# *Object-based Vegetation Mapping in the Kissimmee River Watershed Using HyMap Data and Machine Learning Techniques*

**Caiyun Zhang & Zhixiao Xie**

**Wetlands**
Official Scholarly Journal of the Society
of Wetland Scientists

ISSN 0277-5212

Wetlands
DOI 10.1007/s13157-012-0373-x

ONLINE FIRST
WETLANDS
SWS
Journal of the Society of Wetland Scientists
VOLUME 32 NUMBER 6 DECEMBER 2012 (ISSN 0277-5212)
Springer

Springer

Wetlands
DOI 10.1007/s13157-012-0373-x

ARTICLE

# Object-based Vegetation Mapping in the Kissimmee River Watershed Using HyMap Data and Machine Learning Techniques

Caiyun Zhang · Zhixiao Xie

Received: 24 July 2012 / Accepted: 23 December 2012

**Abstract** Accurate and informative vegetation maps are in urgent demand to support the Kissimmee-Okeechobee-Everglades ecosystem restoration project in South Florida. In this study, we evaluated the applicability of fine spatial resolution hyperspectral data collected from the HyMap sensor for both community- and species-level vegetation mapping. Informative and accurate vegetation maps were produced by combining machine learning methods (Support Vector Machines (SVM) and Random Forest (RF)), object-based image analysis techniques, and Minimum Noise Fraction (MNF) data transformation. An overall accuracy of 90% was obtained in discriminating 14 vegetation communities. Classification of a large number of species is also promising. An overall accuracy of 85% was achieved in identifying 55 species using a SVM model. The results indicate that fine spatial resolution hyperspectral data classification using such automated procedure has great potential to replace the manual interpretation of aerial photos for vegetation mapping in heterogeneous wetland ecosystems.

**Keywords** Wetland vegetation mapping · Hyperspectral remote sensing · Machine learning techniques

## Introduction

The Kissimmee River watershed is located in south-central Florida. Historically, Kissimmee River flowed south from Lake Kissimmee to Lake Okeechobee over a 103-mile path, which created a vast floodplain and supported a diverse mosaic of wetland communities. The Kissimmee River forms the headwaters of Lake Okeechobee and the Everglades. These three together comprise the Kissimmee-Okeechobee-Everglades (KOE) ecosystem, also known as the Greater Everglades, which is the largest subtropical wetland in the United States. In the late 1960s, the Kissimmee River was transformed into a 56-mile-long canal called Canal-38 for flood control. The channelization disrupted the entire riverine ecosystem and imposed a variety of environmental impacts in South Florida. Dominant wetland vegetation types such as broadleaf marsh, wet prairie, and wetland shrub, were replaced by dry pasture and other upland vegetation types after channelization (Bousquin et al. 2005). In 1992, the U. S. Congress approved the Kissimmee River Restoration Project (KRRP) to restore the original river ecosystem and its associated plant communities. Vegetation mapping will document vegetation changes to support the restoration. In 2000, a larger restoration project, the Comprehensive Everglades Restoration Plan (CERP), was also approved to restore the KOE ecosystem (CERP 2012). Similarly, CERP requires accurate vegetation maps at different scales to measure the progress of the project.

Currently, vegetation information in the KOE system is mainly collected through field-based studies and maps generated from the visual interpretation of large scale aerial photographs using stereo plotters (Rutchey et al. 2008; Jones 2011). Both procedures are time-consuming, labor-intensive, and costly. With the emergence of hyperspectral remote sensing techniques, it has been anticipated that these procedures can be superseded by semi-automated or automated digital image analysis. Hyperspectral sensors collect data in hundreds of relatively narrow spectral bands throughout the visible and infrared portions of the electromagnetic spectrum. They are more powerful than traditional

**Electronic supplementary material** The online version of this article (doi:10.1007/s13157-012-0373-x) contains supplementary material, which is available to authorized users.

C. Zhang (✉) · Z. Xie
Department of Geosciences, Florida Atlantic University,
777 Glades Road,
Boca Raton, FL 33431, USA
e-mail: czhang3@fau.edu

Published online: 17 January 2013

multispectral sensors used for vegetation mapping due to their rich spectral content.

The application of hyperspectral analysis has become an important area of research for wetland mapping in the past decade. Such research can be grouped into two categories (Zhang and Xie 2012a). The first is the application of hyperspectral sensors with a coarse spatial resolution (i.e. 20–30 m or larger), such as EO-1/Hyperion and high altitude Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Limitations of these sensors include the coarse spatial resolution and complexity of image processing procedures (Hirano et al. 2003). The second is the employment of hyperspectral data with a fine spatial resolution (i.e. 4 m or smaller), such as imagery collected from low altitude Compact Airborne Spectrographic Imager (CASI), AVIRIS, and HyMap. This type of data has suitable spatial and spectral resolution for vegetation mapping (Zhang and Xie 2012a), but data collection has a higher cost.

Most researchers used endmember-based approaches to classify fine spatial resolution hyperspectral imagery. Examples include the Spectral Angle Mapper (SAM) and linear spectral unmixing, which are specifically designed to extract information from hyperspectral imagery (e.g., Hunter and Power 2002; Schmidt et al. 2004; Artigas and Yang 2005; Li et al. 2005; Jollineau and Howarth 2008). These approaches did not achieve the expected results in the KOE system due to the difficulties inherent in determining hyperspectral endmembers, a shortage of comprehensive spectral libraries for different wetland plants, and the violation of the assumption in the algorithms that only one spectral representative (i.e. the endmember) exists for each vegetation type. The KOE system has a complicated plant community with high community and species diversity, at varying ages, and under varying environmental conditions. As such, multiple spectral signatures often occur for the same community or species. The KOE region also exhibits high spatial heterogeneity. Diverse upland and inland community types or species compositions are present over relatively short distances. These combined factors make the determination of endmembers very challenging. Traditional classifiers such as maximum likelihood and minimum distance method have not generated classifications with reasonable accuracies. Similar to the SAM, the minimum distance assumes a single signature for each class. The Maximum Likelihood (ML) algorithm requires that the spectral response of each class follow a Gaussian distribution, which is not guaranteed for hyperspectral data.

Two emerging machine learning algorithms, Support Vector Machines (SVMs) and the Random Forest (RF) classifier, are promising techniques for hyperspectral image analysis (Waske et al. 2009; Zhang and Xie 2012b). The application of SVMs has been increased significantly in the past decade (Mountrakis et al. 2010), after Gualtieri and Cromp (1998) introduced the methods for remote sensing. A key feature of SVMs is that they are able to produce higher accuracy than traditional classifiers using a small number of training data sets, which makes them particularly appealing in hyperspectral data classification. Several studies have examined the performance of SVMs in hyperspectral data analysis (e.g., Gualtieri and Cromp 1998; Melgani and Bruzzone 2004; Camps-Vallas and Bruzzone 2005; van der Linden et al. 2007; Waske et al. 2009), but previous application in mapping heterogeneous wetlands has been limited.

