Robust Multiple Estimator Systems for the Analysis of Biophysical Parameters From Remotely Sensed Data

Lorenzo Bruzzone, *Senior Member, IEEE,* and Farid Melgani, *Member, IEEE*

***Abstract*—In this paper, an approach based on multiple estimator systems (MESs) for the estimation of biophysical parameters from remotely sensed data is proposed. The rationale behind the proposed approach is to exploit the peculiarities of an ensemble of different estimators in order to improve the robustness (and in some cases the accuracy) of the estimation process. The proposed MESs can be implemented in two conceptually different ways. One extends the use of an approach previously proposed in the regression literature to the estimation of biophysical parameters from remote sensing data. This approach integrates the estimates obtained from the different regression algorithms making up the ensemble by a direct linear combination (combination-based approach). The other consists of a novel approach that provides as output the estimate obtained by the regression algorithm (included in the ensemble) characterized by the highest expected accuracy in the region of the feature space associated with the considered pattern (selection-based approach). This estimator is identified based on a proper partition of the feature space. The effectiveness of the proposed approach has been assessed on the problem of estimating water quality parameters from multispectral remote sensing data. In particular, the presented MES-based approach has been evaluated by considering different operational conditions where the single estimators included in the ensemble are: 1) based on the same or on different regression methods; 2) characterized by different tradeoffs between correlated errors and accuracy of the estimates; 3) trained on samples affected or not by measurement errors. In the definition of the ensemble particular attention is devoted to support vector machines (SVMs), which are a promising approach to the solution of regression problems. In particular, a detailed experimental analysis on the effectiveness of SVMs for solving the considered estimation problem is presented. The experimental results point out that the SVM method is effective and that the proposed MES approach is capable of increasing both the robustness and accuracy of the estimation process.**

***Index Terms*—Biophysical parameters, estimation, multilayer perceptron (MLP) neural networks, multiple estimator systems, radial basis function (RBF) neural networks, regression, remote sensing, support vector machines (SVMs).**

I. Introduction

In THE last years, the remote sensing community has devoted particular attention to the estimation of biophysical parameters via the analysis of remote sensing data. Quantitative analysis of these parameters can be useful in several application domains. For example, at present, a widely investigated application is the estimation of biomass concentration in forest areas. Such estimates allow to better understand the temporal dynamics of activities of reforestation, afforestation, and deforestation over a defined period of time at a global scale [1]. Another important application is the assessment of the soil moisture content, which represents a key parameter in environmental studies characterized by the soil–vegetation–atmosphere trilogy. This estimation task can be accomplished with synthetic aperture radar (SAR) data [2]. Assessments of the ozone and NO_2 concentrations (at a global scale) by analysis of satellite data are other important applications of regression techniques. In fact, ozone concentration is of primary importance for the impact it has on the earth's environment [3], while NO_2 concentrations provide valuable information to monitor the pollution in the atmosphere due to industry or extended forest fires [4]. Another particularly interesting and widely studied estimation problem is the analysis of water quality using multispectral remote sensing sensors. The estimation of water quality parameters, such as concentrations of optically active parameters (i.e., chlorophyll, suspended sediments, and yellow matter) is of great importance for the monitoring of the ocean and lake ecosystems [5]–[10]. However, these are only a few examples of the wide range of possible applications in which the estimation of biophysical parameters from remote sensing images plays a fundamental role.

From a methodological viewpoint, for some kinds of biophysical parameters, it is common to define parametric model-based estimation algorithms [11]. These algorithms rely on a specific model that relates the studied biophysical parameter to measures acquired by the selected satellite sensor. Two different approaches can be considered. One consists of adopting a predefined direct model of the estimated biophysical parameter and inverting it on the basis of the available measurements. In this case, the model usually depends on several variables that require the knowledge of additional information about the analyzed scene (e.g., atmospheric conditions, sun angle, etc.). The other approach is based on the use of simple regression methods, which relate the investigated biophysical parameter to available measures according to interpolation techniques applied to a set of training samples (pairs of *in situ* concentrations and received radiances) [9], [12]. In this case, the regression process is conditioned to the type of interpolation function adopted (e.g., linear, polynomial, etc.). However, in general, it is reasonable to expect that a significant number of biophysical parameters are characterized by more complex relationships with the considered remote sensing data. For instance, when costal waters are considered in water quality estimation problems, the presence

Manuscript received January 22, 2004; revised October 1, 2004. This work was supported by the Italian Space Agency.

The authors are with the Department of Information and Communication Technologies, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it; melgani@dit.unitn.it).

Digital Object Identifier 10.1109/TGRS.2004.839818

of organic detritus and inorganic sediments strongly affects the linearity of water spectral characteristics. The optical behavior of such waters makes the relationship between parameter concentration and reflected radiance highly nonlinear. As a consequence, it is difficult to define both accurate direct models associated with the investigated parameter and proper nonlinear regression functions. In these situations, the use of nonlinear regression techniques based on machine-learning methodologies can represent an effective approach to the solution of estimation problems. In particular, artificial neural networks are a promising tool for an accurate assessment of biophysical parameter concentrations. In [13], it was proved that a multilayer perceptron (MLP) neural network with two layers of weights and sigmoidal activation functions can approximate any nonlinear functional relationship (mapping) with an arbitrary accuracy, provided that enough hidden neurons are available. This justifies the success of this particular class of neural networks in numerous fields of application of regression problems. In [14], it was found that an MLP (with one hidden layer having only two hidden nodes) applied to the three visible Landsat Thematic Mapper channels can model the problem of estimation of the concentrations of optically active water constituents with higher accuracy than what can be obtained by multiple regression analysis. In [15], the effectiveness of MLPs to determine phytoplankton components (such as the chlorophyll concentration and absorption of pigmented matter) was investigated. The authors found that MLPs achieved higher accuracies with respect to reference empirical band-ratio algorithms developed for the dataset they used in the experiments. Another interesting class of learning-by-example model is based on the radial basis function (RBF) neural networks. As opposed to MLPs, which perform a nonlinear combination of the output of hidden nodes, the RBFs derive estimations according to a linear combination of the output of a set of radial basis functions (generally of the Gaussian type). These functions are associated with the hidden units of the network and their parameters are derived in the training phase of the neural estimator. A generalized RBF neural network trained using the regression tree and forward-selection techniques was proposed in [16] for retrieving optically active sea-water parameters. In [17], a comparison between the MLP and RBF neural networks for the analysis of sea water revealed that the MLPs outperform the RBFs, providing a slight increase in estimation accuracy. However, despite their potential effectiveness, MLP and RBF neural networks present three main drawbacks: 1) they require addressing the problem of designing the best neural architecture, which often results in a complex, time-consuming task; 2) they can easily overfit the training data used to represent the considered regression problem (this can be particularly critical in presence of *in situ* measurements affected by some degree of uncertainty); 3) their variables do not apparently exhibit any explicit relation with physical parameters characterizing the estimation process.

A primary goal to achieve in the design of an estimation methodology, especially when the estimation process is carried out at a large geographical scale, is robustness, which is closely related to reliability of results. This is of particular importance when learning-based systems (such as artificial neural networks) are used for parameter estimation, since usually it is not possible to completely validate the learning phase of the network by relating explicitly the estimated network parameters to a physical model of the studied phenomenon. However, this goal is difficult to achieve with a single estimator. An alternative approach is to fuse the estimates obtained with an ensemble of regression algorithms, in order to exploit synergetically estimators with different characteristics. If the ensemble is properly designed, this results in more reliable final estimates (higher robustness) than those obtained with a single regression technique. It is worth noting that the goal of the fusion is not to outperform the single estimators but to obtain accuracies comparable to that of the best single estimator. In the remote sensing literature, no attention at all has been devoted to the use of ensembles of estimation algorithms, and it has also been rather neglected by the general regression literature. Among the few proposed methods, three linear-based combination approaches deserve to be cited. One is the definition of a framework for fusing multiple estimators where each estimator is trained with a different subset of the available samples. Then, the optimal linear combination is defined using a cross-validation method to derive the fusion weights [18], [19]. The other approach consists of finding the optimal combination weights by bootstrap sampling (Monte Carlo simulation). The authors reported results in which the bootstrap-based combination technique provided smaller model errors than classical averaging. In the third approach [20], the weights are regularized by means of the principal components regression, which leads to a decomposition of the original regression models into a new set of independent components. The most significant components are retained to deduce the weights for the original regression models by inverse mapping. This approach proved the most robust with respect to other conventional weight estimation techniques based on least squares and gradient descent procedures. It is worth noting that although ensemble methods have been scarcely considered in the context of regression problems, they have been studied extensively in the pattern recognition literature for the solution of challenging classification problems [21]–[24]. In this context, multiple classifier systems proved robust and accurate in many different remote sensing application domains, such as the classification of multisource data [25] and the updating of land-cover maps without support of ground-truth information [26].

In this paper, we present a novel approach to biophysical parameter estimation based on multiple estimator systems (MESs). Two different methodologies are considered for designing the MESs.

One extends the use of an approach proposed previously in the regression literature to the estimation of biophysical parameters from remote sensing data. This approach integrates the estimates obtained from the different regression algorithms making up the ensemble by a direct linear combination (combination-based approach). The linear combination can be accomplished with two different strategies: one is based on an unsupervised average operation, while the other carries out a weighted averaging after determining (in a supervised way) the weights to be assigned to each member of the ensemble.

The other methodology consists of a novel approach that provides as output the estimate obtained by the regression algorithm (included in the ensemble) characterized by the highest expected accuracy in the region of the feature space associated with the considered pattern (selection-based approach). This ap-

proach is inspired by the idea that different optimal estimators can be associated with different regions of the input feature space. An accurate partition of this space in decision regions (each associated with the expected most accurate single regression algorithm) can be carried out according to two different strategies. One exploits the concept of local accuracy measured on each single estimator, while the other exploits all the training samples available to infer the best partition of the feature space based on a global measure applied to the entire feature space.

A further interesting issue addressed in this paper is related to the choice of the estimation algorithms used for the definition of the ensemble. In particular, we present a detailed assessment of the effectiveness of a promising estimation approach based on support vector machines (SVMs) [27]. The choice to investigate the performances of SVMs in the context of biophysical parameter estimation is motivated by three main reasons: 1) the effectiveness of this approach in terms of accuracy and generalization capability (recently proved in numerous general regression applications [28]–[32]); 2) the limited effort required for architecture design (i.e., SVMs require the definition of few control parameters); 3) the possibility of solving the learning problem according to linearly constrained quadratic programming (QP) methods, which do not suffer from the problem of local minima. It is worth noting that a recent work has shown the effectiveness of such regression approach (as an alternative to MLPs) for the estimation of oceanic chlorophyll concentration [33]. In this paper, we propose to investigate further the performances of SVMs from different viewpoints including: 1) the influence of the kernel type in the SVM regression task; 2) the stability to the parameter settings; and 3) the sensitivity to the available number of training samples.

