Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 107 (2007) 348–361
Remote Sensing of Environment
www.elsevier.com/locate/rse

Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer

Surya S. Durbha *, Roger L. King, Nicolas H. Younan

Department of Electrical and Computer Engineering, GeoResources Institute (GRI), Mississippi State University, Mississippi State, MS 39762-9571, USA

Received 28 February 2006; received in revised form 10 September 2006; accepted 23 September 2006

Abstract

The retrieval of biophysical variables using canopy reflectance models is hindered by the fact that the inverse problem is ill posed. This is due to the measurement, model errors and the inadequacy between the model and reality, which produces similar reflectances for the different combination of the input parameters into the radiative transfer model. This leads to unstable and often inaccurate inversion results. The ill-posed nature of the inverse problem requires some regularization. Regularization means that one tries to consider only those solutions that are in the proximity of the true value. In order to regularize the model inversion, we propose kernel-based regularization by support vector machines regression (SVR) method.

The formulation of the SVR contains meta-parameters C (regularization) and ε-insensitive loss. The SVR generalization performance (estimation accuracy) depends on these two parameters and the kernel parameters. Often the meta-parameters are selected using prior knowledge and/or user expertise. In this paper we adopt methods for the estimation of the meta-parameters from the input data itself instead of relying on any prior information. This paper is focused on the retrieval of leaf area index (LAI) from multiangle imaging spectroradiometer (MISR) data. The proposed methodology was implemented by inverting a 1D canopy reflectance model (PROSAIL) using SVR over MISR data. The results were validated against the LAI retrievals at the Alpilles EOS validation core site. An RMSE of 0.64 was obtained using both near-infrared (NIR) in conjunction with the red band and an RMSE of 0.50 using only the NIR band.

Keywords: Inverse problem; Regularization; Support vector machines; Kernel PCA; Leaf area index

1. Introduction

Radiative transfer models describe the interaction of the sun's electromagnetic radiation with the atmosphere and the Earth's surface accounting for both the scattering and absorption of the radiation. This process is highly nonlinear and numerical simulations are required to understand these complex interactions (Jin and Liu, 1997). In particular, the process in which solar radiation and vegetation interact reveals biome functioning since the analysis of reflectance allows the retrieval of major canopy characteristics. Among all the extraction methods, those relying on physically based models that calculate top-of-canopy reflectances have proved to be a promising alternative to estimate vegetation biophysical parameters (Bacour et al., 2002). The inversion of canopy reflectance models is potentially a more robust and accurate method than the use of empirical relationships for extracting biophysical parameters from images acquired from space. The inversion of canopy models using off-nadir data based on studies of bi-directional reflectance distribution function (BRDF) has been the primary concern for quantitative extraction of biophysical parameters and canopy architecture (Goel, 1988; Hu et al., 2003; Liang, 1993; Myneni et al., 1995). The use of directional information from multiangular measurements, which are likely to become increasingly available either from existing or planned specific space-borne sensors such as MISR, ATSR-2, POLDER, CHRIS-PROBA, etc. (Barnsley et al., 2004; Casa & Jones, 2005; Chen et al., 2003; Deschamps et al., 1994; Diner et al., 1998) should increase the power of model inversions.

* Corresponding author.
E-mail addresses: suryad@gri.msstate.edu (S.S. Durbha), rking@engr.msstate.edu (R.L. King), younan@ece.msstate.edu (N.H. Younan).

0034-4257/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2006.09.031

Traditional inversion method involves numerical optimization techniques that start with an initial guess and search for the optimum parameter set for a specific set of reflectance and is an iterative process that minimizes the error. The commonly used algorithms are the Powell's method, Downhill Simplex method, and Quasi-Newton methods (Press et al., 1986). However, they do not guarantee a stable or optimum inversion, and may be computationally intensive. The look-up table (LUT) approach is a much simpler technique. To construct the LUT, the simulated canopy reflectance is obtained by varying its structural and radiometric properties. In addition, several studies in the recent past used artificial neural networks (ANN) in the forward and inverse modeling of radiative transfer models for retrieving biophysical variables (Atzberger, 2004; Gong et al., 1999; Gopal & Woodcock, 1996; Jensen et al., 1999; Kimes et al., 1997; Knyazikhin et al., 1998a; Pierce et al., 1994; Privette et al., 1996; Schlerf & Atzberger, 2006; Smith, 1993; Weiss & Baret, 1999). A recent addition to the inversion methods is the genetic algorithm (GA) (Fang et al., 2003) which was used to retrieve LAI from field measured reflectance as well as Landsat ETM+ data.

Previous studies have shown that the ANN and LUT approaches generally performed best. However, the lack of good generalization capacity is one of the disadvantages of the LUT and ANN approaches (Fang et al., 2003; Kimes et al., 2000).

In general, the data obtained in empirical modeling are finite and sampled; typically this sampling is non-uniform and due to the high-dimensional nature of the problem, the data will form only a sparse distribution in the input space. Consequently, the problem is nearly always ill posed (Gunn, 1998; Poggio et al., 1985). The problem of biophysical variables estimation by inversion of canopy radiative transfer models is known to be ill posed (Combal et al., 2002). "Ill posed" means that when finding a p that satisfies the equality $Ap = F$, where A is a linear operator, we have large deviations in the resolved p corresponding to small deviation in F. This is due to the measurement, model errors and the inadequacy between the model and reality, which produces similar reflectances for different combination of the input parameters into the radiative transfer model. Also nonlinear interactions between the variables can lead to unstable solutions (Bacour et al., 2002).

The inverse problem can be solved properly only if it is well posed, in the sense defined by Hadamard, i.e., a problem is well posed if and only if its solution exists, is unique, and depends continuously on the data (Combal et al., 2002; Garabedian, 1964). To solve ill-posed problems, regularization techniques can be used to provide stable and reliable solutions. Regularization solves the problem of choosing among the infinite number of functions that all pass through the finite number of data points by imposing a smoothness constraint on the final solution. This results in minimizing the cost functional

$$H[f] = \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda ||f||_k^2 \quad (1)$$

where $||f||_k^2$ is a measure of the deviation from the smoothness of the solution f (Evgeniou et al., 1999; Evgeniou et al., 2000) and the sum is the deviation of the function from the data points (thus we are making a tradeoff between accurately modeling the data points and the smoothness of the learned function). λ is a parameter that tunes the tradeoff between model complexity (l — norm of model parameters, f) and minimization of the training errors (first term on the right hand side of Eq. (1)). Intuitively, regularization is equivalent to finding the estimator which uses the minimum possible energy of the data to estimate the output. Thus it is not necessary to focus on the minimization of an empirical error over existing data. Such minimization is both ill posed and does not necessarily lead to models with good predictive capabilities. Instead, it is necessary to minimize a combination of the empirical error over existing data and a penalty factor that penalizes solutions that are too complex: the smoothness or capacity of the functions considered needs to be controlled (Evgeniou et al., 2002).

Recent methods for regularization in the inversion of canopy reflectance models include:

- Use of prior knowledge to regularize the inversion process in a look-up table (LUT)-based approach (Combal et al., 2002),
- Take into consideration the dynamic evolution of the LAI during the crop cycle (CROMA, 2000), and
- Utilize the neighborhood radiometric information of the pixel actually being inverted (Atzberger, 2004).

A Support Vector Machine (SVM) developed by Vapnik (1995) is based on statistical learning theory and can be regarded as the same type of network, corresponding to exactly the same type of solution f (Eq. (1)) but trained in a different way and therefore with different values of the weights after the training (Evgeniou et al., 1999). SVMs can be used for both classification and regression problems. SVMs provide excellent generalization capabilities, fast, robust to high input space dimension, low number of samples, provide sparse solutions where only the most relevant samples of the training data are weighted resulting in low computational cost and memory requirements.

Some applications of SVMs for classification are isolated handwritten digit recognition (Burges and Scholkopf, 1997), speech recognition (Campbell et al., 2006), and face detection in images (Osuna et al., 1997; Türkan et al., 2006). In the case of regression, SVMs have been applied to benchmark time series prediction tests (Cao, 2003; Muller et al., 1997), the Boston housing problem (Drucker et al., 1997) and financial forecasting (Trafalis & Ince, 2000). In most of these cases SVM generalization performance either matches or is significantly better than competing methods (Burges, 1998).

While much work exists in the domain of retrieving the canopy biophysical variables from model inversion methods using a variety of optimization methods (Baret et al., 1995; Barnsley et al., 2000; Bicheron and Leroy, 1999; Combal et al., 2002; Goel et al., 1984; Jacquemoud & Baret, 1993; Knyazikhin et al., 1998a; Kuusk, 1991), applications of support vector machines in remote sensing problems are focused mainly towards classification (Banerjee et al., 2006; Durbha & King, 2005; Mantero et al., 2005; Mathur and Foody, 2004; Mazzoni et al., 2000; Melgani & Bruzzone, 2004; Wohlberg et al., 2006)

and to the best of our knowledge little work has been reported towards the application of support vector regression (SVR) for the retrieval of biophysical variables from inversion of canopy radiative transfer models.