Random Forest (RF) approach is another promising technique for hyperspectral data analysis. RF is a decision tree based ensemble classifier proposed by Breiman (2001). The premise of an ensemble algorithm is that a combining classifier is often more accurate than any single classifier. Practically, ensemble classifications are not an intuitive choice for analyzing hyperspectral data because they naturally add more computational burden to a procedure already complicated by high dimensional inputs (Chan and Paelinckx 2008). However, since RF is decision tree-based, and the process of building trees is extremely fast, computation cost is not an issue. A key feature of RF is that it can handle high dimensional datasets, which makes it attractive for processing hyperspectral data. Several studies have shown that it performs well for classifying hyperspectral data (Crawford et al. 2003; Ham et al. 2005; Lawrence et al. 2006; Chan and Paelinckx 2008). However, again, previous application in mapping complex wetlands has been limited.

Classified maps can be generated at the pixel level or the object level. Pixel-based mapping methods may lead to a “salt-and-pepper” effect because of the high spatial heterogeneity and diversity of plant communities in the KOE region. Object-based classification removes this effect. These techniques first decompose an image scene into relatively homogeneous objects or areas and then classify these areas instead of pixels. Previous studies have demonstrated that object-based image analysis techniques can produce higher accuracy than pixel-based methods in wetland mapping (Harken and Sugumaran 2005; Kamal and Phinn 2011) and they are effective approaches for analyzing fine spatial resolution imagery (Blaschke 2010).

In this study, we evaluated the performance of SVMs and RF for object-based vegetation mapping at two levels (community- and species-level) in the complex KOE system from fine spatial resolution hyperspectral data. Most previous studies have conducted hyperspectral wetland mapping at pixel level. Examples of object-based mapping from fine spatial resolution hyperspectral data are limited. Only a few studies have combined machine learning techniques with object-based vegetation mapping (Zhang and Xie 2012a & b). Moreover, few studies have examined the applicability of hyperspectral data in classifying a large number of vegetation species (e.g. more than 50) in a complex wetland ecosystem. Therefore, the objectives of this study are: 1) to explore the potential of high spatial resolution hyperspectral data for object-based vegetation mapping and monitoring in the complex KOE system; and 2) to examine the effectiveness of SVMs and RF in classifying wetland vegetation communities as well as large numbers of wetland species.

## Methods

### Study Area

The study area is located in the lower basin of Kissimmee River watershed (Fig. 1). The upper basin of the river contains a series of connected lakes and the lower basin contains Canal-38, remnants of the Kissimmee River, and six water control structures. The river and floodplain slope to the south from an elevation of 15.5 m at Lake Kissimmee to an elevation of 4.6 m at Lake Okeechobee. As the headwaters of Lake Okeechobee and Everglades, the river plays a critical role in the KOE system by controlling the water quality, quantity, and the integrity of the river and floodplain within the landscape. Restoration of the Kissimmee River watershed is one of key components of CERP. The selected study area is the region known as Pool C in the CERP, and consists of a portion of Canal-38 that was backfilled in 2001 with a new river channel carved to allow the river to flow

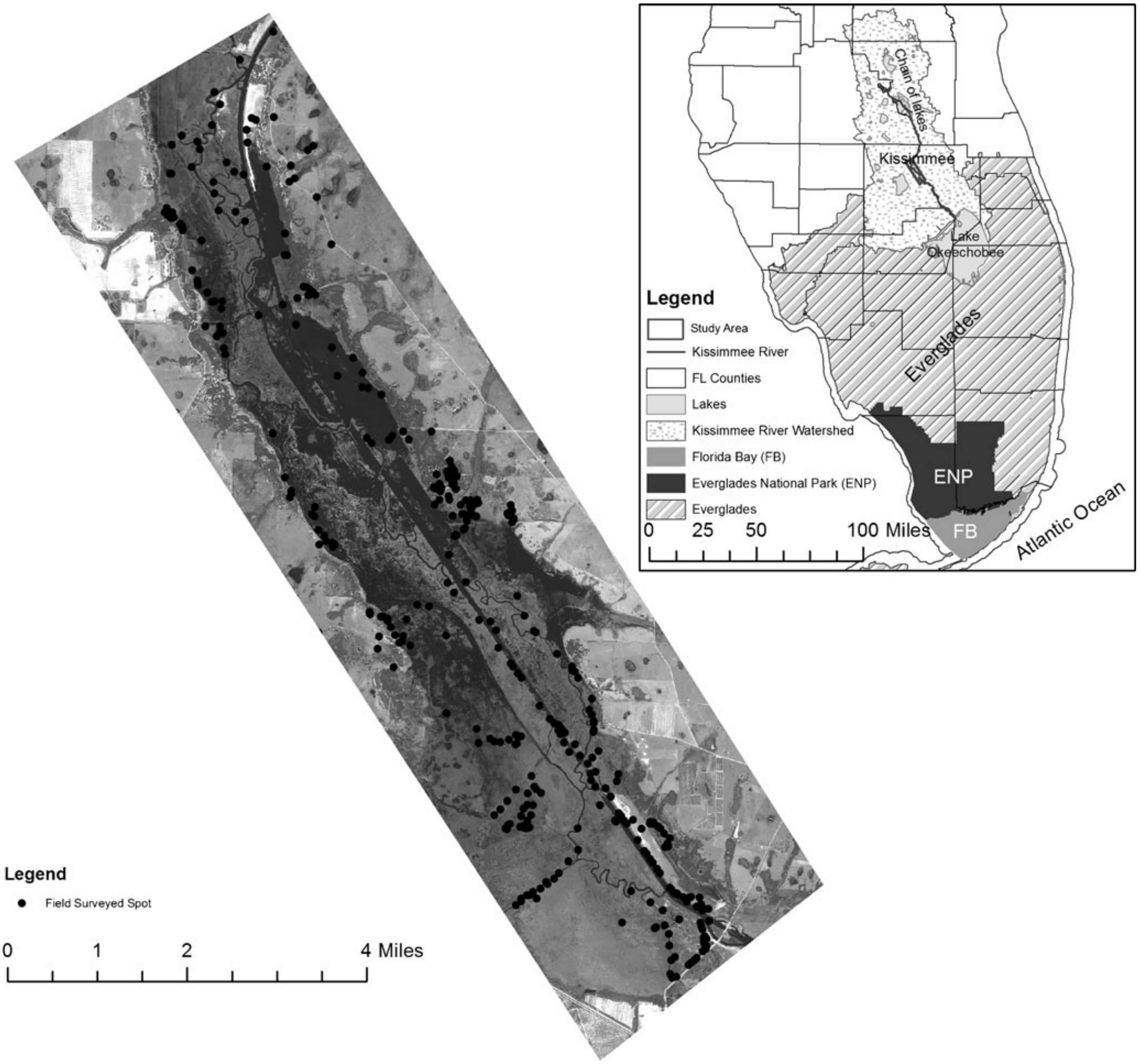

**Fig. 1** Map of the Greater Everglades, Kissimmee River watershed, study area, and field sampling locations

naturally to Lake Okeechobee (Bousquin et al. 2005). This particular region is currently in its post-construction monitoring phase. The degree of change in the river and floodplain vegetation community is one of the factors used to judge the success of the restoration.

