Remote Sens. **2015**, *7*, 1-x manuscripts; doi:10.3390/rs70x000x

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

An Empirical Study of ATCOR vs. FLAASH Atmospheric Correction for Hyperspectral Classification of Tree Species Using SVM

Morteza Shahriari Nia[1,*], Daisy Wang[1], Milenko Petrovic[2], Stephanie Bohlman[3] and Paul Gader[1]

[1] Department of Computer and Information Science and Engineering, University of Florida, 432 Newell Dr., Gainesville, Florida 32611; E-Mails: msnia@cise.ufl.edu; daisyw@cise.ufl.edu; pgader@cise.ufl.edu;

[2] Institute for Human and Machine Cognition, 15 SE Osceola Ave, Ocala, Florida 34471; E-Mail: mpetrovic@ihmc.us;

[3] School of Forest Resources and Conservation, University of Florida, 349 Newins Ziegler Hall, Gainesville, Florida 32611; E-Mail: sbohlman@ufl.edu;

* Author to whom correspondence should be addressed; E-Mail: msnia@cise.ufl.edu.

Received: / Accepted: / Published:

Abstract: Identifying savannah species at ecological scale is a key step in measuring biomass, carbon reserves, drought and invasive species spread predictions. In this paper we perform classification and geo-mapping of tree species from hyperspectral imagery collected using AVIRIS airborne sensors at pixel level. We provide a thorough comparison of the effects of ATCOR and FLAASH atmospheric corrections in prediction accuracy. We exploit Gaussian Filters to eliminate sensor measurements and calibration errors, to the best of our knowledge we are the first in employing Gaussian Filters for hyperspectral species classification. The area of study is Ordway-Swisher Biological Station in north-central Florida. Due to the structure of the collected field data, we realized that applying NDVI and NIR filters do not play constructive roles in classification. Species classification was performed using variety of Support Vector Machine kernels where Radial Basis outperformed others. Our classification produces accurate predictions of about 75%.

Keywords: Species classification; Atmospheric Corrections; FLAASH; ATCOR; Ordway-Swisher Biological Station; National Ecological Observatory Network;

---

1. Introduction

Mapping tree species by remote sensing techniques is an essential step in understanding how planetary species play roles at ecological scale. This will enable us to study land covers, climate change, invasive species, plant competitions, field fire potentials and spreading routes, soil characteristics among others [1,2]. This kind of research has only been possible via the technological advancements in remote sensing facilities such as hyperspectral imagery or Light Detection and Ranging (LiDAR). Various studies have dealt with identifying tree species at both pixel level and crown level and as technology becomes available and economically feasible, studies tend to cover larger areas. Carnegie Airborne Observatory[1] (CAO) is a major pioneer in employing airborne technology for remote sensing at ecology scale where they study large areas in Amazonians, Kruger National Park in South Africa, and Madagascar among others. Colgan et al. [1] uses a two stage Support Vector Machines (SVM) at pixel level and at crown level for tree species classification where LiDAR measurements were used for crown segmentation. Féret and Asner [3] study the accuracy of various parametric/non-parametric supervised classification techniques and observed that there is a clear advantage in using Regularized Discriminant Analysis, Linear Discriminant Analysis, and Support Vector Machines.

Cho et al. [3] compares accuracies when different hyperspectral sensors of CAO, WorldView2 and QuickBird are utilized by convolving the 72 bands of CAO to eight and four multispectral channels available in the WorldView-2 and QuickBird satellite sensors, respectively. Interestingly enough they observed that WorldView-2 produced more accurate classification results then QuickBird and finally CAO. Clark et al. takes on another perspective and compares lab measurements to pixel and to crown level and try to identify important wavelength regions for species discrimination [4]. They observed that optimal regions of the spectrum for species discrimination varied with scale. However, near-infrared (700-1327nm) bands were consistently important regions across all scales. Bands in the visible region (437-700nm) and shortwave infrared (1994-2435nm) were more important at pixel and crown scales [4]. Clark et al. in another work evaluates the applicability of different parameters (indexes, derivatives, signals intensities) for classification [5]. There are other tree species classification efforts such as [3,6-11] that share the same approach with minor variations.

There are other schools of thought that try to identify more context specific features of remote sensing. For example Baldeck and Asner [12] try to measure how similar beta diversity of regions are; they use distance measures such as Euclidean distance and K-means clustering in unsupervised models. Using these clustering techniques one can provide a quick understanding of beta diversities and avoid costly and time consuming field data collections. However, this line of research needs more work as about 50% of pixels are classified as *other*, therefore any conclusion at this scale of uncertainty is not necessarily helpful, the same holds in Baldeck et al.'s latter work in [13]. Sometimes specially tailored tools and methodologies in this field are necessary. As an example, one should note that different bands in a hyperspectral image have different signal to noise ratios, and Principal Components (PC) transform will not always result in components with a steadily increasing noise level. This makes setting a cut-off point difficult. Minimum Noise Fraction (MNF) [14] is a modified

[1] http://cao.stanford.edu/

PC transform which produces a set of principal component images ordered in terms of decreasing signal quality.