This study focuses on the retrieval capacity of SVR on one of the key biophysical variables — vegetation green leaf area index (LAI). LAI defines the size of the interface for exchange of energy and mass between the canopy and the atmosphere (Weiss et al., 2000) and it is a key state variable in land surface models. LAI governs net radiation and its expenditure (energy balance), net primary production (carbon fixation), evapotranspiration and canopy interception (water budget) (Tian et al., 2002). In the development of an SVM predictor, the first important step is feature selection or feature extraction. In an SVM, all the available indicators can be used as the inputs, but irrelevant or correlated features could adversely impact the generalization performance due to the curse of the dimensionality problem. Thus it is critical to perform feature selection or feature extraction in SVM (Bradley & Mangasarian, 1998; Durbha & King, 2005; Guyon et al., 2002; Tay & Cao, 2001; Weston et al., 2001). We perform feature extraction by kernel principle component analysis (KPCA), which is a nonlinear extension to the normal PCA. The extracted features provide more robust estimation of the underlying structure of the data and then are used to perform an SVR.

This paper is organized as follows: in Section 2, we provide a description of the materials and methods used in this work. Specifically, instead of using the exhaustive parameter space realization we propose a method for parameter selection for input into the canopy radiative transfer model that generates a reduced set of parameter combinations. Also the generation of the training and testing data from the forward modeling of the PROSAIL model, the MISR data used in the present study, and the validation data from field derived LAI at the Alpilles validation site are described. Section 3 provides the theoretical background for the kernel-based learning methods adopted in this work; kernel principal components analysis (KPCA) for feature extraction and SVR for model generation and prediction of LAI. This section also presents the proposed parameter selection approaches for the SVR model generation; in particular we provide a brief review of the existing techniques and the methodology we adopted for the parameters selection which were derived from the input training data itself without relying on any prior information thus accounting for the regularization of the solution. The results of the SVR-based approach are presented in Section 4 along with some analysis. We conclude in Section 5 by presenting the advantages of the proposed approach and the limitations therein.

2. Materials and methods

2.1. Radiative transfer model

The SAIL model (Verhoef, 1984) is a 1D turbid medium radiative transfer model. The PROSPECT (Jacquemoud et al., 2000) model has been coupled with the SAIL model to obtain the PROSAIL model and three variables are used to describe the canopy structure, the LAI, the average leaf angle, and a hot spot parameter. For a given input parameter set, both spectral and directional reflectances are calculated after choosing the type of sensor and the number of viewing angles. A number of sensors (e.g., HRV, TM, AVHRR, AVIRIS, and MISR) could be simulated using the PROSAIL model (Jacquemoud et al., 2000). We have used the MISR sensor with 4 bands and 9 viewing angles to simulate the directional reflectance using the PROSAIL model.

Several techniques have been proposed to sample the variables (input parameters); Combal et al., 2002 used 280,000 values of each variable, which were randomly drawn with a distribution specific to each variable. The space of the model input variables was sampled by randomly drawing values within particular distribution functions. Atzberger (2004) proposed the selection of parameter ranges by scaling the parameters intrafield variability; Bacour et al. (2002) used a range of values for each parameter where the lowest (highest) values correspond to the lower (upper) bounds of the ranges of variation increased (decreased) by 5%. The other levels are regularly spaced between these two bounds. We adopted this method and the simulations have been carried out with six free parameters: leaf area index (LAI), mean leaf inclination angle (θ_1), the hot spot parameter (S_1), a spectral soil parameter that controls the soil reflectance levels (α_{soil}), the leaf structure parameter (N), and chlorophyll A and B (C_{ab}). As shown in Table 1, there are six free parameters and each of these free parameters takes 7 values. Therefore, the total number of simulations for the PROSAIL model is 7^6 (i.e., 177,649 simulations). The choice of the seven levels for each parameter is driven by the nonlinear behavior of LAI and C_{ab}, for instance, and is a compromise because more levels would greatly increase the number of simulations. A more detailed description for the choice of these levels is described in Bacour et al. (2002). For the PROSAIL model the soil is characterized by a typical Lambertian soil reflectance spectra multiplied by a brightness parameter S. This permits representing the influence on reflectance of soil roughness and moisture variation (Combal et al., 2002).

2.2. Selection of parameter combinations

As discussed in the previous section, the simulations are enormous when all the combinations of parameters are

Table 1
Input parameters for PROSAIL model and their range of variation

Parameter	Unit	Range of variation	Levels
LAI	$m^2\ m^{-2}$	0–7	0.4, 1.4, 2.5, 3.5, 4.6, 5.6, 6.7
C_{ab}	$\mu g\ cm^{-2}$	1–80	5, 17, 29, 41, 52, 64, 76
θ_1	degrees (°)	5–85	9, 21, 33, 45, 57, 69, 81
s_1		0.01–1	0.06, 0.21, 0.36, 0.51, 0.65, 0.80, 0.95
α_{soil}		0.5–2	0.57, 0.80, 1.02, 1.25, 1.48, 1.70, 1.93
N		2.5	1.1, 1.3, 1.5, 1.8, 2.0, 2.2, 2.4

considered. Bacour et al. (2002) reduced the number of combinations by treating the problem in a similar manner to a fractional factorial experimental design where data from computer runs are collected using a *Hyper Graeco Latin* Geometric Sampling scheme and all factors have the same number of levels. The number of simulations by adopting the above approach has been reduced to 343 from an exhaustive set of 280,000 simulations.

In this paper we adopted an alternative strategy for exhaustive testing; pseudo exhaustive testing attempts to reduce the large test time required for exhaustive testing. The methodology is being commonly used for writing test cases for software testing purposes, but can be used in any scenario, which deals with multiple parameter dimensions. Testing is currently the dominant method for finding and eliminating software errors. Obtaining good test cases is obviously a key requirement to successfully test any software artifact, but many issues (labor intensive, exercise only a restricted subset of the functionality, etc.) complicate this activity. The alternative is to automatically generate a range of test cases, then filter out any test cases that do not satisfy the required input invariants of the system under test.

In this paper we investigate this methodology by generating test cases for an input into the PROSAIL model. Parameters tend to come in dimensions, where a test case for each simulation of the model can choose one parameter from each dimension. For example, LAI is a dimension and the variable can have only one of several ranges of values. Each test case chooses one feature from each dimension, and often chooses them independently of each other. We used a software package named Jenny for this purpose (Jenkins, 2005). Exhaustive testing grows exponentially with the number of dimensions to be tested together, but by constraining the testing space by covering all pairs (triples, quadruples, *n*-tuples) of features and allowing every test case to cover many *n*-tuples, the number of test cases required grows only with the log of the number of dimensions (Jenkins, 2005). The test cases for input into the PROSAIL model have been generated by covering all triples of features and resulted in 567 test cases.

2.3. Validation dataset

A field campaign over a 3×3 km^2 agricultural area near Alpilles in France (43.810°N, 4.750°E) was performed from February 26 to March 15, 2001 (Baret et al., submitted for publication). More than 95% of this site was composed of young and mature wheat and grasses (biome 1). Leaf area index was measured with a LAI-2000 Plant Canopy Analyzer (Jiannan et al., 2007).

A subset of an ETM+ image from March 15, 2001 (path 196, row 90) containing the Alpilles site was selected for the purpose of generating a fine resolution LAI map of the site. The image was atmospherically corrected using the 6S radiative transfer code (Vermote et al., 1997). Various techniques, including an empirical regression based on the simple ratio (SR) and the fine resolution MODIS LAI/FPAR algorithm, were evaluated to identify the most accurate method for generating the fine resolution LAI maps.

The fine resolution MODIS algorithm and the SR relationship were the best candidates for this site and thus were used to generate a 30 m LAI map of a 20×20 km area centered on the Alpilles site (Tan et al., 2005). This map was reprojected from UTM WGS84 projection into the Space Oblique Mercator (SOM) projection first and then degraded to a 1.1-km resolution reference map. This map (Jiannan et al., 2007) was made available to us for this study. Pixels belonging to Biome 1 (grasses and cereal crops) were selected for further analysis.

The MISR instrument on the Earth Observing System (EOS) Terra platform orbits the Earth about 15 times each day. There are 233 distinct orbits, called paths, which are repeated every 16 days, and since the paths overlap, near global coverage is obtained in 9 days. The MISR instrument views symmetrically about the nadir in forward and aftward directions along the spacecraft flight track. Image data are acquired with nominal view zenith angles relative to the surface reference ellipsoid of 0.0°, 26.1°, 45.6°, 60.0° and 70.5° in four spectral bands (446, 558, 672, and 866 nm).

The MISR data from path 196, orbit 6598 (March 15, 2001) were used to validate the proposed SVR-based LAI retrieval algorithm. A 16.5×16.5 km area (15×15 MISR pixels) which coincided with the reference map was extracted from this path.

2.4. Training and testing data

The PROSAIL model was simulated for the 567 test cases obtained by the application of pseudo exhaustive testing method as described in Section 2.2. The generated reflectances in the nine view angles for the red and NIR bands and the corresponding canopy variable (LAI) were partitioned into training and testing data. The training data consisted of PROSAIL simulated directional reflectance for all the nine MISR view angles using 400 test cases and the simulated directional reflectance from 167 test cases was used as the testing dataset.