### Hyperspectral Data

The South Florida Water Management District (SFWMD) provided the hyperspectral data and field surveyed data for this study. The Jacksonville District of the United States Army Corps of Engineers (USACE) and SFWMD executed a research project in 2002 to assess the applicability of hyperspectral techniques for monitoring the restoration of the KOE system. This pilot project collected the hyperspectral data (three flightlines, 107 km$^2$) over the Pool C area on September 21, 2002 using the HyMap sensor. HyMap is an airborne imaging spectrometer that provides 128 spectral bands across the wavelength region of 0.45–2.5 μm with bandwidths between 15 and 20 nm. The collected imagery has a spatial resolution of 3.5 m. Images were atmospherically and geometrically corrected by the provider (Lowe Engineers and SAIC 2003). The data were provided for this study in the projection of Universal Transverse Mercator (UTM) Zone 17N and aligned well to the USGS orthophotos. A natural color composite generated from the HyMap data in grayscale for the study area is displayed in Fig. 1.

### Reference Data

An extensive on-foot and airboat field survey was conducted from September 23 to October 6, 2002. A stratified random sampling scheme augmented with several rules was adopted for the survey. Sampling points are required to: 1) not be within 25 m of each other, 2) exhibit major variation within each class, and 3) capture larger and more homogenous areas. It was also important that the personnel understand the variation within each class. A total of 361 samples were collected and screened using the field photos and georeferenced image data. Suspect samples were removed, leading to a final sample of 340 points. The locations of these points are shown in Fig. 1. Each sample represented a 5-m radius circle (78.5 m$^2$) within which the location, plant community, and other information (e.g. date and time) were recorded. These samples were labeled based on B-Codes, the finest level of a vegetation classification system developed by the SFWMD (Appendix Table). They represent the percent cover of dominant species and co-occurring species or other land covers (Bousquin et al. 2005) and were used as reference data for the species-level classification. A total of 57 B-Codes were found, with 55 of them represented by 4 or more samples. The number of field samples for each of the 55 B-Codes is listed in Table 5. We extracted the mean spectrum of the image objects that the field samples were within to construct the reference signatures, rather than constructing signatures based on the spectra of single pixels where the samples were located. This strategy can reduce the positional discrepancy between the image data and the field data. Additionally, it can increase the robustness of object-based image classification since the reference data collected from an image object is more representative than any pixel within this object. The concept of hierarchical image segmentation was employed to obtain the object-based reference data and classification. The segmentation procedure is detailed in Hierarchical Segmentation subsection.

The species (B-Codes) inventory provides detailed vegetation information; however, some projects require a coarser community-level classification. In this study, we adopted the SFWMD modified Florida Land Use, Land Cover Classification System because it frequently serves as the standard for other GIS data in CERP. The SFWMD has reference data for years 1999 and 2004, which were derived through manual interpretation of large scale aerial photos and validated by field surveys. Unfortunately, there is no reference data for 2002. Large changes may have occurred from 1999 to 2002 and from 2002 to 2004. The high spatial resolution of HyMap images allows the visual interpretation of vegetation communities with assistance from reference data and related large scale digital orthophotos collected in 1999 and 2004. The selected study area included 14 observed vegetation communities and land covers: improved pastures, unimproved pastures, woodland pastures, citrus groves, dry prairies, upland shrub and brush, upland hardwood forests, Brazilian pepper, mixed wetland hardwoods, mixed shrub, cypress-mixed hardwoods, freshwater marshes/Graminoid prairie-marsh, wet prairies, and spoil area. We randomly selected and manually interpreted 986 image objects as the reference data for community-level classification and accuracy assessment. The reference data collection followed a stratified random sampling strategy in which a fixed percentage of samples are randomly selected for each class. The number of samples for each class was roughly estimated based on the results of image segmentation and the 1999 and 2004 reference data. The number of selected reference objects for each community is listed in Table 4. The reference data at the community-level and species-level were split into two halves with one used for training and the other used for accuracy assessment.

### Data Preprocessing

Visual examination revealed that bands 1, 63–65, 125–128 needed to be removed due to water absorption and a low signal-to-noise ratio. The remaining 119 bands were used for further analysis. We masked out the non-vegetated areas, such as open water, to focus on vegetation classification.

Hyperspectral data contains a tremendous amount of redundant spectral information. The Minimum Noise Fraction (MNF) method (Green et al. 1988) is commonly used to reduce the high dimensionality and inherent noise of hyperspectral data. We conducted the MNF transformation in ENVI 4.7 and selected the first 20 MNF eigenimages which are most useful and spatially coherent. To examine the effects of the MNF transformation on classification, we used both the original and transformed data for classification.

### Hierarchical Segmentation

Hierarchical segmentation was adopted to generate image objects at both vegetation levels. Hierarchical segmentation is defined as set of segmentations of the same image at different levels of spatial resolution in which coarser levels can be produced by merging regions at finer levels (Beaulieu and Goldberg 1989). We produced image objects using the multiresolution segmentation algorithm in eCognition Developer 8.64.1 (Trimble 2011). The algorithm starts with single pixel image segments, and merges neighboring segments until a heterogeneity threshold is reached (Benz et al. 2004). The heterogeneity threshold is determined by a user-defined scale parameter, as well as color/shape and smoothness/compactness weights. Image segmentation is scale-dependent and the quality of segmentation and overall object-based classification are heavily dependent on the scale (Liu and Xia 2010). In order to find the optimal scale for image segmentation, we used an unsupervised image segmentation evaluation approach (Johnson and Xie 2011). This approach conducts a series of segmentations using different scale parameters to identify the optimal scale using an unsupervised evaluation method that takes into account global intra-segment and inter-segment heterogeneity measures. A global score combining a normalized weighted variance and Moran's *I* value was used to determine the optimal scale for the segmentation. Preliminary analyses revealed that the scale parameter larger than 10 generated many under-segmented objects and smaller than 2 produced many over-segmented objects. We thus carried out a series of segmentations with the scale parameter ranging from 2 to 10, at an interval of 2. A scale of 6 was found to produce the optimal segmentation. The weights of the MNF layers were set based on their eigenvalues. Color and shape weights were set to 0.9 and 1.0 so that spectral information would be considered more heavily for segmentation. Smoothness and compactness weights were set to 0.5 and 0.5 so as to not favor either compact or non-compact segments.

Species-level image objects were produced using a finer scale based on the result of community-level segmentation. A heuristic method was designed to find the optimal scale for the species-level segmentation. We randomly selected 10% of the community-level segments and handled each segment as an individual image. We then conducted a series of multiresolution segmentations using five different scale parameters (1–5 at an interval of 1) for each selected segment. Global scores were calculated and the optimal scale was determined as the one with the highest frequency to generate the lowest global score among these selected segments. For this case, a scale parameter of 3 was found to be the optimal scale for species-level segmentation. The other parameters were the same as those used for the community-level segmentation.