There has been plenty of research investigating the effects of different atmospheric correction methods for Landsat TM [15,16], QuickBird [17], and many other imaging spectrometer data [18,19]. But unfortunately none of them addresses the impact of atmospheric correction for plant species classification. The main goal of such papers are usually bound to low level signal manipulation, estimation and at most comparison to simple ground data such as asphalt or gravel. As it turns out, the final goal of atmospheric correction being species classification (in this context) is neglected in this regard. There are several approaches at atmospheric correction, some include scene-derived adjustments, in which in-scene statistics are used, such as the Darkest Pixel method [17,20], or purely empirical methods where ground-recorded spectral data are required, e.g., the Empirical Line [21,22]. Some involve radiative transfer algorithms such as the 6S code and MODTRAN, and others use in situ spectral data (model based) such as ATCOR and FLAASH. Radiative transfer models proved to be a cost and time effective approach, wherever ground data are missing, but for our purpose as such data is collected, we focus on ATCOR and FLAASH. Our results demonstrate similar performance by state-of-the-art comparison of the two as in [23], where they focused on basic endmember classification (asphalt, gravel, rocky areas, reddish soil, agricultural areas, grass/dry grass, mequis, and phrygana) but we look into the actual problem of tree species classification.

In this paper we perform species classification and geo-mapping. In our approach we take on a novel perspective at hyperspectral species classification by taking into account the impact of atmospheric correction and Gaussian filtering of reflectance intensities for noise reduction. Airborne data is part of a pilot data collection phase from National Ecological Observatory Network (NEON). NEON is a long-term ecology-monitoring project for discovering, understanding and forecasting impacts of climate change, land use change, and invasive species at continental-scale. National Science Foundation provides funding for NEON as a 30-year project starting 2016. Local ecological measurements at sites distributed within 20 eco-climatic domains across the contiguous United States, Alaska, Hawaii, and Puerto Rico will be coordinated with high resolution, regional airborne remote sensing observations [24]. NEON will have an additional 40 re-locatable terrestrial sites (every five to seven years) as well as 36 aquatic sites[2]. Airborne Observation Platform (AOP) is the remote sensing platform with equipment of meter/sub-meter resolution for hyperspectral and LiDAR measurements. This paper is a pilot study on the pre-mission airborne hyperspectral data collection. No operation at the scale and time span predicted to be fulfilled for NEON (neither geographical and timely span) has ever been carried out before [25].

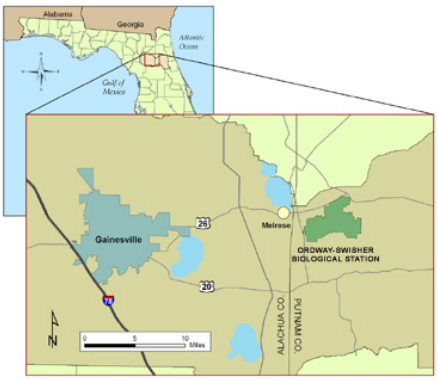
2. Data Collection

The NEON Southeast Domain 3 contains the southern portions of the Gulf Coast states, half of South Carolina, and all of Florida except for the southern tip. The candidate core site for Domain 3 is located at the Ordway-Swisher Biological Station (OSBS)[3] which is an 37 km^2 area in Putnam County

[2] http://www.neoninc.org/science/domains#sthash.aW1THj1N.dpuf

[3] http://ordway-swisher.ufl.edu/index.htm

Figure 1. Location of Ordway-Swisher Biological Station



in north-central Florida and is managed jointly by the University of Florida and the Nature Conservancy (Figure 1). OSBS features diverse natural forests, small pine plantations nearby, a range of wildlife species that reflects the area's ecological communities along with a 75-year history of low human impact. Nine major plant communities exist within the region (defined by the Florida Natural Areas Inventory) and these diverse targets are populated by sandhill, xeric hammock, upland mixed forest, baygalls, basin swamp, basin marsh, marsh lake, clastic upland lake and sandhill upland lakes. The sandhill community is managed using prescribed burning on a scheduled 3-year rotation. The ground sampling part of this campaign focused on a sandhill ecosystem dominated by Long-Leaf Pine (*Pinus Palustris*) and Turkey Oak (*Quercus Laevis*) [25,26].

Since the instrumentation slated for deployment on the eventual AOP remote sensing payloads were not yet available, airborne spectroscopic and LiDAR measurements were made during this campaign using existing systems that exhibit similar performance characteristics as the instrumentation under development [26]. It is important to note that the actual system for NEON will have better conformance of hyperspectral/LiDAR integrations with better spatial resolution. However here we only focus on hyperspectral data.

AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) operated by the Jet Propulsion Laboratory (JPL) deployed on a Twin Otter DeHavilland DHC-6-300 aircraft in partnership with the National Aeronautics and Space Administration Terrestrial Ecology Program was used to collect data. JPL has flown on two separate days over OSBS: morning of September 4, 2010 and mid-day of September 10, 2010. Both of the flights were flown at approximately 4000m AGL at approximately 90 knots with zenith angle of 180.0 and azimuth angle of 0.0 with Altitude of 13kft (SOG ~ 65-91kts)[4] at mostly clear with some haze for Sept. 4th and some puffy clouds for Sept. 10th. Details of these flights can be seen in Figure 2. Dependent on flight line, pixel size ranges from 3.3m to 3.6m. Hyperspectral data was atmospherically corrected using FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) [27] and ATCOR (ATmospheric CORection) [28] algorithms. There are altogether 8 flight lines the total size of which is 20.5GB for FLAASH and 10.2GB for ATCOR data. There are 224 bands recorded with wavelengths from 365.93 nm to 2496.24 nm.

