k \in Z$ to cluster $i, i = 1 \ldots P$. Fuzzy sets $A_{i,j}$ are obtained by projecting the rows of the partition matrix U onto the input variables $\bar{R}_j$ and approximating the projections by pre-fixed membership functions. In the experiments, we used triangular functions defined as

$$\mu(\bar{R}_j; a, b, c) = \max\left(0, \min\left(\frac{\bar{R}_j - a}{b - a}, \frac{c - \bar{R}_j}{c - b}\right)\right) \quad (2a)$$

with $a < b < c$ real numbers on the domain of definition of $\bar{R}_j$. In the cases of $a = b < c, a < b = c$ and $a = b = c$, formula (2a) is not applicable and is replaced by the following three formulas (2b), (2c), and (2d), respectively:

$$\mu(\bar{R}_j; a, b, c) = \begin{cases} \frac{c - \bar{R}_j}{c - b}, & \text{if } b \le \bar{R}_j < c \\ 0, & \text{otherwise} \end{cases} \quad \text{if } a = b < c \quad (2b)$$

$$\mu(\bar{R}_j; a, b, c) = \begin{cases} \frac{\bar{R}_j - a}{b - a}, & \text{if } a < \bar{R}_j \le b \\ 0, & \text{otherwise} \end{cases} \quad \text{if } a < b = c \quad (2c)$$

$$\mu(\bar{R}_j; a, b, c) = \begin{cases} 1, & \text{if } \bar{R}_j = a = b = c \\ 0, & \text{otherwise} \end{cases} \quad \text{if } a = b = c. \quad (2d)$$

We computed the parameter b, which corresponds to the abscissa of the vertex of the triangle, as the weighted average of the $\bar{R}_j$ components of the training patterns, the weights being the corresponding membership values. Parameters a and c were obtained as intersection of the $\bar{R}_j$ axis with the lines obtained as linear regression of the membership values of the training patterns, respectively, on the left and the right sides of b. Obviously, if a or c are beyond the extremes of the definition domain of variable $\bar{R}_j$, the sides of the triangles are truncated in correspondence to the extremes. The use of triangular functions allows easy interpretation of the fuzzy sets in linguistic terms. This characteristic will be useful to associate a meaning with the rules, as explained in Section III. We note that formula (2d) is used when a cluster is composed of a unique point (singleton): this case is however extremely improbable.

The most used fuzzy clustering algorithms to determine the number of rules are the FCM algorithm [9] and the Gustafsson and Kessel's version (GK) of FCM [13]. The FCM algorithm has the disadvantage of not being very flexible since it can determine only spherical clusters approximately of the same size [14]. The GK algorithm overcomes this problem by modeling arbitrary hyper-ellipsoidal clusters, but the derived fuzzy rules are in general not very coherent with the fuzzy clusters. The Cartesian product (determined by the derived rules) of the projections of a hyper-ellipsoid yields the smallest hyper-rectangle that contains the hyper-ellipsoid. If the axes of the hyper-ellipsoid are parallel to the coordinate axes, the difference between the original cluster and the hyper-rectangle is small; otherwise, this difference can be significant causing a remarkable loss of information. To come to a compromise between flexibility and loss of information, Klawonn and Kruse modified the GK algorithm [14], [15]: in the computation of the distance $d_i(\boldsymbol{z}_k, \boldsymbol{v}_i)$ between a generic vector $\boldsymbol{z}_k$ and a cluster prototype $\boldsymbol{v}_i$, they forced the square symmetric positive matrix W_i of the induced norm $d_i^2(\boldsymbol{z}_k, \boldsymbol{v}_i) = \|\boldsymbol{z}_k - \boldsymbol{v}_i\|^2 = (\boldsymbol{z}_k - \boldsymbol{v}_i)^T W_i (\boldsymbol{z}_k - \boldsymbol{v}_i)$, used in GK, to be diagonal. Using the same constraint as in GK on matrices W_i's, that is, $\det(W_i) = \rho_i$, with ρ_i a fixed real number typically equal to 1, Klawonn and Kruse defined a modified GK capable of determining hyper-ellipsoidal clusters with the axes parallel to the coordinate axes. As W_i's are diagonal, we observe that $\det(W_i) = \rho_i$ means that the product of the diagonal elements of W_i is equal to ρ_i.

In this paper, to determine the number of rules, we use a fuzzy clustering algorithm that, like in Klawonn and Kruse's approach, forces matrices W_i's of the induced norm to be diagonal. Unlike [14] and like [16], [17], however, we impose a constraint on the sum of the diagonal elements of W_i rather than on the product. We denote the algorithm weighted fuzzy C-means (WFCM) algorithm.

B. WFCM Algorithm

The WFCM algorithm determines a partition of the data set Z minimizing the same criterion function as in classical FCM or in GK

$$J_m(U, V, W) = \sum_{i=1}^{P} \sum_{k=1}^{N} u_{i,k}^m d_i^2(\boldsymbol{z}_k, \boldsymbol{v}_i) \quad (3)$$

where $U = [u_{i,k}]$ is the matrix of the membership degrees of $\boldsymbol{z}_k \in Z$ to cluster $i, i = 1 \ldots P, V = [v_1, \ldots, v_P]$ is the set of the P prototypes, $W = [W_1, \ldots, W_P]$ is the set of the P square symmetric positive matrices W_i's, m is the fuzzification constant and $d_i^2(\boldsymbol{z}_k, \boldsymbol{v}_i) = (\boldsymbol{z}_k - \boldsymbol{v}_i)^T W_i (\boldsymbol{z}_k - \boldsymbol{v}_i)$. In the classical FCM proposed by Bezdek, W_i's are identity matrices and the distance $d_i(\boldsymbol{z}_k, \boldsymbol{v}_i)$ is the Euclidean distance. In GK, $W_i = [\rho_i \det(Q_i)]^{1/p} Q_i^{-1}$, where Q_i is the fuzzy covariance matrix of cluster i, defined as $Q_i = \left(\sum_{k=1}^{N} u_{i,k}^m (\boldsymbol{z}_k - \boldsymbol{v}_i)^T (\boldsymbol{z}_k - \boldsymbol{v}_i)\right) / \sum_{k=1}^{N} u_{i,k}^m$.

In WFCM, $W_i = [w_{i,f,s}]$ are diagonal matrices whose diagonal elements are expressed as $w_{i,f,f} = g_{i,f}^h$, with $h \in \Re_{>1}$. Then, the distance can be rewritten as $d_i^2(z_k, v_i) = \sum_{f=1}^{F} (z_{k,f} - v_{i,f})^2 g_{i,f}^h$, where each $g_{i,f}$ represents the weight associated with the component f with respect to cluster i, and F is the number of components (in our case, $M + 1$). We impose the constraint $\sum_{f=1}^{F} g_{i,f} = F$. Thus, if $g_{i,f} = 1$ for each $i \in [1 \ldots P]$ and $f \in [1 \ldots F]$, then $d_i^2(\boldsymbol{z}_k, \boldsymbol{v}_i)$ is the Euclidean distance. The exponent h influences the weights $g_{i,f}$ in the same way as the fuzzification constant m influences the membership values $u_{i,k}$. When $h \to 1$, for each cluster i, a weight $g_{i,f}$ tends to be equal to F and the

others equal to 0. Thus, a reflectance has either unrestricted influence or no influence. On the other hand, when $h \to \infty$ all the reflectances tend to assume the same value, i.e., for each cluster i and each reflectance $f, g_{i,f} = 1$, and therefore the weighted distance becomes the Euclidean distance.

Let $R_{P,F}$ be the set of real $P \times F$ matrices. Then, the weights $g_{i,f}$ can be considered as elements of a weight matrix G which belongs to the set $D_{P,F} = \{G \in R_{P,F} | g_{i,f} \in [0,F]\ \forall i,f; \sum_{f=1}^{F} g_{i,f} = F\ \forall i\}$. Each row $\boldsymbol{g}_i$ of G corresponds to the weights associated with the components f for cluster i. Then, the criterion function $J_m(U,V,W)$ in formula (3) becomes $J_{m,h}(U,V,G) = \sum_{i=1}^{P} \sum_{k=1}^{N} u_{i,k}^m \sum_{f=1}^{F} (z_{k,f} - v_{i,f})^2 g_{i,f}^h$.

Since each term of $J_{m,h}(U,V,G)$ is proportional to $(z_{k,f} - v_{i,f})^2$, $J_{m,h}(U,V,G)$ is a squared-error clustering criterion, and solutions to the minimization of $J_{m,h}(U,V,G)$ are least squared error stationary points of $J_{m,h}(U,V,G)$. To minimize the objective function we can use an iterative method based on the successive minimization of the following functions.

A) $J_{m,h}(\cdot,V,G)$, where V and G are fixed.
B) $J_{m,h}(U,\cdot,G)$, where U and G are fixed.
C) $J_{m,h}(U,V,\cdot)$, where U and V are fixed.

