

Biophysical Parameter Estimation With a Semisupervised Support Vector Machine

Gustavo Camps-Valls, *Senior Member, IEEE*, Jordi Muñoz-Marí, Luis Gómez-Chova, *Student Member, IEEE*, Katja Richter, *Member, IEEE*, and Javier Calpe-Maravilla, *Member, IEEE*

Abstract—This letter presents two kernel-based methods for semisupervised regression. The methods rely on building a graph or hypergraph Laplacian with both the available labeled and unlabeled data, which is further used to deform the training kernel matrix. The deformed kernel is then used for support vector regression (SVR). Given the high computational burden involved, we present two alternative formulations based on the Nyström method and the incomplete Cholesky factorization to achieve operational processing times. The semisupervised SVR algorithms are successfully tested in multiplatform leaf area index estimation and oceanic chlorophyll concentration prediction. Experiments are carried out with both multispectral and hyperspectral data, demonstrating good generalization capabilities when a low number of labeled samples are available, which is usually the case in biophysical parameter retrieval.

Index Terms—Biophysical parameter, estimation, graph, kernel method, regression, retrieval, semisupervised learning (SSL), support vector machine.

I. INTRODUCTION

IN REMOTE sensing data analysis, estimating biophysical parameters is of special relevance to better understand the environmental dynamics at local and global scales. In the next years, services to users will include production of biophysical parameters at global scales to support the implementation and monitoring of international conventions. This component will be mainly based on low (> 1 km) or medium spatial resolution images (1 km–250 m) with a high time resolution from nearly real time monitoring systems to five yearly products. In this way, Earth observation images will be massively used to estimate either land [crop yield, defoliation, biomass, and leaf area index (LAI)] or water (yellow substance, suspended matter, or chlorophyll concentration) products.

In this context, there is an urgent need for robust and accurate regression methods. The inversion of analytical models intro-

duces a higher level of complexity and induces an important computational burden. In addition, with such an approach, ill-posed problems are usually encountered and sensitivity to noise becomes an important issue [1]. Consequently, the use of *empirical models* adjusted to learn the relationship between the acquired spectra and actual ground measurements has become very attractive. The original attempts introduced general linear models, but they produced poor results since biophysical parameters are commonly characterized by more complex (nonlinear) relationships with the measured reflectances [1]. More sophisticated models were also developed, including exponential or polynomial terms, but these models are often too simple to capture the relationships between remote sensing reflectance and the investigated biophysical parameters. *Parametric* models have some important drawbacks, which could lead to poor prediction results on unseen (“out-of-sample”) data. For instance, they assume explicit relationships among variables, and an explicit noise model is adopted. As a consequence, *nonparametric* and potentially *nonlinear* regression techniques have been effectively introduced for the estimation of biophysical parameters from remotely sensed images [1]. 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 [2], [3]. However, neural networks offer poor performance when working with low-sized labeled data sets. A promising alternative to neural networks is the support vector regression (SVR), a kernel method for regression and function approximation [4]. The SVR is a nonparametric, regularized, and nonlinear regression tool, which has yielded good results in modeling some biophysical parameters, [5]–[7].

Unfortunately, in biophysical parameter estimation, few ground measurements are typically available (in contrast to the wealth of unlabeled samples in the image), and also very high levels of noise and uncertainty are present in the data. This letter exploits the regularization framework proposed in [8] to develop a semisupervised version of the SVR method. Two methods with increasing sophistication are proposed: the information coming from the unlabeled samples is included in the standard SVR by means of either the graph or hypergraph Laplacian. The two methods are tested successfully in complex biophysical parameter estimation problems, such as LAI and ocean chlorophyll concentration estimation.

The rest of this letter is organized as follows. Section II presents the proposed formulations for semisupervised SVR. The obtained results are shown in Section III, and Section IV concludes this letter.