3. Kernel-based methods to solve the inverse problem

3.1. Kernel principal component analysis (KPCA)

Kernel principal component analysis (KPCA) is an efficient generalization of the traditional principal component analysis (PCA) (Haykin, 1999; Schölkopf et al., 1998, 1999) which computes the principal components in a high-dimensional feature space F, which is nonlinearly related to the input space that allows for the detection and characterization of the low-dimensional nonlinear structure in the multivariate datasets (Fig. 1). KPCA has been successfully applied for denoising images and extracting features (Girolami, 2002; Rosipal et al., 2001; Schölkopf et al., 1998). Given a set of centered input vectors $x_t=(t=1,\ldots l)$ and $\sum_{t=1}^{l} x_t = 0$, each of which is of m dimension $x_t=(x_t(1), x_t(2), \ldots x_t(m))^T$ (usually $m<l$), the idea

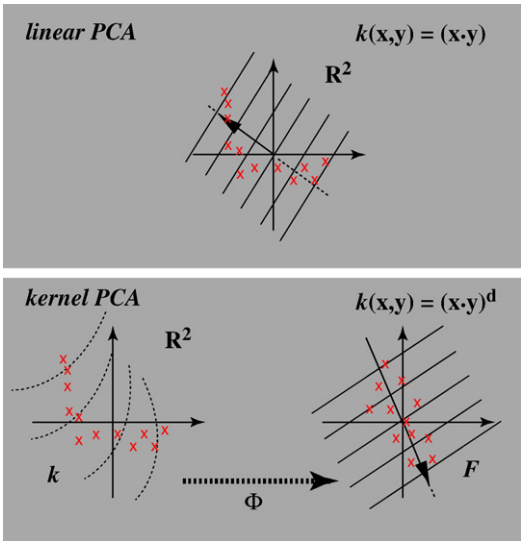

Fig. 1. Kernel PCA implicitly performs a linear PCA in some high-dimensional feature space that is nonlinearly related to input space. Linear PCA in the input space is not sufficient to describe the most interesting direction. By using a suitable nonlinear mapping the resulting nonlinear direction in the input space can find the most interesting direction (adapted from Schölkopf et al., 1998).

of KPCA is to first map the original input vectors x_t into a higher dimensional feature space $\phi(x_t)$ and then to calculate the linear PCA in $\phi(x_t)$ (Fig. 1). By mapping x_t into $\phi(x_t)$ whose dimension is assumed to be larger than l, KPCA solves the eigenvalue problem.

$$\lambda_i u_i = \bar{C} u_i, i = 1,\ldots,l \tag{2}$$

where $\bar{C} = \frac{1}{l}\sum_{i=1}^{l} \varphi(x_t)\varphi(x_t)^T$ is the sample covariance matrix of $\phi(x_t)$, λ_i is the non-zero eigenvalues of $\bar{C}$, and u_i is the corresponding eigenvector.

KPCA was applied to the training dataset and the extracted nonlinear components were projected on to the datasets to produce the training and testing data for SVR. For a complete description of the algorithm, the reader is referred to Smola and Scholkopf (2004) and Durbha and King (2005).

3.2. Support vector regression (SVR)

In recent years, the use of support vector machines (SVMs) on various classification and regression problems has become increasingly popular. SVMs can be applied to both classification and regression problems. In the classification case, we try to find an optimal hyperplane that separates two classes. In order to find an optimal hyperplane, we need to minimize the norm of the vector w, which defines the separating hyperplane. This is equivalent to maximizing the margin between two classes (Fig. 2).

In regression, the goal is to estimate an unknown continuous-valued function based on a finite number set of noisy samples. We perform feature extraction by KPCA as described in Section 3.1, then the training points can be expressed as (s_1, y_1), (s_2, y_2), … (s_n, y_n) where $s_i \subseteq R^n$ is the KPCA transformed input vector, $y_i \subseteq R$ is the target value.

Assumed that the statistical model for data generation has the following form:

$$y = r(s) + \delta \tag{3}$$

where $r(s)$ is an unknown target function (regression), and δ is additive zero mean noise with noise variance σ^2 (Cherkassky & Mulier, 1998; Cherkassky et al., 1999).

In SVM regression, the input s is first mapped onto an m-dimensional feature space using some fixed (nonlinear) mapping, and then a linear model is constructed in this feature space (Smola & Scholkopf, 2004; Vapnik, 1995, 1998). The linear model (in the feature space) $f(\mathbf{x}, w)$ is given by

$$f(\mathbf{x},w) = \sum_{j=1}^{m} \omega_j \varphi_j(s) + b \tag{4}$$

where $\phi_j(s)$, $j=1, \ldots m$ represents a high-dimensional feature space, which is nonlinearly mapped from the input space s, and b is the "bias" term. Often the data are assumed to be zero mean (this can be achieved by preprocessing), so the bias term in expression (4) is dropped.

The quality of estimation is measured by the loss function $L(y, f(s, \omega))$. SVM regression uses a new type of loss function called ε-insensitive loss function proposed by Vapnik (1995):

$$L_\varepsilon(y,f(s,\omega)) = \begin{cases} 0 & \text{if } |y - f(s,\omega)| \le \varepsilon \\ |y - f(s,\omega)| - \varepsilon & \text{otherwise} \end{cases} \tag{5}$$

The empirical risk is:

$$R_{\text{emp}}(\omega) = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(y_i, f(s,\omega)) \tag{6}$$

SVM regression performs linear regression in the high-dimension feature space using ε-insensitive loss and, at the same time, tries to reduce model complexity by minimizing

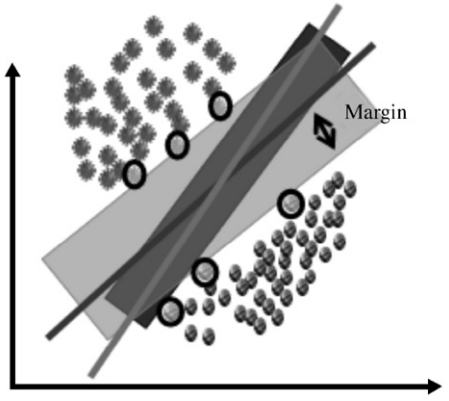

Fig. 2. Linearly separable case; only support vectors (dark circled) are required to define the optimally defined hyperplane.

$\|\omega\|^2$. This can be described by introducing (non-negative) slack variables ξ_i, $\xi_I^* i=1, \ldots n$, to measure the deviation of training samples outside ε-insensitive zone. Thus SVM regression is formulated as minimization of the following functional:

$$\min \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)$$

$$s.t.\begin{cases} y_i - f(s_i,\omega) \le \varepsilon + \xi_i^* \\ f(s_i,\omega) - y_i \le \varepsilon + \xi_i \\ \xi_i,\xi_i^* \ge 0, i = 1,\ldots,n \end{cases} \quad (7)$$

where C is referred to as the regularization constant and ε is the tube size of SVM.

By introducing Lagrange multipliers and exploiting the optimality constraints, the decision function has the following explicit form:

$$f(s) = \sum_{i=1}^{n_{SV}}(\alpha_i - \alpha_i^*)K(s_i,s) \quad s.t.\ 0 \le \alpha_i^* \le C,\ 0 \le \alpha_i \le C, \quad (8)$$

where n_{SV} is the number of Support Vectors (SVs) and the kernel function

$$K(s,s_i) = \sum_{j=1}^{m} \varphi_j(s)\varphi_j(s_i) \quad (9)$$

and α_i^* are the so-called Lagrange multipliers.

The methodology adopted in this paper is illustrated in Fig. 3. We implement a hybrid approach of feature extraction using KPCA and then performing SVR on the resultant components, our previous work has demonstrated that this provides good generalization capabilities (Durbha and King, 2005).

3.3. ε-Support vector regression parameters selection

There are several issues that have to be considered while using ε-support vector regression for inversion applications. The quality of the SVM models depends on a proper setting of the SVM meta-parameters. The main issue for practitioners trying to apply SVM regression is how to set these parameter values (to ensure good generalization performance) for a given dataset. The common technique of resampling for (simultaneously) tuning several SVM regression parameters is very expensive in terms of computational costs and data requirements. There are plenty of contradictory opinions for the selection of optimal parameters for SVR; a brief description of the proposed methods has been summarized in Table 2.

Since noise often occurs in real input datasets, it is very important for us to derive the dependency between the optimal parameters in support vector regression and the noisy input. Kwok and Ivor (2003) and Smola and Scholkopf (2004) established the linear dependency between ε in ε-SVR with the ε-insensitive loss function and the noisy input. Kwok and Ivor (2003), show that there is always a linear relationship between the optimal value of ε and the noise level. Hence we estimate the noise variance in the input training dataset by a *k*-nearest neighbor (KNN) algorithm.

In SVM regression, the parameter C determines the tradeoff between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in the optimization formulation. For example, if C is too large (infinity), then the objective is to minimize the empirical risk without regard to model complexity part in the optimization formulation. The parameter ε controls the width of the ε-insensitive zone, used to fit the training data (Cherkassky & Mulier, 1998; Vapnik, 1995, 1998) and can affect the number of support vectors used to construct the regression function. The bigger the ε is, the fewer support vectors will be selected. On the other hand, bigger ε values result in more 'flat' estimates. Hence, both C and ε values affect the model complexity.