The proposed MES was applied to the problem of estimating water quality parameters. In particular, simulated data from the multispectral Medium Resolution Imaging Spectrometer (MERIS) sensor [mounted onboard the Environmental Satellite (Envisat)] related to the presence of chlorophyll in sea-water were considered [16]. It is worth noting that the use of these simulated data is particularly attractive for the purposes of this paper, since they do not only make it possible to assess the robustness and accuracy of the proposed regression system in a precise quantitative way, but also provide important hints on the effectiveness of the proposed system in estimating water quality parameters from the recently launched MERIS sensor.

The rest of this paper is organized in four sections. Section II presents the proposed multiple estimator system implemented with different combination-based and selection-based strategies. Section III describes the regression methodology based on SVMs, which is used to design each estimator of the ensemble. Section IV deals with the experimental phase of the work. Finally, Section V summarizes the material presented, discusses the results, and draws the conclusions of this paper.

II. Proposed Multiple Estimator System

A. Problem Formulation

Let us consider a set of N training samples (*in situ* measurements) $\mathbf{x}_i$ $(i = 1,2,\ldots,N)$ represented in the d-dimensional feature space $\Re^d$. Let us assume that a target $y_i \in \Re$ $(i = 1,2,\ldots,N)$ is associated with each vector $\mathbf{x}_i$,

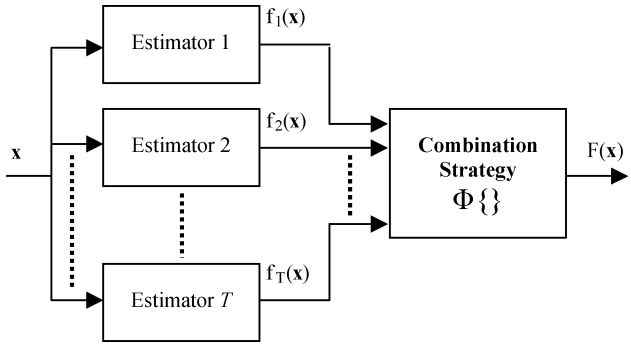


Fig. 1. General block diagram of the proposed MES.

where y_i is a real value representing the concentration of the considered biophysical parameter. Let us consider a set of T estimators $f_i(\mathbf{x})$ $(i = 1,2,\ldots,T)$ trained independently on the available training samples. It is worth noting that the T estimators can be based on the same regression method (with different parameter values) or on different methods. As depicted in Fig. 1, the problem is to define a combination strategy $\Phi\{\}$ such that the resulting estimate $F(\mathbf{x})$ (obtained after fusing the different single estimators) for a given unknown sample $\mathbf{x}$ is given by

$$F(\mathbf{x}) = \Phi\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_T(\mathbf{x})\right\}. \tag{1}$$

As stated in the introduction, we propose to design the combination strategy according to two different conceptual approaches. One extends to the remote sensing data analysis a method previously proposed in the regression literature, in which the final estimate is computed by a linear combination of the estimates obtained by the different single estimators (combination-based approach). The other approach consists of a novel method that derives the final estimate by selecting the output (estimate) of the "best single estimator" according to a strategy that defines a proper partition of the feature space (selection-based approach).

B. Combination-Based Approach

In this approach, two combination strategies are considered: the average combination strategy and the weighted combination strategy.

1) Average Combination Strategy: ACS is a simple unsupervised strategy in which the combination is based on the average operator. The rationale behind this strategy is that, from a statistical viewpoint, the different estimators can be considered as different random processes that model the same (biophysical) parameter. Optimal estimation based on first-order statistics can be obtained by a classical average operation. Accordingly, for a given sample $\mathbf{x}$, the resulting estimate $F(\mathbf{x})$ can be written as

$$F(\mathbf{x}) = \frac{1}{T}\sum_{i=1}^{T} f_i(\mathbf{x}). \tag{2}$$

Theory and experiments show that the average operator is effective if all estimators are unbiased and incur in uncorrelated errors with similar variances or, in general, when all the single estimators exhibit similar accuracies [18], [34].

2) Weighted Combination Strategy: By contrast with the ACS approach, WCS is a supervised combination strategy. The idea of WCS is to exploit available prior knowledge about the data (*in situ* measurements) in order to derive a weighted linear combination of the outputs of the estimators. Thanks to the assignment of a weight to each estimator, the linear combination model can be tuned in order to optimize the robustness of the estimation process. It is worth noting that ACS is a particular case of WCS in which all estimators have the same weight. Similarly to what is done in the pattern classification context with the Hybrid Consensus Theory [35], [36], the weight associated with each estimator can be seen as a "reliability factor" and each estimator can be considered a different information source. Reliability factors represent a way of expressing the degree of confidence of each information source and of weighting its influence in the combination process accordingly. The final estimate provided in output from WCS is given by

$$F(\mathbf{x}) = \sum_{i=1}^{T} \beta_i f_i(\mathbf{x}) \tag{3}$$

where β_i represents the reliability factor (weight) assigned to the ith estimator.

The problem in WCS is the determination of weight values. This problem can be addressed in different ways. A simple solution (which is widely used in the literature) is based on the minimum square error (MSE) pseudoinverse technique [37]. It consists in solving the following system of N linear equations with T unknown variables ($N > T$):

$$\begin{bmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \dots & f_T(\mathbf{x}_1) \\ f_1(\mathbf{x}_2) & f_2(\mathbf{x}_2) & \dots & f_T(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots \\ f_1(\mathbf{x}_N) & f_2(\mathbf{x}_N) & \dots & f_T(\mathbf{x}_N) \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_T \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}$$
$$\Leftrightarrow \overline{F} \cdot \overline{\beta} = \overline{Y}. \tag{4}$$

The estimate of the optimal weight vector $\overline{\beta}^*$ is given by the following equation based on the pseudoinverse $\overline{F}^{\#}$ of the matrix $\overline{F}$:

$$\overline{\beta}^* = \left(\overline{F}^{\mathrm{t}} \cdot \overline{F}\right)^{-1} \cdot \overline{F}^{\mathrm{t}} \cdot \overline{Y} = \overline{F}^{\#} \cdot \overline{Y}. \tag{5}$$

It is worth noting that more complex alternative methods could be used for estimating $\overline{\beta}^*$ (e.g., see [20]).

An interesting property of WCS is that, if proper values of the reliability factors are used, the combination of the estimates obtained from the different single estimators is less sensitive to their respective bias and variance than in the ACS strategy.

C. Selection-Based Approach

In this approach, we propose to analyze the accuracy of each single estimator included in the MES in different portions of the d-dimensional feature space. This is equivalent to making a partition of the d-dimensional feature space, in which each point is associated with the estimator of the ensemble that provides the minimum estimation error (Fig. 2). In other words, the MES behaves like an ideal selector of the most accurate estimate achieved by the set of available estimators. In this way, it is possible to better exploit the peculiarities of the different estimators

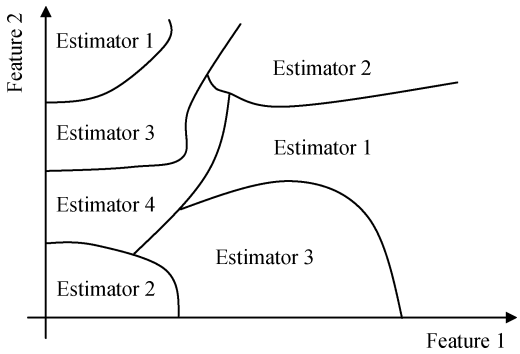


Fig. 2. Example of partition of a two-dimensional input feature space for an ensemble made up of four estimators (selection-based approach). Each decision region indicates the regression algorithm (included in the ensemble) that provides the expected most accurate estimation of the considered biophysical parameter in the corresponding portion of the feature space.

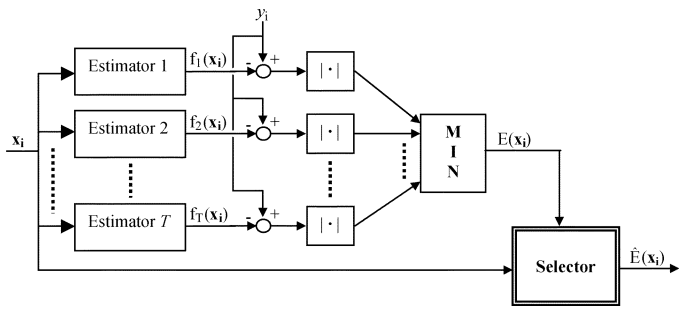


Fig. 3. Block diagram representing the training procedure of a MES implemented through the proposed selection-based approach. The expected best estimator $E(\mathbf{x_i})$ among the set of single estimators is identified according to the minimum value of the absolute error. This information is exploited to supervise the selection module in learning the optimal partition of the d-dimensional feature space.

in order to increase the robustness (and possibly the accuracy) of the estimation process in the entire feature space.

From an operative viewpoint, the selection-based approach can be applied according to the use of the prior knowledge contained in the available *in situ* measurements. It consists of two phases: the training phase and the estimation phase. As depicted in Fig. 3, the training phase includes the identification (among the set of available single estimators) of the expected best single estimator $\hat{E}(\mathbf{x})$ for each point $\mathbf{x}$ of the feature space. The optimal partition of the d-dimensional feature space in a set of regions is obtained according to the analysis of the training samples $\mathbf{x}_i$ ($i = 1, 2, \dots, N$), each being assigned to a single given estimator. The aim of the selection module is to model such an optimal partition in the best possible way. In our case, the concept of optimality is expressed in terms of minimum absolute error (MAE)

$$\hat{E}(\mathbf{x}_i) = \underset{k=1,2,\dots,T}{\arg\min} \left\{ \left| f_k(\mathbf{x}_i) - y_i \right| \right\}. \tag{6}$$

It is worth noting that any other monotonic error function (e.g., the minimum square error) could be adopted. However, this would not change the obtained partition as the result of the "min" operator is not affected by monotonic functions. During the estimation phase, each unknown sample $\mathbf{x} \in \Re^d$ (for which the true value of the investigated biophysical parameter is not known *a priori*) is given as input to the selector (classifier), which provides as output the estimate $\hat{E}(\mathbf{x}) \in \{1, 2, \dots, T\}$

of the most accurate estimator for the considered sample. The estimate $F(\mathbf{x})$ resulting from the MES can be written as

$$F(\mathbf{x}) = f_{\hat{E}(x)}(\mathbf{x}). \tag{7}$$

We propose two strategies of selection, which exploit prior knowledge in different ways: the local selection strategy and the global selection strategy.