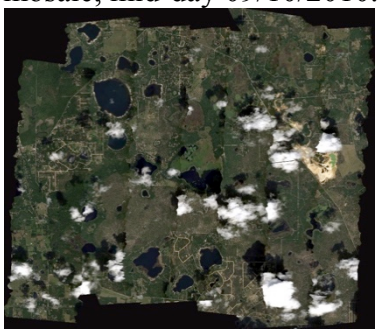
[4] NASA JPL AVIRIS flight f100904t01 details:
http://aviris.jpl.nasa.gov/cgi/flights_10.cgi?step=view_flightlog&flight_id=f100904t01
http://aviris.jpl.nasa.gov/cgi/flights_10.cgi?step=view_flightlog&flight_id=f100910t01

Figure 2. JPL AVIRIS flights over OSBS [25]

(**a**) Flights ground tracks.

(**b**) Hyperspectral true-color mosaic, morning 09/04/2010.

(**c**) Hyperspectral true-color mosaic, mid-day 09/10/2010.



Atmospheric characterization relied on measurements of a CIMEL sun photometer in coordination with the NASA Aerosol Robotic Network[5]. Measurements were collected on September 4, 2010 and the derived atmospheric information was used to improve the atmospheric correction of the AVIRIS spectrometer data. Detailed measurements such as aerosol optical thickness, water vapor, and etc are available online[6].

3. Species Classification

Essential steps after performing flights were ortho-rectification and atmospheric correction which were done bye NEON. With resolution of images being about 3 by 3 square meters we do not get pure pine or oak signals and there is lots of linear/nonlinear mixture of endmembers in each pixel (e.g. road, shadow, branch, under-score vegetation, multiple tree species and etc.). The timing of fights (September) adds to the challenge: leafs might not be as green or some trees might have already started to lose leaves and potentially leads to getting more branch or dead leaf signals. Unlike broadleaf trees where signal returns can be more accurate and wholesome, Longleaf Pine is a conifer (needle leaf), which adds to the complexity.

2.1. Field Data

Geo-identifying tree species as ground data was collected on February 28th, 2014. A laptop preloaded with ArcMap and flight data loaded in ENVI software package was used in conjunction to a professional grade GPS. ArcMap read GPS coordinates and mapped the polygons in ENVI. In this way, we marked several geo-polygons that had similar plant species in ENVI. Later on we overlaid the identified polygons with proper JPL AVIRIS flight, which had the least amount of clouds. This approach works pretty accurate if you are not in a dense forest such as tropical forests where GPS signal under the tree canopy has high deviations due to NLOS (no-line of sight) of GPS signals. One should note that even with these considerations, commercial GPS have sub-meter accuracy and when combined with error accumulated in ortho-rectification of flight images we still need to mark several

[5] NASA AERONET: http://aeronet.gsfc.nasa.gov/cgi-bin/bamgomas_interactive

[6] OSBS Aero Measurements:
http://aeronet.gsfc.nasa.gov/cgi-bin/type_one_station_opera_v2_new?site=Ordway-Swisher

land marks such as roads, large trees or other land marks to be able to re-verify marked points in the map and avoid shifts in coordinates. Altogether we identified species for 1269 pixels. In Table 1 you can find the details of identified Regions of Interest (ROIs) along with their details. The closer ROI Ids are collected at a closer vicinity (geographically/temporally), and the more distant ROI ids means that canopies are more apart. You can see that some species have lots of abundance (334 for Turkey Oak) and some have few (81 for Laurel Oak) this bias in population size inadvertently affect classification accuracy. Figure 3 depicts the abundance of each canopy respectively.

The Oak (other) species represent generic oak specie (*specie details unknown*). Live Oak is specifically Sand Live Oak (*Quercus Geminata*), Laurel Oak represents *Quercus Hemisphaerica*, Turkey Oak is *Quercus Laevis*, Longleaf Pine is *Pinus Palustris*, and Pine (other) represents a mixture of different varieties of pines: Longleaf Pine (*Pinus Palustris*), Loblolly Pine(*Pinus Taeda*) or Slash Pine (*Pinus Elliotti*).

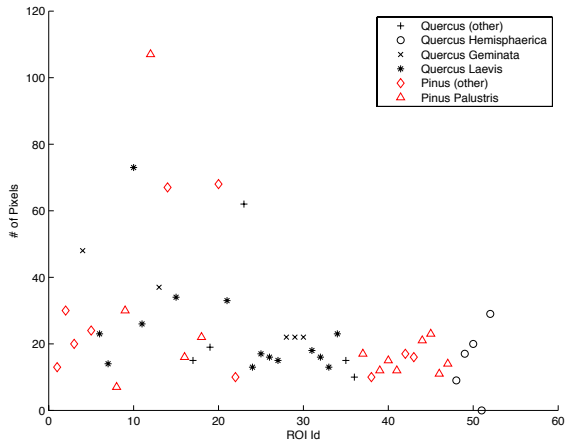
2.1. Atmospheric Correction

NEON performs the atmospheric correction and you can find the details of measurements and in *OSBS Aero Measurements* as mentioned above. Ground spectral measurements were performed concurrently with flight lines. FLAASH atmospheric correction is approximated as in Equation 1 [29]:

Table 1. Field Data Specifications

	Broadleaf								**Conifer**			
Specie	Oak (other)		Laurel Oak		Live Oak		Turkey Oak		Pine (other)		Longleaf Pine	
	ROI Id	Pixels	ROI Id	Pixels	ROI Id	Pixels	ROI Id	Pixels	ROI Id	Pixels	ROI Id	Pixels
	23	62	52	29	4	48	10	73	20	68	12	107
	19	19	50	20	13	37	15	34	14	67	19	30
	17	15	49	49	28	22	21	33	2	30	45	23
	35	15	48	9	29	22	11	26	5	24	18	22
	36	10	53	6	30	22	6	23	3	20	44	21
							34	23	42	17	37	17
							31	18	43	16	16	16
							25	17	1	13	40	15
							26	16	22	10	47	14
							32	16	38	10	39	12
							27	15			41	12
							7	14			46	11
							24	13			8	7
							33	13				
Total	5	121	5	81	5	151	14	334	10	275	13	307
Partial Total	29 ROIs – 687 Pixels								23 ROIs – 582 Pixels			
Grand Total	52 ROIs – 1269 Pixels											

Figure 3. Field Data Distribution of ROIs.



$$L_e \approx \left(\frac{(A+B)\rho_e}{1-\rho_e S} + L_a\right) \tag{1}$$

where ρ_e is an average surface reflectance for the pixel and a surrounding region, S is the spherical albedo of the atmosphere, L_a is the radiance back scattered by the atmosphere, and A and B are coefficients that depend on atmospheric and geometric conditions but not on the surface.

ATCOR performs atmospheric correction as follows [30]:

$$\rho = \frac{\pi\{d^2(c_0 + c_1 DN) - L_{path}\}}{\tau E_g} \tag{2}$$

where τ is the atmospheric (direct or beam) transmittance for a vertical path through the atmosphere, d is the earth-sun distance in astronomical units, c_0 , c_1 and DN are the radiometric calibration offset, gain, and digital number, respectively. ρ is the surface reflectance and E_g is the global flux on the ground.

2.1. Preprocessing

We load hyperspectral images in Matlab using an in-house upgraded version of `enviread`, initially developed by Dr. Ian Howat at Ohio State University [31]. We check for the consistency of calibration and uniformness of pixel sizes and noticed a range of 3.3m to 3.6m pixel sizes due to various flight and measurement conditions. As different flights have different altitudes and hence pixel resolutions, this is an essential step to account for. Here we define a hyperspectral image I with dimensionality (x,y,w,z), where $x \in X = [167000, 833000]$ represents the range of UTM Easting values, $y \in Y = [0, 9400000]$ represents UTM Northing values, $w \in W = \{1, \cdots, 224\}$ is the index of the reflectance wavelengths, and $z \in Z = \{1, \cdots, 60\}$ is the UTM zone of the image. Based on our observations, we take constant $\xi = 10000$ as a cut-off point to avoid erroneous sensor readings. There are some noises in JPL AVIRIS measurements such as negative reflectance values; the range of hyperspectral reflectance values is in range $[-32762, 32724]$: one should note that reflectance is the proportion of sun radiance signals which should be a positive value, but in normalized form reflectance is between zero and one. Below you can see the normalization process:

$$I_{xywz} = \begin{cases} 0 & for\ I_{xywz} < 0 \\ 1 & for\ I_{xywz} > \xi \\ \sqrt{\frac{I_{xywz}}{\xi}} & otherwise \end{cases} \quad (3)$$

To normalize we set negative reflectance values to zero and values greater than 10,000 to 10,000. To enhance the intensity of readings we take the square-root of signal returns. This procedure is due to the following facts: a) There is no standard output of reflectance data and even reflectance of a single crown at different pixels can be quiet different. Unlike minerals that have fixed and known reflectance values, trees can different signal returns based on generation, number of leaf layers for the crown, leaf formation and orientation, leaf area index, condition of growth (water quality, climate, soil) and etc which can make this task challenging. b) Due to the resolution of images signals are all mixtures of several endmembers. Due to these reasons, empirically obtained thresholds and ranges are inevitable. Regarding square root, one should note that without taking the square root the images lack proper day-light intensity and appear dark.

We exclude wavelengths corresponding to strong water vapor absorption bands in the atmosphere. This includes 1333.2 nm to 1482.7 nm, 1791.6 nm to 1967.4 nm, and 2406.9 nm to 2496.2 nm. Due to strong absorption at those wavelengths, a small radiance signal is measured by the instrument.

2.1.1. Non-vegetated/Shaded Pixels

A filter of NIR < 0.33 excludes heavily shaded samples which usually have distorted reflectance signals [1]. Normalized Difference Vegetation Index (NDVI) is an index which shows how green a pixel is and is usually used to remove material that does not belong to planetary material such as roads, clouds and any not-vegetated area, even grass and so on. NDVI is defined as below:

$$NDVI = \frac{NIR - VIS}{NIR + VIS} \quad (4)$$

where is NIR the reflectance in the reflective near-infrared wavelengths (725-1100 nm) and VIS is the reflectance in the visible (red) wavelengths (580-680 nm). The principle behind this is that VIS is in a part of the spectrum where chlorophyll causes considerable absorption of incoming radiation, and the NIR is in a spectral region where spongy mesophyll leaf structure leads to considerable reflectance [32,33]. For this purpose we chose the band at 665.6 nm for red and 734.1 nm for near-infrared. By filtering out pixels with NDVI < 0.4 we are essentially removing pixels that are not green. By properly applying an NDVI filter globally to a whole flight line, you will end up with major tree crowns. However, in our scenario we observed that preserving low NDVI/NIR pixels from ground data actually helps prediction accuracy. As you can see in Table 1, we have tens of pixels per ROI and considering that each pixel is about 3m wide and the fact that species of this study do not have wide crowns; this means that ROIs are generously marked and span across several neighboring tree crowns. This includes the empty spaces between crowns, shady pixels, understory grass, and etc. The initial expectation is that to properly classify tree species we should get rid of low NDVI pixels from such large canopies. Contrary to this belief due to the mixing nature of hyperspectral pixels and aggregate

operation of classification techniques in multi-dimensional space, removing low NDVI pixels from *tree canopies* degraded classification performance by about 10%. So it is better to preserve low NDVI pixels in this scenario as they are mixed in with actual green pixels of trees and provide added value in the aggregate operation of SVM.