### Classification Methods

We examined two machine learning classification algorithms, Support Vector Machines (SVMs) and Random Forest (RF). SVM is a non-parametric supervised learning classifier. The goal of SVMs is to find a hyperplane that can separate the input dataset into a discrete predefined number of classes in a fashion consistent with the training samples (Vapnik 1995). Detailed descriptions of SVM algorithms are given by Huang et al. (2002) and Melgani and Bruzzone (2004) in the context of remote sensing. Kernel based SVMs are commonly used for remote sensing image classification, among which the radial basis function (RBF) kernel and the polynomial kernel are frequently employed. Both kernels were tested to find the best model. The RBF requires specification of the kernel width ($\gamma$), and the polynomial kernel requires specification of the degree ($p$). Both kernels require definition of a penalty parameter ($C$) that controls the degree of misclassification. Those parameters can be optimized by a grid search strategy which tests possible combinations of $C$ and $\gamma$ in a user-defined range (Hsu et al. 2010). For our dataset, we found that the polynomial kernel with $p=2$ and $C=2.0$ generated the best accuracy. The original SVM algorithm was designed for binary classification. Several strategies including one-against-one and one-against-all have been developed to solve multiclass problems. We selected the one-against-one strategy because the one-against-all strategy may result in the estimation of complex discrimination functions and generate unexpected classification results (Melgani and Bruzzone 2004). The SVM classifier was implemented using a SVM software (LIBSVM) developed by Chang and Lin (2011).

RF is a decision tree ensemble classifier. Decision trees split the training samples into smaller subdivisions at "nodes" based on decision rules. For each node, tests are performed on the training data to find the most useful variables and variable values for the split. The RF consists of a combination of decision trees where each decision tree contributes a single vote for assigning the most frequent class to an input vector. RF increases the diversity of decision trees by changing the training set using the bagging aggregating method (Breiman 2001). Bagging creates training data by randomly resampling the original dataset with replacement. A key feature of RF is that the computational complexity is simplified by reducing the number of input features at each node. Different algorithms can be used to generate the decision trees. The RF often adopts the Gini Index (Breiman 2001) to measure the best split selection. More detailed description of RF can be found in Breiman (2001) or in a remote sensing context in Chan and Paelinckx (2008) and Rodriguez-Galiano et al. (2012). The RF classification was implemented using Weka 3.7, an open-source data mining program (Hall et al. 2009). Two parameters must be defined: the number of decision trees to create (*k*) and the number of randomly selected variables (*m*) considered for splitting each node in a tree. RF is not sensitive to *m* and it is often blindly set to $\sqrt{\mathrm{M}}$ (Gislason et al. 2006). The computational complexity of the algorithm can be reduced by selection a smaller *m*. *k* is often set based on trial and error. For our dataset, *m* was set to 10 and 3 for original data and MNF transformed data respectively. Trials using different number of trees (50–500 at an interval of 50) revealed that $k=100$ produced the best accuracy.

### Accuracy Assessment

We compared the accuracies of the object-based vegetation maps by constructing an error matrix and calculating the Kappa statistic (Congalton and Mead 1983). Overall accuracy is defined as the ratio of the number of validation samples that are classified correctly to the total number of validation samples irrespective of the class. The Kappa value describes the proportion of correctly classified validation samples after random agreement is removed. We used the nonparametric McNemar test (Foody 2004) to evaluate the statistical significance of differences in accuracy between different classifications. The difference in accuracy between a pair of classifications is viewed as being statistically significant at a confidence of 95% if the calculated z-score in McNemar test is larger than 1.96.

Note that the number of species-level field samples is limited for each B-Code, which may cause statistical problems for accuracy assessment. It is difficult to expand the reference data at this level through the photo interpretation technique. To solve this problem, an artificial sample interpolation approach (Demir and Ertürk 2009) was adopted. The first new sample of a class was generated using the mean spectrum calculated from the first sample and all other samples belonging to this class. The second new sample was generated using the mean spectrum calculated from the second sample and the remaining samples with the first sample excluded. The procedure was repeated until the last sample of the corresponding class was processed. Previous samples were not considered in the derivation of new samples. By this method, a total of $N(N\text{-}1)/2$ new samples could be produced for the corresponding class, where $N$ is the initial number of samples for this class. Demir and Ertürk (2009) have shown that the interpolation approach is effective for hyperspectral data classification if a limited number of training samples are available. The accuracy assessment was performed on the first 55 B-Codes which had at least 4 original samples. After interpolation, each B-Code class had at least 10 samples.

## Results and Discussion

### Impact of MNF Transformation on Vegetation Classification

We classified both original data (i.e. 119 bands) and MNF transformed data (i.e. 20 MNF layers) to examine the impact of MNF transformation on classification. The processing time in seconds and results of the two classifiers are displayed in Table 1. Both classifiers obtained increased accuracies after MNF transformation for the community-level classification. For the SVM, MNF transformation increased the accuracy from 83% to 89%, while for RF, MNF transformation improved the accuracy from 74% to 90%. These improvements were significant based on the McNemar tests. Between the two classifiers, there is no significant difference with MNF transformed data. For the species-level classification, MNF transformation increased the accuracy from 73% to 85% using the SVM, and from 36% to 79% using the RF. McNemar tests showed these increases are significant. RF was more sensitive to MNF transformation. At the species-level, SVM produced higher accuracy than RF using both datasets. However, RF is more efficient than SVM in terms of processing time, which is expected because RF only uses a random subset for each split. In both cases, the processing time was significantly reduced using MNF transformed data.

The effect of MNF transformation on hyperspectral classification is controversial. Some studies have found that MNF transformation resulted in decreased accuracy (e.g. Pal and Mather 2006), while others demonstrate that MNF increased classification accuracy (e.g. Belluco et al. 2006; Yang et al. 2009; Zhang and Xie 2012a). Our study illustrates that the MNF transformation can significantly improve classification accuracy and decrease the computational time. Few studies have examined the impact of MNF transformation on the performance of RF in hyperspectral data analysis. Most studies have applied RF to classify original hyperspectral data (e.g. Ham et al. 2005; Lawrence et al. 2006; Waske et al. 2009). Breiman (2001) demonstrates that RF is not sensitive to output noise, which was modeled by changing the class labels of the training data. This is confirmed by Rodriguez-Galiano et al. (2012). However, it is clear that the “output noise” in these studies is not the “noise” existing in the original hyperspectral

**Table 1** Comparison of classification using original data and MNF transformed data

| | Original | | | MNF | | |
|---|---|---|---|---|---|---|
| Community-level | | | | | | |
| Classifier | Accuracy | Kappa | Time (s) | Accuracy | Kappa | Time (s) |
| SVM | 83% | 0.80 | 45.7 | 89% | 0.87 | 0.65 |
| RF | 74% | 0.70 | 1.47 | 90% | 0.89 | 0.55 |
| Pairwise statistical test | | | | | | |
| | McNemar Test | | | | McNemar Test | |
| SVM (Original/MNF) | 3.6* | | Original (SVM/RF) | | 4.4* | |
| RF (Original/MNF) | 7.8* | | MNF (SVM/RF) | | 1.3 | |
| Species-level (B-Codes) | | | | | | |
| Classifier | Accuracy | Kappa | Time (s) | Accuracy | Kappa | Time (s) |
| SVM | 73% | 0.72 | 472.6 | 85% | 0.85 | 13.7 |
| RF | 36% | 0.34 | 5.6 | 79% | 0.79 | 1.59 |
| Pairwise statistical test | | | | | | |
| | McNemar Test | | | | McNemar Test | |
| SVM (Original/MNF) | 6.0* | | Original (SVM/RF) | | 13.7* | |
| RF (Original/MNF) | 15.2* | | MNF (SVM/RF) | | 3.4* | |

*SVM* support vector machine

*RF* random forest

*Original* original hyperspectral data with 119 bands

*MNF* MNF transformed data with 20 bands

*significant with 95% confidence

imagery. Based on our results, the RF classifier is very sensitive to the latter noise which severely impacts the RF performance. Therefore the MNF transformation is necessary to effectively apply the RF classifier for vegetation mapping.

