Robust Support Vector Regression for Biophysical Variable Estimation From Remotely Sensed Images

Gustavo Camps-Valls, *Member, IEEE*, Lorenzo Bruzzone, *Senior Member, IEEE*, José L. Rojo-Álvarez, *Member, IEEE*, and Farid Melgani, *Member, IEEE*

Abstract—This letter introduces the ε-Huber loss function in the support vector regression (SVR) formulation for the estimation of biophysical parameters extracted from remotely sensed data. This cost function can handle the different types of noise contained in the dataset. The method is successfully compared to other cost functions in the SVR framework, neural networks and classical bio-optical models for the particular case of the estimation of ocean chlorophyll concentration from satellite remote sensing data. The proposed model provides more accurate, less biased, and improved robust estimation results on the considered case study, especially significant when few *in situ* measurements are available.

Index Terms—Biophysical parameter estimation, Medium Resolution Imaging Spectrometer (MERIS), ocean chlorophyll concentration, regression, robust cost function, Sea-viewing Wide Field-of-view Sensor (SeaWiFS)/SeaWiFS Bio-Optical Algorithm Mini-Workshop, support vector machine (SVM).

I. INTRODUCTION

IN REMOTE sensing data analysis, the estimation of biophysical parameters is of special relevance in order to understand better the environment dynamics at local and global scales [1]. For instance, remotely sensed images can be used to estimate crop yield, defoliation, biomass, leaf area index, water content, yellow substance, pollution, and chlorophyll concentration. In order to relate the image acquired by the satellite sensor to biophysical parameters, model-based estimation algorithms are commonly used. Two different approaches can be considered. In *physical modeling*, predefined direct models of the estimated biophysical parameters are adopted. These models are designed to account for all parameters affecting the radiometric characteristics of the remote sensing data, such as atmospheric conditions, sun angle, sensor gain and offset, and viewing geometry. In *empirical modeling*, regression techniques are commonly developed. These techniques relate the remotely sensed data with the investigated biophysical parameter according to interpolation methods applied over a training set constituted by pairs of *in situ* measurements and collected radiances.

G. Camps-Valls is with the Grup de Processament Digital de Senyals, Departamento Enginyeria Electrènica, Escola Tècnica Superior d'Enginyeria, Universitat de València, 46100 Burjassot (València) Spain (e-mail: gustavo.camps@uv.es).

L. Bruzzone and F. Melgani are with the Department of Information and Communication Technology. University of Trento, I-38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

J. L. Rojo-Álvarez is with the Departmento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, 28911 Madrid, Spain.

The simplification assumptions intrinsic to physical models and the complexity of the inversion process (e.g., large number of variables and physical processes, and complex mathematical formulations) resulted in an increased interest in the use of empirical models. Many different empirical models have been proposed in the literature (e.g., linear, exponential, polynomial), but these models are often too simple to capture the relationships between remote sensing measurements and investigated biophysical parameters, and also to encode in a set of parameters the spatio-temporal variability of the related processes. For these reasons, recently nonlinear regression techniques based on neural networks have been introduced in the remote sensing literature [2]. Different models and architectures of neural networks have been considered for the estimation of biophysical parameters [2]–[5]. However, despite their potential effectiveness, neural networks present some important drawbacks: 1) design and implementation often results in a complex, time-consuming task; 2) following the minimization of the empirical risk, rather than the structural risk, can lead to overfitting the training data; and 3) performance can be degraded when working with low-sized datasets. A promising alternative to neural networks are the support vector machines (SVMs) [6], [7], which alleviate the aforementioned problems of neural networks [5], [8]–[10]. The support vector regression (SVR) is the SVM implementation for regression and function approximation [7], which has yielded good results in estimating biophysical parameters [8]–[10].

Despite the good performance shown by the standard SVR, a main problem still remains. Standard SVR exploits a fixed-cost function, which implicitly assumes a specific density model for the errors (or residuals), independently of their statistical nature. This is not appropriate when working with data containing uncertainty and different kinds of noise, as occurs in many remote sensing problems. In order to address the aforementioned problem, in this letter, we propose to include the robust ε-Huber cost function (presented in [11]) into the standard SVR formulation. This loss function has the advantage that it can properly take into account different kinds of noise, and proved effective in terms of accuracy and robustness to noise in SVM-based system identification [12].