The minimization of functions A., B., and C. is attained, respectively, by applying the following formulas (for functions A. and C., the formulas are obtained by applying the Lagrange multipliers and setting the gradient to 0; for function B., the formulas are derived from setting the derivative of the function to 0):

$$u_{r,k} = \frac{1}{\sum_{i=1}^{P} \left(\frac{d_r(\boldsymbol{z}_k, \boldsymbol{v}_r)}{d_i(\boldsymbol{z}_k, \boldsymbol{v}_i)} \right)^{\frac{2}{m-1}}} \quad \forall r \in [1 \ldots P] \quad \text{and} \quad \forall k \in [1 \ldots N] \tag{4a}$$

$$v_{s,t} = \frac{\sum_{k=1}^{N} u_{s,k}^m z_{k,t}}{\sum_{k=1}^{N} u_{s,k}^m} \quad \forall s \in [1 \ldots P] \quad \text{and} \quad \forall t \in [1 \ldots F] \tag{4b}$$

$$g_{i,t} = \frac{F}{\sum_{f=1}^{F} \left(\frac{\sum_{k=1}^{N} u_{i,k}^m (z_{k,t} - v_{i,t})^2}{\sum_{k=1}^{N} u_{i,k}^m (z_{k,f} - v_{i,f})^2} \right)^{\frac{1}{h-1}}} \quad \forall i \in [1 \ldots P] \quad \text{and} \quad \forall t \in [1 \ldots F]. \tag{4c}$$

The WFCM algorithm can be summarized as follows.

Fix $P, 2 \le P < N, m, m \in (1,\infty)$ and $h, h \in (1,\infty)$. Initialize U^0 randomly. Then, at step $l, l = 0, 1, \ldots,$:

1) Calculate $V^{(l)}$ with (4.b), using $U^{(l)}$.
2) Calculate $G^{(l)}$ with (4.c) using $V^{(l)}$ and $U^{(l)}$.
3) Calculate $U^{(l+1)}$ with (4.a) using $V^{(l)}$ and $G^{(l)}$.
4) Compare $U^{(l)}$ and $U^{(l+1)}$ using the matrix norm: $\|U^{(l+1)} - U^{(l)}\| = \max_{i,k} |u_{i,k}^{l+1} - u_{i,k}^{l}|$; if $\|U^{(l+1)} - U^{(l)}\| \le \varepsilon$ then stop; otherwise go to 1. with $l = l + 1$. ε is the admissible error. Typically, $\varepsilon = 10^{-5}$.

As the WFCM algorithm finds the optimal fuzzy partition starting from a fixed number of clusters, and the number of clusters determines the number of rules which compose the fuzzy model, a criterion has to be adopted to determine the optimal number of clusters. The most common approach is to identify an interval of possible values of the number P of clusters and execute the clustering algorithm for each value in the interval. Each execution is therefore assessed against a validity index. Several different validity indexes have been proposed in the literature [18]. Unfortunately, there does not exist a validity index suitable for all data sets [19]. In the next section, we introduce some of the best known validity indexes used to assess fuzzy partitions and discuss the possible modifications to make these indexes suitable for our fuzzy clustering algorithm.

C. Validity Indexes

The validity indexes proposed to assess a fuzzy partition can be classified as indexes involving only the membership values and indexes involving both the membership values and the data set. In the first category, the most popular are the partition index and the partition entropy index [9].

Though the indexes involving only the membership functions are easy to compute, they show the following drawbacks: 1) monotonous dependency on the number of clusters and 2) lack of direct connection with the distribution of the data [20]. To overcome these problems, several different indexes based on both membership values and the data set have been proposed in recent years [18]. The most used among these indexes, also in application domains similar to ours [8], are the Xie and Beni's index [21], the Fukuyama and Sugeno's index [22], the Gath and Geva's index [23], and the Rezaee, Lelieveldt and Reiber's index [24]. As it is well-known in the literature, there does not exist a validity index good for each data set [22]. Thus, to assess which index provided the best solution, we executed the WFCM with $h = 2$ and with increasing values of the number P of clusters between the estimated lower and upper bounds $2 \le P \le 15$. Since the validity indexes depend on the fuzzification parameter m [19], for each value of P we repeated the execution of WFCM for $m = 1.6, m = 1.8, m = 2.0$, and $m = 2.2$. We used the original versions of the indexes, which adopt the Euclidean distance, and the modified versions which, as suggested in [25], replace the Euclidean distance with the norm used for measuring distances in the clustering space (in our case, the weighted distance). The assessment was based on building the TSK model with the number of rules determined by the indexes and measuring the mse produced both before and after the execution of the GA. We classified the validity indexes on the basis of the number of rules and mse. At the end of this experimentation, we realized that the best behavior is provided by the Fukuyama and Sugeno's index (FS) with weighted distance defined as follows:

$$\text{FS}(U,V,Z) = \sum_{i=1}^{P} \sum_{k=1}^{N} u_{i,k}^m \times \left(\sum_{f=1}^{F} (z_{k,f} - v_{i,f})^2 g_{i,f}^h - \sum_{f=1}^{F} (v_{i,f} - \bar{v}_f)^2 g_{i,f}^h \right) \tag{5}$$

where $\bar{\boldsymbol{v}} = \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{z}_k$. The first term in parenthesis measures the compactness of the clusters and the second one measures the distances of the cluster prototypes. For compact and well-separated clusters we expect small values of FS. Thus, we chose,

as optimal number of clusters, the value of P corresponding to the first distinctive local minimum [26]. Once the antecedent membership functions have been fixed, the consequent parameters $p_{i,q}, q = 1 \ldots M+1$ of each individual rule are obtained as a local least squares estimate. At this point, all the parameters are available to define the three fuzzy systems. Due to the process followed in this first phase of system identification, the rule base obtained so far is characterized by a number of interesting properties: moderate number of rules, membership functions well separated and distinguishable from each other, and coverage of the input-output data space. The performance of this system could be however poor because the fuzzy rules were identified by the WFCM algorithm, and not by minimizing the model error. To improve the model, we apply a GA to tune simultaneously the parameters in the antecedent and consequent parts of each rule in a global optimization.

D. The GA

As it is well known, GAs are probabilistic search algorithms that emulate the genetic processes of biological organisms and are capable of finding optimal or near-optimal solutions to complex problems. To preserve the good properties, which characterize the fuzzy model, the GA is tailored to the specific application. Each real-coded chromosome codifies the parameters, which define the membership functions of the fuzzy sets in the antecedent part of the rules, and the consequents. To maintain space coverage we impose that no gap exists in the partition of each input variable. Further, to preserve membership function distinguishability we allow the parameters that define the fuzzy sets to vary within a specified range around their initial values. More precisely, given a generic fuzzy set $A_{i,j}$ and a real number t, the parameters describing $A_{i,j}$ can vary only within a bound of $\pm t|\varphi_i|$ around their initial values, where $|\varphi_i|$ indicates the length (range) of the domain on which the fuzzy set $A_{i,j}$ is defined. Choosing a low value of t, we can avoid the risk of generating domain-wide and multiple overlapping fuzzy sets. This also enhances interpretability of the fuzzy rule base by domain experts.

The initial population of the GA is composed of 50 chromosomes generated as follows: The first chromosome codifies the system generated by the WFCM, the others are obtained by perturbing the first chromosome randomly within the ranges fixed to maintain distinguishability. Each chromosome represents the entire fuzzy system, rule by rule, with the antecedent and consequent parts (see Fig. 1). Each rule antecedent consists of a sequence of M triplets (a, b, c) of real numbers representing triangular membership functions, whereas each rule consequent contains $M+1$ real numbers corresponding to the consequent parameters. The fitness value is the inverse of the mse between the predicted output and the desired output over the training set.

To generate a new population, we apply the arithmetic crossover and the uniform mutation operators with probabilities 0.9 and 0.1, respectively [27]. Chromosomes to be mated are chosen by using the well-known roulette wheel selection method, which associates with each chromosome a probability proportional to its fitness value [27]. At each generation, the offspring are checked against the aforementioned space-coverage criterion. To speed up the convergence of the algorithm without significantly increasing the risk of premature convergence to local minima, we adopted the following acceptance mechanism: only 80% of the new population is composed of offspring, whereas 20% consists of the best chromosomes of the previous population. When the average of the fitness values of all the individuals in the population is greater than 99% of the fitness value of the best individual, or a prefixed number of iterations has been executed, the GA is considered to have converged.

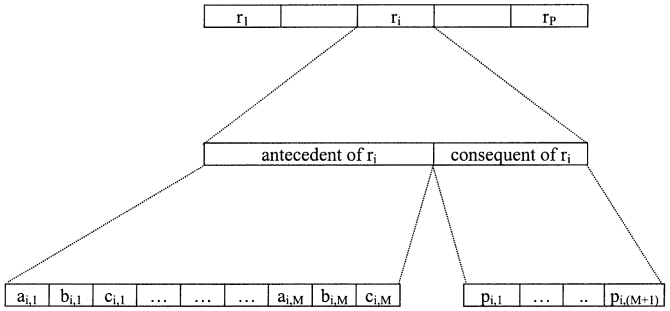

Fig. 1. Chromosome representation.

