Multioutput Support Vector Regression for Remote Sensing Biophysical Parameter Estimation

Devis Tuia, *Member, IEEE*, Jochem Verrelst, Luis Alonso, Fernando Pérez-Cruz, *Senior Member, IEEE*, and Gustavo Camps-Valls, *Senior Member, IEEE*

***Abstract*—This letter proposes a multioutput support vector regression (M-SVR) method for the simultaneous estimation of different biophysical parameters from remote sensing images. General retrieval problems require multioutput (and potentially nonlinear) regression methods. M-SVR extends the single-output SVR to multiple outputs maintaining the advantages of a sparse and compact solution by using an ε-insensitive cost function. The proposed M-SVR is evaluated in the estimation of chlorophyll content, leaf area index and fractional vegetation cover from a hyperspectral compact high-resolution imaging spectrometer images. The achieved improvement with respect to the single-output regression approach suggests that M-SVR can be considered a convenient alternative for nonparametric biophysical parameter estimation and model inversion.**

***Index Terms*—Biophysical parameter estimation, model inversion, regression, support vector regression (SVR).**

I. Introduction

In REMOTE sensing data analysis, the estimation of biophysical parameters is of special relevance in order to better understand the environment dynamics at local and global scales [1]. Leaf chlorophyll content (Chl), leaf area index (LAI), and fractional vegetation cover (fCover) are among the most important vegetation parameters [2], [3]. To date, a large number of spectral vegetation indices have been developed for the study of Chl based on leaf reflectance [4]. Since the launch of imaging spectrometers, such as the Compact High Resolution Imaging Spectrometer (CHRIS) on board the Project for Onboard Autonomy (PROBA) spacecraft, these vegetation indexes have been applied at the canopy level on ecosystems across the globe [5], [6]. Nevertheless, the majority of these indices make use of only two up to five spectral bands and are typically limited to simple relations, such as band ratios or polynomials. Parametric models have the important limitation of assuming explicit (and typically simple) relationships among variables, which could lead to poor prediction results when these assumptions are not met. As a consequence, nonparametric and potentially nonlinear regression techniques have been effectively introduced for the estimation of biophysical parameters from remotely sensed images [7]. Nonparametric models do not assume a rigid functional form, they rely on the available data, and no *a priori* assumptions on variable relations are made. Different models and architectures of neural networks have been considered for the estimation of biophysical parameters [8], [9]. However, despite their potential effectiveness, neural networks offer a poor performance when working with few labeled data points. Support vector regression (SVR) [10] is a promising alternative to neural networks, which yields good results for retrieving some biophysical parameters [11].

Despite its potential usefulness, the standard formulation of the SVR cannot cope with multioutput problems. The usual procedure considers developing a different SVR to learn each parameter individually. However, this approach ignores the (potentially nonlinear) cross relations among biophysical parameters. For example, if one is interested in retrieving the LAI and the fCover, as both parameters are a measure of the plant density and thus related, a model should consider not only the underlying relations between the inputs (spectral channels) and the corresponding outputs (parameters to be predicted) but also the relations between the outputs.

In this letter, we study the applicability of a multioutput SVR (M-SVR) in the context of remote sensing biophysical parameter estimation. The method introduced in this letter was first applied to a complex biomedical problem [12] and was successfully used in MIMO channel estimation [13]. Here, we analyze its capabilities to deal with biophysical parameter estimation and how cross relations can be exploited for nonparametric retrieval. The multioutput approach can be of particular interest within operational processing chains (e.g., the Sentinel missions), as they are aimed at retrieving several interdependent biophysical parameters at once. Section II reviews the formulation of the single-output standard SVR and the proposed formulation for the M-SVR. The data collected for the experiments are presented in Section III. Section IV shows and discusses the experimental results. This letter is concluded in Section V.

II. M-SVR

In this letter, we introduce a generalization of SVRs to solve the problem of regression estimation for multiple variables. Thus, we refer to the proposed method, which is based on a previous contribution [12], as the M-SVR.

