The effect of seasonal spectral variation on species classification in the Panamanian tropical forest

Michael Hesketh[1], G. Arturo Sánchez-Azofeifa[1,2]

[1]Centre for Earth Observation Science, University of Alberta, Edmonton, AB. Canada T6G2E3 (hesketh, asanchez)@ualberta.ca

[2]Smithsonian Tropical Research Institute, Republic of Panama, Panama

Abstract-We explore the effect of inter-seasonal spectral variation on the potential for automated classification methods to accurately discern species of trees and lianas from high-resolution spectral data collected at the leaf level at two tropical forest sites. Through the application of data reduction techniques and classification methods to leaf-level spectral data collected at sites in Panama, we found that in all cases the structure and organization of spectrally-derived taxonomies varied substantially between seasons. We further found that the classification accuracy dropped by a factor of 10 when seasonality was not considered. This study represents one of the first systematic investigations of leaf-level specro-temporal variability, an appreciation for which is crucial to the advancement of species classification methods, with broad applications within the environmental sciences.

Key words- leaf optical properties; seasonality; spectro-temporal variability; dry tropical forest

1. INTRODUCTION

Classification of plant species at the leaf level has been typically attempted using direct spectral-taxonomic relationships (Cochrane 2000). Variations of this concept have been applied in boreal (Fuentes *et al.* 2001) and chaparral (Ustin *et al.* 2004) environments, though the increased complexity and variability of tropical systems suggests some need for caution in their application. More recently, an approach has been put forth to aid in species-level mapping of tropical forest environments coupling hyperspectral reflectance measurements with chemical signatures developed using the relationship between leaf traits and species (Asner and Martin 2008).

While previous research has addressed leaf properties (Asner 1998), and variability between environmental (Asner *et al.* 2009; Sanchez-Azofeifa *et al.* 2009) and structural groups (Castro-Esau *et al.* 2004; Kalacska *et al.* 2007), the question of temporal variation in leaf traits has been largely unexplored at the leaf scale. It is clear from previous studies (Castro-Esau *et al.* 2004; Kalacska *et al.* 2007) that interseasonal spectral variation is a major limiter to our ability to accurately classify forest species in an unsupervised or automated setting, and that a better understanding of the nature and extent of this variation will be critical in the refinement of existing classification techniques and the development of new ones. Our objective, therefore, is to evaluate the effect of seasonality on leaf optical data collected at both wet and dry tropical forest sites. We address here both the effect that this variation has on the structure and organization of classified data, as well as the effect that interseasonal spectral variation may have on overall classification accuracy in both wet and dry tropical environments.

2. METHODOLOGY

Data for this study were collected at two sites in Panama, taking advantage of canopy cranes operated by the Smithsonian Tropical Research Instate to obtain access to the top of the forest canopy. The first site, Parque Natural Metropolitano (PNM) is located just outside of Panama City and is representative of a dry forest. A second set of data was collected at Fort Sherman (FS) representing a wet forest environment. Leaf spectra (350-2500 nm) were collected twice at each site: once during the peak of the rainy season (May 2005), and once at the beginning of the dry season, but before complete leaf loss (March 2007).

Principal components analysis (PCA) was applied to the hyperspectral signatures to reduce the dimensionality and redundancy inherent in these data (Schowengerdt 1996) and the first 4 components were retained such that >97% of the expressed variation in the raw data was represented. Additionally, a set of spectral vegetation indices (SVIs) were calculated from each input spectra. Two resulting datasets were used in subsequent analyses. The first contained only leaf spectra, averaged to the species level. The second contained the retained principle components as well as the three SVIs, this time averaged to the leaf (sample) level.

To visualize the effects of seasonal spectral variation, the raw spectral data was classified using an agglomerative hierarchical clustering. The resulting dendrograms not only illustrate the impact of seasonality on species-level clustering, but allow evaluation of the difference between clustering of spectral data collected during opposing seasons, directly addressing a fundamental requirement for automated species identification: that each species' spectral "fingerprint" is insensitive to seasonal variation. Evaluation of seasonal variation in the structure of the dendrograms was by comparison of the bifurcation ratio (R_b). The value for R_b at a given order μ is the ratio of the number of branches at that order (*N*) to the number at the next order higher (Equation 1).

$$R_b = N_\mu / N_{\mu+1} \qquad (1)$$

To quantify the impact of seasonal variation on classification accuracy, we adapted the method of Kalacska *et al.* (2007). As inputs, we used the four retained principal components and 3 SVIs, averaged to the sample level. For the classifier, we chose the non-parametric decision tree classifier See5 (Rulequest Reseach 2008). The classifier was applied twice to the data from each site: First, the dataset from each season was split in half, with one half used to train the classifier and the other half used to test its accuracy. Second, to test the impact of seasonal variation on the accuracy of the classifier, the entire dataset from the wet season was used as the training set then tested on the entire dry season dataset, and vice versa. Classification accuracy was defined as the percentage of data in the testing set classified correctly.

