Classification of Contamination in Salt Marsh Plants Using Hyperspectral Reflectance

Machelle D. Wilson, Susan L. Ustin, *Member, IEEE*, and David M. Rocke

***Abstract*—In this paper, we compare the classification effectiveness of two relatively new techniques on data consisting of leaf-level reflectance from five species of salt marsh and two species of crop plants (in four experiments) that have been exposed to varying levels of different heavy metal or petroleum toxicity, with a control treatment for each experiment. If these methodologies work well on leaf-level data, then there is hope that they will also work well on data from air- and spaceborne platforms. The classification methods compared were support vector classification (SVC) of exposed and nonexposed plants based on the spectral reflectance data, and partial least squares compression of the spectral reflectance data followed by classification using logistic discrimination (PLS/LD). The statistic we used to compare the effectiveness of the methodologies was the leave-one-out cross-validation estimate of the prediction error. Our results suggest that both techniques perform reasonably well, but that SVC was superior to PLS/LD for use on hyperspectral data and it is worth exploring as a technique for classifying heavy-metal or petroleum exposed plants for the more complicated data from air- and spaceborne sensors.**

***Index Terms*—Heavy metals, hyperspectral, logistic discrimination (LD), partial least squares (PLS), petroleum, reflectance, remote sensing, support vector machines (SVMs).**

I. Introduction

NEW REMOTE sensing technologies that provide high spatial and spectral resolution are technologically mature and could address critical environmental monitoring issues, but the ability to analyze and interpret the data lags behind the technology. When robust analytical methods are developed, they can be applied to a wide range of environmental problems for which remote detection is the best method, e.g., inaccessible areas, sites with potentially hazardous contamination, sites that need routine and frequent monitoring, and where information about the spatial context of such conditions is critical to understanding the location, distribution, or spread of adverse conditions. These methods will, in turn, have a large impact on the policy and economics of environmental monitoring for such industries as petroleum, chemical, waste management, transportation, as well as the military.

Manuscript received February 12, 2003; revised December 2, 2003. This work was supported in part by Chevron and the University of California, in part by the National Science Foundation under Contracts ACI 96-19020 and DMS 98-70172, in part by the National Institute of Environmental Health Sciences, National Institutes of Health under Contract P43 ES04699, and in part under Contract DE-FC09-96SR18546 between the U.S. Department of Energy and the University of Georgia's Savannah River Ecology Laboratory.

M. D. Wilson is with the Savannah River Ecology Laboratory, of Georgia, Aiken, SC 29802 USA (e-mail: wilson@srel.edu).

S. L. Ustin and D. M. Rocke are with the University of California, Davis, CA 95616 USA.

Digital Object Identifier 10.1109/TGRS.2003.823278

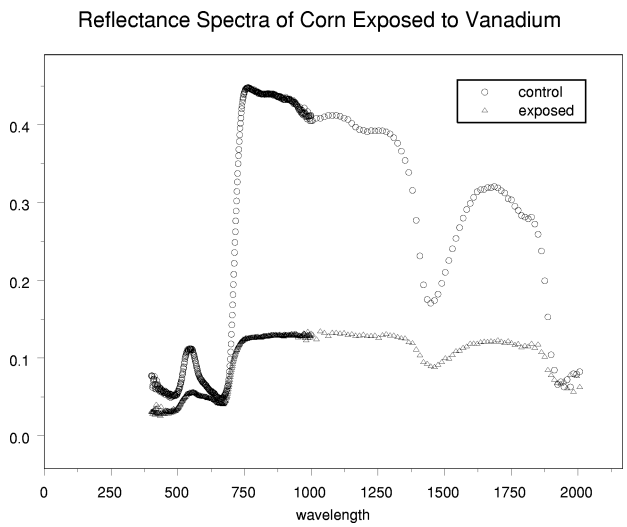

Fig. 1. Reflectance spectra for (circles) a contaminated and (triangles) uncontaminated leaf. Wavelength is given in nanometers.

There has been much research in remote sensing at the leaf level to ascertain the level of stress in plants based on physiological changes and how these changes alter the interaction of light with the foliar medium [1], [2], [13]. The most common and widespread change occurs in the proportion of light-absorbing pigments, most notably chlorophylls a and b, which absorb light in the 430–660-nm region [12]. Investigators have observed differences in reflectance due to stress-induced differences in pigment concentration in the *green peak* and along the *red edge* [7], [16], [25], [31]. Additionally, heavy metal stress is typically manifest as cell membrane damage, alteration of enzyme activities, changes in water content and leaf temperature, and interference with respiratory gas exchange, as well as with the photosynthetic apparatus [3], [14], [15]. This paper seeks to determine if such responses to stress can be detected in reflectance at the leaf level, and it explores the use of two important statistical techniques.

Reflectance spectra can be seen to change under heavy metal stress. Fig. 1 shows an example of this shift. Notice the depression in reflectance in the exposed plant. The fact that differences are detectable by the human eye lends confidence that the algorithms tested here will be able to use the reflectance data to distinguish exposed from unexposed plants.

Interpretation of remotely sensed data for monitoring purposes can be seen as a decision or classification problem, i.e., how can we decide at a given level of certainty that these data have been generated by plants or ecosystems that are experiencing anthropogenic stress? Traditional methods of classification usually rely on assumptions of normality or asymptotic nor-

mality that cannot always be validated. Hence, nonparametric, small-sample methods are needed in order for investigators to make reliable decision and classification rules based on a an available dataset, often called the *training set*. Additionally, the large dimensionality of the data often poses difficulty. Furthermore, remote sensing data present issues in calibration: what is the relationship between changes in reflectance and changes in plant health and the degree of plant exposure to a toxin?