Manuscript received August 30, 2008; revised October 1, 2008. This work was supported in part by the Spanish Ministry for Education and Science under Projects DATASAT ESP2005-07724-C05-03 and CONSOLIDER/CSD2007-00018.

G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, and J. Calpe-Maravilla are with the Department Ingeniería Electrónica, Escola Tècnica Superior d'Enginyeria, Universitat de València, 46100 València, Spain (e-mail: gustavo.camps@uv.es; http://www.uv.es/~gcamps; jordi.munoz@uv.es; luis.gomez-chova@uv.es; javier.calpe@uv.es).

K. Richter is with the Dipartimento di Ingegneria Agraria ed Agronomia del Territorio, Faculty of Agraria, University of Naples “Federico II,” 80055 Portici (Na), Italy (e-mail: katja.rich@gmail.com).

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.2008.2009077

II. SEMISUPERVISED SVR

This section reviews the fundamentals for performing regression and function approximation with support vector machines and how unlabeled samples can be included in the model to improve the performance of the method.

A. SVR Method

In regression and function approximation problems, we are given a set of l training samples with measured parameter, $\{\mathbf{x}_i, y_i\}_{i=1}^l$, and a set of u unlabeled (test) samples $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$, where $\mathbf{x}_i \in \mathbb{R}^N$ is the measured spectra and $y_i \in \mathbb{R}$ represents the variable to be estimated. Supervised algorithms are based on selecting model weights using only labeled data in order to obtain good performance in the out-of-sample (test) data set. The SVR method is a supervised tool that has demonstrated good performance in remote sensing applications [7]. Notationally, given a nonlinear mapping φ to a higher dimensional feature space, the SVR computes the weights \mathbf{w} to obtain the estimation, $\hat{y}_i = \varphi(\mathbf{x}_i)^\top \mathbf{w} + b$, by minimizing $\|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i$ subject to $|y_i - \varphi(\mathbf{x}_i)^\top \mathbf{w} - b| \leq \varepsilon + \xi_i, \forall i = 1, \dots, l$ [4]. Essentially, those errors ξ_i greater than the permitted error ε are penalized with constant C linearly.

The SVR is a kernel-based method whose approximating function $f(\cdot)$ is expressed as a linear expansion over the most relevant training samples, $f(\mathbf{x}_*) = \sum_{i=1}^l \alpha_i K(\mathbf{x}_i, \mathbf{x}_*) + b$, where K represents the *kernel* function. The Gram matrix formed with all kernel evaluations must be a positive semidefinite matrix (i.e., fulfill Mercer's theorem) for the formulation to be valid and is expressed as the dot product of mapped instances $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^\top \varphi(\mathbf{x}_j)$. It should be stressed here that the method's performance strongly depends on the adequate definition of the kernel structural form, which can be cast as a similarity (or distance) measure among samples. Traditionally, the kernel form has been chosen to be either linear, polynomial, or radial basis function (RBF) to fulfill Mercer's conditions.

B. Learning the Kernel From Unlabeled Samples

Despite the good performance of these types of kernels, it is obvious that both the unlabeled information (data structure) and the geometrical relationship between labeled and unlabeled samples are obviated. Including the unlabeled information in the regression method may improve the results, which is the focus of SSL. A simple yet effective way to estimate the marginal data distribution and then include this information into any kernel method consists of "deforming" the structure of the kernel matrix according to the unlabeled data structure. The deformation can be designed either with *cluster kernels* computed from the solution offered by clustering algorithms [9] or by deforming a valid kernel with *graph-based methods* that account for the geometrical relations between labeled and unlabeled [8]. In this letter, we focus on the second strategy.