In this work we adopt the strategy proposed by Cherkassky & Ma (2004) for the selection of the optimal values for the parameters C and ε. Their proposed approach (to parameter selection) is based on a well-known theoretical understanding of SVM regression that provides the basic analytical form of dependencies for parameter selection, below we give only the expressions to calculate the parameters (for detailed derivation the reader is referred to the publication Cherkassky & Ma, 2004).

(i.) Parameter C

$$C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \quad (10)$$

where $\bar{y}$ is the mean of the training responses (outputs), and σ_y is the standard deviation of the training response values. The above estimation of C can effectively handle outliers in the training data.

(ii.) Parameter ε. It is well known that the value of ε should be proportional to the input noise level, that is $\varepsilon \propto \sigma$ (Cherkassky & Mulier, 1998; Kwok & Ivor, 2003; Smola et al., 1998; Vapnik, 1998) where

$$\varepsilon = \tau\sigma\sqrt{\frac{\ln n}{n}} \quad (11)$$

Based on empirical tuning, the constant value $\tau=3$ gives good performance for various dataset sizes, noise levels, and target functions for SVM regression. Thus expression (11) was used in our work.

We use the radial basis function (RBF) kernel that contains a free parameter, γ which is the Parzen window width for the RBF kernel. The proper choice for γ is usually determined by hyper-tuning on an external validation set. The precise choice for γ is usually not crucial, because there is generally a relatively broad range for which the model quality is stable. Selecting a particular kernel type and kernel function parameters is usually based on application domain knowledge and also should reflect distribution of input ($\mathbf{x}$) values of the training data (Vapnik, 1998). We used LIBSVM (Chang &

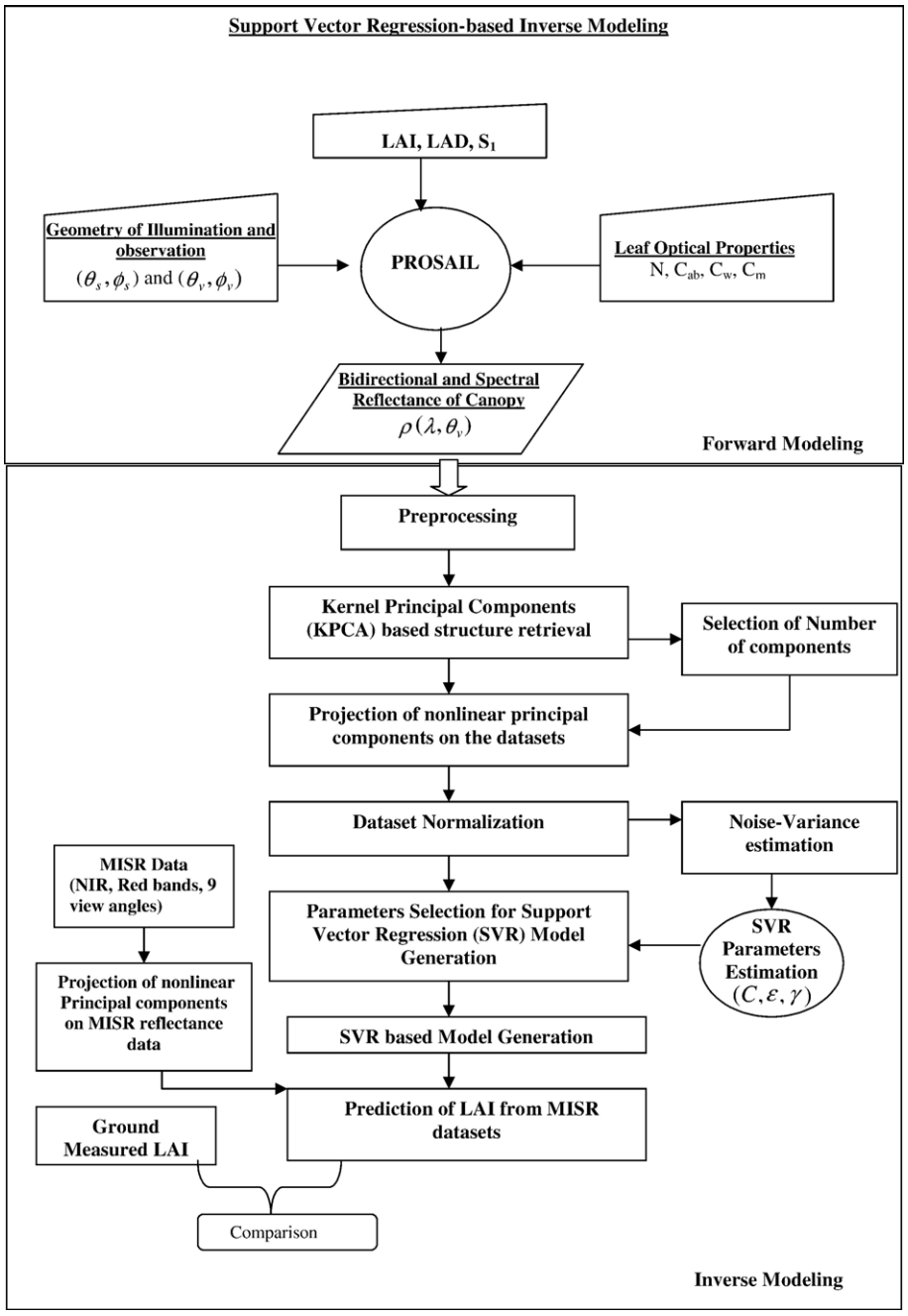

Fig. 3. Flowchart of the approach to estimate LAI with support vector regression algorithm.

Lin, 2001) software for implementing the SVR-based inversion method.

4. Results and analysis

This section presents results from the proposed SVR-based approach and also compares the results with the other prevalent approaches based on ANN and LUT.

4.1. Results from SVR-based methodology

The KPCA method of feature extraction produces nonlinear principal components which are substantially higher (up to the number of data points n) than a normal PCA. This can be nearly always advantageous, especially in the situation where the dimensionality N of the input data points is significantly smaller than the number of data points and the data structure is spread over all eigen directions (Rosipal et al., 2001). Fig. 4a illustrates the number of components extracted by KPCA. Then, the extracted features have then been projected on to the original training and testing datasets.

A model is generated by SVR and it is then tested for its predictive capability on a totally unseen testing data which were also projected on the nonlinear components. The number of components that are used in the training and testing data was determined based on the mean squared error (MSE) on the

Table 2
Proposed methods for parameter selections for SVR

Authors	Parameters	Proposed strategies	Comments
Cherkassky and Mulier, 1998	C and ε	Prior knowledge and/or user expertise	Prior knowledge not available, ordinary users cannot use the methods
Schölkopf et al. (1998)	ε	Control another parameter ν	User defined
Mattera and Haykin (1999)	ε	Percentage of support vectors is 50% of the number of samples	Not true always
Kwok and Ivor (2003)	ε	Asymptotically optimal values proportional to noise variance	Does not reflect sample size.
Mattera and Haykin (1999)	C	C equal to the range of output values	Does not take into account possible effect of outliers in the training data
Schölkopf et al. (1998)	C and ε	Cross-validation for parameter choice	Computation and data-intensive

testing dataset. As can be seen in Fig. 4b for a typical case of the training dataset, the number of components can reach up to 300. The amount of information encoded by the first 24 eigenvectors is illustrated in Fig. 5. The contour lines are values of constant eigenvalue. This figure depicts only datasets that correspond to two view angles (45.6° and 70.5°) of the MISR pixels; however, KPCA was performed on data from all the 9 view angles. The performance for nonlinear components can be improved by using more components than is possible in the linear case. The latter is related to the fact that there are much more higher-order features than there are pixels in an image (Smola & Scholkopf 2004).

The SVR models predictive and generalization capability depends on the optimal selection of the SVR parameters C and ε. The methodology adopted in this work has been described in Section 3.3. Here we present the parameter values used in the training of SVR. As shown in Fig. 6 the KNN algorithm was run with different values of K and for different sizes of the dataset size (N), we found the value of $K=3$ to be a reasonable value and estimated the noise variance for the dataset and finally the value of the ε-insensitive parameter. Table 3, depicts the values of C and ε used in our study. The number of support vectors obtained during the training process represents the extent of fit of the data. Large numbers of free support vectors usually signify over fitting of the data.

The number of free parameters allowed in the inversion process affects the predictive capability of the SVR generated model. However, we have refrained from fixing more than one free parameter (Table 1) as the generalization capability for predicting LAI on a wide range of land cover types would be affected by fixing the free parameters and thus introducing prior information which is generally unknown in many scenarios. Thus the estimation of noise variance and the calculation of the ε insensitive parameter from it implicitly allow the introduction of information about the noise which is a way to introduce prior information on uncertainties in the system. Combal et al., (2002) have introduced a 2.5% noise as a way of including prior information in a neural network-based inversion process.

The SVR training was conducted with training data from both the NIR and red band and in a combination of the NIR and red bands. Several experiments were conducted with both bands (NIR and RED) separately and as a combination and the predictive capability of the model was evaluated on test datasets, whose results are depicted in Table 3.