In this study, we grew the salt marsh plants *Frankenia grandifolia*, *Salicornia virginica*, and *Scirpus robustus* and two crop species, *Zea mays* (corn) and *Phaseolus vulgaris* (kidney beans), under controlled conditions at varying levels of heavy metal stress and *Spartina alterniflora*, *Spartina foliosa* under varying levels of petroleum contamination, then gathered data at the leaf level to test the efficacy of two relatively new classification methods, and to ascertain the mathematical model most appropriate for prediction (i.e., a calibration relationship).

A. Statistical Learning and Multivariate Calibration

There are many statistical data analysis techniques that fall into the category of statistical learning. The two classification techniques considered here, logistic discrimination and support vector machines (SVMs), are among them. The most commonly used statistical learning method is linear regression. We refer to it as "learning" because we have data that we want to use to discover the relationship between a quantity that we would like to predict and one or several other quantities, called predictors. That is, we would like to "learn from examples." The data we have consists of measurements of both the response (the quantity we would like to predict in the future), along with corresponding measurements of the predictors, in what is called the "training set." We then use our data to "train" a statistical algorithm, which produces a mathematical relationship via parameter estimation, from the known data, which we can then use to predict the quantity of interest when all we have measured in the future are the predictors. In this sense, the machine or algorithm, and presumably we ourselves have "learned" from our training data to predict future observations. Here and below boldface characters represent matrices.

In the binary classification scheme, in our case exposed or normal, we have a set of l examples, i.e., observations and their labels (or response), with p variables

$$(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), \ldots, (\mathbf{x_l}, y_l) \in \mathbf{R}^p \times \{1, -1\}. \quad (1)$$

In the general case, i.e., classification *or* regression, let $\mathbf{X}$ be the matrix of all observations and $\mathbf{Y}$ be the corresponding labels (or response). Let α be a parameter vector indexing a function space f. Then, statistical learning involves finding the best function $f_{\boldsymbol{\alpha}}(\mathbf{X})$, out of the set $\{f_\alpha \colon \alpha \in \Lambda\}$ (called the *hypothesis space*), where Λ is the set of all possible parameters for f, such that, with high probability, $f_{\boldsymbol{\alpha}}(\mathbf{X})$ is close to $\mathbf{Y}$ for all $\mathbf{X}$ and the corresponding $\mathbf{Y}$. By "best" we mean a function that minimizes the prediction error. Methods of estimating the prediction error include leave-one-out cross validation, separating the data into a training set and a test set, which is usually not performed unless there are a lot of data, and model-based estimates such as the empirical risk, which estimates the quality of the decision rule [8], [30]. For both methods we use leave-one-out cross validation, which involves removing one data point, fitting the model and using the left-out point as a test point to see how well the classification technique performs on a given observation that has not been included during training. This is done for each observation in the dataset. The number of misclassified test points plus the number of misclassified training points divided by twice the sample size is an estimate of the prediction error. This technique will be discussed in more detail in Section II [10], [23].

Another advantage of both techniques is that neither requires us to make any questionable distributional assumptions. Support vector machines require only the assumption that the observations are independent and identically distributed. Partial least squares is similar to least squares in that we are merely attempting to fit the best linear equation of the data and we are not implicitly making any distributional assumptions. In logistic discrimination, we are making only the assumption that the response is binomial, which can hardly be questioned given that it is binary.

Inherent in the problem of interpreting (and learning from) remotely sensed data is the issue of calibration. In simple calibration, the investigator performs what is basically an univariate linear regression to establish the relationship between the response of the measurement technology and the quantity of interest, for example, the linear relationship between absorbance at a given wavelength and the concentration of an analyte in a sample. However, if the quantity being measured is produced by complex processes, or if there is a high degree of error in the measurement process, we often require the information contained in several different measurements. For example, we may require measurements of absorbance at several wavelengths to determine the concentration of an analyte in a sample. Multivariate calibration, most notably principal components and partial least squares, has been used quite successfully for some time in analytical chemistry [21]. Recently, a combination of partial least squares (PLS) data compression and logistic discrimination as a classification technique has been used for analysis of DNA arrays [24]. We test the same technique below on remotely sensed data. There are many parallels between the issues and problems in analytical chemistry and those of remote sensing. Reflectance and aborbance data in both disciplines are often high dimensional, they presumably have similar error structures, and interest in how changes in chemical composition affect the interaction of light with a substance plays a large role in both fields.

Support vector machines have been attempted in a few remote sensing applications, for problems such as land cover classification [19], detection of soil contamination and discrimination of healthy and diseased plants [17], [18], [22]. Partial least squares regression has been applied to soil spectroscopy for use in classifying contaminated areas in river floodplains [20] with some success. However, the use of support vector machines on hyperspectral data is still relatively untested, and we are not aware of any other investigators who have used PLS/logistic discrimination (LD) or support vector classification (SVC) on hyperspectral data for the detection of heavy metal or petroleum contamination in plants.

B. PLS

As per the above discussion, one good candidate for the calibration of remotely sensed data is partial least squares. PLS is a method of data dimension reduction, similar to principal components, to find the most relevant factors for both prediction and interpretation, and is derived from Herman Wold's development of iterative fitting of bilinear models [32], [33]. Once the data have been compressed, analysis techniques that require the number of predictor variables to be smaller than the number of observations, such as least squares regression, can then be used. Partial least squares regression (PLSR) improves upon principal components analysis by actively using the response variables during the bilinear decomposition of the predictors. By balancing the information in both the predictors and the response, PLS reduces the impact of large, but irrelevant predictor variations. Estimation of prediction error is achieved using cross validation.