Let us define the evaluation map of the function $S(\mathbf{f}) = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_{l+u})]^\top$ and its seminorm $\|S(\mathbf{f})\|^2 = \mathbf{f}^\top \mathbf{M} \mathbf{f}$ given by a symmetric semidefinite matrix \mathbf{M} . The explicit form of the corresponding reproducing kernel $\tilde{K}(\mathbf{x}_i, \mathbf{x}_j)$ can be defined as [8]

$$\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_i^\top (\mathbf{I} + \mathbf{M} \mathbf{K})^{-1} \mathbf{M} \mathbf{K}_j \quad (1)$$

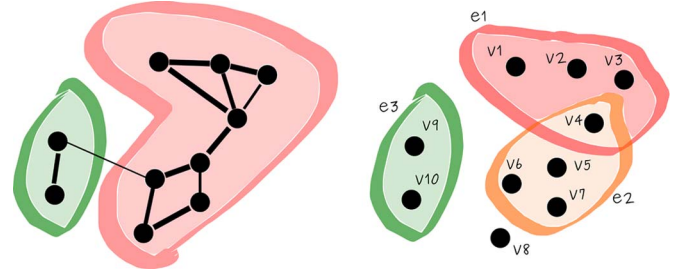


Fig. 1. (Left) Toy graph. The thickness of the edges represents the similarity between samples. Graph methods group the unlabeled samples according to the weighted distance. The two clusters are intuitively correct, even being connected by a thin weak edge. (Right) Toy hypergraph. The same ten vertices v_i and three sets e_i containing different subgraphs are potentially interconnected.

where $i, j \in \{1, \dots, l+u\}$, \mathbf{K} is the (complete) kernel matrix, \mathbf{I} is the identity matrix, and $\mathbf{K}_i = [K(\mathbf{x}_1, \mathbf{x}_i), \dots, K(\mathbf{x}_{l+u}, \mathbf{x}_i)]^\top$. The geometry of the data is included through \mathbf{M} , which is usually defined in terms of a graph. In this letter, we propose a further improvement by using hypergraphs.

1) *Graph Laplacian*: A graph $G(V, E)$ is defined with a set of nodes V connected by a set of edges E . The edge connecting nodes i and j has an associated weight W_{ij} [10]. The nodes are the samples, and the edges represent the similarity among samples in the data set [see Fig. 1 (left)]. A proper definition of the graph is the key to accurately introduce data structure in the regression machine. For computing W_{ij} , the k nearest neighbor algorithm is used in this letter.

The common choice takes $\mathbf{M} = \gamma \mathbf{L}$, where $\gamma \in [0, \infty)$ is a free parameter that controls the deformation of the kernel and \mathbf{L} is the graph Laplacian, which measures the (smooth) variation of the function \mathbf{f} along the graph [10]. The graph Laplacian is given by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where \mathbf{D} is the diagonal degree matrix of \mathbf{W} , i.e., $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$. A graph-based method for remote sensing image classification was proposed in [11].

2) *Hypergraph Laplacian*: A hypergraph is a generalization of a graph, where edges can connect any number of vertices [12]. While graph edges are pairs of nodes, hyperedges are arbitrary sets of nodes and can therefore contain an arbitrary number of nodes [see Fig. 1 (right)]. A hypergraph is also called a *set system* or a *family of sets* drawn from the universal set. Hypergraphs can be viewed as incidence structures and *vice versa*.

Notationally, a hypergraph can be represented by a matrix $H(V, E)$ with entries $h(v, e) = 1$ if $v \in e$ and 0 otherwise, called the *incidence matrix*. The hypergraph Laplacian is given by $\mathbf{L} = (1/2)(\mathbf{I} - \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2})$, where \mathbf{D}_v and \mathbf{D}_e denote the diagonal matrices containing the vertex and hyperedge degrees, respectively. In this letter, we built the hypergraph by first performing k -means clustering and then taking the centroids as hyperedges. The weights for all hyperedges were simply set to 1.

C. Algorithms for Efficient Implementation

Solving the new semisupervised problem is computationally equivalent to the original supervised SVR. Therefore, the ε insensitivity zone, the regularization parameter C , and the width of the RBF kernel σ must be tuned. However, two more parameters need to be adjusted: the number of neighbors in the graph Laplacian k and the amount of introduced deformation γ .