As can be seen, using only the NIR band data, an R^2 value of 0.86 was achieved. A combination of NIR and Red bands did not improve the previous result but rather the R^2 value has decreased to 0.84. The free parameters N, θ_1, C_{ab}, and S_1 have profound effect on the retrieval of LAI. Previous studies (Fang et al., 2003) have fixed these parameters one by one in the course of the retrieval which are mainly based on the knowledge about the field conditions. However, such fixing of the parameters would result in a model that is very specific to a particular land cover type and would loose the generalization capability.

In our first set of experiments we developed SVR-based models with training data consisting of 6 free parameters. This resulted in MSE values of 0.68 (NIR band) and 0.99 (NIR and red

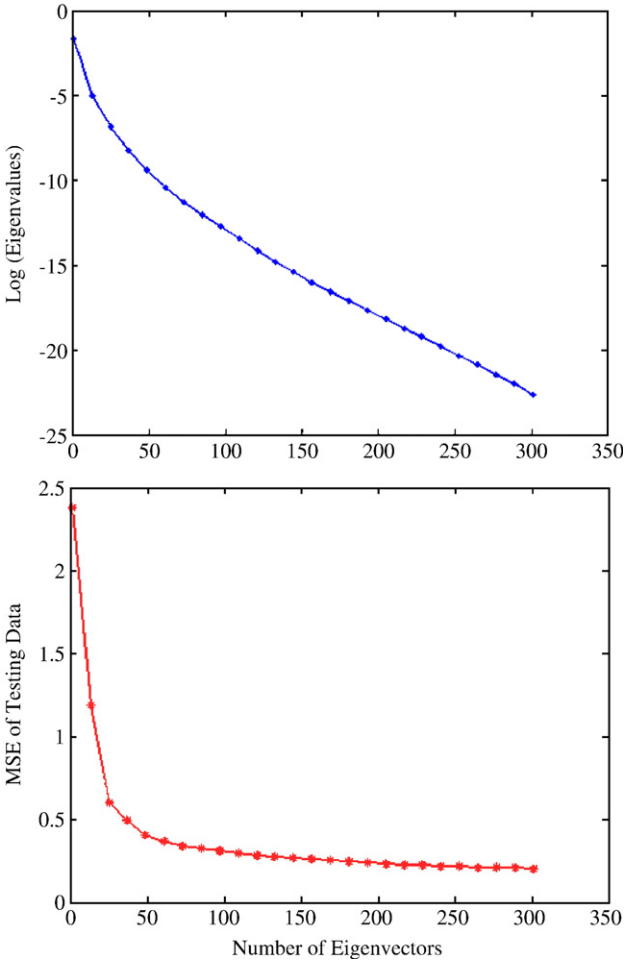

Fig. 4. (a) Eigenvalues corresponding to the number of components. (b) Relationship between MSE of testing data and the number of eigenvectors.

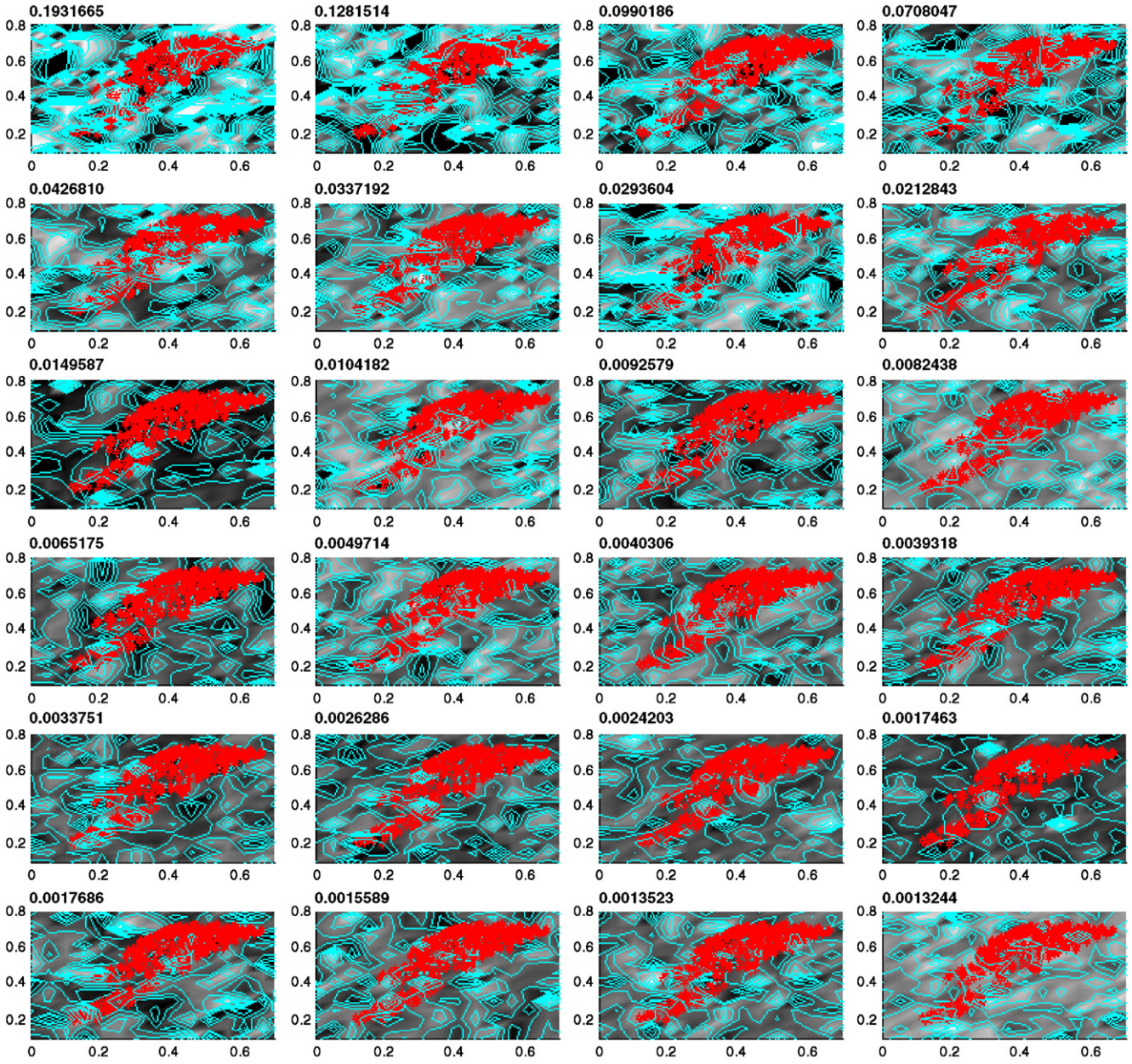

Fig. 5. The first twenty-four eigenvectors are shown in order of decreasing eigenvalue size. The figures contain lines of constant principal component value (contour lines).

band) respectively. The second set of experiments was designed to observe the effect of fixing the parameters. We fixed only the value of the PROSAIL input parameter N to 1.5 (Combal et al., 2002; Jacquemoud et al., 2000) which corresponds to the average value in nature (Bacour et al., 2002b). The MSE error has decreased to 0.20 (NIR) and 0.28 (NIR and red) and R^2 values of 0.96 and 0.94 were achieved respectively.

The models thus generated were applied on the Alpilles validation dataset. An RMSE of 0.72 was achieved using the model generated with only the NIR band data (Fig. 7a) and when N was fixed the RMSE has decreased to 0.50 (Fig. 7b). Similarly the SVR model generated with both NIR and red bands has produced an RMSE of 0.82 (Fig. 8a) which is higher than when only the NIR band was used. The SVR model with both NIR and red bands and N fixed reduced the RMSE to 0.64 (Fig. 8b) at the Alpilles validation site.

In order to test the generalization capability of the SVR model, we also applied the model on MISR data corresponding to May 18, 2001, acquisition at the Alpilles site (a time lag of nearly 2 months from the actual availability of field data at Alpilles site, whose LAI was measured between February 26, 2001 and March 15, 2001). Hence, the SVR retrieval on this date is not strictly comparable to the field LAI. However, it gives a general understanding about the retrieval and generalization capability of the SVR method. An RMSE of 0.51 (Fig. 9) was achieved in this case.

Further, we also applied the model to test the predication capability at a different biome type (temperate mixed forest).

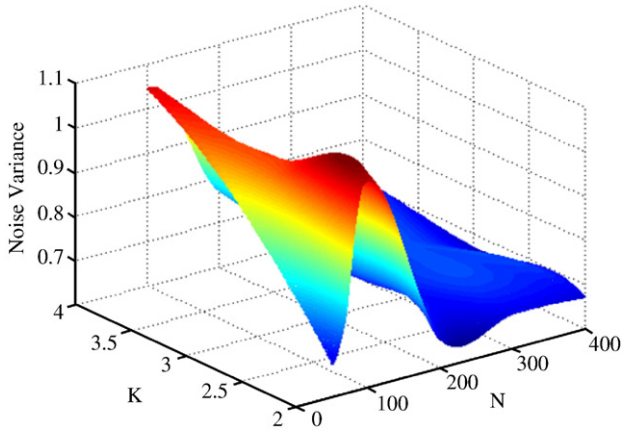

Fig. 6. Using *k*-nearest neighbor method for estimating noise variance for training data with different *k* and *n* values.