Although, under a pure Gaussian perspective, the estimation of each component can be individually addressed without the

Manuscript received September 17, 2010; revised December 14, 2010; accepted January 10, 2011. This work was supported in part by the projects of the Spanish Ministry of Education and Science and in part by the Spanish Ministry for Science and Innovation under Project AYA2008-05965-C04-03 and Project CSD2007-00018. The work of D. Tuia was supported by the Swiss National Science Foundation under Grant PBLAP2-127713/1. The work of J. Verrelst was supported by the Marie Curie Intra-European Fellowships of the European Commission under Grant 252237.

D. Tuia, J. Verrelst, L. Alonso, and G. Camps-Valls are with the Image Processing Laboratory, Universitat de València, 46003 València, Spain (e-mail: gustavo.camps@uv.es).

F. Pérez-Cruz is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Madrid, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2011.2109934

loss of accuracy, the use of a multidimensional regression tool helps in exploiting the dependences between variables and makes each retrieval less vulnerable to noise and measurement errors. Treating all the variables together may allow accurately estimating each of them when only scarce data are available, and the ε-insensitive cost function, which is introduced below, improves the scheme robustness when a different kind of noise and nonlinearities appears in the system.

A. Regression With the Standard SVR Formulation

The unidimensional regression estimation problem is regarded as finding the mapping between an incoming vector $\mathbf{x} \in \mathbb{R}^d$ and an observable output $y \in \mathbb{R}$, from a given set of independent and identically distributed samples, i.e., $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$. The standard SVR [10] solves this problem by finding regressor $\mathbf{w}$ and b that minimizes $\|\mathbf{w}\|^2/2 + C\sum_{i=1}^{l} L_v(y_i - (\boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{w} + b))$, where $\boldsymbol{\varphi}(\cdot)$ is a nonlinear transformation to a higher (possibly infinite) dimensional Hilbert space $\mathcal{H}$, also known as the feature space. The SVR can be solved using only inner products between $\boldsymbol{\varphi}(\cdot)$, not needing to know the nonlinear mapping; thus, we only need to specify a kernel function $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^\top \boldsymbol{\varphi}(\mathbf{x}_j)$ that has to fulfill Mercer's theorem [14]. $L_v(\cdot)$ is known as the Vapnik ε-insensitive loss function, which is equal to 0 for $|y_i - (\boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{w} + b)| < \varepsilon$ and equal to $|y_i - (\boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{w} + b)| - \varepsilon$ for $|y_i - (\boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{w} + b)| \geq \varepsilon$. The solution ($\mathbf{w}$ and b) is formed by a linear combination of the training samples in the transformed space with an absolute error equal to or greater than ε (i.e., the support vectors). This model has been successfully used in biophysical parameter estimation [11], [15].

B. M-SVR Formulation

If the observable output is a vector with Q variables to be predicted, i.e., $\mathbf{y} \in \mathbb{R}^Q$, we need to solve a multidimensional regression estimation problem, in which we have to find regressor $\mathbf{w}^j$ and b^j $(j = 1, \ldots, Q)$ for every output. We can directly generalize the 1-D SVR to solve the multidimensional case, leading to the minimization of

$$L_P(\mathbf{w}, \mathbf{b}) = \frac{1}{2}\sum_{j=1}^{Q} \|\mathbf{w}^j\|^2 + C\sum_{i=1}^{l} L(u_i) \tag{1}$$

with respect to $\mathbf{W}$ and $\mathbf{b}$, where $u_i = \|\mathbf{e}_i\| = \sqrt{\mathbf{e}_i^\top \mathbf{e}_i}$, $\mathbf{e}_i^\top = \mathbf{y}_i^\top - \boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{W} - \mathbf{b}^\top$, $\mathbf{W} = [\mathbf{w}^1, \ldots, \mathbf{w}^Q]$, $\mathbf{b} = [b^1, \ldots, b^Q]^\top$.

The Vapnik ε-insensitive loss function can be extended to multiple dimensions, but being based on an L_1 norm, it needs to account for each dimension independently, which makes the solution complexity grow linearly with the number of dimensions. If we instead use an L_2-based norm, all dimensions can be considered into a unique restriction yielding a single support vector for all dimensions (see details in [13]). Therefore, we propose to use

$$L(u) = \begin{cases} 0, & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2, & u \geq \varepsilon \end{cases} \tag{2}$$

which is a differentiable version of the loss function proposed in [12].