II. ROBUST ε-HUBER SUPPORT VECTOR REGRESSION

A. SVR Method With Standard ε-Insensitive Cost Function

Let us first recall the standard formulation of the SVR method. Given a labeled training dataset $\{(\mathbf{x}_i, y_i), i = 1, \ldots, n\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, the SVR methodology first maps the input data to a higher dimensional (possibly
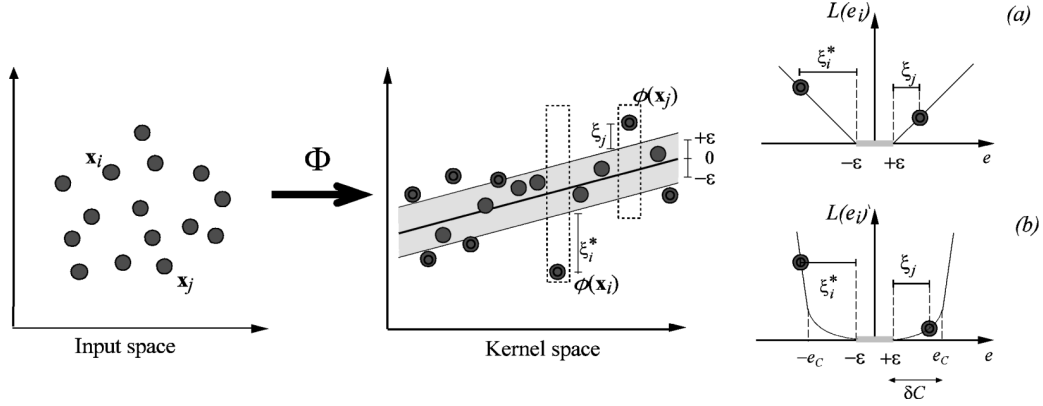
Fig. 1.   Scheme for the SVR. Samples in the original input space are first mapped to a kernel feature space through $\boldsymbol{\phi}$, and a linear regression is performed there. All samples outside a fixed tube of size ε are penalized, and become support vectors (double-circled symbols). Penalization is carried out by applying a cost function. (a) Vapnik's ε-insensitive cost function. (b) ε-Huber cost function proposed in this letter.

infinite) kernel feature space by means of a nonlinear mapping $\boldsymbol{\phi} : \mathbb{R}^N \longrightarrow \mathcal{H}$ and then solves a linear model there

$$\hat{y}_i = f(\mathbf{x}_i, \mathbf{w}) = \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b \qquad (1)$$

where $\hat{y}_i$ are the estimations of y_i, $\mathbf{w}$ is a weight vector in the feature space, and b is the bias term in the regression (see Fig. 1). The samples generated by the underlying system are assumed to follow an additive noise model $y_i = \hat{y}_i + e_i$, where $e_i \in \mathbb{R}$ are the committed errors (or residuals) by the model.

The SVR minimizes a cost function of the committed errors, $L(e_i)$, which is usually regularized using the squared norm of the model coefficients in order to provide smoother solutions and enhanced generalization capabilities

$$F(\mathbf{w}, e_i) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_i L(e_i). \qquad (2)$$

In the standard SVR formulation, one commonly uses Vapnik's ε-insensitive cost function, in which errors e_i lower than ε are not penalized [see Fig. 1(a)], and all further deviations should incur in a linear penalization, given by

$$L_\varepsilon(e_i) = \max(|e_i| - \varepsilon, 0). \qquad (3)$$

However, this cost function results in a biased estimator when is combined with a regularization term [13]. Therefore, in many applications, its use can lead to suboptimal results [6], [7]. In addition, from a maximum-likelihood viewpoint, using a fixed-cost function implies assuming a specific density model for the residuals, which usually is not a good option in remote sensing estimation applications (cf. Section I).

B. Proposed Robust ε-Huber-SVR

The latter problem can be alleviated by considering the loss function originally presented in [11], which is defined as follows:

$$L(e_i) = \begin{cases} 0, & |e_i| \leq \varepsilon \\ \frac{1}{2\delta}(|e_i| - \varepsilon)^2, & \varepsilon \leq |e_i| \leq e_C \\ C(|e_i| - \varepsilon) - \frac{1}{2}\delta C^2, & |e_i| \geq e_C \end{cases} \qquad (4)$$

where $e_C = \varepsilon + \delta C$, ε is the insensitive parameter, and δ and C tune the tradeoff between the L_2-norm regularization of the coefficients and the losses [Fig. 1(b)].