MISR data from the Harvard validation site (42.529°N, 72.173°W) corresponding to Oct 20, 2000 acquisition are used in this experiment. In this case, an RMSE of 0.52 (Fig. 10) was obtained. However, the LAI at this site could not be validated with actual field LAI as was done for the Alpilles site (due to lack of field data) and was compared with the LAI retrieved from MISR LAI/FPAR algorithm (Knyazikhin et al., 1998a,b; Hu et al., 2003, Jiannan et al., 2007).

4.2. Comparison of the proposed SVR-based approach with earlier studies

In this section, we attempt to compare the proposed approach with similar studies. We limit the details only to the earlier proposed approaches that address the problem of regularization. Two prevalent approaches for retrieving LAI from radiative transfer model inversions are either based on NN or use the LUT technique (we omit the discussion on LAI retrieval based on traditional numerical optimization methods).

In the NN approach, no prior feature extraction was performed and the raw reflectances from model simulations were used as the input data to the network (Gong et al., 1999; Schlerf & Atzberger, 2006). Other studies used the reflectances from model simulations in conjunction with Normalized Difference Vegetation Index (NDVI), Soil Reflectance Index (SRI) (Fang et al., 2003), object signatures (Atzberger, 2004), and addition of noise (Combal et al., 2002).

Given that the cost of training data acquisition is often noted as a concern in remote sensing (Chi and Bruzzone, 2005; Foody & Mathur, 2006; Tadjudin & Landgrebe, 2000), the ability to use small training sets could be advantageous and attractive feature for many applications. Large database sizes were used in other studies (e.g., Atzberger, 2004 uses 100,000 samples in a NNT approach), Combal et al., 2002 use 280,000 values of each variable and using prior information the size is reduced to 8032, consequently reducing the time spent to sort this LUT to about 65 times shorter. Thus, in the absence of prior information, the LUT approach is memory and disk storage consuming. However, the proposed approach provides a method for feature extraction (KPCA) that has the ability to extract structure from the data and can work on small training set sizes. Also, both the earlier studies report results only on synthetic datasets in contrast to the proposed method in this paper that compares the results with actual field derived LAI values.

Further, to test the ability of the Neural Network (NN) method on the reduced database, the NN architecture was

Table 3
Parameters and accuracy for the testing data

Data set	*C*-value	Value	MSE	Squared correlation coefficient (R^2)	Number of support vectors (Nsv)
NIR	9.7872	0.3548	0.6809	0.86	227
NIR (Fixed *N*)	9.7872	0.1946	0.2003	0.96	231
NIR+RED	9.7872	0.3615	0.9900	0.83	201
NIR+RED (Fixed *N*)	9.7872	0.2454	0.2845	0.94	240

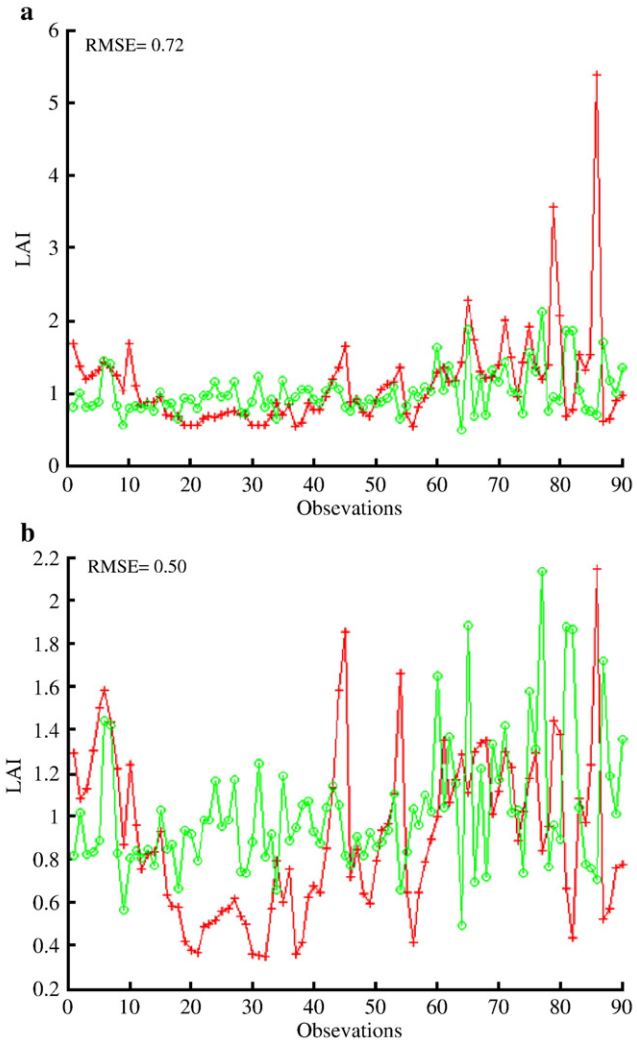

Fig. 7. (a) RMSE of SVR algorithm retrieved LAI values using only NIR band. (b) RMSE of SVR algorithm retrieved LAI values using NIR bands (*N* fixed). (Note: plus (+) sign represents SVR-based retrieval and circle (○) represents actual field derived LAI at Alpilles site. *X*-axis represents observation at each pixel) The MISR data correspond to 03/15/2001 acquisition.

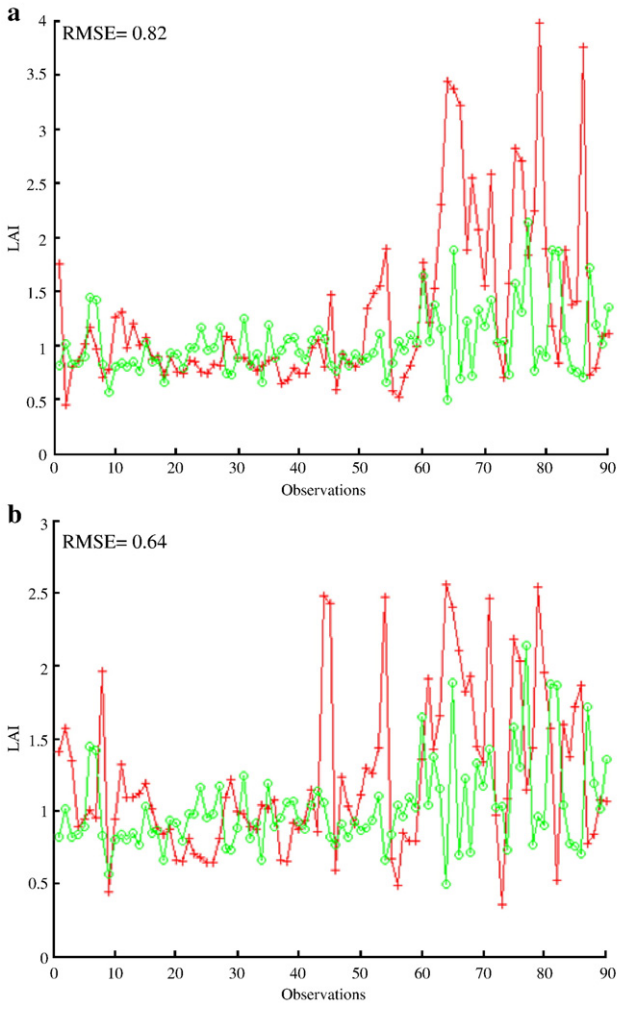

Fig. 8. (a) RMSE of SVR algorithm retrieved LAI values using NIR and red bands. (b) RMSE of SVR algorithm retrieved LAI values using NIR and red bands (*N* fixed). (Note: plus (+) sign represents SVR-based retrieval and circle (○) represents actual field derived LAI at Alpilles site. *X*-axis represents observation at each pixel.) The date of MISR data correspond to 03/15/2001 acquisition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

realized in matlab (www.mathworks.com) (Nørgaard, 2000). A three layer feed-forward network with a tan-sigmoidal transfer function in the hidden layer and a linear transfer function in the output layer was designed. Networks with 2–9 neurons in the hidden layer were evaluated in this study. The Neural Network (NN) was trained with examples obtained directly from the PROSAIL model simulations as was done in the earlier studies. The Levenberg–Marquardt optimization algorithm was implemented to search for the synaptic weights and neuron bias that allows the best fit with the canopy variable (LAI) corresponding to the input BRFs in the training dataset. The following approaches are generally used for regularization:

- Early stopping (requires training, testing and validation data, and is not suitable for small databases),
- Selecting an appropriate weight decay value,
- Jitter (i.e., artificial noise deliberately added to the inputs during training).

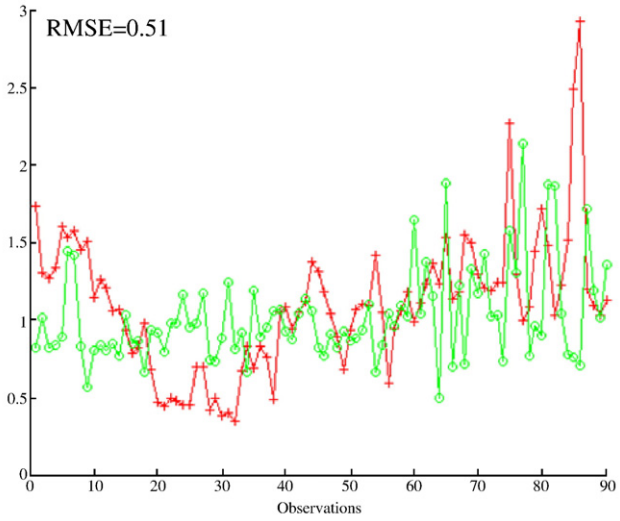

Fig. 9. RMSE of SVR algorithm retrieved LAI values using only NIR band. The MISR data correspond to May 18, 2001 acquisition. This retrieval is strictly not comparable to the actual field data at Alpilles as it is measured between February 26 and March 15, 2001. However, this provides an approximate idea of the generalization capability of the SVR method (Note: plus (+) sign represents SVR-based retrieval and circle (○) represents LAI retrieved by actual field derived LAI at Alpilles site).