III. Experimental Results

A. Data Set Generation

As previously stated, the ocean color in costal areas is usually expressed in terms of the subsurface spectral reflectance $R(\lambda) = f(C, X, Y; \lambda)$, where λ is the wavelength, C is the chlorophyll-a plus phaeophytin-a mass to volume concentration, X the scattering coefficient of nonchlorophyllous particles at 550 nm and Y is the specific absorption coefficient of yellow substance at 440 nm [3]. The parameters C, X, Y can therefore be extracted through an inverse model.

To analyze the potential of the TSK system for the estimation of ocean color, we used a dataset consisting of simulated values of the three optically active parameters and the corresponding mean reflectances over the first eight channels of the MERIS sensor. The eight average subsurface reflectances correspond to the eight channels of MERIS centered around 412.5, 442.5, 490, 510, 560, 620, 665, and 681.25 nm, respectively. As we are interested in determining the concentrations of those substances which affect the ocean color, we use only the channels (eight of the fifteen) in the visible spectrum. These channels are 10 nm width (except for the channel around the chlorophyll fluorescence spectral peak whose bandwidth is 7.5 nm). Here, we briefly recall the ocean color direct model presented in [3] and used to compute the mean reflectance in each channel of the sensor given the triplet $[C, X, Y]$. We used a three component model with a spectral resolution of 2 nm. As proposed in [2], the subsurface spectral reflectance $R(0^-, \lambda)$ is expressed in terms of the spectral absorption a_λ and backscattering $b_{b\lambda}$ coefficients at wavelength λ as follows:

$$R(0^-, \lambda) = 0.33 \frac{a_\lambda}{a_\lambda + b_{b\lambda}}. \tag{6}$$

The two coefficients a_λ and $b_{b\lambda}$ are computed, respectively, as the sum of the absorption coefficients of pure water, phytoplankton, yellow substance, and nonchlorophyllous compo-

TABLE I
PARAMETERS OF THE LOG-NORMAL DISTRIBUTION USED TO GENERATE THE SIMULATED DATA SET

	$\mu_{\mathrm{Log}(y)}$	$\sigma_{\mathrm{Log}(y)}$
Log(C)	0.0	0.5
Log(X)	0.0	0.5
Log(Y)	-0.5	-0.5

nents, and the sum of the backscattering coefficients of pure sea water, phytoplankton, and nonchlorophyllous particles. Coefficients a_λ and $b_{b\lambda}$ are then expressed in terms of C, X and Y.

A set of spectral reflectance curves, which cover the range 400–700 nm, was simulated exploiting a series of N random triples $[C_i X_i Y_i]$, whose statistics represent a typical case-II water scenario. As suggested by Campbell [28], we used a log-normal distribution to generate the set of N triplets $[C_i X_i Y_i]$ which, given as input to the direct model, produce the vector $\boldsymbol{R} = [\bar{R}_1, \ldots, \bar{R}_8]^T$ whose components are the mean reflectances in the eight channels of the sensor. We took into consideration the characteristics of case II waters by appropriately choosing the parameters of the log-normal distribution and by introducing a certain amount of correlation among the optically active parameters, as explained in [3]. Table I reports the mean and standard deviation of the logarithm of C, X, and Y used to initialize the log-normal random number generators. As regards the correlation, we introduced correlation coefficients equal to 0.5 between the pairs of variables $[\mathrm{Log}(C), \mathrm{Log}(X))$ and $(\mathrm{Log}(C), \mathrm{Log}(Y)]$. We randomly generated 5000 pairs $\Gamma \equiv \{[\mathrm{Log}(\bar{R}_{k,1}) \ldots \mathrm{Log}(\bar{R}_{k,8})], \mathrm{Log}(y_k)\}$, where y_k is one of the three optical active parameters C, X or Y. In the experiments, we used the log transformed data, as is usual in this kind of problem [3] to face amplitude range and sensitivity problems. We randomly split the data set into two subsets composed of 1000 and 4000 pairs. The former subset was used to define the fuzzy model, the latter to test the identification capabilities of the model.

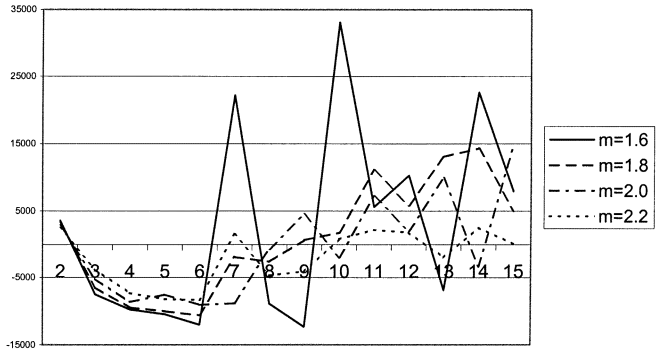

Fig. 2. FS versus number of clusters for $\mathrm{Log}(C)$.

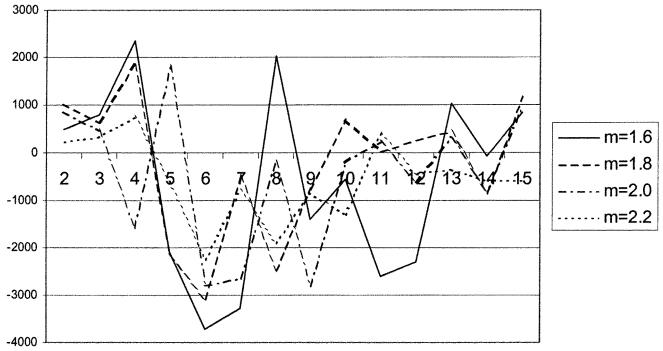

Fig. 3. FS versus number of clusters for $\mathrm{Log}(X)$.

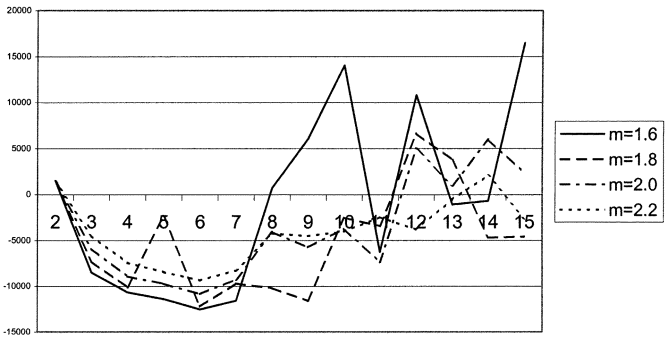

Fig. 4. FS versus number of clusters for $\mathrm{Log}(Y)$.

B. Generation of the TSK Systems

We built three different models to determine, respectively, the parameters C, X and Y from $\boldsymbol{R}$. For each model, we normalized each mean reflectance $\bar{R}_i$ to vary from 0 to 1. We verified that this normalization leads the WFCM algorithm to associate weights with each $\bar{R}_i$ for each cluster based on the actual variations of the mean reflectances $\bar{R}_i$ independently of their dynamics. We would like to point out that this normalization is only used to determine the partition of the data. Since we exploit only the membership values when we determine the TSK model, the model is defined on the original domains of the data. We executed the WFCM clustering algorithm with different values of the number P of clusters. Then, we assessed the optimal number P of rules using the FS as described in the previous section. Figs. 2–4 show the plots of the validity index versus the number of clusters P with different values of m for C, X, and Y, respectively. We can observe that a first distinctive local minimum is located at $P = 6$ for all the parameters. Finally, for each model, we built the initial fuzzy rule base. Figs. 5–7 show the rule bases obtained for C, X, and Y, respectively. Here, we have associated with each fuzzy set a label based on the position of the fuzzy set in the universe of definition. These universes are represented as two thick marks on the horizontal axes. Labels VL, L, ML, MH, H, VH denote, respectively, very low, low, medium-low, medium-high, high, and very high. Each row in the figures corresponds to a rule: the first eight columns represent the antecedent triangular fuzzy sets whose height is equal to 1, whereas the last column shows the values of the parameters $p_{i,1}, \ldots, p_{i,9}$ of each rule.