Note that (4) represents a (slightly) modified Vapnik's ε-insensitive cost function when δ is small enough, and Huber's cost function when $\varepsilon = 0$ [6]. Hence, we refer to this robust cost function as the ε-*Huber loss function*. In this cost function, there are three different regions, allowing us to deal with different kinds of possible situations. The ε-insensitive region does not penalize errors lower than ε. The quadratic cost region takes into account the observation noise with the L_2-norm in this zone (which is appropriate for Gaussian processes). The linear cost zone limits the effect of outliers on the model parameter estimation.

By including (4) into (2), the estimation of the SVR coefficients can be stated as the minimization of

$$\frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2\delta}\sum_{i \in I_1}\left(\xi_i^2 + \xi_i^{*2}\right) + C\sum_{i \in I_2}(\xi_i + \xi_i^*) - \sum_{i \in I_2}\frac{\delta C^2}{2} \quad (5)$$

with respect to $\mathbf{w}$, b, and $\{\xi_i^{(*)}\}$,[1] constrained to

$$y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b \leq \varepsilon + \xi_i \qquad \forall i = 1, \ldots, n \qquad (6)$$

$$\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \varepsilon + \xi_i^* \qquad \forall i = 1, \ldots, n \qquad (7)$$

$$\xi_i, \xi_i^* \geq 0 \qquad \forall i = 1, \ldots, n \qquad (8)$$

where ξ_i and ξ_i^* are slack variables introduced to deal with permitted positive or negative errors, respectively (see Fig. 1). In this formulation, I_1 and I_2 are the sets of samples for which losses are required to have a quadratic and a linear cost, respectively. Note that these sets are not static during the optimization procedure but updated at each iteration.

The procedure for solving this problem is analogous to the one used for optimizing the standard SVR (for full details see [13]). We first procede to include linear restrictions (6)–(8) into (5) by means of Lagrange multipliers $\alpha_i^{(*)}$, $\beta_i^{(*)}$ associated to each sample

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2\delta}\sum_{i \in I_1}\left(\xi_i^2 + \xi_i^{*2}\right) + C\sum_{i \in I_2}(\xi_i + \xi_i^*)$$

$$- \sum_{i \in I_2}\frac{\delta C^2}{2} - \sum_i(\beta_i\xi_i + \beta_i^*\xi_i^*)$$

$$+ \sum_i \alpha_i\left(y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b - \varepsilon - \xi_i\right)$$

$$+ \sum_i \alpha_i^*\left(-y_i + \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - \varepsilon - \xi_i^*\right) \qquad (9)$$

---

[1]Hereafter, $\{\xi_i\}$ and $\{\xi_i^*\}$ will be denoted with $\{\xi_i^{(*)}\}$ for the sake of notation simplicity. The same notation will be used with Lagrange multipliers $\{\alpha_i^{(*)}\}$ and $\{\beta_i^{(*)}\}$.

which is constrained to $\alpha_i^{(*)}, \beta_i^{(*)}, \xi_i^{(*)} \geq 0$. Now (9) has to be minimized with regard to primal variables ($\mathbf{w}$, b, and $\xi_i^{(*)}$) and maximized with regard to the Lagrange multipliers ($\alpha_i^{(*)}$ and $\beta_i^{(*)}$). Therefore, by making zero the gradient of $\mathcal{L}$ with respect to the primal variables, we obtain the following conditions:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_i (\alpha_i - \alpha_i^*) \boldsymbol{\phi}(\mathbf{x}_i) \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_i (\alpha_i - \alpha_i^*) = 0 \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^{(*)}} = 0 \rightarrow \begin{cases} \alpha_i^{(*)} = \frac{1}{\delta} \xi_i^{(*)} - \beta_i^{(*)}, & \text{if } i \in I_1 \\ \alpha_i^{(*)} = C - \beta_i^{(*)}, & \text{if } i \in I_2. \end{cases} \tag{12}$$

If constraints (10)–(12) are included in the Lagrange functional $\mathcal{L}$ (9) in order to remove the primal variables, the dual problem is then obtained, and it corresponds to

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{K} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \frac{\delta}{2} (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*)^T \mathbf{I} (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \right.$$
$$\left. + (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{y} - \varepsilon \mathbf{1}^T (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \right\} \tag{13}$$

constrained to $C \geq \alpha_i^{(*)} \geq 0$ and to (11) where $\boldsymbol{\alpha}^{(*)} = [\alpha_1^{(*)}, \ldots, \alpha_n^{(*)}]^T$, $\mathbf{y} = [y_1, \ldots, y_n]^T$, and $\mathbf{K}$ represents the kernel matrix $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i) \cdot \boldsymbol{\phi}(\mathbf{x}_j)$ satisfying Mercer's condition [14].