### Impact of Training Data Size on Vegetation Classification

We tested the sensitivity of classification accuracy to the number of training samples using the MNF transformed imagery. We randomly selected 20% of samples as the testing data, and then randomly selected different size (20%–80% at an interval of 20%) of training data from the remaining samples. The results from the two classifiers are shown in Table 2. When the number of training samples increased from 20% to 80%, the SVM accuracy increased from 77% to 93%, and the RF accuracy increased from 79% to 92% for the community-level classification. At the species-level, the SVM accuracy increased from 60% to 87%, and the RF accuracy increased from 61% to 85%. The Kappa values also support these improvements in classification.

We found that both classifiers were sensitive to the number of training samples. These findings differ from the results of Waske et al. (2009) who found that both SVM and RF are relatively insensitive to training sample size. Note that their study area was dominated by non-vegetation covers and their pixel-based training samples were more homogeneous. In comparison, our study area is a complex wetland with a high degree of spatial and spectral heterogeneity and our reference data were collected at the object level. Hence our training samples were spectrally more heterogeneous. For the SVM, a smaller number of training samples are required if training data are more homogeneous. A reduction of training samples will not significantly change the classification accuracy. Conversely, for heterogeneous training samples, more support vectors are needed and reduction of training samples may heavily impact the classification results. Rodriguez-Galiano et al. (2012) found that RF shows low sensitivity to the training sample size. Their assessment was based on training samples collected at the pixel level where redundancy is high. Again, due to the high degree of spatial and spectral heterogeneity of our study area, redundancy was rare among our training samples and reduction of training samples may

**Table 2** Accuracies of SVM and RF from different training size

| | SVM | | RF | |
|---|---|---|---|---|
| Community-level | | | | |
| Training Size | Accuracy | Kappa | Accuracy | Kappa |
| 20% | 77% | 0.73 | 79% | 0.74 |
| 40% | 89% | 0.86 | 88% | 0.85 |
| 60% | 86% | 0.84 | 91% | 0.89 |
| 80% | 93% | 0.92 | 92% | 0.90 |
| Species-level (B-Codes) | | | | |
| Training Size | Accuracy | Kappa | Accuracy | Kappa |
| 20% | 60% | 0.59 | 61% | 0.60 |
| 40% | 75% | 0.75 | 73% | 0.72 |
| 60% | 81% | 0.80 | 82% | 0.82 |
| 80% | 87% | 0.87 | 85% | 0.84 |

*SVM* support vector machine

*RF* random forest

The MNF transformed data were used in these tests severely impact the classification. Thus for heterogeneous training samples, larger training sample size should improve or guarantee the performance of both classification algorithms.

According to Landis and Koch (1977), Kappa values larger than 0.81 indicate an almost perfect agreement in classification. For the community-level classification, an overall accuracy of 84% and Kappa value of 0.81 were obtained from both classifiers when 30% of training samples (i.e. a total of 232 samples) were selected. Therefore, 232 samples could serve as a threshold in our case. As for the species-level classification, 60% of training sample size (i.e. a total of 630 samples) produced a Kappa value of around 0.81 from the two classifiers (Table 2). Thus a minimum of 630 samples were required for the species-level classification.

### Impact of the Number of Classes on Classification Accuracy

Few studies have tested the sensitivity of SVM and RF to the number of classes to be identified. We designed a series of experiments using the MNF transformed data to assess the effects of this factor in vegetation classification. Table 4 and Table 5 show the per-class accuracy for two vegetation levels respectively. Classes with relatively lower accuracies may also influence the overall accuracy. We thus excluded classes which had lower accuracies in the experiments. Here, classes with an average of Producer's Accuracy (PA) and User's Accuracy (UA) less than 70% from the SVMs were considered to have lower classification accuracy. Two communities (dry prairies and wet prairies) and seven B-Codes (H.CD, H.HU, H.MxWT, H.NL, H.PST, S.LS, and S. ST) had lower accuracy, which were excluded in the experiments. For the community-level tests, we randomly selected 5 and 10 classes from the 12 remaining communities, leading to the 5-community and 10-community tests. Similarly, for species-level tests, we selected 10, 20, 30, and 40 B-Codes from 48 remaining B-Codes respectively, leading to the 10-species, 20-species, 30-species, and 40-species experiments. Reference data were randomly split into 50% for training and 50% for accuracy assessments. The results from SVM and RF are shown in Table 3. Generally, the classification accuracy decreased as the number of classes increased. For the community-level classification, the accuracy decreased 5% and 3% from SVM and RF respectively when the number of communities changed from 5 to 10. For the species-level classification, the accuracy decreased 7% and 11% respectively from SVM and RF when the number of species varied from 10 to 40.

### Object-based Vegetation Mapping

SVM and RF are promising learning algorithms for vegetation discrimination in the KOE system. Because SVM produced higher accuracy than RF for species-level classification, the final classification was conducted from the MNF transformed data and the SVM algorithm. The produced object-based vegetation maps in grayscale are displayed in Fig. 2. A nested map is displayed in Fig. 2 to illustrate the high diversity of B-Codes and spatial heterogeneity of the study area. The object-based vegetation map is more informative and useful than a traditional pixel-based map which may be noisy due to the high degree of spatial and spectral heterogeneity of the KOE system. Tables 4 and 5 show the producer's and user's accuracies for the final maps. For comparison, the producer's and user's accuracies from RF methods are also displayed.