Therefore, selecting the best model requires trying five parameters ($k, \gamma, C, \varepsilon, \sigma$), which involves high computational burden. Moreover, the new kernel (1) has to be computed, which implies a matrix inversion of size $(l + u) \times (l + u)$. Note that this inversion scales exponentially with the number of used samples, so one pays the cost of including more unlabeled samples to better model the data marginal distribution. However, with an adequate selection of (informative) labeled samples, the need of many unlabeled is limited. There are techniques that make an approximate calculation of the inverse matrix at a lower cost. In this letter, we present two alternative formulations to speed up the matrix inversion and make feasible both training and free parameter tuning for large-scale data sets.

1) *Nyström Method*: Rather than a direct matrix inversion [see (1)], we propose to retain only the first largest p eigenvalues of the *eigendecomposition* of matrix $\mathbf{P} = \mathbf{L}\mathbf{K} = \mathbf{\Lambda}\mathbf{V}^T$, where \mathbf{V} represents the unitary matrix of eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix containing their associated eigenvalues. There are methods to find the first eigenvalues without explicitly solving the whole eigenproblem [13]. However, computational time is drastically reduced only when $p \ll n$, being $n = l + u$.

In order to reduce the computational cost involved, we introduce here the Nyström method [14], which is commonly used to produce an approximate matrix $\tilde{\mathbf{P}}$ by randomly choosing m rows/columns of the original matrix \mathbf{P} and then making $\tilde{P}_{n,n} = P_{n,m}P_{m,m}^{-1}P_{m,n}$, $m \leq n$, where $P_{n,m}$ represents the $n \times m$ block of \mathbf{P} . As a result, the method simplifies the solution of the problem to computing an approximated eigendecomposition of the low-rank kernel matrix $\tilde{\mathbf{P}} = \tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T$, involving $\mathcal{O}(mn^2)$ computational cost.

Therefore, if we approximate the normalized matrix \mathbf{P} with the Nyström method by expanding a small $p \times p$ matrix, $\tilde{\mathbf{P}} = \tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T$, and substitute it into (1), we obtain

$$\tilde{\mathbf{K}}_{ij} = \mathbf{K}_{ij} - \mathbf{K}_i^T (\mathbf{I} + \gamma \tilde{\mathbf{V}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T)^{-1} \mathbf{M}\mathbf{K}_j. \quad (2)$$

Now, by using the Sherman–Morrison–Woodbury formula¹ from linear algebra in our problem statement, it is straightforward to demonstrate

$$\tilde{\mathbf{K}}_{ij} = \mathbf{K}_{ij} - \mathbf{K}_i^T \left(\mathbf{I} - \tilde{\mathbf{V}}(\gamma^{-1}\mathbf{I} + \tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^T \right) \mathbf{M}\mathbf{K}_j \quad (3)$$

which involves inverting a matrix of size $p \times p$ (with $p \leq m \leq n$) with a computational cost $\mathcal{O}(p^2n)$, i.e., linear with the number of $(l + u)$ samples.

2) *ICF*: A positive semidefinite matrix \mathbf{K} can be factorized as $\mathbf{G}\mathbf{G}^T$, where \mathbf{G} is an $n \times n$ matrix. This factorization can be found via direct Cholesky decomposition, which can be cast as a kind of Gaussian elimination. The objective of the proposed method is, nevertheless, to find an approximate matrix $\tilde{\mathbf{G}}$ of size $n \times m$ for a small m such that the difference $\|\mathbf{K} - \mathbf{G}\mathbf{G}^T\| < \eta$, where η is the tolerance or permitted error. Basically, the incomplete Cholesky factorization (ICF) differs from the standard Cholesky decomposition in that all pivots below a certain threshold are skipped. If m is the number of

nonskipped pivots, the lower triangular matrix $\tilde{\mathbf{G}}$ with only m nonzero columns is obtained.