It is straightforward to demonstrate that the dual problem is equivalent to

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \left\{ \frac{1}{2} [\boldsymbol{\alpha}^T \quad \boldsymbol{\alpha}^{*T}] \begin{bmatrix} \mathbf{K} + \delta \mathbf{I} & -\mathbf{K} + \delta \mathbf{I} \\ -\mathbf{K} + \delta \mathbf{I} & \mathbf{K} + \delta \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix} \right.$$
$$\left. + [\varepsilon \mathbf{1}^T - \mathbf{y}^T \quad \varepsilon \mathbf{1}^T + \mathbf{y}^T] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix} \right\} \tag{14}$$

where $\mathbf{1} = [1, \ldots, 1]$ is a unitary vector and $\mathbf{I}$ is the identity matrix. Note that the only difference with the standard ε-SVR formulation is that the term δ is added to the diagonal of the kernel matrix $\mathbf{K} \equiv K(\cdot, \cdot)$. Full theoretical insight on the effect of δ in the solution can be obtained from [11]. The optimization of (14) constitutes a quadratic programming (QP) problem whose final solution then becomes

$$\hat{y} = f(\mathbf{x}) = \sum_i (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \tag{15}$$

where only the training examples whose corresponding Lagrange multipliers are nonzero count in the solution and are called *support vectors* (SVs).

C. Model Development

The solution of a (nonlinear) SVR is obtained by taking the dot product in kernel-generated spaces [see (15)]. Some kernels K are available in the literature. Nevertheless, in this work, we restrict ourselves to the use of the radial basis function (RBF) kernel, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$, due to theoretical [7], [15] and computational convenience. We followed a

nonexhaustive iterative search strategy to select the free parameters σ, C, ε and δ. Basically, at each iteration (T iterations), a sequential search on every parameter domain is performed by splitting the range of the parameter in M linearly or logarithmically equally spaced points. Values of $T = 3$ and $M = 50$ exhibited good performance in our simulations.

III. EXPERIMENTAL RESULTS

In this section, we compare three types of loss-based SVRs: 1) the standard ε-insensitive SVR (ε-SVR); 2) the squared loss SVR; and 3) the proposed ε-Huber-SVR. In addition, we include results obtained with neural networks and classical empirical models. We benchmark models in terms of accuracy and bias of the estimations, and robustness when low number of training samples are available in two different datasets related to the estimation of chlorophyll concentration. However, these are only examples of applications, as the proposed methodology is general and can be applied to any kind of estimation problem.

A. Data Description

Performance is illustrated in the particular problem of modeling the nonlinear relationship between chlorophyll concentration and marine reflectance. This is a specially well-suited case study because chlorophyll concentration measurements are subject to high levels of uncertainty depending on: 1) the difficulties in ground-truth data acquisition; 2) the complexity to obtain a precise geometrical association between *in situ* measurements and satellite-derived data, which is mainly due to the time mismatch between the acquired image and the recorded measurements (e.g., critical for coastal water monitoring); and 3) the noise that affects the acquisition phase. In this work, we use the following two different datasets for illustration purposes.

1) The first dataset simulates data acquired by the Medium Resolution Imaging Spectrometer (MERIS) onboard Envisat (MERIS dataset), and in particular the spectral behavior of chlorophyll concentration, CC, in the subsurface waters. We selected the eight channels in the visible range (412–681 [nm]), following the work in [16]. The range of variation of the chlorophyll concentration is from 0.02–25 mg/m^3. The data were generated according to a fixed (noise-free) model, and thus uncertainty is not encountered. The total number of samples (pairs of *in situ* concentrations and received radiances) available for our experiments was equal to 5000. These samples were randomly divided into three sets: a training set (500 samples) used to build the model, a validation set (500 samples) used to select the best combination of free parameters, and a test set (4000 samples) used to assess the accuracy of the estimator.