2.1.2. Gaussian Filter

Real-life sensor measurements are far from perfect and there are many noisy readings along different bands. We take advantage of abundance of bands and exploit their local-aggregated information by applying Gaussian filter. We take a Gaussian window w of size $N > 0$, the coefficients of the Gaussian window are computed as below:

$$w(n) = e^{-\frac{1}{2}(\alpha\frac{n}{N/2})^2} \tag{5}$$

where $-\frac{N-1}{2} \leq n \leq \frac{N-1}{2}$, and α is inversely proportional to the standard deviation (σ) of a Gaussian random variable ($\sigma = \frac{N}{2\alpha}$). Once we have Gaussian parameters we perform convolution to apply the smoothing factor. By convolving vectors $u \in R^m$ and $v \in R^n$, we will have vector $w \in R^{m+n-1}$ such that:

$$w(k) = \sum_j u(j)v(k-j+1) \tag{6}$$

2.2. Support Vector Machine

It is well known in literature that Support Vector Machines (SVM) outperforms other algorithms on species classification [1,13,34]. Our focus is on the impact of atmospheric correction (specifically FLAASH vs ATCOR) at pixel level classification using non-linear multi-class SVM. We parameterize SVM with k-fold cross validation where $k = 5$. Classifier non-linearity comes from taking the following non-linear functions as kernel for SVM:

- Polynomial function kernel.
- Radial Basis Function (RBF) kernel

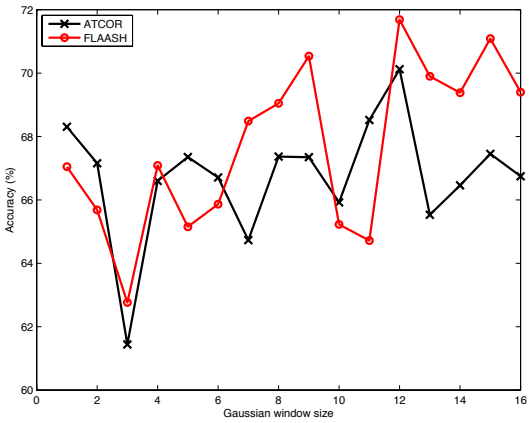
Regarding multi-class classification, we create $\binom{c}{2}$ classifiers where c is the number of classes. In this case $c = 6$, hence we train 15 disjoint binary classifiers at each iteration of k-fold. We train all the classifiers and to decide on the class of a given test pixel, majority voting among classifiers decides which class it belongs to. We avoid aggregating classes of canopy pixels to determine the class of the whole canopy. This process can be performed using well known approaches such as majority of pixels or etc, which is not relevant to the purpose of this paper. Furthermore, the assumption that we always know the exact boundaries of all single-tree canopies does not always hold, specifically in dense vegetation, so we leave this out from our discussion.

3. Results and Discussion

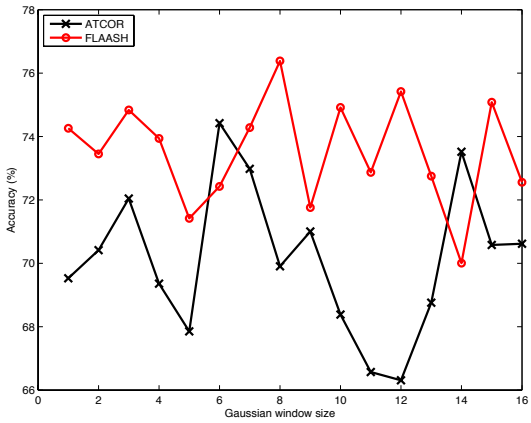
In this part we empirically evaluate the impact of optimizing classifier parameters regarding to FLAASH and ATCOR atmospheric corrections. As atmospheric correction is a low level signal

Figure 4. Impact of Gaussian Window on Prediction Accuracy.

(**a**) Before removing water absorption bands.



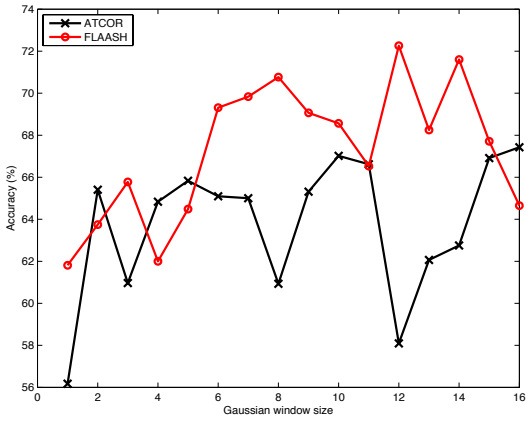
(**b**) After removing water absorption bands.



processing technique, our focus is on the impact of data pre-processing filters on the performance of the species classification while we compare the two. An implementation of a one-vs-one multi-class k-fold cross-validation setup of SVM using non-linear kernels (polynomial and radial basis) functions is used for our model. Majority vote among classifiers determines the specie of a pixel. As for some of our species we only have at most 5 canopies and we cannot take $k > 5$ for a correct implementation of k-fold cross validation therefore we set $k = 5$. Test and train sets are canopy-aware, meaning that pixels of a single canopy are either used for training or test sets and not both. As pixels of one canopy can have many similarities, it is not suggested to use several pixels of a canopy for train and other pixels of the same canopy for test purposes. This could have the same meaning as using the same data both for train and test, which leads to over fitting and erroneous accuracy results. A mixture of grid search and heuristic optimization determines various parameters for SVM kernels, but for brevity we only present 2D diagrams here as other parameters settings are out of scope of this project. Specifically we set a cost of $C = \infty$ for misclassification, meaning that we have little to zero tolerance for improperly classified samples.