For $\varepsilon = 0$, this problem reduces to a kernel ridge regression (KRR) for each component, but for a nonzero ε value, the solution takes into account all outputs to construct each individual regressor. This way, the cross-output relations are exploited, thus leading to possibly more accurate predictions. Note that the M-SVR returns a multidimensional and sparse solution, thus solving the main issues of the SVR and the KRR when dealing with multiple outputs; the SVR cannot handle multiple outputs, whereas the KRR is not sparse and thus accounts for output relations in a dense way. We have devised a quasi-Newton approach in which each iteration has, at most, the same computational complexity as a least-squares procedure for each component. This iterative reweighted least-squares (IRWLS) procedure [16] is a weighted least-squares problem, and the number of iterations needed to obtain the final result is small, making the procedure only slightly more computationally demanding than least-squares regression for each component.

C. Resolution of the M-SVR

Optimization problems are solved using iterative procedures that rely on each iteration k on the previous solution ($\mathbf{W}^k$ and $\mathbf{b}^k$ in our case) to obtain the following one, until the optimal solution is reached. To construct the IRWLS procedure, we approximate (1) using a first-order Taylor expansion of $L(u)$ over the previous solution, leading to

$$L'_P(\mathbf{W}, \mathbf{b}) = \frac{1}{2}\sum_{j=1}^{Q} \|\mathbf{w}^j\|^2 + C\left(\sum_{i=1}^{l} L\left(u_i^k\right) + \left.\frac{dL(u)}{du}\right|_{u_i^k} \frac{\left(\mathbf{e}_i^k\right)^\top}{u_i^k}\left[\mathbf{e}_i - \mathbf{e}_i^k\right]\right) \tag{3}$$

where $u_i^k = \|\mathbf{e}_i^k\| = \sqrt{(\mathbf{e}_i^k)^\top \mathbf{e}_i^k}$ and $(\mathbf{e}_i^k)^\top = \mathbf{y}_i^\top - \boldsymbol{\varphi}(\mathbf{x}_i)^\top \mathbf{W}^k - (\mathbf{b}^k)^\top$, which presents the same value and gradient as $L_P(\mathbf{W}, \mathbf{b})$ for $\mathbf{W} = \mathbf{W}^k$ and $\mathbf{b} = \mathbf{b}^k$ (i.e., $L'_P(\mathbf{W}^k, \mathbf{b}^k) = L_P(\mathbf{W}^k, \mathbf{b}^k)$ and $\nabla L'_P(\mathbf{W}^k, \mathbf{b}^k) = \nabla L_P(\mathbf{W}^k, \mathbf{b}^k)$). $L'_P(\mathbf{W}^k, \mathbf{b}^k)$ is a lower bound of $L_P(\mathbf{W}, \mathbf{b})$ (i.e., $L_P(\mathbf{W}, \mathbf{b}) \geq L'_P(\mathbf{W}, \mathbf{b}), \quad \forall \quad \mathbf{W} \in \mathbb{R}^{\mathcal{H}} \times \mathbb{R}^Q$) and $\forall \mathbf{b} \in \mathbb{R}^Q$, because $L'_P(\mathbf{W}, \mathbf{b})$ is a first-order Taylor expansion of a convex function.

Now, we construct a quadratic approximation from (3), i.e.,

$$\begin{aligned} L''_P &= \frac{1}{2}\sum_{j=1}^{Q} \|\mathbf{w}^j\|^2 + C\left(\sum_{i=1}^{l} L\left(u_i^k\right) + \left.\frac{dL(u)}{du}\right|_{u_i^k} \frac{u_i^2 - \left(u_i^k\right)^2}{2u_i^k}\right) \\ &= \frac{1}{2}\sum_{j=1}^{Q} \|\mathbf{w}^j\|^2 + \frac{1}{2}\sum_{i=1}^{l} a_i u_i^2 + \tau \end{aligned} \tag{4}$$

where

$$a_i = \frac{C}{u_i^k}\left.\frac{dL(u)}{du}\right|_{u_i^k} = \begin{cases} 0, & u_i^k < \varepsilon \\ \frac{2C\left(u_i^k - \varepsilon\right)}{u_i^k}, & u_i^k \geq \varepsilon \end{cases} \tag{5}$$

and τ is a sum of constant terms that do not depend on either $\mathbf{W}$ or $\mathbf{b}$, which also presents the same value and gradient as $L_P(\mathbf{W}, \mathbf{b})$ for $\mathbf{W} = \mathbf{W}^k$ and $\mathbf{b} = \mathbf{b}^k$. It is shown that (4) is a weighted least-squares problem in which the weights depend on the previous solution, incorporating the knowledge of all

the components of each $\mathbf{y}_i$. To optimize (1), we construct a descending direction using the optimal solution of (4), and then, we compute the next step solution using a line search algorithm [17]. The IRWLS procedure can be summarized in the following steps:

1) Initialization: Set $k=0$, $\mathbf{W}^k=\mathbf{0}$, $\mathbf{b}^k=\mathbf{0}$, and compute u_i^k and a_i.
2) Compute the solution to (4), and label it as $\mathbf{W}^s$ and $\mathbf{b}^s$. Define a descending direction for (1) as
$$\mathbf{P}^k=\begin{bmatrix}\mathbf{W}^s-\mathbf{W}^k\\(\mathbf{b}^s-\mathbf{b}^k)^\top\end{bmatrix}.$$
3) Obtain the next step solution, i.e.,
$$\begin{bmatrix}\mathbf{W}^{k+1}\\(\mathbf{b}^{k+1})^\top\end{bmatrix}=\begin{bmatrix}\mathbf{W}^{k}\\(\mathbf{b}^{k})^\top\end{bmatrix}+\eta^k\mathbf{P}^k$$
computing the step size η^k using a backtracking algorithm.
4) Compute u_i^{k+1} and a_i, set $k=k+1$, and go back to step 2 until convergence.

Before actually computing $\mathbf{W}^s$ and $\mathbf{b}^s$, note that $\mathbf{P}^k$ is not a vector but a matrix in which each column is a descending direction for each regressor. Therefore, one should see it as an aggregate of descending directions for each component to be estimated. The value of η^k is computed using a backtracking algorithm [17] that initially sets $\eta^k=1$ and checks $L_P(\mathbf{W}^{k+1},\mathbf{b}^{k+1})<L_P(\mathbf{W}^k,\mathbf{b}^k)$. If the condition is not met, η_k is multiplied by a positive constant lower than 1, and the procedure is repeated until a decrease is achieved in the minimizing functional.

To obtain $\mathbf{W}^s$ and $\mathbf{b}^s$, one needs to solve the weighted least-squares problem in (4), in which each component is decoupled. Therefore, each component is independently solved by equating its gradient to zero, i.e.,

$$\nabla_{\mathbf{w}^j}L_P''=\mathbf{w}^j-\sum_i\boldsymbol{\varphi}(\mathbf{x}_i)a_i\left(y_{ij}-\boldsymbol{\varphi}(\mathbf{x}_i)^\top\mathbf{w}^j-b^j\right)=\mathbf{0} \quad (6)$$

$$\nabla_{b^j}L_P''=-\sum_i a_i\left(y_{ij}-\boldsymbol{\varphi}(\mathbf{x}_i)^\top\mathbf{w}^j-b^j\right)=0 \quad (7)$$

for $j=1,\ldots,Q$, which can be expressed as a linear system of equations, i.e.,

$$\begin{bmatrix}\boldsymbol{\varphi}^\top\mathbf{D}_a\boldsymbol{\varphi}+\mathbf{I} & \boldsymbol{\varphi}^\top\mathbf{a}\\ \mathbf{a}^\top\boldsymbol{\varphi} & \mathbf{1}^\top\mathbf{a}\end{bmatrix}\begin{bmatrix}\mathbf{w}^j\\ b^j\end{bmatrix}=\begin{bmatrix}\boldsymbol{\varphi}^\top\mathbf{D}_a\mathbf{y}^j\\ \mathbf{a}^\top\mathbf{y}^j\end{bmatrix} \quad (8)$$

where $\boldsymbol{\varphi}=[\boldsymbol{\varphi}(\mathbf{x}_1),\ldots,\boldsymbol{\varphi}(\mathbf{x}_n)]^\top$, $\mathbf{a}=[a_1,\ldots,a_n]^\top$, $(\mathbf{D}_a)_{ij}=a_i\delta(i-j)$, $\mathbf{y}^j=[y_{1j},\ldots,y_{nj}]$, and $\mathbf{1}$ is an all-one column vector. It is shown that the matrix in the previous linear system does not depend on j; therefore, it is identical for all components, and the difference on the linear systems associated to each pair $(\mathbf{w}^j,b^j)$ is due to the independent term in (8). Each column of $\mathbf{W}^s$ and $\mathbf{b}^s$ is constructed with the Q solutions of (8).