For the community-level classification, the producer's accuracy varied from 40% to 100% and the user's accuracy varied from 57% to 100% using the SVM classifier. The RF classifier resulted in the producer's accuraces in the range of 30%–100% and user's accuracies in the range of 60%–100%. A relatively low accuracy was obtained for discriminating dry prairies (class 5) and wet prairies (class 13) due to confusion of these two communities. Higher accuracies were achieved for the other communities. It is worthy to note that a good result was obtained for Brazilian Pepper (class 8), an exotic species in South Florida. Accurate identification of exotic species is critical because tremendous time, effort, and expense have been devoted to their control and removal. For species-level classification, the producer's accuracy using SVM was in the range of 36%–100, and user's accuracy was in the range of 41%–100%. For the RF classifier, both the producer's and user's accuracies ranged from 0% to 100%. The RF completely failed to identify H.MxM (class 19, miscellaneous marsh vegetation) that was misclassified as S.MCF (class 48, floating mat shrubland) or as S.MxFS (class 49, miscellaneous floating mat shrubland). For most of the

**Table 3** Classification accuracy using different number of classes

| | SVM | | RF | |
|---|---|---|---|---|
| Community-level (Excluded Classes with Lower Accuracies) | | | | |
| No. of Classes | Accuracy | Kappa | Accuracy | Kappa |
| 5 | 93% | 0.90 | 95% | 0.93 |
| 10 | 88% | 0.86 | 92% | 0.91 |
| Species-level (B-Codes) (Excluded Classes with Lower Accuracies) | | | | |
| No. of Classes | Accuracy | Kappa | Accuracy | Kappa |
| 10 | 99% | 0.98 | 97% | 0.96 |
| 20 | 95% | 0.95 | 92% | 0.91 |
| 30 | 94% | 0.94 | 88% | 0.87 |
| 40 | 92% | 0.92 | 86% | 0.86 |

*SVM* support vector machine

*RF* random forest

The MNF transformed data were used in these tests

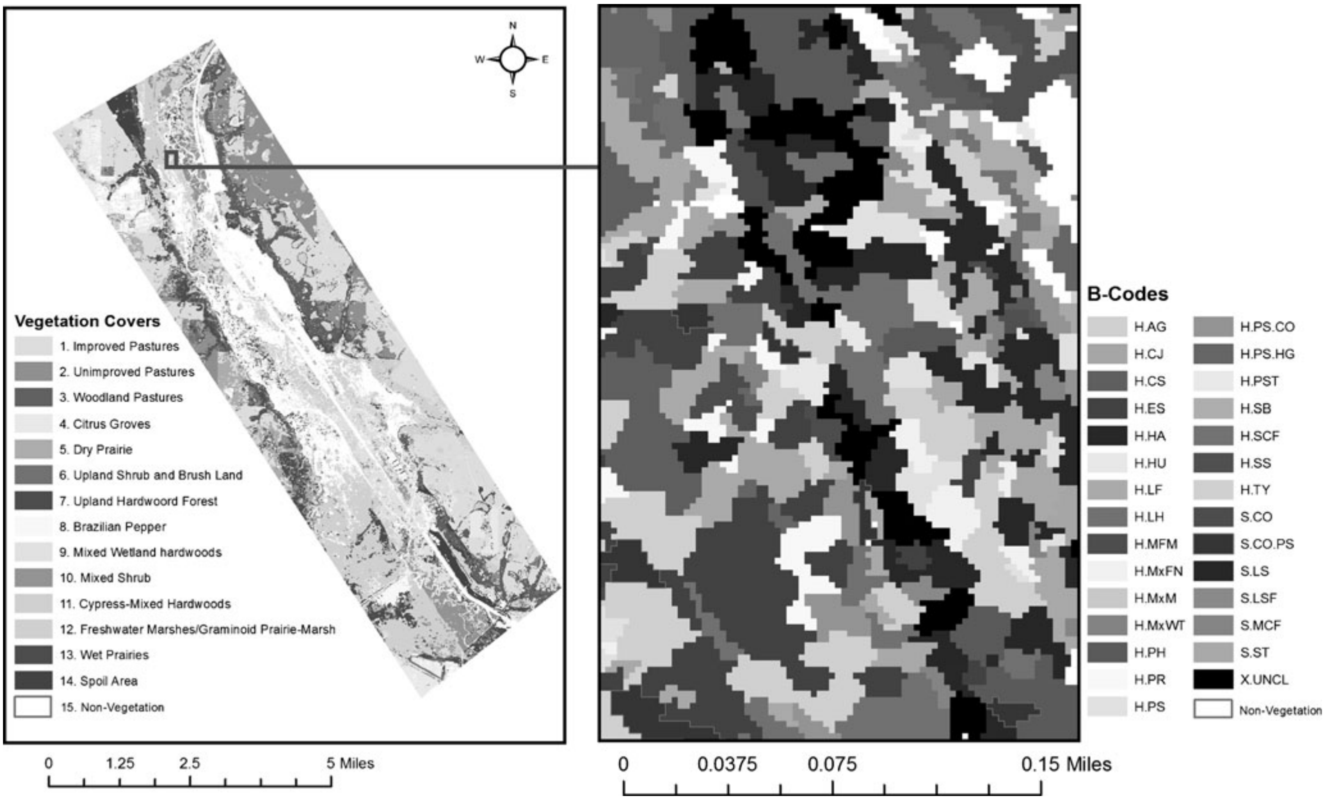

**Fig. 2** Vegetation community map (*left*) and vegetation B-Code map (*right*) from the SVM models

B-Codes, both SVM and RF performed well. Comparison of the SVM and RF results shows that some vegetation communities or species are more easily discriminated by RF, indicating an ensemble of these two classifiers may improve the mapping results.

### Applicability of Fine Spatial Resolution Hyperspectral Data in the KOE Ecosystem

It is difficult to map heterogeneous wetlands, especially at the species-level. We achieved high classification accuracy

**Table 4** Classification accuracy for each community from SVM and RF classifiers

| Community | NR | SVM | | RF | |
|---|---|---|---|---|---|
| | | PA | UA | PA | UA |
| 1. Improved pastures | 126 | 89 | 85 | 97 | 87 |
| 2. Unimproved pastures | 90 | 87 | 83 | 82 | 90 |
| 3. Woodland pastures | 112 | 93 | 95 | 95 | 96 |
| 4. Citrus groves | 32 | 94 | 94 | 94 | 94 |
| 5. Dry prairies | 20 | 40 | 57 | 30 | 60 |
| 6. Upland shrub and brush land | 10 | 100 | 100 | 80 | 100 |
| 7. Upland hardwood forests | 34 | 82 | 78 | 94 | 84 |
| 8. Brazilian pepper | 34 | 82 | 78 | 94 | 94 |
| 9. Mixed wetland hardwoods | 46 | 83 | 83 | 80 | 82 |
| 10. Mixed shrub | 42 | 91 | 87 | 81 | 90 |
| 11. Cypress-mixed hardwoods | 44 | 87 | 86 | 91 | 95 |
| 12. Freshwater marshes/Graminoid Prairie-Marsh | 310 | 95 | 96 | 98 | 92 |
| 13. Wet prairies | 54 | 69 | 83 | 67 | 82 |
| 14. Spoil area | 32 | 100 | 94 | 100 | 100 |
| Overall accuracy and Kappa value | | Accuracy: 89%<br>Kappa: 0.87 | | Accuracy: 90%<br>Kappa: 0.89 | |

*SVM* support vector machine
*RF* random forest
*PA* Producer's accuracy (%)
*UA* User's accuracy (%)
*NR* number of references