2) The second dataset (SeaBAM) [17], [18] gathers 919 *in situ* measurements of chlorophyll concentration around the U.S. and Europe related to five different multispectral remote sensing reflectance that correspond to some of the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) wavelengths (412, 443, 490, 510, and 555 nm). The chlorophyll concentration values span an interval between 0.019 and 32.787 mg/m^3. The available data were randomly split into three sets: 230 samples for training, 230 samples for validation, and the remaining 459 samples for testing the performance. Note that with this partition, we intentionally generate a more difficult problem than the one considered in [9], in which no validation set was used

TABLE I
MEAN ERROR, ROOT MEAN-SQUARED ERROR, MEAN ABSOLUTE ERROR,
AND CORRELATION COEFFICIENT (r) OF MODELS IN THE TEST SET

	ME	RMSE	ABSE	r
MERIS dataset				
NN-BP, 4 hidden nodes	-7.68e-4	0.024	0.016	0.998
ε-SVR	-2.36e-4	0.015	0.061	0.998
L₂-loss SVR	-9.96e-4	0.031	0.018	0.998
Proposed ε-Huber-SVR	-3.26e-6	0.011	0.004	0.999
SeaBAM dataset				
Morel-1†	-0.023	0.178	0.139	0.956
Morel-3	-0.025	0.182	0.143	0.954
CalCOFI 2-band cubic	-0.051	0.177	0.142	0.960
CalCOFI 2-band linear	0.079	0.325	0.256	0.956
NN-BP, 4 hidden nodes	-0.046	0.143	0.111	0.971
ε-SVR in [9]	–	0.138	–	0.973
ε-SVR	-0.070	0.139	0.105	0.971
L₂-loss SVR	-0.034	0.140	0.107	0.971
Proposed ε-Huber-SVR	-0.020	0.137	0.104	0.972

†The result provided in this table for Morel and CalCOFI models slightly differ from the ones given in Ref. [17] since they are shown for the test set. In addition, models in [17] used all available data to fit the models and hence no validation procedure was followed.

and hence the model was trained using a richer (459 samples) training set. In both cases, we transformed the concentration data logarithmically, $Y_{CC} = \log(CC)$, according to [16]. Hereafter, units of all accuracy and bias measurements are referred to Y_{CC} [log (mg/m³)] instead of CC [mg/m³].

B. Model Accuracy and Comparison

Table I presents results in the test set for all SVR models and datasets. For comparison purposes, we include results obtained with a feedforward neural network trained with backpropagation (NN-BP) for both datasets. In the case of the SeaBAM dataset, we also include results with the models Morel-1, Morel-3, and CalCOFI two-band (cubic and linear), as they performed best among a set of 15 *empirical* estimation models of the Cl-a concentration in [17], and results from the ε-SVR obtained in [9]. We show the following measures for the prediction errors: mean error (ME) as a measure of bias; the root mean-square error (RMSE) and the mean absolute error (ABSE) as a measure of accuracy, and the correlation coefficient (r) between the desired output and the output offered by the models as a measure of fit. From Table I, one can observe that the proposed technique is more accurate (RMSE, ABSE) and unbiased (ME) than the rest of the models, specially in the MERIS data. All the models achieved high values of the correlation coefficient. We did a one-way analysis of variance (ANOVA) to compare the means of the three models. For the MERIS data, appreciable numerical (see Table I) and statistical differences between ε-Huber-SVR and the rest of the models were observed for both bias ($F = 2.46$, $p < 0.05$) and accuracy ($F = 767.8$, $p < 0.001$). For the SeaBAM dataset, the proposed method showed an improved numerical performance even if no statistical differences were observed in bias ($F = 0.08$, $p = 0.92$) or accuracy ($F = 0.04$, $p = 9.86$). Similar results are obtained to those previously reported in [9], but we used half of the training samples (this is particularly interesting considering the few *in situ* measures usually available in real biophysical parameter estimation problems).