It is usual to work with the feature-space kernel (the inner product of the transformed vectors, i.e., $\kappa(\mathbf{x}_i,\mathbf{x}_j)=\boldsymbol{\varphi}(\mathbf{x}_i)^\top\boldsymbol{\varphi}(\mathbf{x}_j)$), instead of the whole nonlinear mapping [14]. Here, the representer's theorem [14] is used by which, under fairly general conditions, a learning problem can be expressed as a linear combination of the training samples in the feature space, i.e., $\mathbf{w}^j=\sum_i\boldsymbol{\varphi}(\mathbf{x}_i)\beta^j=\boldsymbol{\varphi}^\top\boldsymbol{\beta}^j$. If we replace this expression into (6) and (7), the linear system in (8) can be expressed as follows:

$$\begin{bmatrix}\mathbf{K}+\mathbf{D}_a^{-1} & \mathbf{1}\\ \mathbf{a}^\top\mathbf{K} & \mathbf{1}^\top\mathbf{a}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}^j\\ b^j\end{bmatrix}=\begin{bmatrix}\mathbf{y}^j\\ \mathbf{a}^\top\mathbf{y}^j\end{bmatrix} \quad (9)$$

for $j=1,\ldots,Q$, where $(\mathbf{K})_{ij}=\kappa(\mathbf{x}_i,\mathbf{x}_j)$ is known as the kernel matrix. The line search algorithm can be readily expressed in terms of $\boldsymbol{\beta}^j$, as it was presented for $\mathbf{w}^j$.

III. Data and Experimental Setup

A. Data Collection and Campaigns

The data used in this letter were obtained during the European Space Agency SPectra bARrax Campaign (SPARC) 2003 and 2004 campaigns in Barrax, Spain. The test area is an agricultural research facility with an extent of 5 km $\times$ 10 km. It is characterized by a flat morphology and large uniform land-use units of irrigated and dry lands. The measuring strategy was based on an elementary sampling unit (ESU) of 10-m radius, where a large number of samples were taken and averaged for different parameters, obtaining a local characterization. For each crop type, several ESUs were defined to be randomly distributed within the total cultivated area in order to characterize the global variation of the biophysical parameters. The Chl was measured with a calibrated Minolta CCM-200 from 50 samples per ESU. The LAI was derived from canopy measurements made with a LiCor LAI-2000 at 24 locations per ESU. The fCover was derived from hemispherical photographs taken at the same locations as the LAI measurements. All parameters present standard errors between 3% and 10% [18].

During the 2003 campaign carried out in July 12–14, a total of 121 ESUs were measured for different crops, whereas in 2004 (July 15–16), a total of 14 ESUs were sampled. Selected crops include garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard, and wheat. The large variety of crop types and phenological stages provide a very complete sample of possible combinations of parameters and almost the widest range of values in them, i.e., values of the LAI that vary between 0.4 and 6.3, that of the Chl between 2 and 55 μg/cm^{-2}, and that of the fCover between 0 and 1. This results in a very representative database for agricultural sites. Collecting such a database is not an easy task since many resources are needed, which are not always available. This makes the data set representative and well suited to multioutput regression studies.

Simultaneously to the ground sampling, hyperspectral images were collected by the CHRIS/PROBA spaceborne sensor. The data provided have 62 bands in the visible and near-infrared (NIR) region (400–1000 nm) at a spatial resolution of 34 m. The images selected for this letter were those acquired from the nadir view sharing similar observation configuration in order to minimize angular and atmospheric effects. The images were geometrically corrected, followed by atmospheric correction using the official CHRIS/PROBA Toolbox for BEAM (CHRIS-Box) at http://www.brockmann-consult.de/beam/chris-box/.