To interpret the rules, we followed this procedure: for each pattern $z_k = ([\mathrm{Log}(\bar{R}_{k,1}), \ldots \mathrm{Log}(\bar{R}_{k,8})], \mathrm{Log}(y_k))$ in the training set, we input the values of the eight reflectances to the TSK model and measured the activation degree of each rule. We aimed to discover whether there exists a relation between the activation of a rule and the values of the parameters C, X,

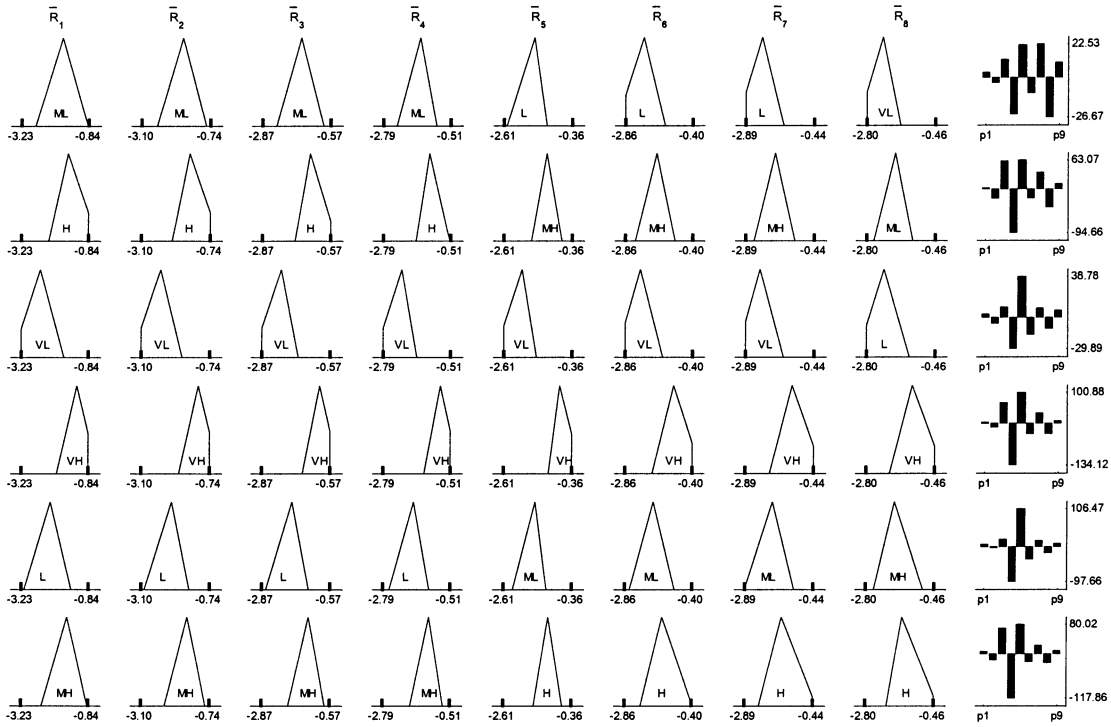

Fig. 5. Rule base of the fuzzy system used to identify C before GA.

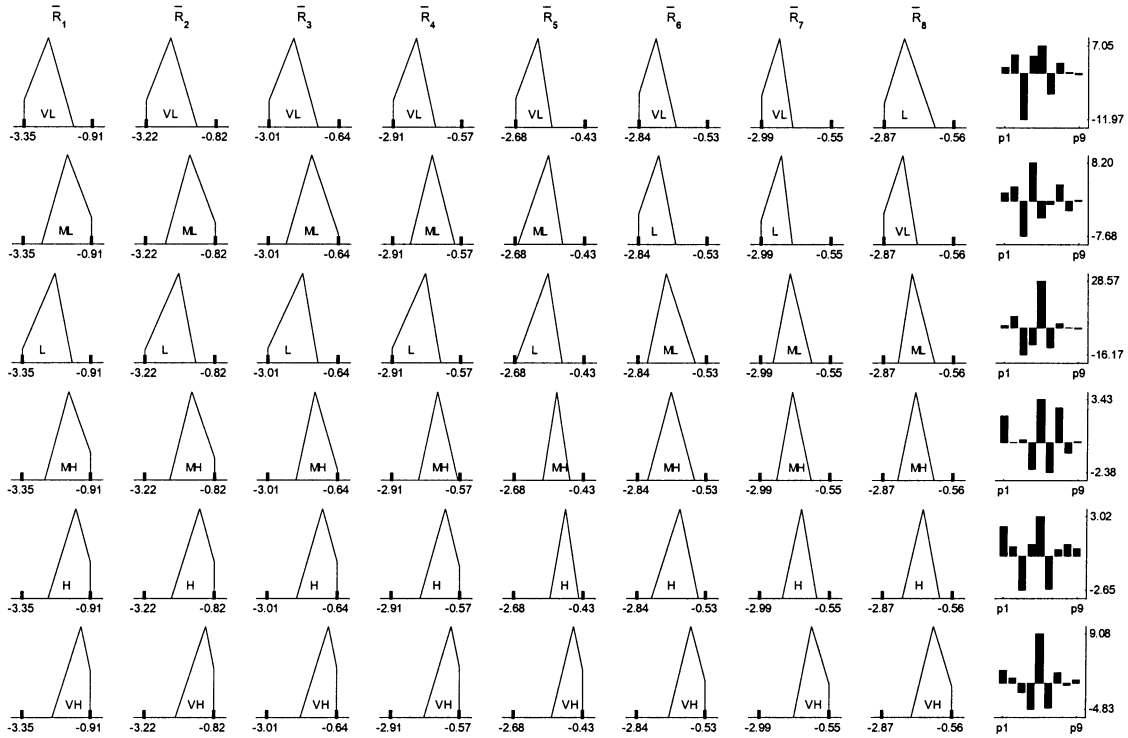

Fig. 6. Rule base of the fuzzy system used to identify X before GA.

and Y. We verified that this relation exists. Table II shows, for each rule, the mean value of $\mathrm{Log}(C), \mathrm{Log}(X)$ and $\mathrm{Log}(Y)$ of the patterns which activate this rule more than the other rules. The rules in the table are listed according to increasing mean values of the parameters. This association between rules and parameter values helps us interpret the meaning of the rules. For instance, if we observe rule r_1 in Fig. 5, and taking the first row in Table II into account, we can conclude that medium-low values of $\mathrm{Log}(\bar{R}_1), \ldots, \mathrm{Log}(\bar{R}_4)$ together with low values of $\mathrm{Log}(\bar{R}_5), \ldots, \mathrm{Log}(\bar{R}_7)$ and very low value of $\mathrm{Log}(\bar{R}_8)$ imply the minimum mean value of $\mathrm{Log}(C)(-0.4290)$. A global analysis of the rule base in Fig. 6 highlights that increasing

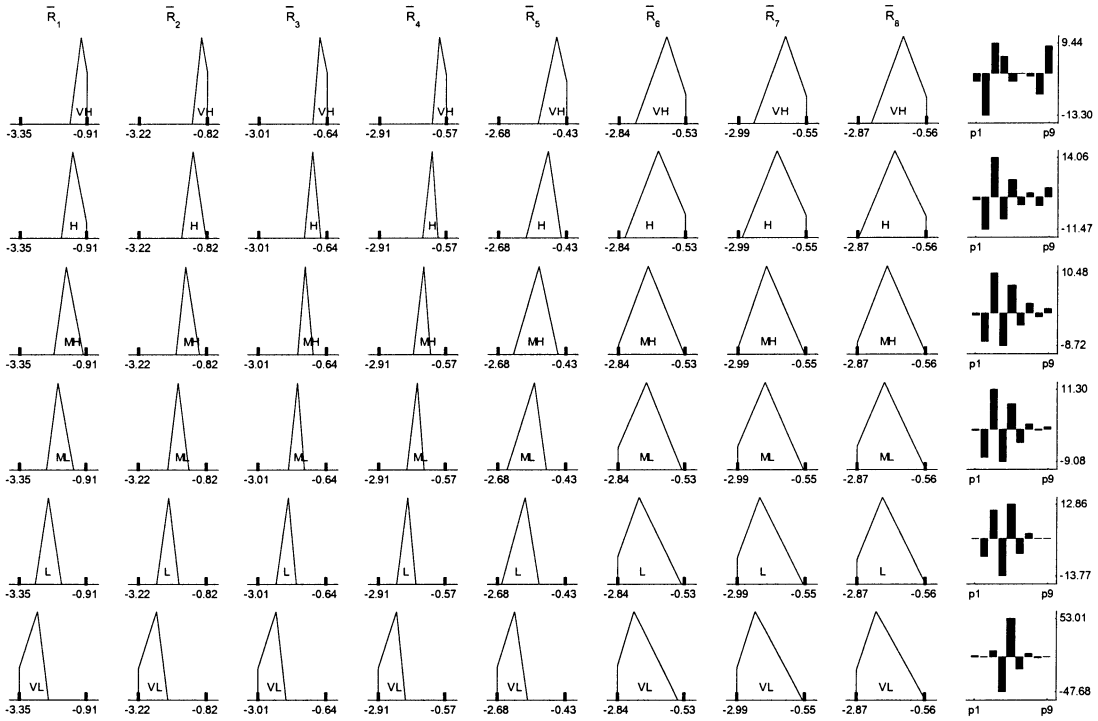

Fig. 7. Rule base of the fuzzy system used to identify Y before GA.