In our experiments, we chose to investigate the selection of the weight decay parameter, as the network performed poorly in the absence of its setting. An earlier study (Combal et al., 2002) has investigated the effects of adding jitter to the input datasets to perform regularization.

The generalization ability of the network can depend crucially on the decay constant, especially with small training sets as was used in this experiment. Weight decay is a subset of regularization methods. The penalty term in weight decay, by definition, penalizes large weights. However, the selection of a specific value for weight decay is difficult. One of the often used approaches is to train the network with a range of values and then assess its generalization capacity. The weight decay

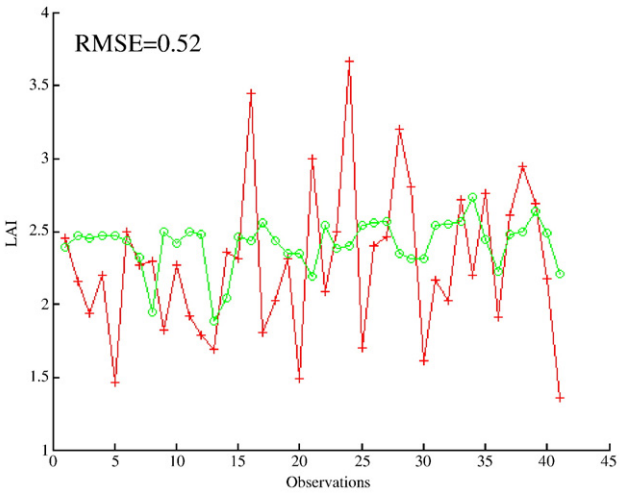

Fig. 10. RMSE of SVR algorithm retrieved LAI values using only NIR bands. (Note: plus (+) sign represents SVR-based retrieval and circle (○) represents LAI retrieved by actual field derived LAI at Alpilles site).

value was varied in the range of 1e-2 to 1e-4. An RMSE of 1.7 was obtained for the combined NIR and Red bands database and an RMSE of 1.55 using only NIR band data. More rigorous studies are required to address the regularization aspects when using NN for canopy radiative transfer model inversion.

5. Conclusion

The solution for ill-posed problems requires robust regularization techniques. A first attempt was carried out to test the usefulness of kernel-based regularization by SVR for the inversion of a 1D canopy radiative transfer model. A procedure for selecting the parameter combinations has been introduced, which considerably reduces the number of simulations of the model, saving computational resources and time. The SVR method provides a very attractive alternative to the current optimization methods used in the inversion, wherein it enables to generate predictive models that have good generalization capabilities and can handle linearly non-separable data with multiple hyper planes, which is difficult to accomplish with comparable methods like artificial neural networks. The SVR method is firmly grounded in statistical theory, and since the SVR is a convex Quadratic Programming (QP) it can always find a global optimum whereas, artificial neural networks are local solutions (e.g., gradient descent search). Also, there is no need for an initial guess, as required for other optimization algorithms like the Powell and Quasi-Newton methods.

Two sets of data (only NIR, and NIR in conjunction with the red band) were investigated. Once the model has been generated, the prediction of LAI on validation data is almost instantaneous. The SVR-based retrieval of LAI provides a reasonably good agreement with the field derived LAI.

The present study could be validated only at one biome type (grasses and cereal crops). More rigorous studies are required to apply this method at several different biome types to test the generalization capability of the SVR method. Currently the SVR method provides the ability to retrieve only one parameter at a time. Multi-output support vector regression is still a very active area of research and we intend to pursue it in the future for simultaneously retrieving multiple parameters from the inversion of a canopy radiative transfer model.

Acknowledgement

The authors thank Knyazikhin Y. and Jiannan H. for making available to us the field derived LAI map (Alpilles) used for comparison in this study.

References

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment*, *93*, 53−67.

Bacour, C., Jacquemoud, S., Tourbier, Y., Dechambre, M., & Frangi, J. -p.(2002a) Design and analysis of numerical experiments to compare four canopy reflectance models. *Remote Sensing of Environment*, *79*, 72−83.

Bacour, C., Jacquemoud, S., Leroy, M., Hautecoeur, O., Weiss, M., Prévot, L., et al. (2002b) Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne polder data. *Agronomie*, *22*, 555−565.

Banerjee, A., Burlina, P., & Diehl, C. (2006). A support vector method for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geosciences and Remote Sensing*, *44*(8), 2282−2291.

Baret, F., Cleavers, J. G. P. W., & Steven, M. D. (1995). The robustness of canopy gap fraction estimates from red and near-infrared reflectances: A comparison of approaches. *Remote Sensing of Environment*, *54*, 141−151.

Baret, F., Weiss, M., Garrigue, S., Allard, D., Leroy, M., Jeanjean, H., et al. (submitted for publication). VALERI: A network of sites and a methodology for the validation of medium spatial resolution satellite products. *Remote Sensing of the Environment*.

Barnsley, M. J., Lewis, P., O'Dwyer, S., Disney, M. I., Hobson, P., Cutter, M., et al. (2000). On the potential of CHRIS/PROBA for estimating vegetation canopy properties from space. *Remote Sensing Reviews*, *00*, 1−19.

Barnsley, M. J., Settle, J. J., Cutter, M. A., Lobb, D. R., & Teston, F. (2004). The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral, multiangle, observation of the Earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1512−1520.

Bicheron, P., & Leroy, M. (1999). A method of biophysical parameter retrieval at global scale by inversion of vegetation reflectance model. *Remote Sensing of the Environment*, *67*, 251−266.

Bradley, P. S., & Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. *Proc. of the 15th International Conference on Machine Learning, Madison, WI, USA* (pp. 82−90).

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*, 121−167.

Burges, C. J. C., & Scholkopf, B. (1997). Improving the accuracy and speed of support vector learning machine. *Advances in neural information processing systems*, *vol. 9*. (pp. 375−381). Cambridge, MA: MIT Press.

Campbell, W. M., Campbell, J. P., Reynolds, D. A., Singer, E., & Torres-Carrasquillo, P. A. (2006). Support vector machines for speaker and language recognition. *Computer and Speech Language*, *20*, 210−229.

Cao, L. (2003). Support vector machines experts for time series forecasting. *Neurocomputing*, *51*, 321−339.

Casa, R., & Jones, H. G. (2005). LAI retrieval from multiangular image classification and inversion of a ray tracing model. *Remote Sensing of Environment*, *98*, 414−428.

Chang, C.-C., Lin, Chih-Jen. 2001. LIBSVM: A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chen, J. M., Liu, J., Leblanc, S. G., Lacaze, R., & Roujean, J. L. (2003). Multiangular optical remote sensing for assessing vegetation structure and carbon absorption. *Remote Sensing of Environment*, *84*, 516−525.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*(2), 113−126.

Cherkassky, E., & Mulier, F. (1998). *Learning from data: Concepts, theory, and methods.* John Wiley and Sons.

Cherkassky, V., Shao, X., Mulier, F., & Vapnik, V. (1999). Model complexity control for regression using VC generalization bounds. *IEEE Transaction on Neural Networks*, *10*(5), 1072−1089.

Chi, M., & Bruzzone, L. (2005). A semilabeled-sample-driven bagging technique for ill-posed classification problems. *IEEE Geoscience and Remote Sensing Letters*, *2*, 69−73.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al. (2002). Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, *84*(1), 1−15.

CROMA. (2000). *Crop reflectance operational models for agriculture. Description of work.* Energy, environment and Sustainable Development work Programme. EF5/PhD/0035.00 (http://www.astrium-space.com/croma).

Deschamps, P. -Y., Breon, F. -M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J. -C., et al. (1994). The POLDER mission: Instruments characteristics and scientific objectives. *IEEE Transactions on Geoscience and Remote Sensing*, *32*, 586−615.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., et al. (1998). Multi-angle imaging spectroradiometer (MISR)—instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1072−1087.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, *9*, 155−161.

Durbha, S. S., & King, R. L. (2005). Semantics enabled framework for knowledge discovery from earth observation data archives. *IEEE Trans. on Geosci. Remote Sensing*, *43*(11), 2563−2572.

Evgeniou, T., Poggio, T., Pontil, M., & Alessandro, V. (2002). Regularization and statistical learning. Theory for data analysis. *Computational Statistics & Data Analysis*, *38*, 421−432.

Evgeniou, T., Pontil, M., & Poggio, T. (1999). A unifed framework for Regularization Networks and Support Vector Machines. A.I. Memo No. 1654, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Evgeniou, T., Pontil, M., & Poggio, T. (2000). Regularization networks and support vector machines. *Advances in Computational Mathematics*, *13*(1), 1−50.