The proposed iterative method involves picking one column of \mathbf{K} at a time. This is done by maximizing a lower bound on the error reduction of the approximation. After t iterations, a good approximation $\mathbf{K}_t = \tilde{\mathbf{G}}_t\tilde{\mathbf{G}}_t^T$ is obtained, where $\tilde{\mathbf{G}}_t$ is $n \times t$. The ranking of the $n - t$ vectors that might be added in the next step is done by comparing the diagonal elements of the remainder matrix $\mathbf{K} - \tilde{\mathbf{G}}_t\tilde{\mathbf{G}}_t^T$. Each of these elements requires $\mathcal{O}(t)$ operations, and the update of $\tilde{\mathbf{G}}_t$ has $\mathcal{O}(tn)$ cost, so the overall complexity is $\mathcal{O}(m^2n)$.

Once the factorization is obtained, the new formulation of the algorithm is obtained by first substituting $\mathbf{L}\mathbf{K} = \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T$ in (1) and then applying the Sherman–Morrison–Woodbury formula

$$\tilde{\mathbf{K}}_{ij} = \mathbf{K}_{ij} - \mathbf{K}_i^T \left(\mathbf{I} - \gamma \tilde{\mathbf{G}}(\mathbf{I} + \gamma \tilde{\mathbf{G}}^T \tilde{\mathbf{G}})^{-1} \tilde{\mathbf{G}}^T \right) \mathbf{M}\mathbf{K}_j. \quad (4)$$

This method has several nice properties: 1) the time complexity is roughly the same as the Nyström method, but only one free parameter is needed; 2) the only elements of \mathbf{K} to be stored in memory are the diagonal elements of the kernel that are equal to one for the RBF, and the rest of the elements can be computed on demand; and 3) the number of m can be chosen to fit the approximation as tight as desired.

III. EXPERIMENTAL RESULTS

In this section, we illustrate the performance of the novel methods proposed in several biophysical parameter estimation problems: LAI estimation and chlorophyll concentration prediction from different sensors. Methods are evaluated in terms of the accuracy as a function of the number of training samples. Finally, the computational efficiency of the two proposed algorithmical alternatives is evaluated.

A. Model Development and Experimental Setup

We used the RBF kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$, where $\sigma \in \mathbb{R}^+$ is the kernel width. The RBF is particularly interesting since it has less numerical difficulties and only the Gaussian width has to be tuned. In addition, the RBF kernel is a universal kernel and includes other valid kernels as particular cases. For both the graph and hypergraph Laplacian, we tuned the k nearest neighbors in the range $\{2, \dots, 50\}$ and computed the edge weights W_{ij} using the Euclidean distance among samples.

The available data were split into labeled l and unlabeled data u . We randomly selected different rates of training samples l , selected the free parameters ($k, \gamma, C, \varepsilon, \sigma$) through a threefold cross-validation in the training set, and show the averaged root mean square error (rmse) over 100 realizations for the unlabeled test set. Complementary material is available at <http://www.uv.es/gcamps/semisvr/>.

B. Experiment 1: LAI Estimation From Hyperspectral Data

This first experiment shows results in the complex problem of estimating the LAI from hyperspectral satellite images, which is characterized by high uncertainty and ill-conditioned data. We used data from the ESA Spectra Barrax Campaigns (SPARC) (<http://gpdps.uv.es/sparc/>). During the campaign, 17 multiangular hyperspectral CHRIS/PROBA images were

¹The Sherman–Morrison–Woodbury formula states that $(\mathbf{C} + \mathbf{A}\mathbf{B})^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{B}\mathbf{C}^{-1}$, where \mathbf{C} is an invertible $n \times n$ matrix, $\mathbf{A} \in \mathbb{R}^{n \times m}$, and $\mathbf{B} \in \mathbb{R}^{m \times n}$.