Summarizing, a total of $n=135$ data points with $d=62$ dimensions and $Q=3$ output variables (Chl, LAI, fCover) constitute the database.

TABLE I
MEAN AND STANDARD DEVIATION OF AVERAGE RESULTS

[%] Train	Model	RMSE		MAE		R	
		μ	σ	μ	σ	μ	σ
50	Single SVR	0.33	0.03	0.25	0.03	0.94	0.01
	M-SVR	0.31	0.02	0.23	0.02	0.95	0.01
80	Single SVR	0.28	0.02	0.22	0.02	0.96	0.01
	M-SVR	0.25	0.02	0.19	0.02	0.97	0.01

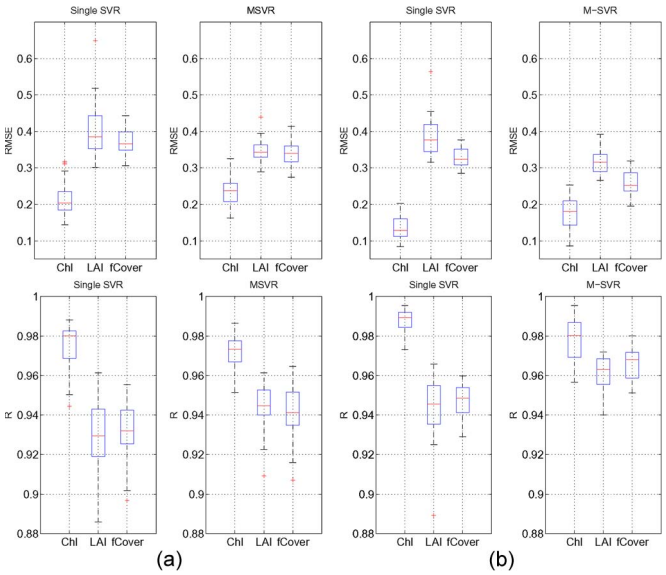

Fig. 1. Boxplots for RMSE and R (40 realizations), for SVRs and the proposed M-SVR, and for different rates of training pixels. (a) 50% and (b) 80% of pixels in train.

B. Experimental Setup

The ESU data set, along with the CHRIS measurements, was divided into different training and testing sets. Spectra were normalized with respect to the maximum NIR shoulder reflectance. Biophysical parameters were reduced to standard scores. Two experiments were conducted using 50% and 80% of the data points for training and the rest for testing the model robustness. To cover the entire variability of the biophysical parameters to be predicted, the training data were selected by regular sampling in the Chl/LAI/fCover-space cube. In addition, in order to avoid skewed results, each model was run 40 times using randomized training/testing data. The model's performance was evaluated with the root-mean-square error (RMSE) and the mean absolute error (MAE) to assess the accuracy. Pearson's correlation coefficient R was used to account for the goodness of fit. Model parameters were optimized using sixfold cross validation in the ranges $\sigma = \{10^{-1}, \ldots, 10\}$, $C = \{1, \ldots, 100\}$, and $\varepsilon = \{10^{-6}, 10^{-3}\}$, and the best combination was selected according to the least RMSE. In all cases, we compare the single SVR with the proposed M-SVR.

IV. RESULTS AND DISCUSSION

A. Numerical Comparison

The numerical results for both the SVR and M-SVR models are compared in Table I for the three variables jointly, while boxplots of the distribution of the RMSE, the MAE, and the R value for each individual variable are illustrated in Fig. 1. The M-SVR shows an overall improvement of the SVR performance with a decrease of 0.03 in the RMSE, 0.03 in the MAE, and an increase of 0.01 in the correlation. Slightly lower improvements are obtained when using 50% of the data for training. Regarding the single outputs, the M-SVR improves the results for the LAI and the fCover, whereas there is a slight decrease in the performance for the Chl (see Fig. 1). This may be related to the fact that, in the M-SVR, all the biophysical parameters share the same model's parameters, whereas in the SVR, each one has been separately optimized. This way, if one of the parameters is not related with the others, its separate optimization may lead to a better solution. The distribution of the results over the 40 realizations of the experiments is shown in Fig. 2, where histograms of the average distance between predictions $\hat{\mathbf{y}}$ and outputs $\mathbf{y}$ are illustrated; for both settings, the M-SVR results in smaller average distances, with a mode at 0.25 versus 0.3 for both experimental settings.