1) Local Selection Strategy (LSS): This strategy is based on the concept of local accuracy of each single estimator in the d-dimensional feature space. It is worth noting that this concept was previously introduced in the framework of multiple classifier systems [38]. We propose to extend this concept to the case of the estimation problem, taking the peculiarities of this specific task into account.

The measure of local accuracy is based on the analysis of the effectiveness of different estimators in the area of the feature space surrounding the considered sample. This analysis can be accomplished by partitioning the feature space according to the behavior of the estimators on the training samples in a neighborhood of each point of the feature space. In other words, the identification of the best single estimator for a given unknown sample $\mathbf{x} \in \Re^d$ is obtained by analyzing the optimal estimators for the training samples nearest to $\mathbf{x}$ (it is worth noting that, in order to avoid overfitting, the training samples used in this phase should be different from the ones used in the learning of the single regression algorithms included in the ensemble). The LSS strategy can be implemented according to the K-nearest neighbors nonparametric estimation method, which results in the well-known K-nearest neighbors decision rule [37]. In our problem of selecting the best estimator $\hat{E}(\mathbf{x})$ of the ensemble for the unknown sample $\mathbf{x}$, this rule is equivalent to the following maximization:

$$\hat{E}(\mathbf{x}) = \underset{i=1,2,\ldots,T}{\arg\max} \{K_i(\mathbf{x})\} \tag{8}$$

where $K_i(\mathbf{x})$ stands for the number of training samples among the K nearest ones to the sample $\mathbf{x}$ that are assigned to the ith estimator (class). $K_i(\mathbf{x})$ $(i = 1, 2, \ldots, T)$ must satisfy the following condition:

$$\sum_{i=1}^{T} K_i(\mathbf{x}) = K. \tag{9}$$

For small values of K, the LSS provides a detailed partition of the feature space but, at the same time, it is subject to high sensitivity to noise (isolated training samples) involving a decrease in estimation reliability (robustness).

2) Global Selection Strategy (GSS): By contrast with the LSS, the GSS exploits the entire training set to partition the d-dimensional feature space among the T estimators included in the ensemble. All the training samples contribute to defining the model that approximates the regions of the feature space to which the single estimators are assigned. Compared to the LSS, this results in a smoother partition of the feature space that involves lower sensitivity to local variations of the MAE criterion adopted to identify the best single estimator. Since the distribution of the patterns in the portions of the feature space that should be associated with different estimators is not known *a priori*, nonparametric partition (classification) approaches must be adopted. In particular, we propose to use RBF neural networks trained with the technique described in [39]. The choice of this kind of classifier is motivated by the high classification accuracies it can obtain with an acceptable processing time. The typical architecture of an RBF neural network is depicted in Fig. 4. In particular, when applied to our partition problem, this neural architecture provides a set of outputs $Z_i(\mathbf{x})$ $(i = 1, 2, \ldots, T)$ that can be written as

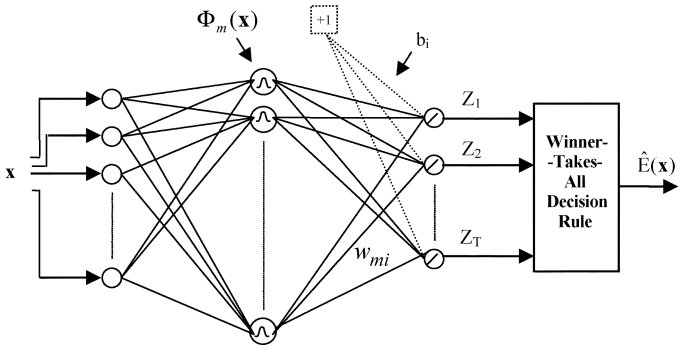


Fig. 4. Typical architecture of a selection module based on RBF neural networks.

$$Z_i(\mathbf{x}) = \sum_{m=1}^{M} w_{mi}\Phi_m(\mathbf{x}) + b_i \tag{10}$$

where M is the number of hidden neurons, w_{mi} represents the weight of the connection between the mth hidden and the ith output neurons, b_i stands for the bias associated with the ith output, and $\Phi_m(\cdot)$ refers to the kernel function associated to the mth hidden neuron. The selection of the best estimator $\hat{E}(\mathbf{x})$ of the ensemble for the unknown sample $\mathbf{x}$ is carried out according to the common winner-takes-all decision rule, i.e.,

$$\hat{E}(\mathbf{x}) = \underset{i=1,2,\ldots,T}{\arg\max} \{Z_i(\mathbf{x})\}. \tag{11}$$

In brief, the training procedure adopted for the considered RBF neural network deals with the estimation of both the kernel-function parameters and the weights and biases associated with the hidden and output neurons, respectively. The kernel-function parameters include the kernel center and width, which are computed by applying the k-means clustering algorithm separately to each class. This results in a reduced effect of the problem of mixed clusters (i.e., clusters shared by different classes) with respect to the standard case, in which the k-means clustering algorithm is applied simultaneously to all classes [39]. The training of the output layer consists in estimating the weights and the bias that characterize the links of each output neuron. For each output neuron, this task is accomplished by formulating the estimation problem as a linear system of N equations with $M + 1$ unknown variables $(N > M + 1)$ which is solved according to the MSE pseudoinverse technique. For more details, we refer the reader to [39].

III. Design of the Ensemble of Estimators

The MES requires the definition of an ensemble of different estimation algorithms. Such an ensemble can be designed by using the same regression method with different parameter settings (i.e., different architectures in the case of neural regression techniques) or according to different estimation methods.

In this paper, we considered both approaches, with a particular attention to the first one. Such an attention is strengthened by the interest in analyzing in greater detail the effectiveness of the approach based on SVMs for the estimation of biophysical parameters.

In the next subsection, the basics related to SVMs as a regression tool are briefly recalled.

A. Regression With SVMs

Let us refer to the estimation problem defined in Section II. In the ε-SVM regression approach introduced in [27], the goal is to find a function $f(\mathbf{x})$ that has at most ε deviation from the desired targets y_i and, at the same time, is as smooth as possible. This is obtained by mapping the data from the original d-dimensional domain to a higher dimensional feature space, i.e., $\Phi(\mathbf{x}) \in \Re^{d'}$ $(d' > d)$, both in order to increase the flatness of the function and to approximate it in a linear way as follows:

$$f(\mathbf{x}) = \mathbf{w}^* \cdot \Phi(\mathbf{x}) + b^*. \tag{12}$$

The optimal linear function in the higher dimensional feature space is the one that minimizes a cost function, which expresses a combination of two criteria: Euclidean norm minimization (which is equivalent to maximizing flatness) and error minimization. The cost function is defined as

$$\Psi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)\,. \tag{13}$$

This cost function minimization is subject to the following constraints:

$$\begin{cases} y_i - (\mathbf{w}\cdot\Phi(\mathbf{x}_i) + b) \le \varepsilon + \xi_i \\ (\mathbf{w}\cdot\Phi(\mathbf{x}_i) + b) - y_i \le \varepsilon + \xi_i^* \end{cases}, \qquad i = 1, 2, \ldots, N \tag{14}$$

and

$$\xi_i,\ \xi_i^* \ge 0, \qquad i = 1, 2, \ldots, N \tag{15}$$

where the ξ_i and ξ_i^*'s are the so-called *slack variables* introduced to account for samples that do not lie in the ε-deviation tube. Constant C represents a regularization parameter that allows to tune the tradeoff between the flatness of the function $f(\mathbf{x})$ and the value up to which deviations larger than ε are accepted. The formulation of the error function is equivalent to dealing with a so-called ε-insensitive loss function $|\xi|_\varepsilon$ typically defined as

$$|\xi|_\varepsilon = \begin{cases} 0, & \text{if } |\delta| \le \varepsilon \\ |\delta| - \varepsilon, & \text{otherwise} \end{cases} \tag{16}$$

where δ represents the deviation with respect to the desired target. This means that the differences between the targets and the estimated values are tolerated inside the ε-tube (error smallest than ε), while a linear penalty is assigned to estimates lying outside the ε-insensitive tube (see the example reported in Fig. 5).

The above optimization problem can be reformulated through a Lagrange functional. The Lagrange multipliers can be found by a dual optimization leading to a QP solution [40], [41]. The final result is a function of the data conveniently expressed in the original (lower) dimensional feature space as

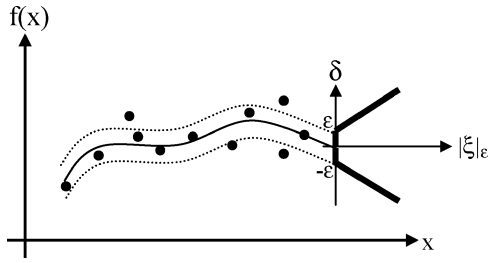


Fig. 5. Example of ε-insensitive tube and error function used in the SVM-based regression technique.

$$f(\mathbf{x}) = \sum_{i\in S}(\alpha_i - \alpha_i^*)\,K(\mathbf{x}_i, \mathbf{x}) + b^* \tag{17}$$

where $K(\cdot,\cdot)$ is a kernel function. S is the subset of indices $(i = 1, 2, \ldots, N)$ corresponding to the nonzero Lagrange multipliers α_i or α_i^*'s. The Lagrange multipliers weight each training sample according to its importance in determining a solution. The training samples associated to nonzero weights are called *support vectors*. In S, margin support vectors that lie within the ε-insensitive tube and nonmargin support vectors that correspond to errors coexist. The kernel $K(\cdot,\cdot)$ must satisfy the condition imposed by the Mercer's theorem so that it can correspond to some type of inner product in the transformed (higher) dimensional feature space [27]. Examples of common kernels that fulfill Mercer's condition are the polynomial kernel functions and the Gaussian radial basis functions.

An important aspect to be pointed out is the intrinsic good generalization capability of SVMs, which stems from the selection of the hyperplane that maximizes the flatness of the considered function in the transformed d'-dimensional feature space. In other words, the obtained solution minimizes the structural risk, i.e., it optimizes the *tradeoff between the quality of approximation function of the given data and the complexity of the approximating function* [27]. In a biophysical parameter estimation context, this maximum margin solution allows to obtain estimates that properly deal with the variability and uncertainty often associated with the few *in situ* measurements available.

We refer the reader to [27], [40], and [41] for greater detail on the SVM estimation theory.