C. Classification Using Logistic Discrimination

For this study, we do not wish to perform a regression, but we wish to classify our data as either exposed or unexposed. Once we have used PLS to compress our data with the optimal compression matrix, we can use more classical techniques that require $l \geq p$. One such classification technique is logistic discrimination.

Let $\mathbf{x}$ be the column vector of p reflectance measurements (or their components) observed on a plant and $y = \{0, 1\}$, where a 0 value indicates that the plant was not experiencing any stress, and a 1 value indicates that the plant was. In logistic regression, the conditional class probability, $\pi = \text{Prob}(y = 1 \,|\, \mathbf{x}) = \text{Prob}$(the plant has been exposed given the reflectance $\mathbf{x}$), is modeled by

$$\pi = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)}. \tag{2}$$

The estimated response probabilities, $\hat{\pi}$ are obtained by replacing the parameter β with its maximum-likelihood estimate (MLE) $\hat{\beta}$. The predicted class of each sample (as either exposed or not) is $\hat{y} = I(\hat{\pi} \geq 1 - \hat{\pi})$, where $I(\cdot)$ is the indicator function. That is, $I(A) = 1$ if A is true and zero otherwise. Note that if $\pi = \text{Prob}$ (plant has been exposed given the reflectance $\mathbf{x}$), then $1 - \pi = \text{Prob}(y = 0 \,|\, \mathbf{x}) = \text{Prob}$(plant has not been exposed given $\mathbf{x}$). So, we classify a sample as exposed $(\hat{y} = 1)$ if the estimated conditional probability that it is exposed is greater than the estimated conditional probability that it is not. In this case, we are using the components of the reflectance data, rather than the reflectance data itself.

To estimate the prediction error of this classification method, we use leave-one-out cross validation. For PLS/LD, each data point was left out sequentially, the remaining data were used to train the full PLS/LD procedure, and the left out point fed back into the procedure to obtain the predicted class and any errors noted.

D. SVMs

Support vector classification is based on a particular type of statistical learning machine, with supporting theory well developed by Vapnik and others [4]–[6], [11], [28], [29]. Support vector classification requires no assumptions on the distribution of the underlying population, other than that the data are independent and identically distributed. Furthermore, support vector machines exploit theorems bounding the actual risk in terms of the empirical risk [10], [29], rather than estimating error using asymptotic convergence to normality. Hence, an upper bound on the prediction error of the decision rule can be calculated even for small samples, while making no distributional assumptions. The optimal machine seeks a balance between consistency in the training set and generalization to future datasets. That is, the optimal machine uses the information in the data to predict future observations, while avoiding overfitting. Additionally, support vector machines allow us to avoid the degradation of both computational and generalization performance that often occurs in high dimensions [10], [29]. For an excellent and more thorough introduction to SVMs, see [26]. Because of these important qualities, support vector classification is a good candidate for analyzing spatial and spectral data, and could provide a solution to many of the problems encountered when trying to use these data in environmental monitoring.

Support vector machines find the optimal separating hyperplane in R^p, where p is the number of predictors and "optimal" is defined as the hyperplane which minimizes the prediction error, i.e., support vector machines find a linear decision rule such that we classify the plant as exposed if it lies on one side of the hyperplane and as unexposed if it lies on the other. In this paper, we use soft margin hyperplanes and cross validation to find the optimal decision rule. The decision rule, or the equation for the separating hyperplane, is derived from a process called *structural risk minimization*, as well as standard optimization theory, and finds a balance between correct classification of the training set and predictive ability. Any given hyperplane can be written

$$\{\mathbf{x} \in \mathbf{F}\text{: } (\mathbf{w} \cdot \mathbf{x}) + b = 0\} \tag{3}$$

for some $\mathbf{w} \in \mathbf{F}$ and $b \in \mathbf{R}^1$. Then, for $\alpha = (\mathbf{w}, b)$, if f_α is a separating hyperplane, then the decision rule is: classify $\mathbf{x}$ as exposed if $f_\alpha(\mathbf{x}) \geq 0$ and as not exposed if $f_\alpha(\mathbf{x}) \leq 0$. The vector $\mathbf{w}$ is a vector of weights that is optimized during training. The decision rule then has the form

$$f_{\mathbf{w},b} = \text{sgn}((\mathbf{w} \cdot \mathbf{x}) + b). \tag{4}$$

However, if the points in our dataset lie very close to the separating hyperplane, we may expect that there will be a good deal of error in our prediction of future data. One measure of predictability (both intuitively and mathematically, though the mathematics will not be shown here) is the size of the margin, defined as the distance between the separating hyperplane and the nearest data point. Finding the best balance between predictability and correct classification of the training set amounts to find the separating hyperplane with the widest margin, while still correctly classifying as much of the training set as possible. The data points that lie on, and define, the margin are called the *support vectors*. In the results shown below, the parameter C is inversely related to the size of the margin. That is, a large C implies a small margin, and vice versa.