An easy way to show results graphically is to plot predicted versus observed representations. A plot of the best model for the ε-Huber-SVR on the test set is shown in Fig. 2 (left). In the same figure, we also show the predicted values against the residuals for this model [Fig. 2 (right)]. These goodness-of-fit
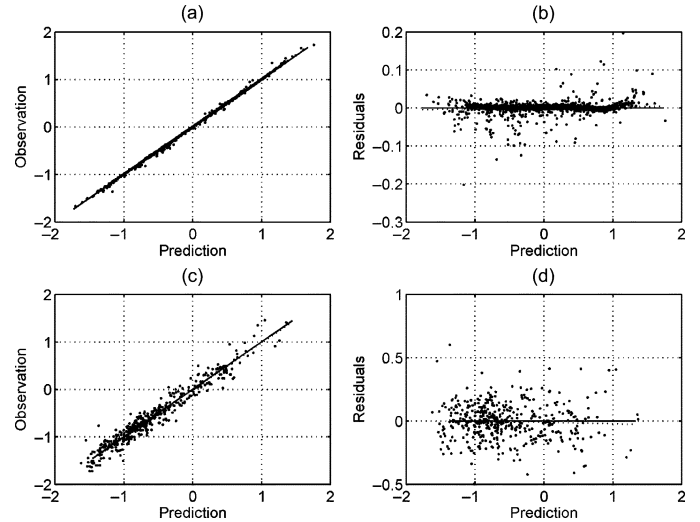

Fig. 2. Performance of the ε-Huber-SVR in the test set. (Top row) (a) Predicted versus observed concentrations and (b) predicted versus residuals for the MERIS data. (Bottom row) (c) Predicted versus observed concentrations and (d) predicted versus residuals for the SeaBAM data. The solid line represents the line of identity, and the dotted one is the regression line.

graphics confirm the good predictive performance of the model. Excellent coefficient of determination ($R^2 = 0.999$ for MERIS and $R^2 = 0.945$ for SeaBAM, test set) and unbiased estimations (Slope $\pm 95\%$CI : 0.991 ± 0.011, intercept $\pm 95\%$CI : $3e-6 \pm 4e-7$) or errors are observed. This is an important issue in this problem, in which unbiased estimations are desirable.

C. Model Flexibility and Uncertainty

In addition to model accuracy demonstrated in the previous section, the ε-Huber cost function provides some relevant information about the reliability of the available data, which is extremely important when dealing with uncertainty. Fig. 3 shows the histograms for obtained residuals with the ε-Huber-SVR model. In the case where no uncertainty is present in data (MERIS data, synthetic data generated according to a physical model), one can see that the δC parameter covers a wide range of residuals and mainly penalizes them with a squared loss function, assuming that they belong mainly to a Gaussian distribution, and that the (small) tails in the residuals distribution are penalized linearly, assuming that (few) outliers are present in the dataset. Note that the data we are dealing with holds an inherent normal nature [16]. When working with uncertainty in the data (SeaBAM dataset, real data including several amounts and noise sources), the algorithm adapts itself to provide tighter $(\varepsilon, \delta C)$-pairs to accommodate the underlying noise distributions. These histograms are plotted using the test data, which confirms that models have captured the underlying nature of noise from a reduced dataset, and can extrapolate this knowledge to an independent dataset.

D. Model Robustness to Reduced Datasets

An important problem when designing an estimation study is the number of the available *in situ* measurements available, which is often scarce. In this experiment, we test the capabilities of the models to deal with low-sized datasets. Fig. 4 shows the behavior of the RMSE versus the number of training samples in the SeaBAM data. We show the average value among 100
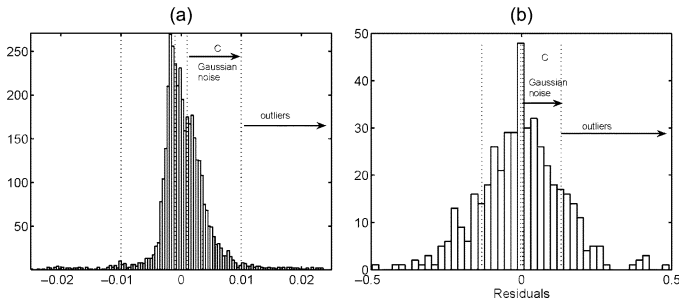
Fig. 3. Histograms of obtained residuals with the proposed ε-Huber-SVR model for (a) the MERIS and (b) SeaBAM datasets.
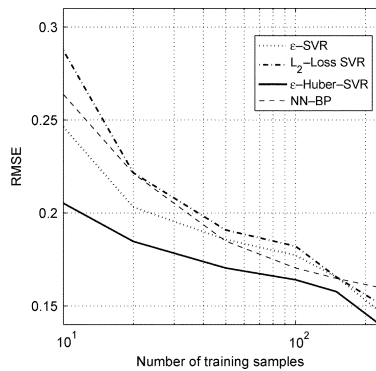