IV. Experimental Results

A. Dataset Description

The proposed MES was assessed using a set of synthetic multispectral data that simulate the spectral behavior of the chlorophyll concentration in the subsurface coastal waters. The data simulate the first eight channels (412–681 [nm]) of MERIS mounted onboard the European Space Agency's Envisat satellite launched on March 2002. These channels are the most useful for sea color applications and, in particular, for the analysis of chlorophyll concentration. For greater details on the simulation procedure adopted to generate these data, we refer the reader to [16]. It is worth recalling that the MERIS sensor is a passive imaging spectrometer with 15 spectral bands

TABLE I
BEST MSE VALUES ON THE TEST SET AND COMPUTATIONAL TIME EXHIBITED BY SVMS AND MLPS WITH TWO HIDDEN LAYERS

ESTIMATION METHOD	BEST MSE	TIME [s]
SVM-Linear (C=500)	0.0262	4747
SVM-Polynomial (Order=3, C=40)	0.0013	3707
SVM-RBF (γ=0.2; C=1000)	0.0041	1743
MLP (5×5)	0.0013	6228
MLP (10×10)	0.0021	11124

(programmable in width and position within the visible and near-infrared range) and a ground spatial resolution of 300 m. The primary mission of MERIS is the measurement of sea color in the oceans and coastal areas in order to monitor the ocean carbon cycle and the thermal regime of the upper ocean, and manage coastal zones and fisheries.

In our experiments, the total number of available samples (pairs of *in situ* concentrations and received radiances) is equal to 5000. The range of variation of the chlorophyll concentration is from 0.02–25 mg/m^3. In order to make smoother the multidimensional function to be approximated, both concentration and simulated radiance values were converted to the logarithmic scale. The samples, defined in an *eight*-dimensional feature space, were divided into three sets: two training sets (each made up of 500 samples), and a test set (with 4000 samples). The first training set was used for the learning of the single estimators of the ensemble, while the second was necessary to train the different combination/selection supervised strategies proposed for the MES (i.e., the WCS, LSS, and GSS). It is worth noting that most samples were included in the test set to obtain small-size training sets with numbers of samples comparable to those typically available in operational applications of remote sensing image regression. Assessment of the effectiveness of the single SVM estimators and the MES was carried out based on the samples of the test set by computing different validation criteria, such as the MSE, MAE, and the minimum and maximum values of the absolute error. For the sake of brevity, in this paper, we report the results in terms of MSE, which represents the most commonly used validation criterion (the other results confirm the conclusions obtained by analyzing the behavior of the MSE).

B. Results Obtained With SVM Estimators

As stated in Section III, the choice of this kind of estimator is motivated by the interest in a detailed and complete assessment of the effectiveness of the SVM regression approach when applied to the biophysical parameter estimation problem. In particular, we considered three different kinds of SVMs: a linear SVM (SVM-Linear) (which corresponds to an SVM without kernel transformation), a nonlinear SVM with polynomial kernels (SVM-Polynomial), and a nonlinear SVM with Gaussian radial basis functions (SVM-RBF). This allowed us to evaluate the influence of the kernel type in the SVM regression process, and to obtain useful indications for choosing the estimators appropriate to implement the different scenarios defined to evaluate the robustness of the MES.

For the three SVM-based regression techniques, it was necessary to derive the value of the regularization parameter C,

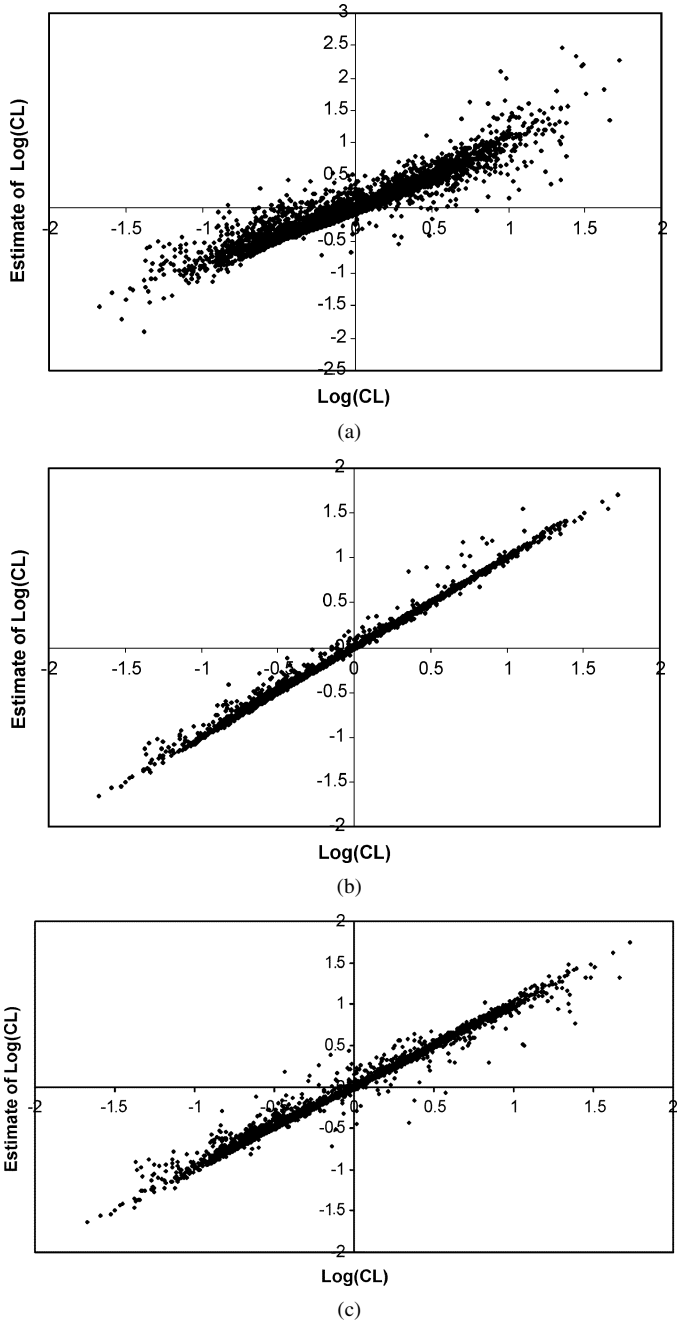


Fig. 6. Scatterplot of the best estimate of the chlorophyll concentration versus the true concentration Log(CL) obtained on the test set by (a) SVM-Linear estimator ($C = 500$), (b) SVM-Polynomial estimator (Order $= 3$; $C = 40$), and (c) SVM-RBF estimator ($\gamma = 0.2$; $C = 1000$).

since data are not ideally contained in the ε-insensitive tube (ε was fixed to a value of 0.001). By contrast with the linear SVM, the nonlinear SVMs required the determination of additional parameters, i.e., the order of the polynomial and the γ parameter for the SVM-Polynomial and the SVM-RBF, respectively. Concerning the SVM-polynomial, on the one hand, by increasing the order of polynomial kernels we can obtain more accurate regression potentialities. On the other hand, the generalization capabilities of the estimator decrease. This becomes critical in operational situations where the number of training samples is very limited and a high polynomial degree is considered (large number of coefficients to estimate). The γ parameter

TABLE II
ANALYSIS OF THE STABILITY OF THE MSE AND OF THE COMPUTATIONAL TIME VERSUS THE SETTING OF THE PARAMETERS OF THE CONSIDERED SVM REGRESSION TECHNIQUE (WITH DIFFERENT KERNEL FUNCTIONS)

ESTIMATION METHOD	PARAMETER RANGE	MSE		AVERAGE TIME [s]
		Average	Std. Deviation	
SVM-Linear	$C \in [1, 1000]$	0.0310	0.0069	804
SVM-Polynomial	$C \in [1, 1000]$; Order = 2	0.0038	0.0019	9114
SVM-Polynomial	$C \in [1, 1000]$; Order = 3	0.0017	0.0005	5660
SVM-Polynomial	$C \in [1, 1000]$; Order = 4	0.0031	0.0013	5769
SVM-RBF	$C \in [1, 1000]$; $\gamma = 1$	0.0258	0.0080	3005
SVM-RBF	$\gamma \in [0.05, 2]$; C = 1000	0.0149	0.0120	2241

of the SVM-RBF is related to the width of the Gaussian radial basis kernels, and consequently, tunes the smoothing of the approximating function.

Several experiments were carried out in order to identify empirically (on the basis of the test samples) the best parameter(s) associated with each of the three considered types of SVM (see Table I). As expected, the linear SVM proved the least effective among the three kernel types considered. The smallest MSE value found for the linear SVM was equal to 0.0262 ($C = 500$). This confirms that the linear regression often does not meet user's accuracy requirements. By contrast, with the nonlinear SVMs, one order of magnitude was gained in terms of MSE. In greater detail, the smallest MSEs obtained by the SVM-Polynomial and by the SVM-RBF were 0.0013 ($C = 40$; third-order polynomial) and 0.0041 ($C = 1000$; $\gamma = 0.2$), respectively. In Fig. 6, the scatterplots of the output of the different SVM-based regression techniques (SVM-Linear, SVM-Polynomial and SVM-RBF) illustrate graphically the accuracy of the obtained estimates of the chlorophyll concentration over the whole range of variation considered in the simulation (i.e., $[-2, +2]$ in the logarithmic scale). Such plots show clearly the very good estimation accuracies achieved by the nonlinear SVMs (and in particular the SVM-Polynomial) with respect to the linear SVM estimator. For the latter, one can observe higher deviations of the estimates from the true values of the chlorophyll concentration, which lead to strong over- and underestimations of the considered biophysical parameter [see Fig. 6(a)]. The better accuracies of the SVM-Polynomial with respect to the SVM-RBF can be motivated by the higher fitting flexibility of the polynomial kernel with respect to the symmetric Gaussian kernel. However, the higher number of parameters it requires results in a computationally more demanding training process (see Tables I and II).

For the sake of comparison, we also trained two multilayer perceptron neural networks characterized by two hidden layers with 5-5 and 10-10 neurons, respectively. The MSE obtained on the test samples and the computational time are reported in Table I. These results confirm that SVMs are a good alternative to the well-established neural regression method based on MLPs, since they achieve a very similar accuracy with (in average) a lower computational time.

In order to assess the sensitivity of each SVM-based estimator to the parameter settings, we derived some statistics by looking at the MSE and at the computational time as random realizations obtained varying the parameters in a predefined range of values. The results reported in Table II confirm the superiority of the nonlinear SVM based on polynomial kernels in terms of

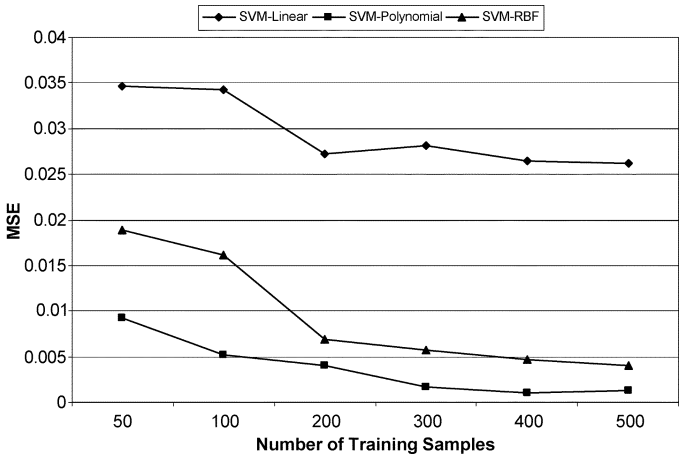


Fig. 7. Best MSE values obtained on the test set from the three considered SVM-based estimators versus the number of training samples used.