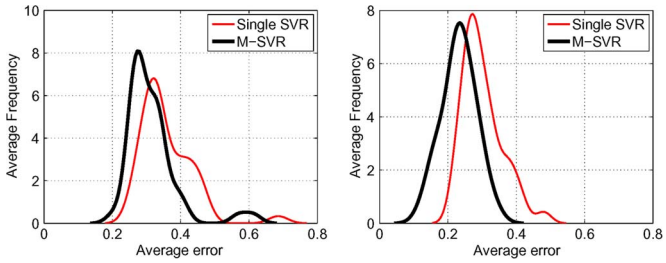

Fig. 2. Histograms of the average errors for single and multiple SVRs for (left) 50% and (right) 80% training data.

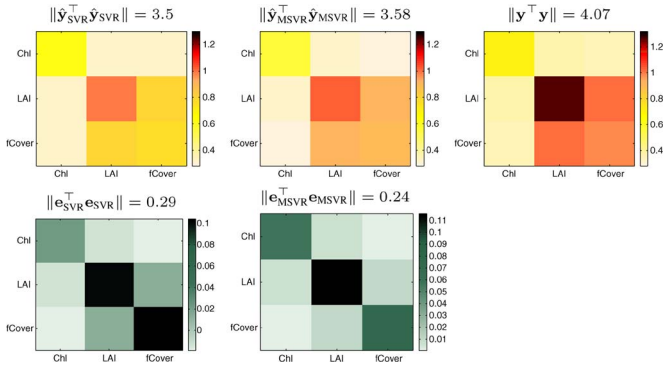

Fig. 3. Covariance matrices of (top) predictions $\hat{\mathbf{y}}$ and actual $\mathbf{y}$ signals and (bottom) residuals for the SVR and the M-SVR in a representative run.

B. Statistical Comparison

The M-SVR exploits nonlinear relations between variables and hence provides a more consistent nonredundant solution. The relationships between the outputs are actually better described and closer to the one observed in the test data. Fig. 3 illustrates the covariance matrices for the obtained output predictions and residuals with the SVR and the M-SVR. The Hilbert–Schmidt norm of the covariance matrices of predictions and residuals, i.e., $\|C\|_{\mathrm{HS}} = \sqrt{\mathrm{trace}(C^{\top}C)}$, summarizes the degree of second-order dependences and shows that the M-SVR provides more independent and uncorrelated residuals (0.24 versus 0.29 for the SVR). For higher order dependence estimation, we considered the Hilbert–Schmidt independence criterion (HSIC) proposed by Gretton *et al.* [19], [20]. The HSIC exploits kernels to estimate cross covariances in higher

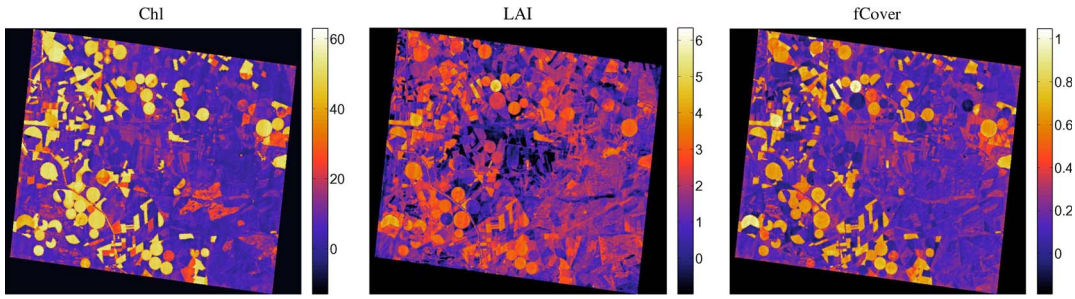

Fig. 4. Prediction maps of the M-SVR for the three biophysical parameters using 80% of available train pixels. (Left) Chl $[\mu g/cm^{-2}]$; (middle) LAI $[m^2/m^2]$; (right) fCover [%].

dimensional spaces. In all cases, the radial-basis-function kernel with σ estimated with the average mean distance was used. Here, HSIC values are higher for the M-SVR (77.24) than for the SVR (75.55), thus showing higher nonlinear dependences with respect to the true outputs.