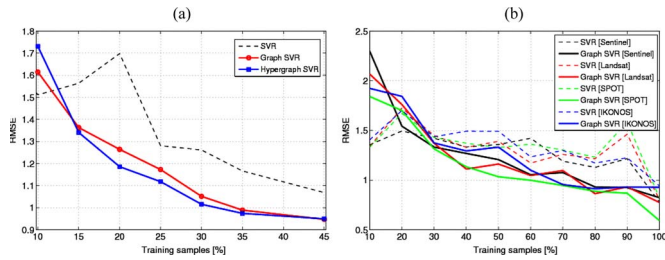


Fig. 2. RMSE in the test set as a function of the rate of training samples for LAI estimation from (a) CHRIS/PROBA (SPARC) and (b) from different sensors: Landsat-5 TM, SPOT5, IKONOS, and Sentinel-2 (PLEIADeS).

acquired. Field nondestructive measurements of LAI were made by means of the digital analyzer LI-COR LAI-2000. The LAI parameter corresponds to the effective plant area index (PAI_{eff}), since neither corrections of leaf clumping were done and the influence of additional plant components was removed. The final database consists of 139 LAI measurements and their associated surface reflectance spectra from the 62 CHRIS reflectance bands [15].

Fig. 2(a) compares the standard SVR and the proposed graph and hypergraph SVR methods. The proposed graph-based methods clearly improve the results of the supervised SVR (between a 10% and 50% improvement in rmse) when a sufficient number of labeled training samples are used ($> 12\%$, i.e., 16 samples). In this application, the use of hypergraphs shows only a slight improvement over the graph. This could be due to the fact that data may be governed by simple one-to-one cluster relations. Nevertheless, we advocate that the low number of available data and the high uncertainty derived from the acquisition could bias the results, and thus, these conclusions should be taken carefully.

C. Experiment 2: LAI Estimation From Multispectral Data

In this letter, we evaluate the performance of the algorithms on data acquired during a field campaign in the framework of the PLEIADeS project (<http://www.pleiades.es>) in the Cuga river basin of Sardinia (Italy). Multitemporal hyperspectral measurements on chicory, maize, alfalfa, and vineyard were collected in July and August 2007 using an ASD Field-Spec field spectroradiometer operating in the spectral range from 350 to 1050 nm. Along with the spectral acquisitions, $l = 45$ ground measurements of LAI were collected (alfalfa: 20, maize: 22, vineyard: 2, and chicory: 1) by means of the LAI-2000 (LiCOR) instrument at the same plots as the acquired spectral signatures. The same measurement conditions as in the first experiment hold here. Configurations of three commonly used sensors (Landsat-5 TM, SPOT5, and IKONOS), as well as the future ESA Sentinel-2 satellite, were simulated from the hyperspectral data. In this way, the data set presents only differences due to the spectral characteristics of the sensors, but not due to varying atmospheric conditions or different spatial resolutions [16].

Fig. 2(b) shows test rmse obtained with the standard SVR and the proposed graph-based methods for all four types of sensors. In this experiment, hypergraph results are not shown as very similar results to graphs were obtained. The graph methods show a higher estimation quality than the supervised SVR in all cases (up to 50% improvement of rmse) when more than 20%

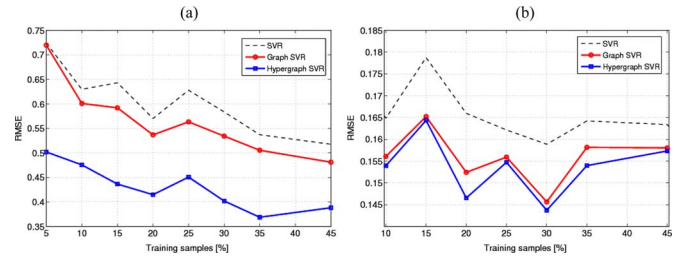


Fig. 3. RMSE in the test set as a function of the rate of training samples for chlorophyll concentration estimation from (a) the MERIS and (b) the SeaWiFS sensor.