TABLE III
MSE YIELDED ON THE TEST SET IN THE FIRST TWO CONSIDERED EXPERIMENTS BY THE THREE SINGLE ESTIMATORS INCLUDED IN THE ENSEMBLE AND THE FOUR DESCRIBED MES STRATEGIES. THE BEST THEORETICAL ESTIMATION THAT CAN BE ACHIEVED BY THE SELECTION-BASED STRATEGY (ORACLE) IS ALSO GIVEN FOR COMPARISON

Estimation Method	Mean square error (MSE)	
	Experiment 1	**Experiment 2**
Estimator 1	SVM-Linear (C = 500) 0.0262	SVM-Polynomial (Order = 2 ; C= 1000) 0.0018
Estimator 2	SVM-Polynomial (Order = 3 ; C= 40) 0.0013	SVM-Polynomial (Order = 3 ; C= 40) 0.0013
Estimator 3	SVM-RBF (γ = 0.2 ; C = 1000) 0.0041	SVM-Polynomial (Order = 4 ; C= 10) 0.0014
ACS	0.0052	0.0011
WCS	0.0015	0.0012
LSS	0.0025	0.0013
GSS	0.0017	0.0013
Oracle	**0.0008**	**0.0005**

average MSE and stability (it provided the lowest standard deviations) even though it proved to be the most computationally demanding. It is worth noting that nonlinear SVMs based both on the polynomial and the RBF kernels seem to be more sensitive to the choice of the polynomial order and kernel width value γ, respectively, than to the regularization parameter C. The linear SVM showed the worst average MSE (equal to 0.0310), since a

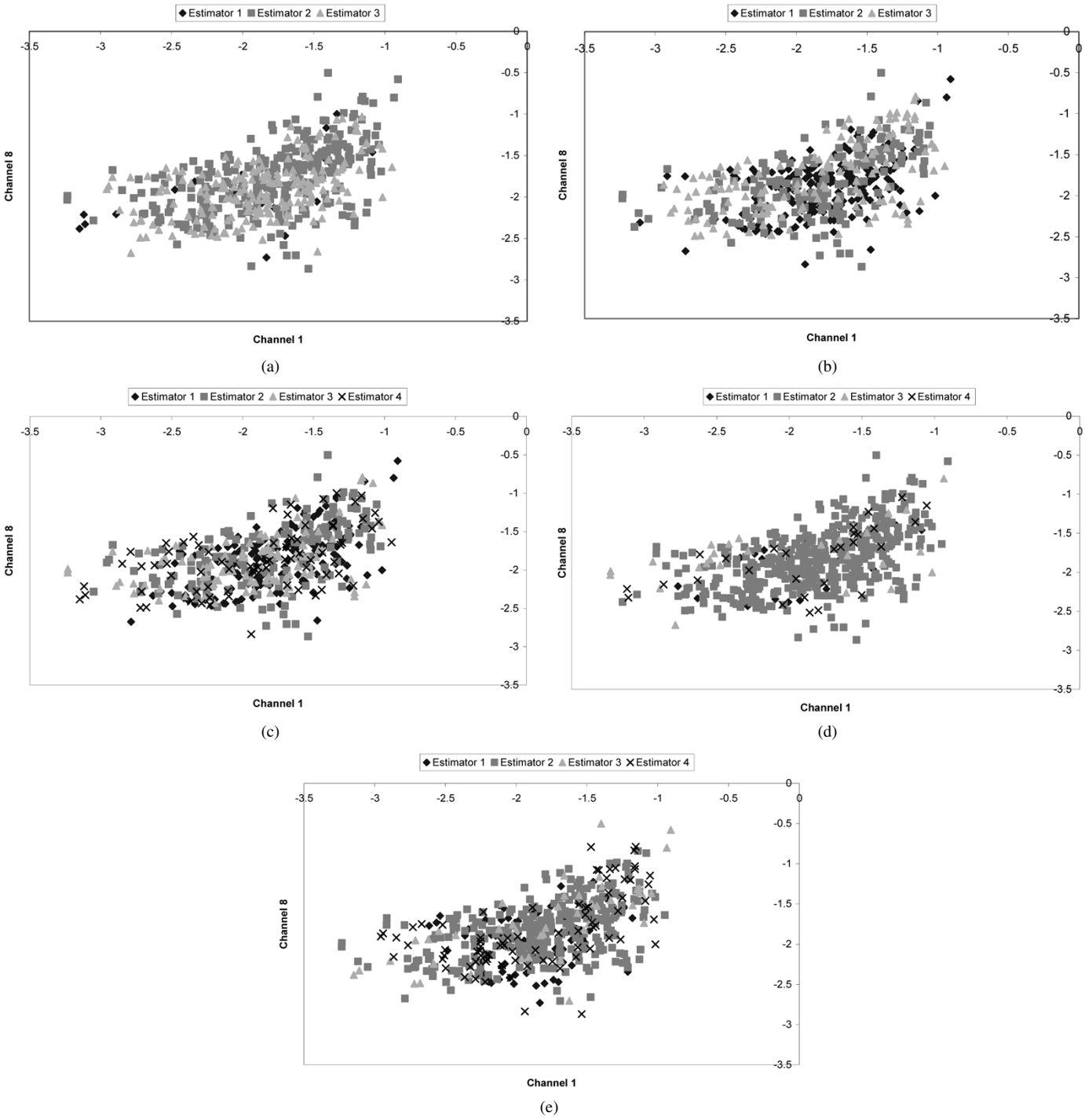


Fig. 8. Best possible (optimal) association between training patterns and estimators included in the ensemble in the bidimensional input feature space defined by channels 1 and 8 (represented in logarithmic scale). (a) Experiment 1. (b) Experiment 2. (c) Experiment 3. (d) Experiment 4. (e) Experiment 5. The description of the estimators is provided in Table III for Experiments 1 and 2 and in Table V for Experiments 3–5.

linear kernel applied to a nonlinear estimation problem involves a large number of error samples lying outside the ε-insensitive tube. This significantly decreases the effectiveness of the regularization mechanism implemented in the SVM approach to take into account the presence of error samples.

Finally, to provide a complete assessment of the performances of the SVM regression approach, we analyzed its sensitivity to the number of samples used in the training phase. We generated different training sets by decreasing the number of training samples starting from 500 down to 50 samples. For each set, we trained the three different SVM-based regression techniques (SVM-Linear, SVM-Polynomial and SVM-RBF) until the lowest possible MSE value was reached (empirically) on the test set. Fig. 7 depicts the behavior of the MSE versus the number of training samples for each kernel type. From 500 to 50 training samples, the ranges of variation of the MSE were [0.0262, 0.0347], [0.0013, 0.0092], and [0.0041, 0.0189] for the SVM-Linear, the SVM-Polynomial, and the SVM-RBF, respectively. These results clearly confirm the superiority of the SVM-Polynomial over the two other techniques. In addi-

tion, they point out the low sensitivity of SVMs (in particular of the SVM-Polynomial) to the number of training samples used in the learning phase. This is explained by the fact that, by exploiting a margin-based "geometrical" criterion, SVMs define the best solution (i.e., the one that maximizes the fitting of the considered function) by using only a specific subset of the available training samples (i.e., the support vectors). Accordingly, even if few training samples (but representative enough) are used, the SVMs are capable of providing a good approximation of the function to be modeled.

C. Results Obtained With the Proposed Multiple Estimator Systems

Let us now consider the proposed MESs. Five experiments were carried out to assess the effectiveness of the MESs in different operational conditions where the single estimators included in the ensemble are: 1) based on the same or on different regression methods; 2) characterized by different tradeoffs between correlation of errors and estimation accuracy; 3) trained on samples affected or not by measurement errors.

The first group of experiments refers to two scenarios where the same regression method (which is based on SVMs) is used for all the members of the ensemble. In particular, in the first experiment, in order to assess the robustness of the MES, the ensemble included a "poor" estimator and two "good" estimators. The second experiment aimed at studying the ability of the MES to exploit the information present in the ensemble even in presence of correlated estimators.

The second group of experiments represents three different scenarios in which the estimators included in the ensembles are designed by adopting two different regression methods based on SVMs and MLPs. In particular, in the third experiment, the ensemble was composed of two SVM-based and two MLP-based estimators. The fourth and fifth experiments were carried out to assess the robustness of the MES to the problem of measurement errors. To this end, the four estimators adopted in the third experiment were learnt on a training set partially corrupted by measurement errors (with different proportions for the fourth and fifth experiments).

In order to quantify the measure of diversity between members of an ensemble, similarly to what done in [42] for the combination of ensembles of classifiers, we propose to use a pairwise correlation coefficient of errors defined as follows:

$$\rho_{mn} = \frac{\sum_{i=1}^{S} [f_m(\mathbf{x_i}) - y_i] \cdot [f_n(\mathbf{x_i}) - y_i]}{\sqrt{\sum_{i=1}^{S} [f_m(\mathbf{x_i}) - y_i]^2} \cdot \sqrt{\sum_{i=1}^{S} [f_n(\mathbf{x_i}) - y_i]^2}} \tag{18}$$

where ρ_{mn} is the correlation coefficient (defined in the range $[-1, 1]$) between the mth and nth estimators of the ensemble and S stands for the number of test samples. A zero value of the coefficient indicates that the two estimators are uncorrelated, whereas values equal to $+1$ and -1 point out that the estimators incur in completely correlated errors. The global correlation coefficient ρ can thus be approximated by a simple averaging of the pairwise correlation coefficients, i.e.,

$$\rho = \frac{2}{T(T-1)} \sum_{m=1}^{T} \sum_{n>m}^{T} \rho_{mn}. \tag{19}$$

The five experiments carried out are described in detail in the following subsections.

1) Experiment 1: Ensemble With a "Poor" Estimator: This experiment is designed to study the robustness of the MES when the ensemble includes a "poor" estimator (i.e., an estimator with a relatively high MSE value). This is a realistic situation in operational applications because the number of *in situ* samples is often insufficient to guarantee that the evaluation of the effectiveness of the estimator carried out on the test samples is representative of the true MSE on all considered data. It is worth noting that this situation is particularly critical when only one regression algorithm is used to solve the estimation problem. In order to obtain an ensemble with this characteristic, we considered a set of three different kinds of estimators (SVM-Linear, SVM-Polynomial, and SVM-RBF). Each kind of SVM-based estimator was trained to reach its best accuracy (i.e., the lowest MSE value) on the test samples. The optimization of the estimators on the test set allowed to create a condition where high differences in accuracy characterize the estimators included in the ensemble. The MSEs characterizing the set of single estimators were very different, since an order of magnitude separates the MSE achieved by the SVM-Linear from those obtained by the nonlinear SVMs (see Table III). Accordingly, the poor estimator of the considered ensemble is represented by the SVM-Linear estimator. The value of the correlation coefficient ρ on the test samples was found equal to 0.416, indicating thus a relatively weak correlation among the errors incurred by the three members of the ensemble due to the consistent difference of accuracy between them.