1) Support Vectors in Nonlinear Feature Spaces: Often, the data are not linearly separable in the input space (i.e., the untransformed data cannot be separated by hyperplanes in their original dimension), or we may know *a priori* that most of the information is contained in functions of the data, rather than the data themselves. When this is the case, we may need or prefer to work with nonlinear functions of the data which map the data into high-dimensional feature spaces where they are linearly separable, using the map ϕ: $\mathbf{x_i} \rightarrow \mathbf{z_i}$. However, the chance of overfitting and the computational effort grow quite rapidly for certain nonlinear transformations. The mathematics of feature spaces allows us to control the growth of computations, while still gaining an advantage that the increased richness of our hypothesis space offers in terms of classifying the training set. The key is to work exclusively within a dot product space and to find a nonlinear kernel representation that allows us to compute the dot product in the feature space without explicitly mapping into it. A *kernel* is a function K, such that for all $\mathbf{x}, \mathbf{z} \in \mathbf{X}$

$$K(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{z}). \tag{5}$$

If we can find such a kernel, we can maximize the target function and evaluate the decision function in terms of the kernel without calculating the values in the feature space $\phi(\mathbf{x}), \phi(\mathbf{z})$. Since we are not explicitly mapping the data into the new space, we do not have to perform anymore calculations that we would have in the original space, once we have computed the kernel. This yields a decision function of the form

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \lambda_i \cdot k(\mathbf{x}, \mathbf{x_i}) + b\right) \tag{6}$$

for some kernel k where the λ_i are Lagrange multipliers that are found during optimization. Notice the similarity to the linear decision rule. Hence, all of the developed arguments for the linear machine also apply to nonlinear machines by using a valid kernel K instead of the Euclidean dot product. One important aspect of using kernels is that we do not need to know the underlying feature map in order to be able to learn in the feature space. If we can identify that $K(\mathbf{x}, \mathbf{z})$ is a kernel for some map ϕ and in some dot-product space $\mathbf{F}$, we do not need to actually identify the ϕ that produced it. For more information about kernels, see [10] and [26].

This result also holds in general for polynomials of degree d with n variables. Define ϕ as the map which takes $\mathbf{X} \ni \mathbf{x} \rightarrow \mathbf{z} \in \mathbf{F}$ where the entries of the vector $\mathbf{z}$ are all possible dth degree products of the entries of $\mathbf{x}$. Then, the corresponding kernel that computes the dot product in $\mathbf{F}$ is

$$K(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^d. \tag{7}$$

See [10] and [20].

The use of kernels allows us to find hyperplanes in high-dimensional spaces, without the explosion of computational effort normally associated with such spaces. Additionally, for classifiers such as the radial bias function: $k(\mathbf{x}, \mathbf{x_i}) = \exp(-\|\mathbf{x} - \mathbf{x_i}\|^2/c)$ and the neural network kernel: $k(\mathbf{x}, \mathbf{x_i}) = \tanh(\kappa \cdot (\mathbf{x} \cdot \mathbf{x_i}) + \Theta)$ we know that, since these functions satisfy the definition of a kernel, there is a corresponding dot product space where our theory based on linear functions holds and, therefore, that the problem has a solution.

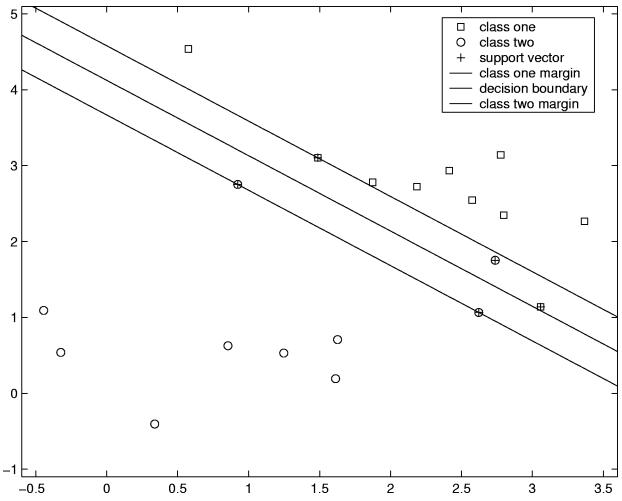

Fig. 2. Two-dimensional example of an optimal decision boundary using the full dataset.

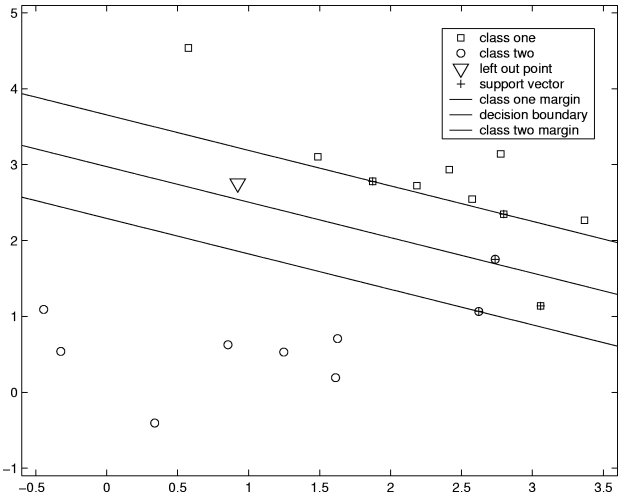

Fig. 3. Two-dimensional example with class one support vector left out.

2) Leave-One-Out Cross Validation: A tighter bound on the prediction error than can be found theoretically is an estimate provided by leave-one-out cross validation. Traditionally, leave-one-out cross validation involves removing one data point, fitting the model, and using the left-out point as a test point to see how well the classification technique performs on a given observation that has not been included during training. This is done for each observation in the dataset. The number of misclassified test points plus the number of misclassified training points divided by twice the sample size is an estimate of the prediction error. When using SVM, only the support vectors are actually relevant in the classification. Only when the left-out point is a support vector will the classification of the remaining data change. Since whatever errors (or not) were made on the training set will remain exactly the same until one of the support vectors is removed as a test point, when performing a leave-one-out cross validation, one only needs to cross validate the support vectors. See Figs. 2 and 3 for an example of this procedure for two-dimensional data.