of labeled training samples are used (i.e., $l > 9$ samples). Even though the curves follow the same trends, there are some small differences between the satellite data. For a number of training samples lower than 30%, Sentinel-2 achieves the lowest rmse. With an increasing number of used training samples, Landsat-5 TM and, particularly, SPOT5 channels demonstrate a higher performance for LAI estimation. However, for agricultural applications, the availability of training samples is usually small and limited, which makes Sentinel-2 a good option in these cases.

D. Experiment 3: Oceanic Chlorophyll Concentration Prediction From MERIS Data

In this experiment, we used simulated data of the Medium Resolution Imaging Spectrometer (MERIS) to predict chlorophyll concentration CC in subsurface waters. We selected the eight channels in the visible range (412–681 nm), as in [17]. The range of variation of the chlorophyll concentration is from 0.02 to 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. We randomly divided the set into a training ($l = 500$) and a test ($u = 4500$) set.

Results in terms of rmse versus rate of training data are shown in Fig. 3(a). In both cases, we transformed the concentration data logarithmically, $Y_{CC} = \log(CC)$, according to the study in [17]. Hereafter, units are referred to $Y_{CC}[\log(\text{mg}/\text{m}^3)]$ instead of $CC[\text{mg}/\text{m}^3]$. Unlike in the previous problem of LAI estimation, here, we observe an excellent performance of the semisupervised SVR methods. A clear gain in rmse is observed, especially in the synthetic MERIS data. These good results suggest that the improvement may come from considering a higher number of unlabeled data in the model (between three and four times more than that for LAI estimation).

E. Experiment 4: Oceanic Chlorophyll Concentration Prediction From SeaWiFS Data

For the last experiment, we used the SeaBAM data set [18], which gathers 919 *in situ* measurements of chlorophyll concentration around the U. S. and Europe related to 5 different multispectral remote sensing reflectance bands that correspond to some of the 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 $l = 460$ and $u = 459$, as in [5].

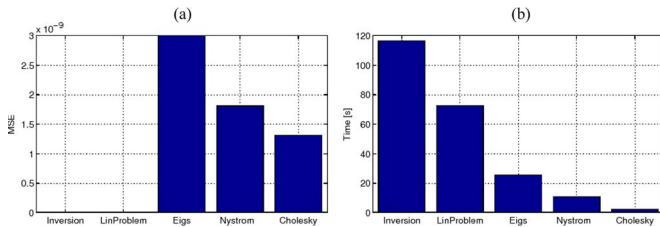


Fig. 4. Analysis of matrix inversion methods to accelerate the semisupervised SVR method. (a) RMSE and (b) CPU time (in seconds).

In this experiment, only a slight improvement of the results is observed [around 7% in rmse, see Fig. 3(b)]. We should note that hypergraphs perform better than simple graphs, thus suggesting that data structure is supradyadic, i.e., it involves more than two “types” of spectra to model chlorophyll concentration, or they can even involve numerous (complex) aspects in the relationship.

F. Analysis of Algorithm Efficacy

In this letter, we have introduced two novel formulations based on the Nyström method and the ICF in order to make it feasible the inversion of a big matrix. For the Nyström method, two free parameters need to be tuned: m is the number of samples used to compute the approximate decomposition of the kernel matrix, and p is the number of demanded largest eigenvalues (and corresponding eigenvectors). For the ICF method, only an error tolerance value η has to be fixed. In this section, we analyze the tradeoff between the accuracy of the approximation and the computational cost. We focus on the experiment in Section III-D for illustration purposes. We used the optimal values of $\sigma = 0.85$, $\gamma = 0.1$, and $k = 6$, computed the corresponding kernel matrix, the graph Laplacian, and run five different inversion methods: 1) direct inversion of the 5000×5000 matrix; 2) solving the linear system instead through standard Cholesky factorization; 3) using a fast implementation with the ARPACK method [13], in which only the largest p eigenvalues and corresponding eigenvectors are returned; and the proposed 4) Nyström and 5) ICF algorithms.