Once the different single estimators were trained using the first training set, we applied the learning phase to the three investigated supervised strategies (WCS, LSS, and GSS) based on the samples of the second training set. Concerning the WCS, the values of the three weights defining the linear combination among the outputs of the SVM-Linear, SVM-Polynomial and SVM-RBF were determined automatically by means of an MSE pseudoinverse technique (see Section II-B2). Differently, the training phase of the two selection-based strategies consisted of defining an effective partition of the feature space according to the minimization of the MAE. In order to better understand the complexity of the problem, Fig. 8(a) shows the best possible association between training patterns and estimators included in the ensemble in the bidimensional feature space defined by channels 1 and 8 (these channels were chosen as they represented the least correlated spectral bands among the eight available ones). This optimal association (which is equivalent to an "oracle") was obtained by a manual supervised analysis of the outputs of the estimators included in the ensemble for the training patterns. In other words, it represents the target-partition that the two strategies LSS and GSS will attempt to model during their learning phase. From the figure, one can make two

considerations. The first is related to the SVM-Linear estimator, which clearly appears to be an unreliable source of information throughout almost the entire bidimensional feature subspace (only few local partitions of the feature space were assigned to this estimator). The second observation concerns the two nonlinear SVM estimators of the ensemble. The partition of the feature space is: 1) dominated by the SVM-Polynomial (since it is the most accurate in the ensemble); 2) in spite of the fact that the feature space is characterized by some large homogeneous regions assigned to the same estimator, the partition seems fragmented/localized, in particular in the central region of the feature space. This second point can be explained by the fact that the high estimation accuracy reached by the two nonlinear estimators renders their estimation error functions more similar, without a dominant behavior of one on the other over the entire feature space. Accordingly, the selection of the best of them depends on the specific portion of the feature space considered and is therefore localized. On the basis of the samples of the second training set, the optimal value of nearest neighbors for the LSS was $K = 1$ and the optimal number of hidden nodes in the Radial Basis Functions neural network [39] for the GSS was found equal to 20. It is worth underlining that the computational time required by all the three supervised MES strategies was very short.

Table III reports the results achieved by the different strategies on the considered 4000 test samples. Despite the presence of a poor estimator in the ensemble, the three supervised MES strategies (WCS, LSS, and GSS) proved very robust since they allowed to maintain a final accuracy comparable to that reached by the best single estimator (SVM-Polynomial) included in the pool of regression algorithms. This means that the MES is effective in reducing the effects of the unpredictable presence of a poor estimator in the ensemble. This important property is explained by the fact that the supervised MESs are capable of discriminating between the most and the least reliable information sources by properly assigning weights (combination-based strategy) or regions of the input feature space (selection-based strategy) to each estimator.[1] In greater detail, the best strategy was the WCS, which reduced the MSE achieved by the best and the worst single estimators, respectively, by a factor of 0.87 (which means a slight increase in the MSE) and 17.47. As expected, a situation in which the accuracies of the single estimators of the ensemble are not balanced strongly penalizes the ACS. Nonetheless, it is worth underlining that the ACS provided far more accurate estimates than the worst single estimator (SVM-Linear) of the ensemble, and that it is characterized by an MSE of the same order of magnitude as that provided by the best single estimator (SVM-Polynomial).

The lower bound (oracle) achievable by the two selection-based strategies on the test samples is also reported in Table III. This bound is useful to understand the capability of the LSS and GSS strategies in capturing the best estimate available in the ensemble. It is computed by selecting the best single estimator for each sample of the test set, using the prior knowledge on the errors incurred by each member of the ensemble. In this experiment, it clearly appears that the LSS and GSS did not reach the best achievable performances (MSE $=$ 0.0025 and 0.0017, respectively, against 0.0008 for the oracle) due to the relative complexity of the selection task [Fig. 8(a)].

[1]It is worth noting that, in this experiment, all the members of the ensemble were optimized on the test samples (i.e., they represent the best possible estimators with each kind of kernel on the test set) whereas the MES strategies were not optimized on these samples. This penalized the MES approach with respect to single estimators.

TABLE IV
(a) BIAS AND (b) VARIANCE OF THE ESTIMATION ERRORS OBTAINED ON THE TEST SET BY THE THREE CONSIDERED SINGLE ESTIMATORS AND THE FOUR DESCRIBED MES STRATEGIES FOR THE FIRST TWO REPORTED EXPERIMENTS. THE BEST THEORETICAL ESTIMATION THAT CAN BE ACHIEVED BY THE SELECTION-BASED STRATEGY (ORACLE) IS ALSO GIVEN FOR COMPARISON

a

Estimation Method	Error bias	
	Experiment 1	Experiment 2
Estimator 1	-0.0300	-0.0072
Estimator 2	-0.0058	-0.0058
Estimator 3	-0.0090	-0.0022
ACS	-0.0149	-0.0051
WCS	-0.0049	-0.0062
LSS	-0.0088	-0.0062
GSS	-0.0073	-0.0068
Oracle	**-0.0046**	**-0.0034**

b

Estimation Method	Error variance	
	Experiment 1	Experiment 2
Estimator 1	0.0253	0.0017
Estimator 2	0.0013	0.0013
Estimator 3	0.0040	0.0014
ACS	0.0049	0.0011
WCS	0.0014	0.0012
LSS	0.0024	0.0013
GSS	0.0016	0.0013
Oracle	**0.0008**	**0.0005**

Finally, we analyzed the performances of the different estimation strategies by comparing the bias and variance of their estimation error on the test samples (see Table IV). These values motivate the MSE obtained by the ACS, which is higher than that exhibited by the WCS. In particular, one can observe that the three different single estimators of the ensemble are relatively biased (-0.0300, -0.0058, and -0.0090) and have different variances (0.0253, 0.0013, and 0.0040). Accordingly, they do not meet the theoretical conditions required by the unsupervised linear average operator [18], [34]. Nonetheless, the supervised linear WCS decreased the variance and the MSE of the final estimate with respect to the ACS (0.0014 against 0.0049, and 0.0015 against 0.0052 for the variance and the MSE, respectively), as it explicitly considers the reliability information of each member of the ensemble. This information is useful to give less weight to unreliable estimators (like Estimator 1, which is characterized by the highest bias and variance values) and, accordingly, to attenuate the problem of the biases and disparities between the variances of the single estimators. The results also show that the selection-based approach only partially overcomes the constraint on the biases and variances of the single estimators, providing performances that are better than the ACS but slightly worse with respect to the WCS. This different tradeoff depends on the fact that it is not based on a direct combination of the estimates yielded by all the

estimators, but on a selection-based mechanism that identifies the expected best available estimate.

2) Experiment 2: Ensemble With Correlated Estimators: The second experiment was aimed at assessing the effectiveness of an ensemble made up of correlated estimators. To this end, we defined a set of three SVM estimators with the same kernel model (SVM-Polynomial) by changing the polynomial order (*second*, *third*, and *fourth*-order) that controls the fitting flexibility. It is worth underlining that the limited number of training samples (500) penalizes high-order polynomials. This is confirmed by the obtained results, in which the MSE on the test set slightly increases by moving from a *third*-order to a *fourth*-order polynomial (see Table III). As expected, the fact that the three estimators are based on the same type of kernel function and have high and very similar accuracies increased the global correlation coefficient to 0.616.

In the MES training phase, it was found empirically that for the LSS the best K value was 3, whereas for the GSS the optimal number of hidden nodes in the RBF-based selector was 30. Fig. 8(b) confirms what was partially observed in Experiment 1, i.e., estimators with comparable high accuracies result in similar error functions, without a dominant behavior of one on the other over the entire feature space. This involves a strong fragmentation of the optimal partition of the feature space, which is difficult to model with the selection-based approach. This is confirmed by the experimental results, which show that, in spite of the fact that the LSS and GSS strategies proved satisfactorily robust (i.e., they provided MSE values lower than those of the worst estimator of the ensemble), they were not capable of reducing the MSE value of the final estimate (0.0013) with respect to that of the best estimator of the ensemble (0.0013). Since the MSE value obtained by the oracle (lower bound achievable for the considered ensemble of estimators) was lower (0.0005) than those of LSS and GSS, we can conclude that from a theoretical viewpoint the selection-based strategy can improve the estimation accuracy also in this case, but the considered strategies could not fully exploit the available information to model the complexity of the selection process. By contrast, in such a situation (characterized by accurate and correlated single estimators), the combination-based approach (CBA) proved to be the most effective. In particular, the best MSE value was yielded by the ACS, which resulted in a reduction of the MSE obtained by the best and the worst single estimators by a factor of 1.18 and 1.64, respectively (see Table III). This is motivated by the fact that the three estimators are almost unbiased and have very similar variances (see Table IV).

3) Experiment 3: Ensemble With Different Regression Methods: The third experiment was carried out in order to assess the effectiveness of the MES with an ensemble composed of estimators based on different regression methods. This setup, which is widely adopted in the context of multiple classifier systems, allows one to define members of the ensemble characterized by significantly uncorrelated estimation errors. In particular, we designed an ensemble composed of two SVM-based and two MLP-based estimators exhibiting comparable accuracies (see Table V). The global correlation coefficient characterizing this ensemble is equal to 0.555.

TABLE V
MSE YIELDED ON THE TEST SET IN THE LAST THREE CONSIDERED EXPERIMENTS BY THE FOUR SINGLE ESTIMATORS INCLUDED IN THE ENSEMBLE AND THE FOUR DESCRIBED MES STRATEGIES. THE BEST THEORETICAL ESTIMATION THAT CAN BE ACHIEVED BY THE SELECTION-BASED STRATEGY (ORACLE) IS ALSO GIVEN FOR COMPARISON

Estimation Method	Mean square error (MSE)		
	Experiment 3	**Experiment 4**	**Experiment 5**
Estimator 1	SVM-Polynomial (Order = 2 ; C= 1000) 0.0018	SVM-Polynomial (Order = 2 ; C= 1000) 0.0024	SVM-Polynomial (Order = 2 ; C= 1000) 0.0947
Estimator 2	SVM-Polynomial (Order = 3 ; C= 40) 0.0013	SVM-Polynomial (Order = 3 ; C= 40) 0.0014	SVM-Polynomial (Order = 3 ; C= 40) 0.0120
Estimator 3	MLP (5x5) 0.0013	MLP (5x5) 0.0156	MLP (5x5) 0.0751
Estimator 4	MLP (10x10) 0.0021	MLP (10x10) 0.0205	MLP (10x10) 0.0814
ACS	0.0011	0.0059	0.0395
WCS	0.0011	0.0077	0.1260
LSS	0.0012	0.0014	0.0127
GSS	0.0014	0.0073	0.0216
Oracle	**0.0004**	**0.0007**	**0.0047**

In the learning phase of the MES, the best parameter values empirically derived (on the basis of the second training set) for the LSS and GSS were $K = 17$ and 60 hidden nodes, respectively. Fig. 8(c) confirms what was expected from the previous experiment (Experiment 2), i.e., accurate and comparable single estimator accuracies result in a high risk of local fragmentation of the feature space in randomly distributed small decision regions. This effect renders particularly difficult the task of modeling the various decision boundaries.