Fig. 2 shows the full dataset. There is one classification error (one circle on the wrong side of the decision boundary) and one margin error (one square on the correct side of the decision boundary but within the margin). The support vectors (which include the errors) are the data with $+$'s inside the square or circle.

Fig. 3 shows the change in the decision boundary when the class one support vector is removed. Notice the widening of the margin.

II. Materials and Methods

To test and compare these techniques, we obtained leaf-level reflectance data on several salt marsh species and two crop plants. The plants were grown for six weeks in 1X Hoaglands solution before the addition of the metal salt directly into the hydroponic container and fresh Hoaglands. Plants were grown under metal stress for six more weeks before the final spectral reflectance data were taken. Water that evapotranspirated from the containers was replaced with distilled water. We grew the two species of *Spartina* together under varying levels of petroleum contamination, the two crop species together, and the *Salicornia* and *Scirpus* together under varying levels of different heavy metal stress. This was done to introduce some of the variability that would be found in the field, but under well-known conditions.

A. Plant Species and Growing Methods

Frankenia grandifolia plants were grown from rhizomes collected at Suisun Marsh. The rhizomes were covered in vermiculite and watered daily until sprouts were large enough and with sufficient new root mass to transplant into hydroponic containers. Sprouts that erupted from the same rhizome were placed in different treatments to avoid violations of the assumption of independence. The *Frankenia* treatments were copper at 0, 1.56, 7.79, and 15.6 ppm in the form of $CuSO_4$. There were ten replicates for each treatment. Some replicates were lost upon data collection, resulting in a total of 34 samples for the analysis.

Zea mays (corn) and *Phaseolus vulgaris* (kidney beans) were grown in several levels of Vanadium contamination. These crops plants are vigorously growing and genetically identical within a species, with large leaves. The corn and kidney beans were grown from seed by germinating in potting soil for four weeks before being transplanted into the hydroponic containers. The crop species treatments were vanadium at 0, 5, 10, 20, and 50 ppm in the form of $VOSO_4$. There were ten replicates for each of the crop plant species, but some were lost during the growing period, resulting in a total sample size of 83 plants spread among five treatments.

Two species of *Spartina*, *alternaflora*, and *foliosa* were grown together in treatments consisting of varying levels of petroleum at 0 ppm with 100 replicates, 60, 80, 100, and 120 ppm with 30 replicates each, in the form Alba crude, a class A crude, which is a light volatile oil that penetrates soil. Eleven plants died during the experiment, yielding a total of 209 replicates.

Salicornia and *Scirpus* were grown together in treatments consisting of varying levels of cadmium in the form of $CdSO_4$, 0 ppm with 45 replicates, 20, 40, 60 ppm with 20 replicates each, and 80 ppm with 30 replicates for a total of 135 plants.

B. Data Collection

Every week subsequent to the introduction of the metal salt, leaf-level measurements of chlorophyll fluorescence were taken on dark adapted excised leaves to determine the level of stress using the Waltz PAM-2000 fluorimeter. At the end of the growing period, reflectance was measured at a high number (445 for *Frankenia* and crop species and 950 for the others) of different wavelengths ranging from 400.95–1506.84 nm using a GER 1500 and a Licor Integrating Sphere. The *Frankenia*, *Spartina*, *Salicornia*, and *Scirpus* leaves were smaller than the sample port in the Licor Sphere. Hence, we devised an adaptor to hold the leaf and prevent excess light from entering the sphere from around the sides of the leaf. The adaptor consisted of a cardboard square covered with optically black material, with a disc removed whose diameter was no larger than the largest diameter of the leaf. The effect of the device was compensated for by subtracting out its reflectance. That is, normally reflectance is calculated by dividing the radiance of the target (leaf) by the radiance of the reference panel. To correct for the effect of the adaptor, we measured the radiance of the adaptor with no leaf, and we measured the radiance of the reference panel through the adaptor. Then, we subtracted the radiance of the adaptor from the radiance of the leaf, and we divided this quantity by the radiance of the reference panel minus the radiance of the adaptor. This device produced spectra as expected for healthy vegetation when tested on such, except that there was a good deal more noise. The crop plant leaves did not require the use of an adaptor.

For the metal experiments, the leaves were harvested at the end of the treatment period, and the metal concentrations in the leaves were analyzed for each individual plant in order to assess plant uptake of the metal.

C. Data Analysis

For the all datasets other than the *Frankenia*, we analyzed both species together as independent observations. This was done in order to keep the sample size high and to introduce some variability that would normally be present in the field.

The PLS analyses and LD classifications were performed using the software package SAS (SAS Institute, Cary, NC: http://www.sas.com/) using a macro to perform the leave-one-out cross validation on the full procedure. SAS chooses the number of components to include in the model by choosing the model with the smallest predicted residual sums of squares (PRESS). However, this model can be only negligibly better than a smaller model. SAS also performs van der Voet's test [27]. This test chooses the model with the fewest components whose PRESS statistic is insignificantly larger than that of the model with minimum PRESS. We applied van der Voet's test in choosing the number of components to be used during compression. We used continuous data (the response was the concentration) during the PLS procedure to compress the data, and then we used binary data during LD procedure to classify the data as exposed or unexposed.