Fang, H., Liang, S., & Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm with a canopy Radiative transfer model. *Remote Sensing of Environment*, *85*(3), 257−270.

Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, *103*, 179−189.

Garabedian, P. (1964). *Partial differential equations.* Wiley.

Girolami, M. (2002). *Orthogonal series density estimation and the kernel eigenvalue problemNeural Computation*, *vol. 14 no. 3*. (pp. 669−688). MIT Press.

Goel, N. S. (1988). Models of vegetation canopy reflectance and their use in the estimation of biophysical parameters from reflectance data. *Remote Sensing Review*, *4*, 1−222.

Goel, N. S., Strebel, D. E., & Thompson, R. L. (1984). Inversion of vegetation canopy reflectance models for estimating agronomic variables II—use of angle transforms and error analysis as illustrated by the Suit's model. *Remote Sensing of Environment*, *14*, 77−111.

Gong, P., Wang, D., & Liang, S. (1999). Inverting a canopy reflectance model using a neural network. *International Journal of Remote Sensing*, *20*(1), 111−122.

Gopal, S., & Woodcock, C. (1996). Remote sensing of forest change using artificial neural networks. *IEEE Transactions on Geosciences and Remote Sensing*, *34*(2), 398−403.

Gunn, S. (1998). Support vector machines for classification and regression. *ISIS technical report, Image Speech and Intelligent Systems Group University of Southampton*.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389−422. doi:10.1023/A: 1012487302797

Haykin, S. (1999). Neural Network—A Comprehensive Foundation (2nd ed.). *Englewood Cliffs* NJ: Prentice Hall.

Hu, J., Tan, B., Shabanov, N., Crean, K. A., Martonchik, J. V., Diner, D. J.et al. (2003). Performance of the MISR LAI and FPAR algorithm: A case study in Africa. *Remote Sensing of Environment*, *88*, 324−340.

Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J. -P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance—direct and inverse mode. *Remote Sensing of Environment*, *74*, 471−481.

Jacquemoud, S., & Baret, F. (1993). Crop structure and microclimate: Characterizations and applications. *Estimating vegetation biophysical parameters by inversion of a reflectance model on high spectral data, INRA Edition, Paris, France* (pp. 339−350)..

Jenkins, B., 2005. Jenny: An alternative to exhaustive testing. (www.burtleburtle.net/bob/math/jenny.html).

Jensen, J., Qiu, F., & Jim, M. (1999). Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *International Journal of Remote Sensing*, *20*(14), 2805−2822.

Jiannan, H., Su, Y., Tan, B., Huang, D., Yang, W., Bull, M. A., et al. (2007). Analysis of the MISR LAI/FPAR product for spatial and temporal coverage, accuracy and consistency. *Remote Sensing of Environment*, *107*(1), 334−347.

Jin, Y., & Liu, C. (1997). Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks. *International Journal of Remote Sensing*, *18*(4), 971−979.

Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgaism, A. A., & Gao, F. (2000). Inversion methods for physically-based models. *Remote Sensing Reviews*, *18*, 381−440.

Kimes, D., Ranson, K., & Sun, G. (1997). Inversion of a forest backscatter model using neural networks. *International Journal of Remote Sensing*, *18* (10), 2181−2199.

Knyazikhin, Y., Martonchik, J., Diner, D., Myneni, R., Verstrate, M., Pinty, B., et al. (1998a) Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. *Journal of Geophysical Research*, *103*(D24), 32239−32256.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., & Running, S. (1998b) Synergistic algorithm for estimating vegetation canopy leaf are index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, *103*(D24), 32257−32276.

Kuusk, A. (1991). The inversion of the Nilson-Kuusk canopy reflectance model, a test case. *Proceedings of the International Geosciences and Remote Sensing Symposium (IGARSS'91)* (pp. 1547−1550).

Kwok, J. T., & Ivor, W. T. (2003). Linear dependency between ε and the input noise in ε-support vector regression. *IEEE Transaction on Neural Networks*, *14*(3), 544−553.

Liang, S., Radiative transfer modeling of coupled atmosphere and plant canopy and BRDF retrieval. *PhD thesis*. Boston University, Massachusetts, USA.

Mantero, P., Moser, G., & Serpico, S. B. (2005). Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), 559−570.

Mathur, A., & Foody, G. M. (2004). Land Cover classification by support vector machine: Towards efficient training. *IEEE Geoscience and Remote Sensing Symposium*, 742−744.

Mattera, D., & Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In B. Schölkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in Kernel Methods: Support Vector Learning* (pp. 211−242). Cambridge, MA: MIT Press.

Mazzoni, D., Horvath, A., Garay, J., Tang, B., & Davies, R. (2000). A MISR cloud-type classifier using reduced support vector machines. *Proceedings of the 2000 IEEE Geosciences and Remote Sensing. Symposium (IGRASS 2000)* http://wwwaig.jpl.nasa.gov/public/mls/papers/mazzoni/mazzoni-msd05-misr.pdf

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(8), 1778−1790.

Muller, K. -R., Smola, A., Ratsch, G., Scholkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. In W. Gerstner, A. Germond, M. Hasler, & J. -D. Nicoud (Eds.), *Lectures Notes in Computer Science: Vol. 1327. Artificial Neural Networks ICANN'97* (pp. 999−1004). Berlin: Springer.

Myneni, R. B., Maggion, S., Iaquinto, J., Privette, J. L., Goboron, N., Pinty, B., et al. (1995). Optical remote sensing of vegetation-modeling, caveats, and algorithms. *Remote Sensing of Environment*, *51*, 169−188.

Nørgaard, M., 2000. "Neural network based system identification toolbox" Tech Report. 00-E-891. Department of Automation, Technical University of Denmark.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: An application to face detection. *IEEE conference on computer vision and pattern* (pp. 130−136).

Pierce, L., Sarabandi, K., & Ulaby, F. (1994). Application of an artificial neural network in canopy scattering inversion. *International Journal of Remote Sensing*, *15*(16), 3263−3270.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*(26), 314−319.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1986). *Numerical recipes.* New York: Cambridge University Press.

Privette, J. L., Emery, W. J., & Schimel, D. S. (1996). Inversion of a vegetation reflectance model with NOAA AVHRR data. *Remote Sensing of Environment*, *58*(2), 187−200.

Rosipal, R., Girolami, M., Trejo, L., & Cichoki, A. (2001). Kernel PCA for feature extraction and de-noising in non-linear regression. *Neural Computing and Applications*, *10*(3), 231−243.

Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, *100*, 281−294.

Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*, 1299−1319.

Schölkopf, B., Smola, A., & Müller, K. R. (1999). Kernel principal component analysis. In B. Schölkopf, C. Burges, & A. Smola (Eds.), *Advances in kernel methods—support vector learning* (pp. 327−352). Cambridge, MA: MIT.

Smith, J. (1993). LAI inversion using a back-propagation neural network trained with a multiple scattering model. *IEEE Trans on Geoscience and Remote Sensing*, *31*(5), 1102−1107.

Smola, A., Murata, N., Schölkopf, B., & Muller, K. (1998). Asymptotically optimal choice of –loss for support vector machines. *Proceedings of the International Conference on Artificial Neural Networks,* (pp. 105−110). Berlin: Springer. http://citeseer.ist.psu.edu/26614.html

Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 199−222, 0.1023/B:STCO.0000035301.49549.88.

Tadjudin, S., & Landgrebe, D. A. (2000). Robust parameter estimation for mixture model. *IEEE Transactions on Geoscience and Remote Sensing*, *38*, 439−445.

Tan, B., Hu, J., Zhang, P., Huang, D., Shabanov, N. V., Weiss, M., et al. (2005). Validation of MODIS LAI product in croplands of Alpilles, France. *J. Geophys. Res.*, *110*. doi:10.1029/2004JD004 860

Tay, F. E. H., & Cao, L. J. (2001). A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines. *Intelligent Data Analysis*, *5*, 191−209.

Tian, Y., Wang, Y., Zhang, Y., Knyazikhin, Y., Bogaert, J., & Myneni, R. B. (2002). Radiative transfer based scaling of LAI/FPAR retrievals from reflectance data of different resolutions. *Remote Sensing of Environment*, *84*, 143−159.

Trafalis, T. B., & Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. Neural Network. IJCNN 2000. *Proceedings of the IEEE-INNS-ENNS Intemational Joint Conference*, *6*, 348−353.

Türkan, M., Dülek, B., Onaran, I., & Çetin, A. E. (2006). Human face detection in video using edge projections. *Proceedings of SPIE-6246*. doi:10.1117/12.666704

Vapnik, V. (1995). *The Nature of Statistical Learning Theory.* New York: Springer-Verlag.

Vapnik, V. (1998). *Statistical Learning Theory.* New York: Wiley.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, *16*, 125−141.

Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M., & Morcrette, J. J. (1997). Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, *35*, 675−685.

Weiss, M., & Baret, F. (1999). Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment*, *70*, 293−306.

Weiss, M., Baret, F., Myneni, R. B., Pragnère, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional data. *Agronomie*, *20*, 3−22.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. (2001). Feature selection for SVMs. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 668−674). Cambridge, MA: MIT Press.

Wohlberg, B., Tartakovsky, D. M., & Guadagnini, A. (2006). Subsurface characterization with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(1), 47−57.