TABLE II
CORRESPONDENCE BETWEEN RULES AND PARAMETER VALUES BEFORE GA

	Log(C)	Log(X)	Log(Y)
r_1	-0.4290	-0.6325	-0.8057
r_2	-0.1992	-0.5545	-0.7027
r_3	-0.0069	-0.0571	-0.6817
r_4	0.1668	-0.0035	-0.4536
r_5	0.2435	0.3611	-0.1868
r_6	0.3293	0.7908	0.2222

TABLE III
MSE AND CORRELATION COEFFICIENT ρ BEFORE GA

	Training Set		Test Set	
	MSE	**ρ**	**MSE**	**ρ**
Log(*C*)	$4.6430\text{x}10^{-3}$	0.9910	$4.9354\text{x}10^{-3}$	0.9897
Log(*X*)	$9.2194\text{x}10^{-5}$	0.9999	$1.4862\text{x}10^{-4}$	0.9995
Log(*Y*)	$6.1288\text{x}10^{-4}$	0.9988	$9.3327\text{x}10^{-4}$	0.9979

values of parameter X correspond approximately to increasing values of the eight mean reflectances. On the contrary, the rule base in Fig. 7 shows an inverse relation between the eight mean reflectances and the values of parameter Y. As regards C, no evident relation seems to exist in the rule base of Fig. 5.

In the following, we denote the desired and the predicted parameters as T and T', (where $T = C, X, Y$, in turn), respectively. Table III shows the mse and the correlation coefficient ρ between $\mathrm{Log}(C)$ and $\mathrm{Log}(C')$, $\mathrm{Log}(X)$ and $\mathrm{Log}(X')$, and

TABLE IV
MSE AND CORRELATION COEFFICIENT ρ FOR THE RADIAL BASIS FUNCTION NEURAL NETWORK

	Training Set		Test Set	
	MSE	**ρ**	**MSE**	**ρ**
Log(*C*)	$3.7768\text{x}10^{-4}$	0.9993	$2.300\text{x}10^{-3}$	0.9954
Log(*X*)	$9.7271\text{x}10^{-6}$	0.9999	$1.7435\text{x}10^{-4}$	0.9996
Log(*Y*)	$1.0388\text{x}10^{-4}$	0.9998	$4.6450\text{x}10^{-4}$	0.9991

TABLE V
MSE AND CORRELATION COEFFICIENT ρ FOR THE MLP NEURAL NETWORK

	Training Set		Test Set	
	MSE	**ρ**	**MSE**	**ρ**
Log(*C*)	$0.9700\text{x}10^{-3}$	0.9990	$1.5802\text{x}10^{-3}$	0.9978
Log(*X*)	$1.0023\text{x}10^{-5}$	0.9998	$1.8343\text{x}10^{-4}$	0.9996
Log(*Y*)	$1.1443\text{x}10^{-4}$	0.9998	$4.566\text{x}10^{-4}$	0.9991

$\mathrm{Log}(Y)$ and $\mathrm{Log}(Y')$ on the training and test sets obtained by the TSK model before applying the GA. We can note that the three models show quite good performance. To assess the goodness of our results, we compare them with the ones obtained on the same type of data by applying RBF [3] and MLP neural networks [12]. Tables IV and V show the mse and correlation coefficient ρ between $\mathrm{Log}(C)$ and $\mathrm{Log}(C')$, $\mathrm{Log}(X)$ and $\mathrm{Log}(X')$, and $\mathrm{Log}(Y)$ and $\mathrm{Log}(Y')$ on the training and test sets obtained, respectively, by the RBF and the MLP networks. Considering the test sets, we note that the TSK model is better than RBF and MLP in the estimation of X even before applying the GA,

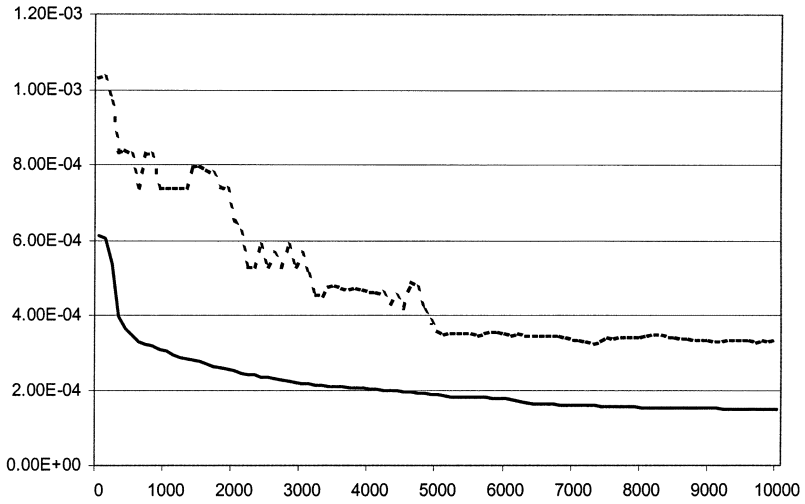

Fig. 8. Mse versus number of generations.

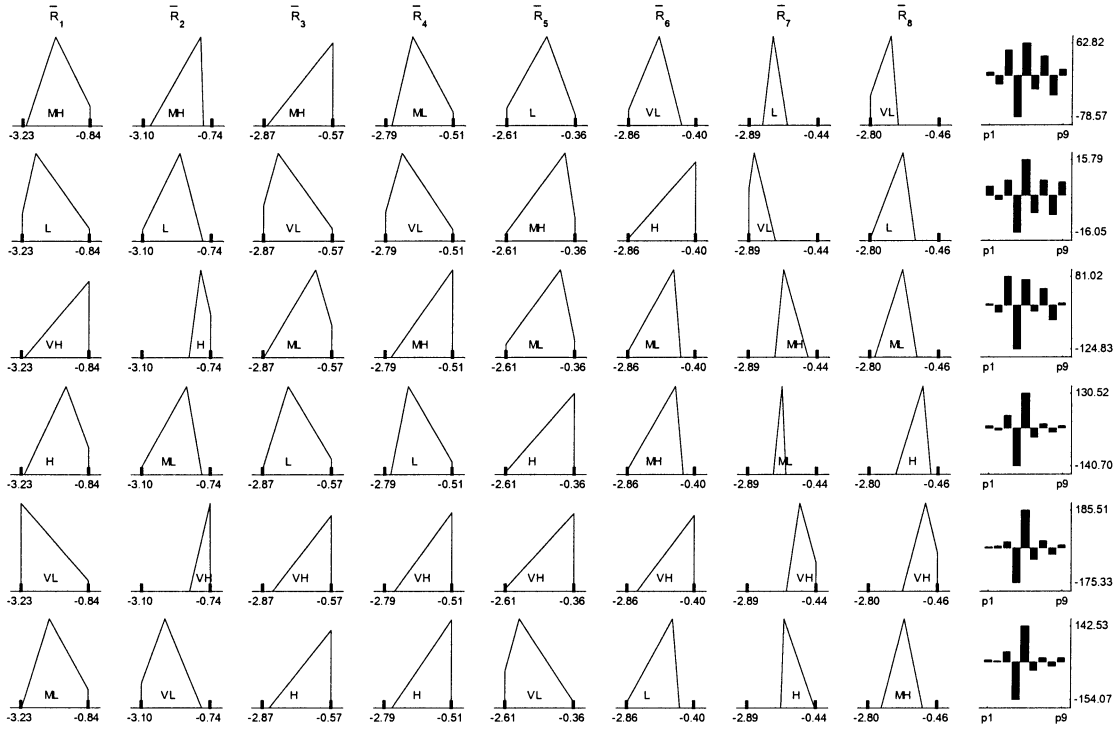

Fig. 9. Rule base of the fuzzy system used to identify C after GA.

whereas the mse on the estimation of C and Y is worse though of the same order of magnitude.

C. Optimization of the TSK Systems

To further reduce the mse of the TSK model, we applied the GA. Fig. 8 shows the trend of the mse corresponding to the average fitness versus the number of generations (continuous line) for parameter Y. We can observe that the GA strongly reduces the mse on the training set after approximately 2000 generations. To verify whether the GA suffers from overfitting problems, we randomly extracted 1000 patterns from the test set to produce a validation set. The dashed line in Fig. 8 shows the trend of the mse corresponding to the average fitness versus the number of generations computed on the validation set. We can observe that the mse computed on the validation set follows the mse computed on the training set, thus confirming no evidence of overfitting problems. Figs. 9–11 show the rule bases obtained for C, X, and Y by the optimization process performed by the GA. We note that the membership functions are slightly changed with respect to the ones shown in Figs. 5–7. Thanks to the constraints enforced on the genetic operators, however, the fuzzy sets are still linguistically interpretable.