The SVM classifications were performed in Matlab using a toolbox written by Gavin Cawley [9], which was adapted to perform the cross validation on our data. The toolbox employs soft-margin hyperplanes. The parameter C was chosen from values between 0.01 and 1000 such that the prediction error was minimized, using a binary search algorithm. For purposes of classification, the data were divided into two groups—exposed and unexposed, i.e., all treatments except the control were labeled "exposed." The results could be inspected to determine if the classification scheme had difficulty classifying the lowest concentrations. Additionally, since plants exposed to a toxin in the field will not all experience that toxin at the same level, this experimental design adds some realistic variability to the experiment. We tested a linear kernel and polynomial kernels of degree 2 and 3.

III. Results

The chemical analyses of leaf metal content showed that uptake occurred in all treatments for the copper experiment and in all treatment excepting the control for vanadium and cadmium. Heavy metal concentrations were below the detection limit for the controls, except for copper, which is an essential nutrient. Metal concentrations were significantly higher in the treatment groups for all treatments measured $(p < 0.05)$.

The chlorophyll fluorescence measurements showed significant $(p < 0.05)$ increases in stress between the control and the treatments for all experiments by the end of the six weeks of growth.

A. PLS Classification

Plots of the components' weightings (not shown) for all datasets showed an almost equal weighting of all wavelengths for the first component and a contrasting of high and low wavelengths for the second and third components, with contrasting occurring when higher numbers of components were included as well.

1) Frankenia: The leave-one-out cross validation performed by SAS to choose the model that minimized the PRESS statistic included five PLS components. The PRESS statistic was 0.7863. However, applying van der Voet's test, we chose a model with three components and a PRESS statistic of 0.904 03 to compress the data.

The leave-one-out cross-validation test error for the *Frankenia* experiment was 8.82% and a training error of about 15%, yielding an estimate of the prediction error of 11.91%. Most of the mistakes were made on either the zero or lowest concentration, suggesting that better results may be possible for datasets with a higher minimum concentration above the control.

2) Crop Species: During compression, the leave-one-out cross validation performed by SAS to choose the model that minimized the PRESS statistic included five PLS components. However, according to van der Voet's test, a model with zero components was not significantly better than that with five. We need to compress the data in order to use logistic discrimination, so we chose the model with the minimum PRESS statistic, even so. The PLS/LD procedure had a leave-one-out cross-validation estimate of the prediction error of 15.67%, and a training error of 12.2%, yielding a total error of 13.9%.

3) Salt Marsh Species: There were over 1000 wavelengths in the dataset taken on the *Spartina*, *Salicornia*, and *Scirpus* datasets. This high number of wavelengths was too high for the SAS algorithm to handle. Hence, the wavelengths below 400 and those above 1500 were removed, leaving a total of 950 wavelengths.

For the *Spartina*, the leave-one-out cross validation and van der Voet's test performed by SAS to choose the model that minimized the PRESS statistic included 13 PLS components. However, during the PLS/LD procedure, using this many factors led to failure of convergence due to no classification error during logistic discrimination. Therefore, we reduced the number of factors until convergence occurred, which resulted in the inclusion of the first ten components in the final model. The PLS/LD procedure had a leave-one-out cross-validation estimate of the prediction error of 0.4% and a training error of 2.87%, yielding an estimate of the prediction error of 1.63% for this method.

For the *Scirpus* and *Salicornia* dataset, the leave-one-out cross validation and van der Voet's test resulted in the inclusion of eight PLS components. However, as with the *Spartina* dataset, this model failed to converge, due to no classification error during logistic discrimination. As above, we reduced the number of factors until convergence occurred, which resulted in the inclusion of seven factors. For this model, the PLS/LD procedure had a training error of 4.9% and a CV error of 4.4%, yielding an estimated prediction error of 4.64%.

B. Support Vector Classification

1) Frankenia: For this very noisy dataset, the leave-one-out estimate of the prediction error was minimized with a choice of $C = 45$ for the linear kernel. The training error was 2.9%, with CV test error was 11.76%, yielding an estimate of the prediction error of 7.33%. There were 13 support vectors out of a dataset of 34 observations.

Interestingly, most of the prediction errors were made on the unexposed plants. The machine only classified a left-out unexposed plant correctly twice. The remaining errors were made on the lowest concentration treatment. This may imply that the lowest concentration is below the detection limit for this method for this dataset.

When a degree 2 polynomial kernel was used, the total error was minimized with a choice of $C = 2.5$. The training error was 3.0% with a CV error of 14.1%, yielding a total estimated prediction error of 8.55%. There were 13 support vectors.

When a degree 3 polynomial kernel was used, the total error was minimized with a choice of $C = 0.01$. The training error was 3.0%, with a CV error of 17.65%, yielding a total estimated prediction error of 10.31%. There were 15 support vectors.

2) Crop Species: The choice of C, which minimized the prediction error, was $C = 680$, which is a very narrow margin. The CV error was 15.66% with a training error of 3.6%, yielding an estimated prediction error of 9.63%.

For the polynomial of degree 2 kernel, the choice of C, which minimized the prediction error, was $C = 6.5$, which is a reasonably wide margin. The training error was 3.7%, with a CV error

TABLE I
SUMMARY OF ESTIMATED PREDICTION ERROR FOR SVM AND PLSLD

species	sample size	linear SVM	polynomial 2 SVM	polynomial 3 SVM	PLSLD
Frankenia	34	**7.33**	8.85	10.32	8.82
Crop	83	9.63	**7.87**	8.92	15.00
Spartina	209	**1.43**	2.39	2.87	1.63
Salicornia/Scirpus	135	**2.95**	3.70	4.82	4.44

of 12.05%, yielding an estimated prediction error of 7.87%. There were 38 support vectors.