In Figure 4 we show how removing water absorption bands helps us improve prediction accuracy. Before removing these bands (Figure 4.a) prediction accuracy starts at about 66% and rises to about 71% as we increase Gaussian window size (about 5% performance improvement). There is a positive trend across different window sizes, which means we can expect better accuracy the bigger the window size gets. This stems from the fact that water absorption bands add to randomness and as we increase window size this randomness is dissolved to the majority of neighbors and would have less and less impact as window size increases. After removing water absorption bands (Figure 4.b) prediction accuracy starts at 74% and rises to about 76.5% at a window size of 8. Beyond that there is no significant change in accuracy. As window size increases you can see that there is no predictable impact of Gaussian smoothing in accuracy; This is due to the fact that all the currently available data signals has some level of inherent entropy (less noisy) and increasing window size beyond certain point just takes data away from their original meaning which can either help or harm accuracy in unpredictable ways. Before removing water absorption bands FLAASH shows minor advantages over

Figure 5. Impact of removing low NDVI and NIR pixels among field data



ATCOR, but after removal this difference becomes a meaningful 4% better prediction accuracy for FLAASH.

Next, we look at the impact of removing low NDVI-NIR pixels from dataset. As you can see in Figure 5 there is a general degradation of performance from the previous scenario that we used the whole dataset. In FLASSH dataset we are at about 70% and ATCOR yields 66% accuracy. This is due to the fact that pixel size is large (3m) and canopies are large as well. According to Table 1, canopy sizes are in the range of [6,107] pixels. This generous marking of canopies includes areas with little greenness, shadows, branches, gravel and etc (all with low NDVI and low NIR values). But you should note that due to the mixing nature of reflectance values, even low NDVI/NIR pixels of a continuous canopy, still contains signals from the underlying species, which might not be as green. Here we can see that removing low NDVI/NIR pixels of a continuous canopy actually degrades the performance of the classification model with an impact of about 4%. So in similar scenarios where *field data* consists of large continuous canopies we advise on preserving low NDVI/NIR pixels. It is important to realize that this is only relevant in the context of field data, as we already know the species are there, and is only useful in training the classifier. In global application of NDVI/NIR filters to whole flight lines we suggest as literature does which is removal of low values to get rid of roads, grass, etc. Here we can still observe the benefit of FLAASH data versus ATCOR on average by 2%-3%.

Finally, we look at the effect of atmospheric correction on prediction accuracy vs classification model parameters. C, σ, P, MaxIter and o*ptimization method* are the knobs of SVM that we tune in a mixture of grid and heuristic search. C stands for cost or penalty of misclassification against simplicity of the decision surface, σ (or also known as γ in literature) defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close' in RBF function. P is the polynomial degree for polynomial function as kernel. MaxIter is the maximum number of iterations the optimization function is supposed to run; and *optimization method* defines what optimization method we select. Here we set C=+∞, MaxIter=10,000 and use quadratic programming as *optimization method*. In what follows we evaluate the impact of P and σ respectively as demonstrated in Figure 6.

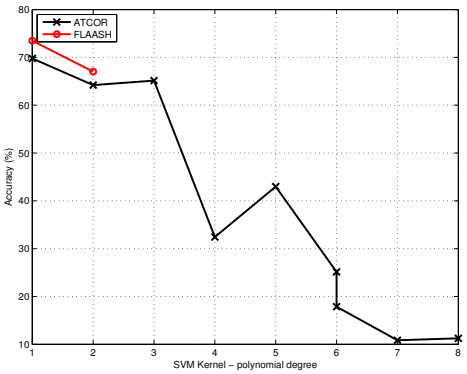
Figure 6.a demonstrates the impact of polynomial degree P on prediction accuracy using a polynomial kernel in SVM. FLAASH atmospherically corrected data yields 73.5% of accuracy

whereas ATCOR results in 69.8%. The simpler the polynomial, the better the performance; with more complicated polynomial we get a high bias classification model which performs poorly when evaluated on test data. For FLAASH atmospheric correction ($P \geq 3$) the optimization function does not converge. On the other hand, ATCOR data performs as predicted. Accuracy gets as low as 10.8% and 11.3% with polynomial degrees of 7 and 8 (due to extreme over-fitting).

The Radial Basis kernel has better performance yields than polynomial. As shown in Figure 6.b, with a peak at 75.3% we achieve the best results using FLAASH data while ATCOR comes close at 74.9%. RBF does not show good performance at either too low or too high σ values. σ is the inverse of the width of the RBF kernel (roughly defining the area of influence of a support vector); in other terms, it defines how much influence a single training example has. The larger σ is, the closer other examples must be to be affected. As RBF takes data to a higher dimensionality a small σ gives you a pointed bump in the higher dimensions, and a large σ gives you a softer, broader bump. So neither extreme shows a good fit and a middle point of σ =10,000 provides best results. On the negative side, FLAASH begins with an accuracy of 0% and ATCOR at 19.9%, but they quickly get to a stable region close by to each other while FLAASH demonstrates superior performance in most of the cases. There is a good range of σ values ([10, 10000]) where we reach a plateau in prediction accuracy, which implies a more robust performance than polynomial.