Table VI shows, for each final rule, the mean values of $\mathrm{Log}(C), \mathrm{Log}(X)$ and $\mathrm{Log}(Y)$ of the patterns which activate this rule more than other rules. As it is expected, the mean values are slightly different from the ones shown in Table II: in order to reduce the mse the rules are modified and now are activated by

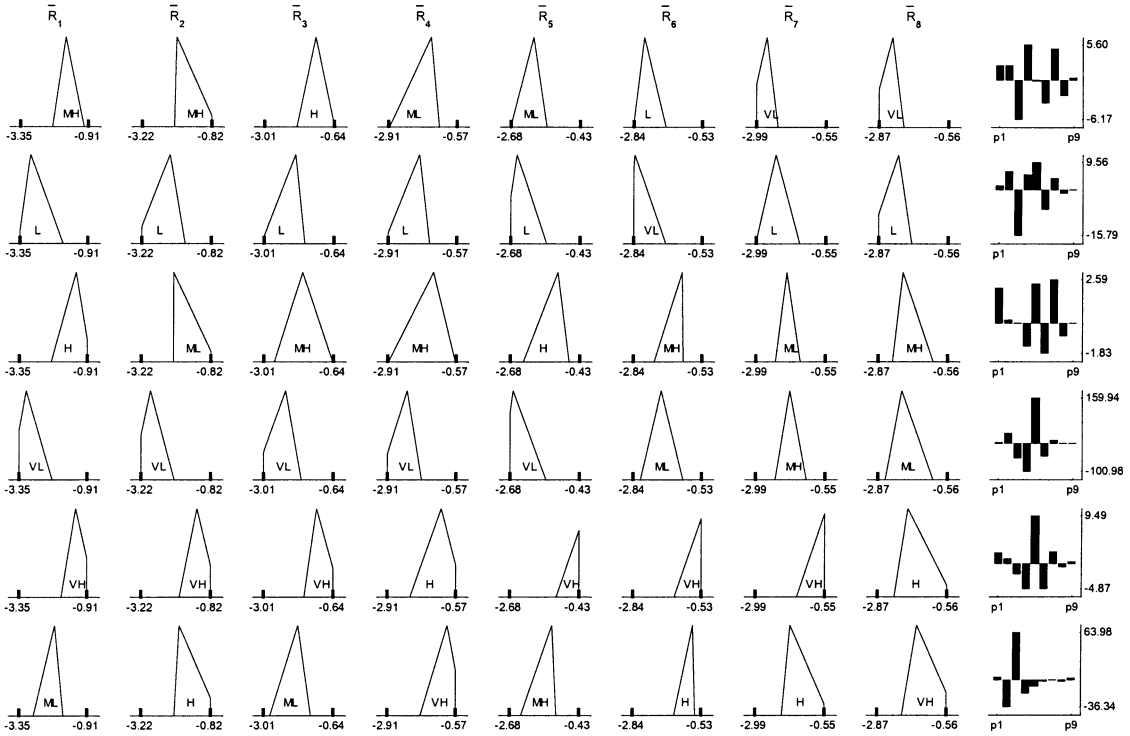

Fig. 10. Rule base of the fuzzy system used to identify X after GA.

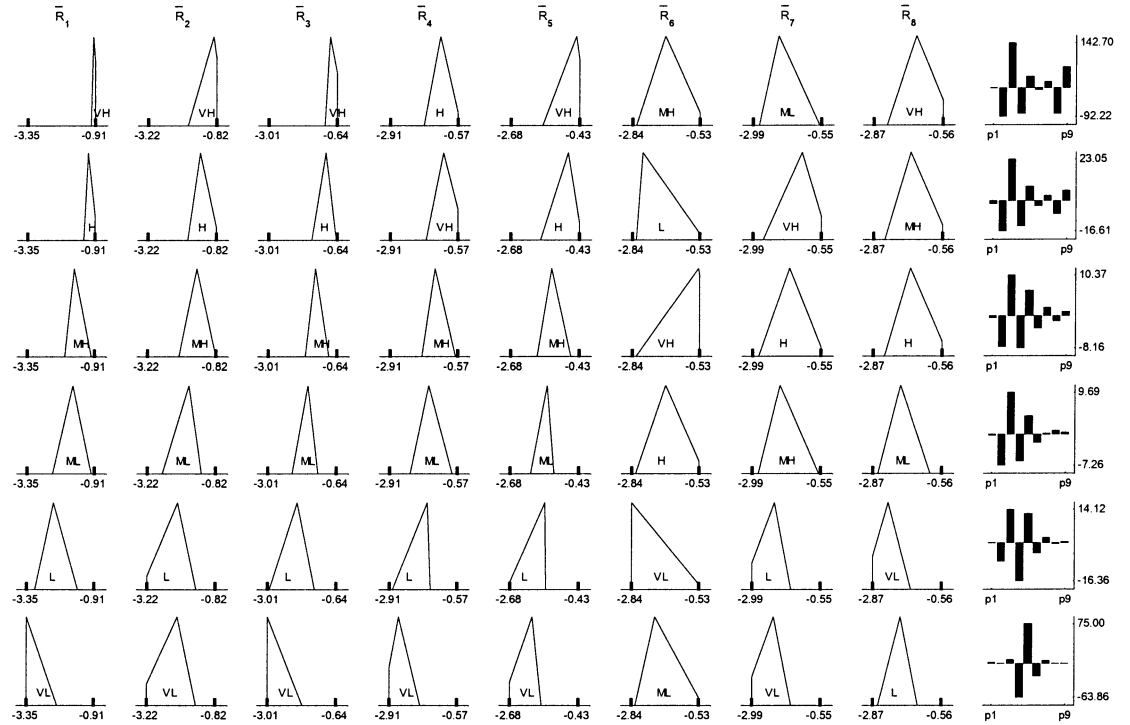

Fig. 11. Rule base of the fuzzy system used to identify Y after GA.

different sets of patterns. Based on Figs. 9–11 and Table VI, we can qualitatively determine the relation between reflectances and parameter values. For instance, rule r_1 of Fig. 9 shows that medium-high values of $\mathrm{Log}(\bar{R}_1), \ldots, \mathrm{Log}(\bar{R}_3)$ together with medium-low value of $\mathrm{Log}(\bar{R}_4)$, low values of $\mathrm{Log}(\bar{R}_5)$ and $\mathrm{Log}(\bar{R}_7)$, and very low values of $\mathrm{Log}(\bar{R}_6)$ and $\mathrm{Log}(\bar{R}_8)$ imply the minimum mean value of $\mathrm{Log}(C)(-0.2571)$. The same kinds of relation between the values of either X or Y, and the reflectances identified before applying the GA still exist in Figs. 10 and 11, but in a looser and less evident way than in Figs. 6 and 7.

Table VII shows the mse and the correlation coefficient ρ between $\mathrm{Log}(C)$ and $\mathrm{Log}(C')$, $\mathrm{Log}(X)$ and $\mathrm{Log}(X')$, and $\mathrm{Log}(Y)$ and $\mathrm{Log}(Y')$ on the training and test sets obtained by

TABLE VI
CORRESPONDENCE BETWEEN RULES AND PARAMETER VALUES AFTER GA

	Log(C)	Log(X)	Log(Y)
r_1	-0.2571	-0.6626	-1.1575
r_2	-0.1630	-0.5573	-0.8675
r_3	-0.0178	0.0318	-0.5954
r_4	0.2840	0.0624	-0.4419
r_5	0.3576	0.7195	-0.3298
r_6	0.4895	0.7288	0.5176

TABLE VII
MSE AND CORRELATION COEFFICIENT ρ AFTER GA

	Training Set		Test Set	
	MSE	**ρ**	**MSE**	**ρ**
Log(C)	3.0242×10^{-3}	0.9941	3.3731×10^{-3}	0.9928
Log(X)	3.4148×10^{-5}	0.9999	1.4714×10^{-4}	0.9997
Log(Y)	1.4952×10^{-4}	0.9997	4.3323×10^{-4}	0.9991

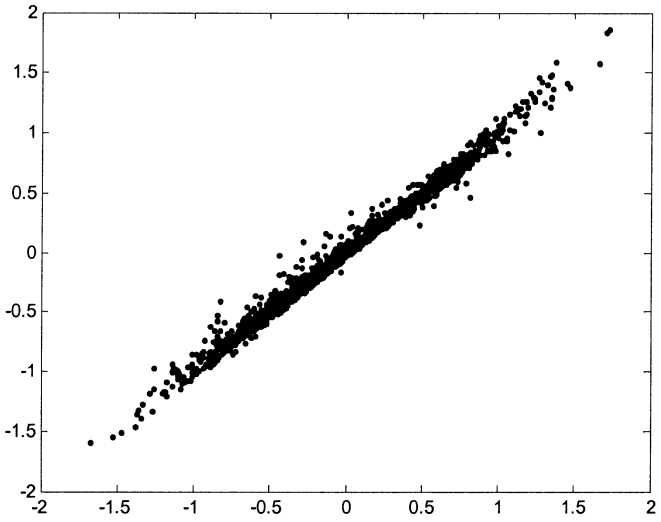

Fig. 12. Scatterplot of predicted output $\mathrm{Log}(C')$ versus desired output $\mathrm{Log}(C)$ over the training set.

the TSK model after applying the GA. By comparing Table VII with Table III, we note that the GA reduces the mse of about 35%, 63%, and 76% on, respectively, C, X, and Y in the training set, and of about 32%, 1%, and 46% on, respectively, C, X, and Y in the test set. As expected, the reduction of the mse is significant in all cases on the training set. Good results are also obtained on the test set except for parameter X whose mse is already optimal, as observed previously. Comparing our results with those obtained with the RBF and the MLP neural networks shown in Tables IV and V we note that the mse we obtained for X and Y in the test set is slightly lower, whereas the mse for C is slightly higher.