Fig. 4. Evolution of the RMSE in the test SeaBAM set as a function of the number of training samples.

different trials with different random sample selection, which represents a good confidence interval for the least measured test RMSE. The gain of the proposed algorithm is noticeable, obtaining an average improvement of 8.5% to 13% in RMSE with respect to the other models. This effect is especially significant when working with very reduced training sets, when an average gain of 16.67% in RMSE is obtained with respect to the standard ε-SVR method in the extreme case of using only ten training samples. It is also worth noting that the NN-BP model closely follows the L_2-Loss SVR, as they use the same (quadratic) cost function. However, the use of L_1-loss (ε-SVR) is optimal under the maximum-likelihood estimation theory, and more appropriate when *a priori* knowledge about noise content is not available. The standard ε-SVR uses a L_1 cost function for the residuals and hence it becomes theoretically more robust than models using the squared cost function. However, further improvements are obtained with the proposed ε-Huber-SVR, which can deal efficiently with different noise sources simultaneously.

We can conclude that when the number of available samples is low and samples are affected by uncertainty factors, the versatility to accommodate different noise models to the available data makes ε-Huber-SVR an efficient and robust model.

IV. CONCLUSION

A robust ε-Huber SVR technique has been presented for the estimation of biophysical parameters extracted from remotely sensed data. The good performance obtained in the experiments

reported in this letter (related to the estimation of chlorophyll concentration in sea water) confirm the effectiveness of the proposed approach.

Our future work is tied to: 1) extend the test of the proposed general method to other applications of biophysical parameter estimation; and 2) improve estimation results by including the ε-Huber-SVR model in multiple estimator systems.

REFERENCES

[1] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed. New York: Wiley, 2004.

[2] D. S. Kimes, Y. Knyazikhin, J. Privette, A. Abuelgasim, and F. Gao, "Inversion methods for physically-based models," *Remote Sens. Rev.*, vol. 18, pp. 381–439, 2000.

[3] 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 on Geo-Spatial Knowledge Processing for Natural Resource Management*, Varese, Italy, 2002, pp. 54–58.

[4] E. Balaguer-Ballester, G. Camps-Valls, J. L. Carrasco-Rodriguez, E. Soria-Olivas, and S. del Valle-Tascon, "Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks," *Ecol. Model.*, vol. 156, pp. 27–41, 2002.

[5] L. Bruzzone and F. Melgani, "Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 159–174, Jan. 2005.

[6] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[7] B. Schölkopf and A. Smola, *Learning With Kernels. Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.

[8] E. J. Kwiatkowska and G. Fargion, "Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2844–2860, Dec. 2003.

[9] 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.

[10] L. Ji and A. J. Peters, "Forecasting vegetation greenness with satellite and climate data," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 1, pp. 2844–2860, Jan. 2004.

[11] J. L. Rojo-Álvarez, M. Martínez-Ramón, A. R. Figueiras-Vidal, A. García-Armada, and A. Artés-Rodríguez, "A robust support vector algorithm for nonparametric spectral analysis," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 320–323, Nov. 2003.

[12] J. L. Rojo-Álvarez, M. Martínez-Ramón, M. dePrado Cumplido, A. Artés-Rodríguez, and A. R. Figueiras-Vidal, "Support vector method for robust ARMA system identification," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 155–164, Jan. 2004.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, pp. 199–222, 2004.

[14] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York: Wiley, 1953.

[15] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.

[16] P. Cipollini, G. Corsini, M. Diani, and R. Grass, "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, pp. 1508–1524, 2001.

[17] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103, no. C11, pp. 24 937–24 953, Oct. 1998.

[18] J. E. O'Reilly and S. Maritorena, "SeaBAM evaluation data set," The SeaWiFS Bio-optical Algorithm Mini-Workshop (SeaBAM) Univ. California. Santa Barbara, 1997 [Online]. Available: http://seabass.gsfc.nasa.gov/seabam/seabam.html.