For the degree 3 polynomial, the choice of C, which minimized the prediction error, was $C = 0.1$. The training error was 6.1%, with a CV error of 10.84%, yielding an estimated prediction error of 8.92%. There were 37 support vectors.

3) Salicornia/Scirpus: For the *Salicornia/Scirpus* dataset, the leave-one-out estimate of the prediction error was minimized with a choice of $C = 140$ for the linear kernel. The training error was 2.2%, with a CV error of 3.70%, yielding an estimated prediction error of 2.95%. There were 47 support vectors out of a dataset of 135 observations.

For the degree 2 polynomial machine, the leave-one-out estimate of the prediction error was minimized with a choice of $C = 2$. Both the training error and the CV error were not improved from the linear kernel, yielding the same estimated prediction error. However, the number of support vectors was reduced to 37.

For the degree 3 polynomial, the value of C, which minimized the prediction error, was $C = 0.01$. The training error was 1.5%, with a CV error of 8.15%, yielding an estimated prediction error of 4.82%. There were 37 support vectors.

4) Spartina: For the *Spartina* dataset, the leave-one-out estimate of the prediction error was minimized with a choice of $C = 955$ for the linear kernel. The training error was 0.0%, with a CV error of 1.43%, yielding an estimated prediction error of 1.43%. There were 69 support vectors out of a dataset of 209 observations.

For the degree 2 polynomial, the estimated prediction error was minimized with a choice of $C = 5$. The training error was 0%, with a CV error of 2.4%, yielding an estimated prediction error of 2.4%. There were 57 support vectors.

For the degree 3 polynomial, the estimated prediction error was minimized with a choice of $C = 0.05$. The training error was 0%, with a CV error of 2.87%. There were 50 support vectors.

C. Summary of Results

The results for all datasets are summarized in Table I. Figures shown are percent total error.

IV. CONCLUSION AND DISCUSSION

Using the cross-validation total error as a method of comparison, the classification technique with the lowest prediction error was support vector classification. The linear kernel provided the best decision rule for all datasets except the crop plant experiment. For the crop plants, the degree 2 polynomial provided the lowest prediction error.

While the leave-one-out method for estimating prediction error is well accepted, another proposed estimate for SVM is the number of support vectors divided by the sample size. Since the support vectors lie on the margin, this has some intuitive appeal. We do not use this method here. However, we feel it should be mentioned because the number of support vectors is high for some of these datasets, and the low estimated prediction error might be questionable on this basis.

These results suggest that good predictability can be achieved when classifying exposed plants based on hyperspectral data using support vector machines. Additional difficulties are to be expected for air- and spaceborne platforms, since atmospheric noise, differing angles, and the lower resolution all contribute to the noisiness in the reflectance data. However, since low error was found for these data, it might be expected that even with additional noise, reasonable predictability is achievable.

ACKNOWLEDGMENT

The authors would like to thank M. Lay, T. Yu, S. Gutman, and E. Cadwell for help with the greenhouse work and P. Zarco-Tejada for help with the remote sensing technologies. The authors would also like to thank the TGRS reviewers for their many insightful and helpful comments.

REFERENCES

[1] J. M. Baker, "Responses of salt marsh vegetation to oil spills and refinery effluents," in *Ecological Processes in Coastal Environments: Proc. 19th Symp. British Ecological Soc. and 1st Eur. Ecological Symp. (British Ecological Soc. Symp.)*, R. L. Jefferies and A. J. Davy, Eds, 1979, pp. 529–542.

[2] G. I. Ballew, "Terrestrial higher plants which hyperaccumulate metallic elements—A review of their distributions, ecology, and phytochemistry," in *Proc. 10th Int. Symp. Remote Sensing Environment*, vol. 2. Ann Arbor, MI, 1975, p. 1045.

[3] J. Barcelo and C. H. Poschenrieder, "Responses of plants to heavy metal contamination," in *Proc. 4th National Symp. Mineral Nutrition of Plants*. Alicante, Spain, 1992, pp. 45–61.

[4] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[5] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.

[6] B. E. Boser, I. M. Guyon, and V.N. Vapnick, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. ACM Workshop Computational Learning Theory*, 1992, pp. 144–152.

[7] G. A. Carter, "Ratios of leaf reflectance in narrow wavebands as indicators of plant stress," *Int. J. Remote Sens.*, vol. 15, pp. 697–704, 1994.

[8] G. Casella and R. L. Berger, *Statistical Inference*. Belmont, CA: Wadsworth Brooks/Cole, 1990, ch. 10.

[9] G. Cawley. (2000) MATLAB support vector machine toolbox (v0.54 beta). School of Inform. Syst., Univ. East Anglia, Norwich, U.K. [Online]. Available: http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox.