Figs. 12–17 show the scatterplots of Log (T') versus $\mathrm{Log}(T)$ over the training and test sets, respectively. We can appreciate the good predictive capabilities of the fuzzy model. Moreover,

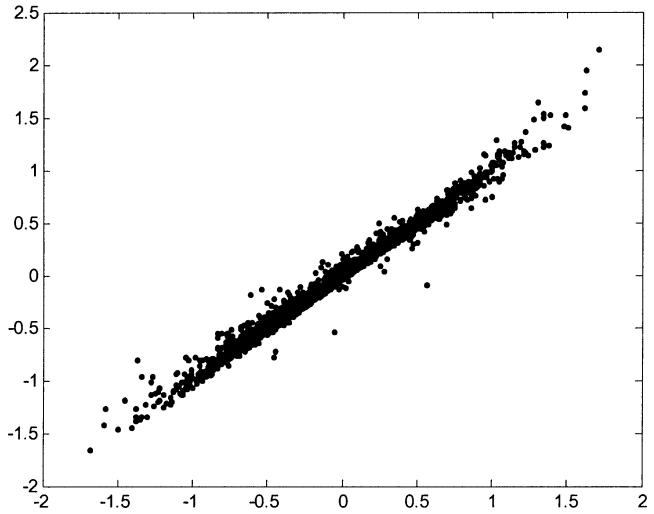

Fig. 13. Scatterplot of predicted output $\mathrm{Log}(C')$ versus desired output $\mathrm{Log}(C)$ over the test set.

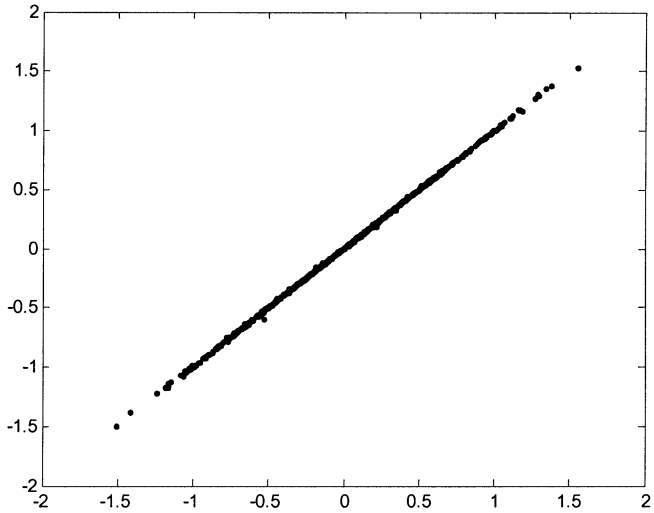

Fig. 14. Scatterplot of predicted output $\mathrm{Log}(X')$ versus desired output $\mathrm{Log}(X)$ over the training set.

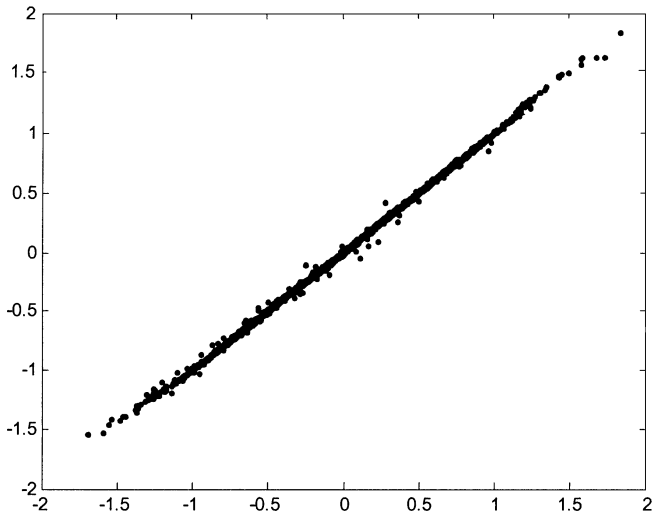

Fig. 15. Scatterplot of predicted output $\mathrm{Log}(X')$ versus desired output $\mathrm{Log}(X)$ over the test set.

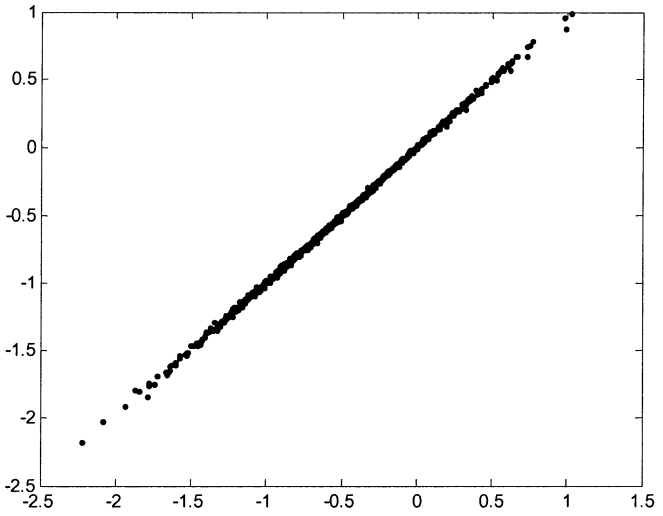

Fig. 16. Scatterplot of predicted output $\mathrm{Log}(Y')$ versus desired output $\mathrm{Log}(Y)$ over the training set.

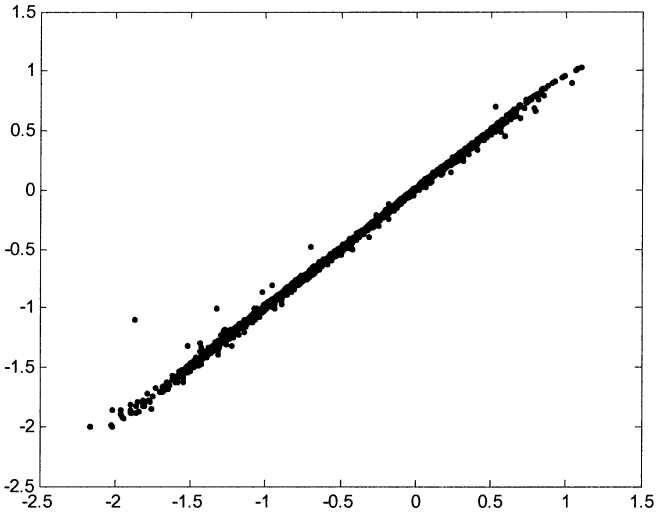

Fig. 17. Scatterplot of predicted output $\mathrm{Log}(Y')$ versus desired output $\mathrm{Log}(Y)$ over the test set.

we note that the minimal dispersion of the points in the figures reveals that the error remains very low along the whole dynamic range of Log (T). The similarities of the MSE values computed on the training and test sets prove the optimal generalization capabilities of the fuzzy model.

D. Generation of a Unique TSK System

As the reflectance values determine a unique set of the optical parameters C, X, and Y, we investigated the opportunity of deriving C, X, and Y from one model rather than from three separate models. To this aim, we built the following TSK model:

$$
\begin{aligned}
&r_i: \textbf{ if } \bar{R}_1 \textbf{ is } A_{i,1} \textbf{ and} \cdots \textbf{and } \bar{R}_M \textbf{ is } A_{i,M},\\
&\textbf{then } o_i^C = p_{i,1}^C \bar{R}_1 + \cdots + p_{i,M}^C \bar{R}_M + p_{i,M+1}^C\\
&\qquad o_i^X = p_{i,1}^X \bar{R}_1 + \cdots + p_{i,M}^X \bar{R}_M + p_{i,M+1}^X \quad i = 1 \ldots P\\
&\qquad o_i^Y = p_{i,1}^Y \bar{R}_1 + \cdots + p_{i,M}^Y \bar{R}_M + p_{i,M+1}^Y.
\end{aligned}
$$

Here, the antecedent of each rule determines a unique region of the input space for the three parameters. The three mathematical functions in the consequent part of the TSK model approximate the behavior of the system for the three parameters in the region fixed by the antecedent. The global TSK model is obtained by using the same method described in Section II. First, we applied the WFCM algorithm to the training set $S = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N]$, where $\boldsymbol{s}_k = [\mathrm{Log}(\bar{R}_{k,1}), \ldots, \mathrm{Log}(\bar{R}_{k,8}), \mathrm{Log}(C_k), \mathrm{Log}(Y_k), \mathrm{Log}(X_k)]$, with values of the number P of clusters in the range $2 \ldots 15$ and values of the fuzzification constant $m = 1.6, 1.8, 2.0$, and 2.2. Second, we computed the optimal number of clusters by applying the FS index and determining the first distinctive local minimum. This minimum was found in correspondence with six clusters. Then, we determined the fuzzy sets $A_{i,j}$ by projecting the rows of the partition matrix U onto the input variables $\bar{R}_j$ and approximating the projections by triangular functions. Once the antecedent membership functions had been fixed, the three sets of the consequent parameters $p_{i,q}^C, p_{i,q}^X, p_{i,q}^Y, q = 1 \ldots 9$, of each individual rule were obtained as a local least-squares estimate. Then, the GA was applied to tune the TSK model. Each chromosome codifies a rule of the TSK model and consists of a sequence of M triplets of real numbers representing the triangular membership functions of the antecedent, and of a sequence of $3M + 3$ real numbers corresponding to the consequent parameters. The fitness value is the inverse of the average of the mse's between the predicted outputs and the desired outputs of the three parameters C, X, and Y over the training set. Tables VIII and IX show the mse and the correlation coefficient ρ between $\mathrm{Log}(C)$ and $\mathrm{Log}(C')$, $\mathrm{Log}(X)$ and $\mathrm{Log}(X')$, and $\mathrm{Log}(Y)$ and $\mathrm{Log}(Y')$ on the training and test sets obtained by the TSK model before and after applying the GA, respectively. We can observe that these results are slightly inferior to those obtained by using three separate models, thus confirming the validity of the approach with a unique model. When, however, performance is the main concern, the solution with three separate models appears more powerful, allowing the antecedent of the rules to model the regions of the input space more accurately.