**Table 5** Classification accuracy for each B-Code from SVM and RF classifiers

| B-Code | NFS | SVM | | RF | | B-Code | NFS | SVM | | RF | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PA | UA | PA | UA | | | PA | UA | PA | UA |
| 1. H.AF | 6 | 100 | 90 | 90 | 81 | 29. H.PR | 6 | 100 | 91 | 90 | 71 |
| 2. H.AG | 5 | 87 | 100 | 87 | 87 | 30. H.PS | 10 | 96 | 83 | 100 | 64 |
| 3. H.CD | 5 | 71 | 41 | 42 | 100 | 31. H.PS-CO | 6 | 90 | 71 | 63 | 63 |
| 4. H.CJ | 5 | 62 | 83 | 25 | 100 | 32. H.PS-HG | 6 | 100 | 90 | 70 | 63 |
| 5. H.CS | 5 | 100 | 100 | 85 | 85 | 33. H.PS-PH | 6 | 100 | 91 | 81 | 90 |
| 6. H.EC | 6 | 81 | 90 | 100 | 84 | 34. H.PS-PH-CO | 8 | 92 | 100 | 85 | 80 |
| 7. H.EC-PST | 4 | 80 | 57 | 100 | 71 | 35. H.PST | 6 | 60 | 42 | 60 | 66 |
| 8. H.ES | 6 | 90 | 100 | 80 | 100 | 36. H.RN | 6 | 90 | 83 | 72 | 80 |
| 9. H.HA | 6 | 90 | 100 | 90 | 90 | 37. H.SB | 6 | 100 | 90 | 80 | 72 |
| 10. H.HG | 6 | 100 | 100 | 100 | 83 | 38. H.SCF | 6 | 90 | 76 | 72 | 80 |
| 11. H.HU | 6 | 36 | 80 | 63 | 77 | 39. H.SS | 6 | 100 | 100 | 100 | 83 |
| 12. H.JEd | 5 | 100 | 100 | 85 | 100 | 40. H.TY | 7 | 92 | 92 | 85 | 85 |
| 13. H.JEp | 5 | 100 | 100 | 100 | 100 | 41. S.CO | 7 | 92 | 92 | 85 | 75 |
| 14. H.LH | 6 | 80 | 72 | 70 | 77 | 42. S.CO-PS | 6 | 90 | 76 | 81 | 64 |
| 15. H.MFM | 6 | 90 | 83 | 63 | 87 | 43. S.CO-PS-PH | 6 | 80 | 66 | 60 | 100 |
| 16. H.MxE | 6 | 70 | 100 | 80 | 100 | 44. S.HF | 6 | 100 | 83 | 90 | 69 |
| 17. H.MxFA | 4 | 80 | 66 | 80 | 100 | 45. S.LS | 7 | 57 | 80 | 85 | 85 |
| 18. H.MxFN | 6 | 90 | 100 | 81 | 90 | 46. S.LSF | 6 | 63 | 77 | 36 | 80 |
| 19. H.MxM | 5 | 42 | 100 | 00 | 00 | 47. S.MC | 8 | 77 | 93 | 88 | 72 |
| 20. H.MxN | 7 | 92 | 81 | 85 | 80 | 48. S.MCF | 6 | 70 | 77 | 60 | 66 |
| 21. H.MxW | 12 | 92 | 90 | 94 | 82 | 49. S.MxFS | 7 | 92 | 86 | 92 | 68 |
| 22. H.MxWP | 7 | 85 | 92 | 78 | 84 | 50. S.SC | 6 | 100 | 78 | 90 | 90 |
| 23. H.MxWT | 6 | 54 | 66 | 27 | 50 | 51. S.SR | 4 | 100 | 100 | 80 | 100 |
| 24. H.NL | 5 | 42 | 60 | 57 | 100 | 52. S.ST | 4 | 40 | 50 | 20 | 100 |
| 25. H.PD | 6 | 100 | 73 | 63 | 70 | 53. V.LM | 7 | 85 | 92 | 85 | 66 |
| 26. H.PH | 6 | 80 | 100 | 80 | 88 | 54. V.MxV | 6 | 70 | 87 | 90 | 81 |
| 27. H.PN | 6 | 63 | 100 | 81 | 75 | 55. XUNCL | 6 | 90 | 100 | 81 | 100 |
| 28. H.PP | 6 | 100 | 83 | 100 | 100 | | | | | | |
| Overall accuracy and Kappa value | | | | | | | | Accuracy: 85% | | Accuracy: 79% | |
| | | | | | | | | Kappa: 0.85 | | Kappa: 0.79 | |

*SVM* support vector machine

*RF* random forest

*PA* Producer's accuracy (%)

*UA* User's accuracy (%)

*NFS* Number of field samples

at both community- and species-level by combining the fine spatial resolution hyperspectral data, MNF data transformation, and machine learning algorithms. Belluco et al. (2006) have indicated that the use of high spatial resolution dataset for vegetation mapping is particularly advantageous in heterogeneous wetland environments where such datasets can reduce the mixed pixel problem in classification. Data transformation is also important in hyperspectral applications. Although hyperspectral techniques are powerful for material identification, the redundant bands and inherent noises in the data may severely reduce classification accuracy. The selection of the most effective classification algorithms is also very important. Previous studies have shown that traditional classifiers such as maximum likelihood and minimum distance method are not effective in regions with high spatial heterogeneity (Zhang and Xie 2012a). Contemporary machine learning techniques are more effective in such environments.

Although overall high accuracies were obtained using hyperspectral vegetation mapping techniques, it is still difficult to detect and map communities or species that are in small or narrow patches with a width less than 3 m using the 3.5-meter resolution of HyMap data. This could be mitigated by fusing high spatial resolution aerial photography and hyperspectral imagery. The performance may also be improved by increasing the number of field samples to account for the heterogeneity of the study area.

Our study demonstrates the potential for fine spatial resolution hyperspectral data to replace aerial photography for vegetation mapping in heterogeneous wetland environments. One key aspect of hyperspectral technology is that the data collected can be dissected and reassembled in many different ways, allowing the same data collection effort to be used to address different questions. For example, data collected to assess the impacts of restoration on vegetation community may also be used for monitoring the spread of exotic vegetation. This would allow cost reduction through economy of scale and cost-sharing.

## Conclusions

For this study, we evaluated the applicability of fine spatial resolution hyperspectral data for vegetation mapping in the complex Kissimmee-Okeechobee-Everglades system. Two machine learning techniques, support vector machines (SVMs) and random forest (RF), were examined in the classification. We draw the following conclusions:

1) Fine spatial resolution hyperspectral data is a promising solution for vegetation mapping of the complex Kissimmee-Okeechobee-Everglades system to support the on-going restoration projects in South Florida. Accurate and informative vegetation maps were produced by combining machine learning classification methods, object-based mapping techniques, and data transformation. An overall accuracy of 90% was obtained for the community-level mapping, and an overall accuracy of 85% was achieved for discriminating 55 species.
2) The minimum noise fraction (MNF) data transformation is an important step in vegetation mapping using hyperspectral data. This preprocess removes the inherent noise in the hyperspectral data, improves the classification accuracy, and reduces the data dimensionality to decrease the computational cost.
3) The number of training samples can largely impact the classification results. In this study, both SVM and RF were sensitive to the number of training data in the classification, which is different from the findings reported elsewhere in the literature. The sensitivity is likely caused by the high spatial and spectral heterogeneity of the study area.
4) The number of classes could impact the classification accuracy. In general, the accuracy decreased when the number of classes increased.
5) The study indicates that integration of fine spatial resolution hyperspectral data and contemporary machine learning techniques shows potential for mapping heterogeneous wetland environments including the Greater Everglades.