Figure 6. Parameter tuning for classification algorithms.

(**a**) Tuning P for SVM with polynomial kernel function.



(**b**) Tuning σ in SVM with RBF kernel function.

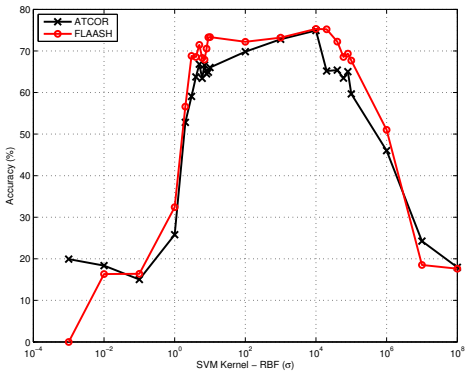


Table 2. Add a descriptive label of the table here.

		Predicted Class						
		Pine (other)	Longleaf Pine	Turkey Oak	Live Oak	Oak (other)	Laurel Oak	Sum
Known Class	Pine (other)	10	10	3	0	0	0	23
	Longleaf Pine	20	36	0	0	0	4	60
	Turkey Oak	1	0	62	0	0	0	63
	Live Oak	0	0	44	4	0	0	48
	Oak (other)	0	0	0	0	1	9	10
	Laurel Oak	1	0	0	0	0	5	6
	Sum	32	46	65	44	5	18	210

Table 2 demonstrates the confusion matrix of the proposed classification model at a near peak setup (75.2%). You can see that the majority of misclassifications are between Pine (other) class and Longleaf Pine. Similar misclassification can be found between Oak (other) and other types of Oak. This is due to the fact that this class of Oak or Pines is a mixture of different types of Oak or Pine respectively, which might include Longleaf Pine or Laurel Oak and such conformance might be unavoidable. The main advantage of this experiment is that there is mere misclassification between oak vs. pine category. The oak category is rarely misclassified as pine, but pines have been mistaken with Turkey Oak and Laurel Oak. One should also note that different species of oak are not misclassified to each other. Laurel Oak, Turkey Oak and Live Oak are well separable from each other, and this shows a good intra broad-leaf species classification. In our experiments we observed better performance of FLAASH atmospherically corrected data versus ATCOR. This is in conformance with recent observations of Manakos et. al. [23] where they focused on basic endmember classification (asphalt, gravel, rocky areas, reddish soil, agricultural areas, grass/dry grass, mequis, and phrygana) but we look into the actual problem of tree species classification.

4. Conclusions

Identifying species using remote sensing technologies such as hyperspectral and LiDAR sensors has a critical utility in studying global warming, bio-mass estimation, carbon preserves, invasive species identification and etc. In this paper we perform species classification using SVM over AVIRIS hyperspectral data available for Ordway Swisher Biological Station in north-central Florida. Our focus is on comparing FLAASH and ATCOR atmospheric corrections while we analyze pre-processing techniques both at signal level (Gaussian Filter, Water Absorption Bands Filter) and pixel level (NDVI and NIR Filters). Our observation shows a clear advantage on using FLAASH vs. ATCOR, by margins of about 2% to 4%. Another interesting observation was the suggestion to preserve even low NDVI/NIR pixels of a *canopy* area; as the mixing nature of hyperspectral images suggests in favor of their aggregate performance in SVM classification. Our classification model was robust in intra broad-leaf (oak) classification, with minor inter conifer/broad-leaf misclassifications.

Acknowledgments

Authors are thankful to NEON Inc. for providing hyperspectral data. Also they are thankful to Ms. Sarah Graves, Ms. Leila Kalantari and Dr. Stephanie Bohlman for providing field data. The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc. The NEON 2010 Pathfinder data set is based on work supported by the National Science Foundation under Grant DBI-0752017.

References

1. Colgan, M.S.; Baldeck, C.A.; Féret, J.-B.; Asner, G.P. Mapping savanna tree species at ecosystem scales using support vector machine classification and brdf correction on airborne hyperspectral and lidar data. *Remote Sensing* **2012**, *4*, 3462-3480.
2. Scholes, R.; Archer, S. Tree-grass interactions in savannas. *Annual review of Ecology and Systematics* **1997**, *28*, 517-544.