[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[11] C. Cortes and V. Vapnik, "Support vector network," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
[12] P. J. Curran, "Estimating green LAI from multispectral aerial photography," *Photogramm. Eng. Remote Sens.*, vol. 49, pp. 1709–1720, 1983.
[13] R. A. Efroymson, M. E. Will, G. W. Suter, and A. C. Wooten. (1997) Toxicological benchmarks for screening contaminants of potential concern for effects on terrestrial plants. Lockheed Martin Energy Systems, Inc., Bethesda, MD. [Online]. Available: http://www.hsrd.ornl.gov/ecorisk/tm85r3.pdf
[14] M. J. Farabee, "Photosynthesis," in *The Online Biology Book*. Avondale, AZ: Estrella Mountain Community College, 1997. [Online]. Available: http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookTOC.html.
[15] R. Ginocchio, "Quantifications of tolerance to copper and sulphate in two woody species of central chile," *Revista Chilena Historia Natural*, vol. 69, pp. 413–424, 1996.
[16] A. A. Gitelson and M. N. Merzlyak, "Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll," *J. Plant Physiol.*, vol. 148, pp. 494–500, 1996.
[17] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," *Proc. SPIE*, vol. 3584, pp. 221–232, 1998.
[18] J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, "Support vector machine classification as applied to AVIRIS data," in *Summaries of the 8th JPL Airborne Earth Science Workshop*. Pasadena, CA, Feb. 1999, JPL Pub. 99-17. [Online]. Available: http://aviris.jpl.nasa.gov/docs/workshops/99, pp. 217–227.
[19] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, 2002.
[20] L. Kooistra, R. Wehrens, L. M. C. Buydens, R. S. E. W. Leuven, and P. H. Nienhuis, "Possibilities of soil spectroscopy for the classification of contaminated areas in river floodplains," *Int. J. Appl. Earth Obs. Geoinform.*, vol. 3, no. 4, pp. 337–344, 2003.
[21] H. Martens and T. Naes, *Multivariate Calibration*. New York: Wiley, 1989.
[22] H. H. Muhammed, "Using hyperspectral reflectance data for discrimination between healthy and diseased plants, and determination of damage-level in diseased plants," in *Proc. 31st Applied Imagery Pattern Recognition Workshop*, Washington, DC, Oct. 2002.
[23] J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models*. Boston, MA: Irwin, 1990, ch. 12.
[24] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using gene expression data," *Bioinformatics*, vol. 18, pp. 39–50, 2002.
[25] B. N. Rock, T. Hoshizaki, and J. R. Miller, "Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline," *Remote Sens. Environ.*, vol. 24, pp. 109–127, 1988.
[26] B. Schölkopf, "Support vector learning," Dissertation, Informatik der Technischen Univ., Berlin, Germany, 1997.
[27] H. van der Voet, "Comparing the predictive accuracy of models using a simple randomization test," *Chemomet. Intell. Lab. Syst.*, vol. 25, pp. 313–323, 1994.
[28] V. Vapnik and A. Chervonenkis, *Theory of Pattern Recognition*. Moscow, Russia: Nauka, 1979. In Russian. German translation: W. Wapnik and A. Tscherwonenkis, *Theorie der Zeichenerkennung*, Berlin, Germany: Akademie-Verlag.
[29] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
[30] ——, *Statistical Learning Theory*. New York: Wiley, 1998.
[31] J. E. Vogelman, B. N. Rock, and D. M. Moss, "Red edge spectral measurements from sugar maple leaves," *Int. J. Remote Sens.*, vol. 14, pp. 1563–1575, 1993.
[32] H. Wold, "Soft modeling: The basic design and some extensions," *Systems Under Indirect Observation, Causality-Structure-Prediction*, 1981.
[33] ——, *Food Research and Data Analysis*, H. Martens and H. Russwurm, Eds. London, U.K.: Applied Science, 1983.

Machelle D. Wilson received the B.A. degree in humanities from Brigham Young University, Provo, UT, in 1980, the M.S. degrees in ecology and statistics from the University of California, Davis (UC-Davis), in 1997 and 1999, respectively, and the Ph.D. degree in ecology from UC-Davis, in 2001.

She is currently an Assistant Professor with the Savannah River Ecology Laboratory, University of Georgia, Aiken, SC, with adjunct status in the Department of Statistics and the Institute of Ecology. Her current research focuses on developing statistical methodologies for environmental and ecological data.

Susan L. Ustin (M'90) received the B.S. and M.S. degrees in biology from the California State University, Hayward, and the Ph.D. degree in botany in the area of plant physiological ecology from the University of California, Davis (UC-Davis).

She is currently a Professor in the Department of Land, Air, and Water Resources, UC-Davis. Since 1983, her research has focused on the use of remote sensing instruments to measure ecological properties and processes. In particular, she has been active in the development of imaging spectrometery from airborne and spaceborne platforms for measuring ecosystem properties. Her research has focused on developing and testing methods for retrieving environmental information from remote sensing imagery at spatial scales from microscopic to regional and encompasses both theoretical and applied research. She has studied many different ecological communities and farmlands in the United States and abroad. She is Director of the UC-Davis California Space Institute Center of Excellence and the Center for Spatial Technologies and Remote Sensing (CSTARS), and is Director of the Department of Energy's Western Center for Global Environmental Change (WESTGEC), Davis.

David M. Rocke received the A.B. degree (with distinction) in natural sciences from Shimer College, Waukegan, IL, in 1966, and the M.S. and Ph.D. degrees in mathematics from the University of Illinois, Chicago, in 1968 and 1972, respectively.

He is currently a Professor in the Division of Biostatistics (Medicine) and the Department of Applied Science (Engineering), Co-Director of the Center for Image Processing and Integrated Computing, and Director of the Center for Digital Security, University of California, Davis. His current research focuses on analysis methods for large, high-dimensional datasets, predominantly related to life and environmental sciences.

Dr. Rocke was awarded the American Society for Quality Control's Youden Prize in 1982 and the Shewell Award in 1987. He also received the American Statistical Association's (ASA) Award for Interlaboratory Testing in 1985 and the Statistics in Chemistry Award in 1997. He is a Fellow of the ASA and an Elected Member of the International Statistical Institute.