Fig. 4 shows the Frobenius norm of the committed error in the matrix inversion (zero error is considered for the direct inversion and the linear system solution) and the computational cost involved in calculating the inverse matrix (in seconds). Results show that the best tradeoff between the accuracy and the computational cost is given by the ICF that is 48 times faster than the direct matrix inversion and 4 times faster than the Nyström method while reduced inverse errors are obtained ($\sim 10^{-9}$). The computational cost and the approximation error can be easily trimmed with a single intuitive parameter η .

IV. CONCLUSION

Two novel semisupervised SVR methods for model inversion have been proposed in this letter. The information from unlabeled samples is included in the standard SVR by means of the graph or hypergraph Laplacian. Two alternative fast formulations were proposed to achieve operational processing times. Results on ocean chlorophyll concentration prediction and LAI

estimation have demonstrated the good performance of the method. Note that there is still room for further improvement since the σ parameter and the hypergraph were not optimized for the semisupervised methods.

ACKNOWLEDGMENT

The authors would like to thank ESA for the availability of the data acquired in the framework of the ESA-SPARC Project ESTEC-18307/04/NL/FF.

REFERENCES

- [1] 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, pp. 381–439, 2000.
- [2] 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 Knowl. Process. Natural Resource Manage.*, Varese, Italy, 2002, pp. 54–58.
- [3] B. Dzwonkowski and X.-H. Yan, “Development and application of a neural network based ocean colour algorithm in coastal waters,” *Int. J. Remote Sens.*, vol. 26, no. 6, pp. 1175–1200, Mar. 2005.
- [4] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
- [5] 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.
- [6] 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.
- [7] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, “Robust support vector regression for biophysical variable estimation from remotely sensed images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 339–343, Jul. 2006.
- [8] V. Sindhwani, P. Niyogi, and M. Belkin, “Beyond the point cloud: From transductive to semi-supervised learning,” in *Proc. 22nd ICML*, 2005, pp. 824–831.
- [9] O. Chapelle, J. Weston, and B. Schölkopf, “Cluster kernels for semi-supervised learning,” in *Advances in Neural Information Processing Systems 15*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2002, pp. 601–608.
- [10] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. Cambridge, MA: MIT Press, 2006.
- [11] G. Camps-Valls, T. Bandos, and D. Zhou, “Semi-supervised graph-based hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
- [12] D. Zhou, J. Huang, and B. Schölkopf, “Learning with hypergraphs: Clustering, classification, and embedding,” in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 1601–1608.
- [13] R. B. Lehoucq and D. C. Sorensen, “Deflation techniques for an implicitly restarted Arnoldi iteration,” *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 4, pp. 789–821, Oct. 1996.
- [14] C. K. I. Williams and M. Seeger, “Using the Nyström method to speed up kernel machines,” in *Advances in Neural Information Processing Systems, NIPS2001*, vol. 13, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. Cambridge, MA: MIT Press, Dec. 2001, pp. 682–688.
- [15] F. Vuolo, G. D’Urso, and L. Dini, “Cost-effectiveness of vegetation biophysical parameters retrieval from remote sensing data,” in *Proc. IEEE IGARSS*, Jul. 2006, pp. 1949–1952.
- [16] K. Richter, F. Vuolo, and G. D’Urso, “Leaf area index and surface albedo estimation: Comparative analysis from vegetation indexes to radiative transfer models,” in *Proc. IEEE IGARSS*, Boston, MA, Jul. 2008.
- [17] 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.
- [18] 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. 24937–24953, Oct. 1998.