In general, in this experiment, the MSE, the estimation error bias and variance achieved by the different MES strategies are very similar to those obtained in the second experiment (see Tables V and VI). This was expected for two main reasons: 1) the single estimators in the two ensembles are characterized by similar and high accuracies; and 2) the global correlation coefficients of the two ensembles are similar. All the four strategies provided accuracies very close to that exhibited from the best single estimator and, in most cases, slightly better. This confirms the capability of the MES to enhance the robustness of the estimation task. In greater detail, the best strategies were the ACS and WCS, which reduced the MSE obtained by the best and the worst single estimators by a factor of 1.18 and 1.91, respectively (see Table V). This confirms the fact that, in an ensemble characterized by accurate and moderately uncorrelated estimators, the combination-based approach is more attractive than the selection-based approach thanks to its effectiveness and lower computational complexity.

4) Experiments 4 and 5: Robustness to Measurement Errors: In order to complete the experimental analysis of the proposed MES approach, we carried out two other experiments (Experiments 4 and 5) to evaluate the robustness of the MES to the problem of measurement errors. Such a problem is typical in regression applications (e.g., estimation of chlorophyll concentrations in subsurface waters) in which the time between the *in situ* measurements and the remote sensing data acquisition may result in changes in the concentrations of the analyzed parameter. Another possible motivation of measurement errors is related

TABLE VI
(a) BIAS AND (a) VARIANCE OF THE ESTIMATION ERRORS OBTAINED ON THE TEST SET BY THE FOUR CONSIDERED SINGLE ESTIMATORS AND THE FOUR DESCRIBED MES STRATEGIES FOR THE LAST THREE REPORTED EXPERIMENTS. THE BEST THEORETICAL ESTIMATION THAT CAN BE ACHIEVED BY THE SELECTION-BASED STRATEGY (ORACLE) IS ALSO GIVEN FOR COMPARISON

a

Estimation Method	Error bias		
	Experiment 3	Experiment 4	Experiment 5
Estimator 1	-0.0073	-0.0173	-0.0872
Estimator 2	-0.0058	0.0100	-0.0387
Estimator 3	-0.0023	-0.1086	-0.2400
Estimator 4	0.0051	-0.1285	-0.2534
ACS	-0.0026	-0.0648	-0.1552
WCS	-0.0043	-0.0728	-0.3242
LSS	-0.0069	-0.0100	-0.0412
GSS	-0.0054	-0.0510	-0.0686
Oracle	**-0.0026**	**-0.0100**	**-0.0245**

b

Estimation Method	Error variance		
	Experiment 3	Experiment 4	Experiment 5
Estimator 1	0.0017	0.0021	0.0871
Estimator 2	0.0013	0.0013	0.0105
Estimator 3	0.0013	0.0038	0.0175
Estimator 4	0.0021	0.0040	0.0172
ACS	0.0011	0.0017	0.0154
WCS	0.0011	0.0024	0.0209
LSS	0.0012	0.0013	0.0110
GSS	0.0014	0.0047	0.0169
Oracle	**0.0004**	**0.0006**	**0.0041**

to the low spatial resolution of several remote sensors (e.g., the MERIS sensor with 300 m), which makes it difficult an accurate *in situ* assessment of the investigated parameter over the entire ground-projected instantaneous field of view.

In these experiments, we generated new training sets partially corrupted by measurement errors with two different proportions for the fourth and fifth experiments. This was done by adding to the real target (expressed in the logarithmic scale) of 1/10 and 1/3 of the two training sets, for the fourth and fifth experiments respectively, a uniform random noise defined in the interval $[-1, +1]$, while the test samples were kept unchanged. For each trial, the four estimators used in the third experiment were trained again with the same parameters on the corresponding new first training set. As expected, their accuracy decreased with increasing proportions (from 0 to 1/10 and 1/3) of corrupted samples (see Table V). In greater detail, with 1/10 of noisy training samples, the MLP-based estimators resulted in a significantly higher MSE (about one order of magnitude) than the SVM-based estimators. This confirms the theoretically expected good generalization capabilities of SVMs. The superiority of SVMs was maintained also with a higher proportion of noisy samples (i.e., 1/3). In this case, the best single estimator was the SVM-Polynomial of third order $(\text{MSE} = 0.0120)$ followed by the MLP with 5×5 hidden neurons (0.0751), the MLP with 10×10 hidden neurons (0.0814), and the SVM-Polynomial of second order (0.0947). By analyzing Table VI, one can observe that the relatively poor results of the MLP are explained by the high bias values it introduces in the estimation process due to the presence of perturbations in the real desired targets. It is worth noting that the global correlation coefficients characterizing the two ensembles in Experiments 4 and 5 were equal to 0.524 and 0.491, respectively.

The three supervised fusion strategies (WCS, LSS, and GSS) were learnt by exploiting the samples of the second training set (which is also corrupted by noise). For the two selection-based LSS and GSS strategies, the best parameter values were $K = 19$ and a number of hidden nodes equal to 30, respectively. Fig. 8(d) and (e) depicts the optimal decision regions achievable on the samples of the second training set (partially corrupted by measurement errors). Since the test samples do not contain any noise perturbation for simulating measurement uncertainty, the target-partitions achieved by using the training samples involve a distortion in the decision regions. Fig. 8(d) shows that the most accurate single estimator of the ensemble (i.e., the SVM-Polynomial of third order) dominated almost the whole feature space, whereas Fig. 8(e) presents a slightly more balanced partition among the estimators.

In general, despite the presence of training samples affected by unreliable measurements of the chlorophyll concentration, the MES strategies allowed to provide more robust performances with respect to those achieved by the single estimators. In greater detail, in Experiment 4, all the four strategies allowed to overcome the strong sensibility to error measurements exhibited by the MLP-based single estimators. The best MES strategy was the LSS, which yielded the same accuracy of the best single estimator (i.e., SVM-Polynomial of third order with $\text{MSE} = 0.0014$). In Experiment 5, a similar behavior can be observed, with the exception that the WCS strategy failed completely in combining the estimates provided by the members of the ensemble. This is motivated by the fact that the pseudoinverse technique adopted to estimate the weights associated with each estimator included in the ensemble is directly affected from the measurement errors (since it minimizes the square error between the estimates and the targets partially affected by noise). In such a situation, it is better to utilize the simple ACS strategy that assigns the same weights to all the members of the ensemble.

These experiments confirm that the MES represents a valid solution to the problem of measurement errors, which can affect, in an uncontrolled manner and with different degrees, single estimation algorithms. In such a situation of high uncertainty in the measurements, MESs, in general, and the LSS, in particular (thanks to the concept of local accuracy) increase significantly the robustness and the reliability of the estimation process.

V. DISCUSSION AND CONCLUSION

In this paper, a novel approach to the estimation of biophysical parameters from remote sensing images based on a multiple estimator system has been presented. The MES aims at exploiting the peculiarities of an ensemble of different estimators to improve the robustness (and in some cases the accuracy) of the estimation process. Four different strategies to implement the MES have been described. The four strategies differ from each other for: 1) the fusion procedure; 2) their supervised or unsupervised properties; and 3) the technique with which the available prior information is exploited.

The ensembles of estimators used in the experiments were based on both support vector machines and multilayer perceptron neural networks. The detailed experimental analysis carried out to evaluate the effectiveness of SVMs for biophysical parameter estimation points out that: 1) they can provide estimation accuracies very similar to what can be yielded by other effective neural regression methods like MLPs; 2) they exhibit low sensitivity to the critical problem of the small number of training samples available; 3) they intrinsically present very good generalization capabilities (confirmed by the experiments carried out with simulated data affected by measurement errors); and 4) they require (in average) a lower computational cost than that involved by the widely used MLPs. It is worth noting that the second and third properties are very important in applications like biophysical parameter estimation, in which the number of *in situ* measurements available is often scarce and characterized by significant uncertainty degrees.

In order to obtain a reliable and detailed assessment of the effectiveness of the proposed MES, five experiments (with different ensembles of SVM and MLP estimators) were designed. These experiments simulated different operational conditions, in which members of the ensemble are: 1) based on the same or on different regression methods; 2) characterized by different tradeoffs between correlated errors and accuracy of the estimates; 3) trained on samples affected or not by measurement errors. The results of the first experiment confirm that, if an ensemble with an unpredictable "poor" estimator is defined, the MESs are capable of overcoming its effects, thus exhibiting significant robustness. The second experiment points out the ability of the MES to exploit the ensemble to slightly improve the estimation accuracy even in presence of correlated estimators. The third experiment underlines that the fusion of single estimators based on different regression methods (in this case, SVMs and MLPs) represents a valid way to enhance the robustness of the estimation process. Finally, the fourth and fifth experiments point out the robustness of MES to measurement errors that often affect the quality of the few available training samples in real situations.

In general, all the experimental results pointed out the ability of the MESs to increase the robustness of the estimation process since in most cases they provided MSE values very close to that of the best estimator included in the ensemble.

Comparisons between the two considered fusion approaches point out that, on the one hand, the selection-based approach proved to be more effective than the combination-based approach in situations where data are corrupted by measurement errors. This depends on its ability to properly exploit the different accuracies of the estimators included in the ensemble in different portions of the input feature space. On the other hand, the combination-based approach is more attractive from the viewpoint of implementation simplicity. Concerning comparisons among the specific fusion strategies presented, one can distinguish two situations. If data are corrupted by measurement errors, which is the most realistic situation, the proposed LSS outperformed the GSS followed by the ACS and the WCS. In absence of measurement errors, the best strategy is the WCS with which competes the GSS followed by the LSS and the ACS. However, only very slight differences can be observed between them in this second situation by contrast to the first one. From the analysis of the obtained results, as expected from the theory, it turns out that the ACS is affected by different values of the biases and variances of the single estimators making up the ensemble. Concerning the WCS, it is effective if the available training samples are characterized by low measurement uncertainties; otherwise it fails completely (see Experiment 5). A comparison between the MSE values obtained by the two proposed selection-based strategies (LSS and GSS) and the minimal achievable MSE (oracle) suggests that future studies should be developed to define more complex strategies capable of further increasing accuracy in the partitions of the input feature space and, accordingly, the robustness of the system. In particular, the selection-based approach could be improved by considering: 1) an adaptive weighted selection (instead of a crisp selection as done in the LSS and GSS) where the weights are adapted according to the error committed by each single estimator in any point of the space; 2) a rejection mechanism to ensure that the system provides a reliable global estimate by identifying regions of the feature space in which all the single estimators are not enough accurate; and 3) a hybrid system that fuses in a supervised way both the selection and combination-based approaches.