3. Féret, J.-B.; Asner, G.P. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *Geoscience and Remote Sensing, IEEE Transactions on* **2013**, *51*, 73-84.
4. Clark, M.L.; Roberts, D.A.; Clark, D.B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote sensing of environment* **2005**, *96*, 375-398.
5. Clark, M.L.; Roberts, D.A. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing* **2012**, *4*, 1820-1855.
6. Dalponte, M.; Ørka, H.O.; Ene, L.T.; Gobakken, T.; Næsset, E. Tree crown delineation and tree species classification in boreal forests using hyperspectral and als data. *Remote sensing of environment* **2014**, *140*, 306-317.
7. Féret, J.-B.; Asner, G.P. Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and lidar. *Remote Sensing* **2012**, *4*, 2457-2476.
8. Ghosh, A.; Fassnacht, F.E.; Joshi, P.; Koch, B. A framework for mapping tree species combining hyperspectral and lidar data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation* **2014**, *26*, 49-63.
9. Immitzer, M.; Atzberger, C.; Koukal, T. Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote Sensing* **2012**, *4*, 2661-2693.
10. Naidoo, L.; Cho, M.; Mathieu, R.; Asner, G. Classification of savanna tree species, in the greater kruger national park region, by integrating hyperspectral and lidar data in a random forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing* **2012**, *69*, 167-179.
11. Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.e., phane; Schaepman, M.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment* **2009**, *113*, S67-S77.
12. Baldeck, C.A.; Asner, G.P. Estimating vegetation beta diversity from airborne imaging spectroscopy and unsupervised clustering. *Remote Sensing* **2013**, *5*, 2057-2071.
13. Baldeck, C.; Colgan, M.; Féret, J.-B.; Levick, S.; Martin, R.; Asner, G. Landscape-scale variation in plant community composition of an african savanna from airborne species mapping. *Ecological Applications* **2014**, *24*, 84-93.
14. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions on* **1988**, *26*, 65-74.
15. Janzen, D.T.; Fredeen, A.L.; Wheate, R.D. Radiometric correction techniques and accuracy assessment for landsat tm data in remote forested regions. *Canadian Journal of Remote Sensing* **2006**, *32*, 330-340.
16. Watmough, G.R.; Atkinson, P.M.; Hutton, C.W. A combined spectral and object-based approach to transparent cloud removal in an operational setting for landsat etm+. *International Journal of Applied Earth Observation and Geoinformation* **2011**, *13*, 220-227.
17. Wu, J.; Wang, D.; Bauer, M.E. Image-based atmospheric correction of quickbird imagery of minnesota cropland. *Remote Sensing of Environment* **2005**, *99*, 315-325.
18. Vaudour, E.; Moeys, J.; Gilliot, J.; Coquet, Y. Spatial retrieval of soil reflectance from spot multispectral data using the empirical line method. *International Journal of Remote Sensing* **2008**, *29*, 5571-5584.
19. Xu, J.-F.; Huang, J.-F. Empirical line method using spectrally stable targets to calibrate ikonos imagery. *Pedosphere* **2008**, *18*, 124-130.
20. Hadjimitsis, D.G.; Papadavid, G.; Agapiou, A.; Themistocleous, K.; Hadjimitsis, M.; Retalis, A.; Michaelides, S.; Chrysoulakis, N.; Toulios, L.; Clayton, C. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. *Natural Hazards and Earth System Science* **2010**, *10*, 89-95.

21. Manakos, I.; Liebler, J.; Schneider, T. Parcel based calibration of remote sensing data for precision farming purposes. *Proceedings: Angewandte Geographische Informationsverarbeitung XII* **2000**, 333-344.
22. Hadjimitsis, D.G.; Clayton, C.; Retalis, A. The use of selected pseudo-invariant targets for the application of atmospheric correction in multi-temporal studies using satellite remotely sensed imagery. *International Journal of Applied Earth Observation and Geoinformation* **2009**, *11*, 192-200.
23. Manakos, I.; Manevski, K.; Kalaitzidis, C.; Edler, D. In *Comparison between atmospheric correction modules on the basis of worldview-2 imagery and in situ spectroradiometric measurements*, 7th EARSeL SIG Imaging Spectroscopy workshop, Edinburgh, 2011; pp 11-13.
24. Kampe, T.U.; Johnson, B.R.; Kuester, M.; Keller, M. Neon: The first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing* **2010**, *4*, 043510-043510.
25. Krause, K.; Kuester, M. Airborne observation platform (aop) pathfinder 2010 data release. http://neoninc.org/pds/files/NEON.AOP.015068.pdf.
26. Kampea, T.; Krausea, K.; Meiera, C.; Barnetta, D.; McCorkela, J. The neon 2010 airborne pathfinder campaign in florida. 2010.
27. Adler-Golden, S.; Berk, A.; Bernstein, L.; Richtsmeier, S.; Acharya, P.; Matthew, M.; Anderson, G.; Allred, C.; Jeong, L.; Chetwynd, J. In *Flaash, a modtran4 atmospheric correction package for hyperspectral data retrievals and simulations*, Proc. 7th Ann. JPL Airborne Earth Science Workshop, 1998; pp 97-21.
28. Richter, R.; Schläpfer, D. Atmospheric/topographic correction for satellite imagery. *DLR report DLR-IB* **2005**, 565-501.
29. Adler-Golden, S.M.; Matthew, M.W.; Bernstein, L.S.; Levine, R.Y.; Berk, A.; Richtsmeier, S.C.; Acharya, P.K.; Anderson, G.P.; Felde, J.W.; Gardner, J.*, et al.* In *Atmospheric correction for shortwave spectral imagery based on modtran4*, SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, 1999; pp 61-69.
30. Richter, J.; Schläpfer, D. In *Atmospheric / topographic correction for satellite imagery*, ATCOR-2/3 User Guide, Version 8.3.1, February 2014, 2014.
31. Howat, I. Envi file reader, updated 2/9/2010. http://www.mathworks.com/matlabcentral/fileexchange/15629-envi-file-reader--updated-2-9-2010.
32. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment* **1979**, *8*, 127-150.
33. Jackson, R.; Slater, P.; Pinter Jr, P. Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. *Remote sensing of environment* **1983**, *13*, 187-208.
34. Cho, M.A.; Mathieu, R.; Asner, G.P.; Naidoo, L.; van Aardt, J.; Ramoelo, A.; Debba, P.; Wessels, K.; Main, R.; Smit, I.P.*, et al.* Mapping tree species composition in south african savannas using an integrated airborne spectral and lidar system. *Remote Sensing of Environment* **2012**, *125*, 214-226.