TABLE VIII
MSE AND CORRELATION COEFFICIENT ρ FOR THE GLOBAL MODEL BEFORE GA

	Training Set		Test Set	
	MSE	ρ	MSE	ρ
Log(*C*)	4.6858×10^{-3}	0.9909	4.9373×10^{-3}	0.9897
Log(*X*)	1.3370×10^{-4}	0.9997	1.3647×10^{-4}	0.9997
Log(*Y*)	9.9526×10^{-4}	0.9980	1.1386×10^{-3}	0.9972

TABLE IX
MSE AND CORRELATION COEFFICIENT ρ FOR THE GLOBAL MODEL AFTER GA

	Training Set		Test Set	
	MSE	ρ	MSE	ρ
Log(*C*)	3.2305×10^{-3}	0.9944	4.2951×10^{-3}	0.9910
Log(*X*)	8.7477×10^{-5}	0.9998	1.7326×10^{-4}	0.9996
Log(*Y*)	7.0231×10^{-4}	0.9985	7.6508×10^{-4}	0.9986

IV. Conclusion

Thanks to the improvement in sensor technology on board satellites, new opportunities have risen for analyzing ocean color. In particular, satellite multispectral and hyperspectral spectrometers can measure with high accuracy the energy level of the part of the sunlight reflected by the sea water and its components. The concentration of optically active components of the sea water can therefore be retrieved from measured reflected energy levels.

In this paper, we have proposed a fuzzy logic-based approach to estimate the concentrations of three optically active constituents, namely chlorophyll, dissolved organic matter, and suspended nonchlorophyllous particles. We tested the approach using a set of synthetic measures of average subsurface reflectances over spectral channels centered around prefixed wavelengths of the MEdium Resolution Imaging Spectrometer (MERIS) on board the ESA-ENVISAT satellite. We developed both three separate models and a unique global model for the three optically active parameters. Each fuzzy model is automatically extracted from the data using a two step procedure. First, a fuzzy clustering algorithm is applied to identify a compact initial rule-based model. Then, the model is optimized by means of a GA which tunes the rules so as to minimize the error between desired and predicted outputs. Though the three separate models achieve results superior to those of the unique global model, the performance obtained by the latter is, however, considerable. From the experiments conducted, we can appreciate the good generalization properties of both the separate fuzzy system and the global system. Furthermore, the modeling power of our approach has proven to be comparable (and often superior) to that of two neural-network-based methods applied to the same domain. Finally, the rules composing the fuzzy model allow easy interpretation of input–output relations between subsurface reflectances and optically active constituents of sea water.

References

[1] A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnol. Oceanogr.*, vol. 22, pp. 709–722, 1977.

[2] S. Sathyendranath, L. Prieur, and A. Morel, "A three-component model of ocean color and its application to remote sensing of phytoplankton pigments in coastal water," *Int. J. Remote Sensing*, vol. 10, no. 8, pp. 1373–1394, 1989.

[3] P. Cipollini, G. Corsini, M. Diani, and R. Grasso, "Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1508–1524, July 2001.

[4] L. Gross, S. Thiria, and R. Frouin, "Applying artificial neural network methodology to ocean color remote sensing," *Ecol. Model.*, vol. 120, pp. 237–246, 1999.

[5] S. B. Hooker, W. E. Esaias, G. C. Feldman, W. W. Gregg, and C. R. McClain, "An overview of SeaWiFS and ocean color," SeaWiFS Tech. Rep., Greenbelt, MD, NASA Tech. Memo 104 566(1).

[6] C. Fonlupt, "Solving the ocean color problem using a genetic programming approach," *Appl. Soft Comput.*, vol. 1, pp. 63–72, 2001.

[7] M. Zhang, L. O. Hall, and D. B. Goldgof, "A generic knowledge-guided image segmentation and labeling system using fuzzy clustering algorithms," *IEEE Trans. Syst., Man, Cybern. B*, vol. 32, pp. 571–581, Oct. 2002.

[8] T. S. Moore, J. W. Campbell, and H. Feng, "A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1764–1776, Feb. 2001.

[9] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[10] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116–132, Feb. 1985.

[11] M. Setnes and H. Roubos, "GA-fuzzy modeling and classification: Complexity and performance," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 509–522, Oct. 2000.

[12] G. Corsini, M. Diani, R. Grasso, M. De Martino, P. Montero, and S. B. Serpico, "Radial basis function and multilayer perceptron neural networks for sea water optically active parameter estimation in case II waters: A comparison," *Int. J. Remote Sensing*, vol. 24, no. 20, pp. 3917–3932, 2003.

[13] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with fuzzy covariance matrix," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, pp. 605–620.

[14] F. Klawonn and R. Kruse, "Constructing a fuzzy controller from data," *Fuzzy Sets Syst.*, vol. 85, pp. 177–193, 1997.

[15] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis*, Chichester, U.K.: Wiley, 1999.

[16] A. Keller and F. Klawonn, "Fuzzy clustering with weighting of data variables," in *Proc. 1999 EUSFLAT-ESTYLF Joint Conf.*, Palma de Maiorca, Spain, 1999, pp. 497–500.

[17] F. Marcelloni, "Recognition of olfactory signals based on supervised fuzzy c-means and k-NN algorithms," *Pattern Recognit. Lett.*, vol. 22, no. 9, pp. 1007–1019, 2001.

[18] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal, *Fuzzy Model and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA: Kluwer, 1999.

[19] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 370–379, Aug. 1995.

[20] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Syst.*, vol. 17, no. 2–3, pp. 107–145, 2001.

[21] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 841–847, Aug. 1991.

[22] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 370–379, Aug. 1995.

[23] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 773–781, July 1989.

[24] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognit. Lett.*, vol. 18, pp. 237–246, 1998.

[25] R. E. Hammah and J. H. Curran, "Validity measures for the fuzzy cluster analysis of orientations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1467–1472, Dec. 2000.

[26] M. Setnes and R. Babuška, "Fuzzy relational classifier trained by fuzzy clustering," *IEEE Trans. Syst., Man, Cybern. B*, vol. 29, pp. 619–625, Oct. 1999.

[27] Z. Michalewicz, Genetic algorithms + data structures = evolution programs, 2nd ed. New York: Springer-Verlag, 1994.

[28] J. W. Campbell, "The lognormal distribution as a model for bio-optical variability in the sea," *J. Geophys. Res.*, vol. 100, no. C7, pp. 13 237–13 254, 1995.

Marco Cococcioni received the Laurea degree in computer engineering in 2000 from the University of Pisa, Pisa, Italy, where he is currently working toward the Ph.D. degree in computer engineering.

His main research interests include neuro-fuzzy modeling, multiple classifier systems, fuzzy clustering, genetic algorithms, statistical signal processing.

Giovanni Corsini (M'89) was born in Grosseto, Italy, in July 1953. He received the Dr. Eng. degree in electronic engineering from the University of Pisa, Pisa, Italy, in 1979.

Since 1983, he has been with the Department of Information Engineering, University of Pisa, where he is currently a Full Professor of Telecommunication Engineering. His main research interests include signal detection and processing, with emphasis on radar application, radar imaging, and remote sensing of the sea surface.

Beatrice Lazzerini (M'98) received the Laurea degree in computer science from the University of Pisa, Pisa, Italy, in 1981.

Currently, she is a Full Professor of Decision Support Intelligent Systemsin the Faculty of Engineering, University of Pisa. Her main research interests lie in the area of knowledge engineering, with particular emphasis on fuzzy systems, neural networks, and evolutionary computation. She has co-authored seven books and has published more than 90 papers in international journals and conferences. She is co-editor of two books.

Mrs. Lazzerini is a member of the IEEE Computer Society.

Francesco Marcelloni received the Laurea degree in electronics engineering and the Ph.D. degree in computer engineering from the University of Pisa, Pisa, Italy, in 1991 and 1996, respectively.

Currently, he is an Associate Professor in the Faculty of Engineering, University of Pisa. His research interests include object-oriented software development processes, object-oriented models, approximate reasoning, fuzzy rule-based systems, fuzzy clustering algorithms, and pattern recognition. He is (co-)author of more than 70 papers in international journals and conferences.