**Acknowledgments** We appreciate the constructive comments and suggestions from two anonymous reviewers to improve this paper. The South Florida Water Management District (SFWMD) provided data for this research.

## References

Artigas FJ, Yang JS (2005) Hyperspectral remote sensing of marsh species and plant vigour gradient in the New Jersey Meadowlands. International Journal of Remote Sensing 26:5209–5220

Beaulieu JM, Goldberg M (1989) Hierarchy in picture segmentation: a stepwise optimal approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 11:150–163

Belluco E, Camuffo M, Ferrari S, Modenese L, Silvestri S, Marani A, Marani M (2006) Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. Remote Sensing of Environment 105:54–67

Benz U, Hofmann P, Willhauck G, Lingenfelder I, Heynen M (2004) Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58:239–258

Bousquin, SG, Anderson DH, Williams GE, Colangelo DJ (2005) Establishing a baseline: Pre-restoration studies of the channelized Kissimmee River. South Florida Water Management District, West Palm Beach, Florida, USA. Technical Publication ERA #432

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65:2–16

Breiman L (2001) Random forests. Machine Learning 45:5–32

Camps-Valls G, Bruzzone L (2005) Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 43:1351–1362

CERP, Comprehensive Everglades Restoration Plan (CERP), 2012. http//www.evergladesplan.org/, accessed on July 24, 2012.

Chan JC-W, Paelinckx D (2008) Evaluation of Random Forest and Adaboost tree based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment 112:2999–3011

Chang C-C, Lin C-J (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, http://www.csie.ntu.edu.tw/~cjlin/libsvm, accessed on May 22, 2012

Crawford MM, Ham J, Chen Y, Ghosh J (2003) Random forests of binary hierarchical classifiers for analysis of hyperspectral data. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 27-28 October 2003, pp. 337–345.

Congalton R, Mead RA (1983) A quantitative method to test for consistency and correctness in photointerpretation. Photogrammetric Engineering and Remote Sensing 49:69–74

Demir B, Erturk S (2009) Increasing hyperspectral image classification accuracy for data sets with limited training samples by sample interpolation, Recent Advances in Space Technologies, pp 367–369.

Foody GM (2004) Thematic map comparison, evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing 70:627–633

Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random Forests for land cover classification. Pattern Recognition Letters 27:294–300

Gualtieri JA, Cromp RF (1998) Support vector machines for hyperspectral remote sensing classification. In Proceedings of the 27th AIPR Workshop, Advances in Computer Assisted Recognition, Washington, D.C, 27 October. SPIE, pp 221–232

Green AA, Berman M, Switzer P, Craig MD (1988) A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing 26:65–74

Hall M, Frank E, Holmes G, Pfahringer B, Reutmann P, Witten I (2009) The WEKA data mining software, an update. SIGKDD Explorations 11:1–18

Ham J, Chen Y, Crawford MM, Ghosh J (2005) Investigation of the Random Forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43:492–501

Harken J, Sugumaran R (2005) Classification of Iowa wetlands using an airborne hyperspectral image: a comparison of the spectral angle mapper classifier and an object-oriented approach. Canadian Journal of Remote Sensing 31:167–174

Hirano A, Madden M, Welch R (2003) Hyperspectral image data for mapping wetland vegetation. Wetlands 23:436–448

Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. International Journal of Remote Sensing 23:725–749

Hunter EL, Power CH (2002) An assessment of two classification methods for mapping Thames Estuary intertidal habitats using CASI data. International Journal of Remote Sensing 23:2989–3008

Hsu C, Chang C, Lin C (2010) A practical guide to support vector classification. Final report, National Taiwan University, Taipei City, Taiwan

Johnson B, Xie Z (2011) Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS Journal of Photogrammetry and Remote Sensing 66:473–483

Jollineau MY, Howarth PJ (2008) Mapping an inland wetland complex using hyperspectral imagery. International Journal of Remote Sensing 29:3609–3631

Jones JW (2011) Remote sensing of vegetation pattern and condition to monitor changes in Everglades biogeochemistry. Critical Reviews in Environmental Science and Technology 41:64–91

Kamal M, Phinn S (2011) Hyperspectral data for mangrove species mapping: a comparison of pixel-based and object-based approach. Remote Sensing 3:2222–2242

Landis J, Koch GG (1977) The measurement of observer agreement for categorical data. Biometics 33:159–174

Lawrence RL, Wood SD, Sheley RL (2006) Mapping invasive plants using hyperspectral imagery and Breiman and Cutler classifications (Random Forest). Remote Sensing of Environment 100:356–362

Li L, Ustin SL, Lay M (2005) Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping, a case study in China Camp, CA, USA. International Journal of Remote Sensing 26:5193–5207

Liu D, Xia M (2010) Assessing object-based classification, advantages and limitations. Remote Sensing Letters 1:187–194

Lowe Engineers, SAIC (2003) Kissimmee River restoration remote sensing pilot study project, Atlanta, GA, USA

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

Mountrakis G, Im J, Ogole C (2010) Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66:247–259

Pal M, Mather PM (2006) Some issues in the classification of DAIS hyperspectral data. International Journal of Remote Sensing 27:2895–2916

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67:93–104

Rutchey K, Schall T, Sklar F (2008) Development of vegetation maps for assessing Everglades restoration progress. Wetlands 28:806–816

Schmidt KS, Skidmore AK, Kloosterman EH, Vanoostern H, Kumar L, Janssen JAM (2004) Mapping coastal vegetation using an expert system and hyperspectral imagery. Photogrammetric Engineering and Remote Sensing 70:703–715

Trimble, 2011. eCognition Developer 8.64.1 reference book.

van der Linden S, Janz A, Waske B, Eiden M, Hostert P (2007) Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines. Journal of Applied Remote Sensing, doi:10.1117/1.2813466, 1, No. 013543.

Vapnik VN (1995) The Nature of Statistical Learning Theory. Springer, New York

Waske B, Benediktsson JA, Árnason K, Sveinsson JR (2009) Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Canadian Journal of Remote Sensing 35:S106–S116

Yang C, Everitt JH, Johnson HB (2009) Applying image transformation and classification techniques to airborne hyperspectral imagery for mapping Ashe juniper infestations. International Journal of Remote Sensing 30:2741–2758

Zhang C, Xie Z (2012a) Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery. Remote Sensing of Environment 124:310–320

Zhang C, Xie Z (2012b) Data fusion and classifier ensemble techniques for vegetation mapping in the coastal Everglades. Geocarto International, in press