ACKNOWLEDGMENT

The authors wish to thank Prof. G. Corsini (University of Pisa, Italy) for providing the simulated data of the MERIS sensor and Prof. S. Rüping for supplying the software used in the experiments (see http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/). The authors are also grateful to the anonymous referees for their constructive comments.

REFERENCES

[1] D. G. Goodenough, A. S. Bhogall, H. Chen, and A. Dyk, "Comparison of methods for estimation of Kyoto protocol products of forests from multitemporal Landsat," in *Proc. IGARSS*, vol. 2, 2001, pp. 764–767.

[2] S. Le Hégarat-Mascle, M. Zribi, F. Alem, A. Weisse, and C. Loumagne, "Soil moisture estimation from ERS/SAR data: Toward an operational methodology," *IEEE Trans. Geosci. Remote. Sens.*, vol. 40, no. 12, pp. 2647–2658, Dec. 2002.

[3] F. Del Frate, A. Ortenzi, S. Casadio, and C. Zehner, "Application of neural algorithms for a real-time estimation of ozone profiles from GOME measurements," *IEEE Trans. Geosci. Remote. Sens.*, vol. 40, no. 10, pp. 2263–2270, Oct. 2002.

[4] L. Bruzzone, S. Casadio, R. Cossu, F. Sini, and C. Zehner, "A system for monitoring NO2 emissions from biomass burning by using GOME and ATSR-2 data," *Int. J. Remote Sens.*, vol. 24, pp. 1709–1721, 2002.

[5] V. Klemas, D. Bartlett, W. Philpot, and R. Roger, "Coastal and estuarine studies with ERST-1 and Skylab," *Remote Sens. Environ.*, vol. 3, pp. 153–177, 1974.

[6] H. Gordon, D. Clark, J. Brown, O. Brown, R. Evans, and W. Broenkow, "Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparison of ship determinations and CZCS estimates," *Appl. Opt.*, vol. 22, pp. 20–36, 1983.

[7] S. Tassan, "Evaluation of the potential of the Thematic Mapper for marine applications," *Int. J. Remote Sens.*, vol. 8, pp. 1455–1478, 1987.

[8] R. Stumpf and J. Pennock, "Calibration of a general optical equation for remote sensing of suspendent sediments in a moderately turbid estuary," *J. Geophys. Res.*, vol. 94, no. C10, pp. 14.363–14.371, 1989.

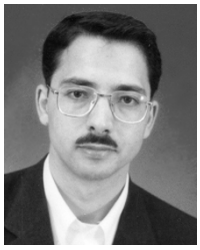
[9] C. Whitlock, C. Kuo, and S. LeCroy, "Criteria for the use of regression analysis for remote sensing of sediment and pollutants," *Remote Sens. Environ.*, vol. 12, pp. 151–168, 1982.

[10] R. Lathrop and T. Lillesand, "Use of Thematic Mapper data to assess water quality in Green Bay and central Lake Michigan," *Photogramm. Eng. Remote Sens.*, vol. 52, pp. 671–680, 1986.
[11] J. O'Reilly, S. Maritorena, B. Mitchell, D. Siegel, K. Carder, S. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103, no. C11, pp. 24.937–24.953, 1998.
[12] K. R. Manjunath, M. B. Potdar, and N. L. Purohit, "Large area operational wheat yield model development and validation based on spectral and meteorological data," *Int. J. Remote Sens.*, vol. 23, pp. 3023–3038, 2002.
[13] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
[14] L. Keiner and X.-H. Yan, "A neural network model for estimating sea surface chlorophyll and sediments from Thematic Mapper imagery," *Remote Sens. Environ.*, vol. 66, pp. 153–165, 1998.
[15] D. D'Alimonte and G. Zibordi, "Phytoplankton determination in an optically complex coastal region using a multilayer perceptron neural network," *IEEE Trans. Geosci. Remote. Sens.*, vol. 41, no. 12, pp. 2861–2868, Dec. 2003.
[16] 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 Sens.*, vol. 39, no. 7, pp. 1508–1524, Jul. 2001.
[17] M. De Martino, P. Mantero, S. B. Serpico, E. Carta, G. Corsini, and R. Grasso, "Water quality estimation by neural networks based on remotely sensed data analysis," in *Proc. Int. Workshop Geo-Spatial Knowledge Processing for Natural Resource Management*, Varese, Italy, 2002, pp. 54–58.
[18] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, pp. 241–259, 1992.
[19] L. Breiman, "Stacked regression," Dept. Statistics, Univ. California, Berkeley, Tech. Rep. TR-367, 1992.
[20] C. J. Merz and M. J. Pazzani, "A principal components approach to combining regression estimates," *Mach. Learn.*, vol. 36, no. 1–2, pp. 9–32, 1999.
[21] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.
[22] K. Woods, K. Bowyer, and W. P. Kegelmeyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.
[23] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[24] L. Hansen and P. Salamon, "Neural networks ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.
[25] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geosci. Remote. Sens.*, vol. 40, no. 10, pp. 2291–2299, Oct. 2002.
[26] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote. Sens.*, vol. 40, no. 9, pp. 1984–1996, Sep. 2002.
[27] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
[28] O. Mangasarian and D. Musicant, "Robust linear and support vector regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 950–955, Sep. 2000.
[29] T. Van Gestel, J. Suykens, D.-E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, and J. Vandewalle, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Trans. Neural Networks*, vol. 12, no. 4, pp. 809–821, Jul. 2001.
[30] G. Camps-Valls, E. Soria-Olivas, J. Perez-Ruixo, F. Perez-Cruz, A. Figueiras-Vidal, and A. Artes-Rodriguez, "Cyclosporine concentration prediction using clustering and support vector regression methods," *Electron. Lett.*, vol. 38, pp. 568–570, 2002.
[31] S. Bengio and J. Mariethoz, "Learning the decision function for speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, 2001, pp. 425–428.
[32] J. D. Martin-Guerrero, G. Camps-Valls, E. Soria-Olivas, A. J. Serrano-Lopez, J. J. Perez-Ruixo, and N. V. Jimenez-Torres, "Dosage individualization of erythropoietin using a profile-dependent support vector regression," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1136–1142, Oct. 2003.
[33] H. Zhan, P. Shi, and C. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Trans. Geosci. Remote. Sens.*, vol. 41, no. 12, pp. 2947–2951, Dec. 2003.
[34] R. A. Jacobs, "Methods for combining expert's probability assessment," *Neural Comput.*, vol. 7, pp. 867–888, 1995.
[35] R. L. Winkler, "The consensus of subjective probability distributions," *Manage. Sci.*, vol. 15, pp. B.61–B.75, 1968.
[36] J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," *IEEE Trans. Syst., Man, Cybern., A*, vol. 22, no. 4, pp. 688–704, Jul./Aug. 1992.
[37] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
[38] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behavior," *Pattern Recognit.*, vol. 34, pp. 1879–1881, 2001.
[39] L. Bruzzone and D. Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," *IEEE Trans. Geosci. Remote. Sens.*, vol. 37, no. 2, pp. 1179–1184, Mar. 1999.
[40] A. Smola and B. Schölkopf, "A tutorial on support vector regression," Royal Holloway College, Univ. London, London, U.K., NeuroCOLT Tech. Rep. NC-TR-98-030, 1998.
[41] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[42] L. I. Kuncheva and C. J. Whitaker, "Ten measures of diversity in classifier ensembles: Limits for two classifiers," in *Proc. DERA/IEE Workshop on Intelligent Sensor Processing*, 2001, pp. 10.1–10.10.

Lorenzo Bruzzone (S'95–M'99–SM'03) received the laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications, both from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is currently Head of the Remote Sensing Laboratory in the Department of Information and Communication Technologies at the University of Trento, Trento, Italy. From 1998 to 2000, he was a Postdoctoral Researcher at the University of Genoa. From 2000 to 2001, he was an Assistant Professor at the University of Trento, where he has been an Associate Professor of telecommunications since November 2001. He currently teaches remote sensing, advanced pattern recognition, and electrical communications. His current research interests are in the area of remote sensing image processing and recognition (analysis of multitemporal data, feature selection, classification, data fusion, and neural networks). He conducts and supervises research on these topics within the frameworks of several national and international projects. He is the author (or coauthor) of more than 110 scientific publications, including journals, book chapters, and conference proceedings. He is a referee for many international journals and has served on the Scientific Committees of several international conferences.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). He is the Delegate in the scientific board for the University of Trento of the Italian Consortium for Telecommunications (CNIT) and a member of the Scientific Committee of the India–Italy Center for Advanced Research. He was a recipient of the *Recognition of IEEE Transactions on Geoscience and Remote Sensing Best Reviewers* in 1999 and was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote sensing images (November 2003). He was the Chair and Co-chair of, respectively, the First and Second IEEE International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (Trento, Italy, September 2001—Ispra, Italy, July 2003). Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing (Barcelona, Spain, September 2003—Maspalomas, Gran Canaria, September 2004). He is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is a member of the International Association for Pattern Recognition (IAPR) and of the Italian Association for Remote Sensing (AIT).



Farid Melgani (M'04) received the State Engineer degree in electronics from the University of Batna, Batna, Algeria, in 1994, the M.Sc. degree in electrical engineering from the University of Baghdad, Baghdad, Iraq, in 1999, and the Ph.D. degree in electronic and computer engineering from the University of Genoa, Genoa, Italy, in 2003.

From 1999 to 2002, he cooperated with the Signal Processing and Telecommunications Group, Department of Biophysical and Electronic Engineering, University of Genoa. He is currently an Assistant Professor of telecommunications at the University of Trento, Trento, Italy, where he teaches pattern recognition, radar remote sensing systems, and digital transmission. His research interests are in the area of processing and pattern recognition techniques applied to remote sensing images (classification, multitemporal analysis, and data fusion). He is coauthor of more than 30 scientific publications and is a referee for several international journals.

Dr. Melgani has served on the Scientific Committee of the SPIE International Conferences on Signal and Image Processing for Remote Sensing since 2000. He is a referee for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.