C. Visual Comparison

Fig. 4 illustrates the prediction maps derived from the 12 July 2003 CHRIS image for the three biophysical parameters. The maps clearly show the irrigated crops (the circles in orange red), the natural areas (light blue), and the bare soil areas (dark blue). A simple visual comparison of the maps shows that the M-SVR is useful for mapping vegetation properties of heterogeneous surface types. While the three parameters show somewhat similar patterns, one can also note the variable-dependent differences, e.g., within the irrigated circles. The low values obtained in the center of the image are due to the presence of dry barley, harvested barley, and bright bare soils, which were underrepresented in the training data, and thus, the model cannot extrapolate to this unknown classes.

V. Conclusion

We have proposed the use of the M-SVR for the joint estimation of different biophysical parameters. The method outperforms the single-output SVR by taking into account the nonlinear relations not only between features but also among the biophysical parameters themselves. This is particularly noticeable when output variables are correlated. The model improves the accuracy but also yields less dependent residuals. The M-SVR model roughly uses the same number of support vectors as the single SVR, but since they are used just once in the prediction phase, the computational load is roughly divided by the number of variables in our experiments. The model is thus accurate and fast. These characteristics make them a convenient alternative for nonparametric biophysical parameter estimation and model inversion, particularly in the common setting of correlated output variables.

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*, 2001, vol. 2, pp. 764–767.

[2] R. H. Whittaker and P. L. Marks, "Methods of assessing terrestrial productivity," in *Primary Productivity of the Biosphere*. Berlin, Germany: Springer-Verlag, 1975, pp. 55–118. s t

[3] H. K. Lichtenthaler, "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," *Methods Enzymol.*, vol. 148, pp. 350–382, 1987.

[4] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, Apr. 2004.

[5] J. Verrelst, M. E. Schaepman, B. Koetz, and M. Kneubühler, "Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2341–2353, May 2008.

[6] S. Stagakis, N. Markos, O. Sykioti, and A. Kyparissis, "Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a phlomis fruticosa mediterranean ecosystem using multiangular chris/proba observations," *Remote Sens. Environ.*, vol. 114, no. 5, pp. 977–994, May 2010.

[7] D. S. Kimes, Y. Knyazikhin, J. L. Privette, A. A. Abuelgasim, and F. Gao, "Inversion methods for physically-based models," *Remote Sens. Rev.*, vol. 18, no. 2–4, pp. 381–439, Sep. 2000.

[8] 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 Know. Process. Nat. Resource Manage.*, Varese, Italy, 2002, pp. 54–58.

[9] B. Dzwonkowski and X. -H. Yan, "Development and application of a neural network based colour algorithm in coastal waters," *Int. J. Remote Sens.*, vol. 26, no. 6, pp. 1175–1200, Mar. 2005.

[10] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[11] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, "Robust support vector regression for biophysical parameter estimation from remotely sensed images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 339–343, Jul. 2006.

[12] F. Pérez-Cruz, G. Camps, E. Soria, J. Pérez, A. R. Figueiras-Vidal, and A. Artés-Rodríguez, "Multi-dimensional function approximation and regression estimation," in *Proc. ICANN*, Madrid, Spain, 2002, pp. 757–762.

[13] M. Sanchez-Fernandez, M. de Prado-Cumplido, J. Arenas-Garcia, and F. Perez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.

[14] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.

[15] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 248–252, Apr. 2009.

[16] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "An IRWLS procedure for SVR," in *Proc. EUSIPCO*, Tampere, Finland, Sep. 2000.

[17] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

[18] I. Jonckheere, S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss, and F. Baret, "Reviews of methods for *in situ* leaf area index determination. Part I. Theories, sensors, and hemispherical photography," *Agricultural Forest Meteorol.*, vol. 121, no. 1/2, pp. 19–35, Jan. 2004.

[19] A. Gretton, O. Bousquet, A. J. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert-Schmidt norms," in *Proc. Algorithmic Learn. Theory*, S. Jain and W. S. Lee, Eds., 2005, pp. 63–77.

[20] G. Camps-Valls, J. Mooij, and B. Schölkopf, "Remote sensing feature selection by kernel dependence measures," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 587–591, Jul. 2010.