3. RESULTS

Dendrograms resulting from hierarchical clustering of full spectral data from both sites show marked variation between wet and dry seasons (Figure 1). While the mean of the R_b values across all orders shows some variation between wet and dry seasons at each site (0.12 and 0.19 difference at FS and PNM respectively), these differences are muted in comparison to the differences found at each of the lower orders. Difference between wet and dry seasons at the level of each order range between R_b 0.2 and 1.0, with the exception of the value for $\mu=2$ at FS site of R_b 2.2.

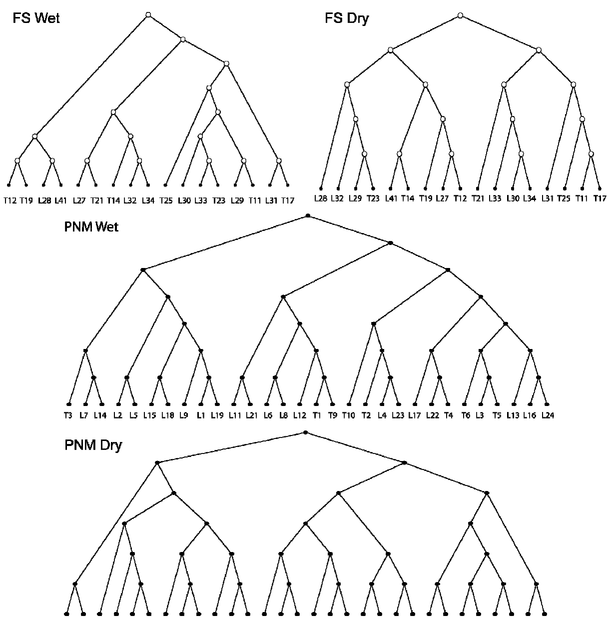

Figure 1. Dendrograms resulting from agglomerative clustering of spectral data.

Seasonal variation in the organization of species within the dendrogram was also considered. Here, the nearest neighbors of each species were compared at each site and each season, and the instances of neighbors common in both seasons noted. A measure of consistency, defined as the percentage of species in the data set with the same neighbor in each season, is provided. Similar groupings are uncommon at both sites, with consistency found to be lower at PNM, with a value of 6.7%, than at FS, where just under one quarter of species shared at least one neighbor between seasons.

Figure 2 shows the impact of seasonal variation in spectral properties on classification accuracy. In all cases, accuracy was dramatically higher where the classification was tested using a classification trained during the same season. The rainforest site (FS) showed accuracy of 80.4-83.5% when training and testing data were from the same season, but dropped to 8.6-10.9% when opposing seasons were evaluated. This decrease was more pronounced at the dry forest (PNM) site, where within-season accuracy was between 80.7 and 83.0%, but between season accuracy fell by more than a factor of 10, to 4.6-7.7%.

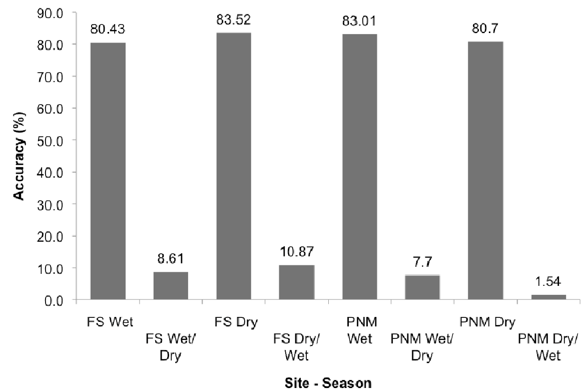

Figure 2. Classifiaction accuracy. Slashed seasons indicate training/testing datasets.

4. DISCUSSION

Throughout this investigation, we found consistent evidence that leaf spectral properties vary sufficiently between seasons that clustering results for wet and dry seasons were measurably dissimilar and classification accuracy was dramatically affected. Methodologies using spectral data for the classification of species and assessment of biodiversity are largely in their infancy (Asner and Martin 2008), and this strong evidence of the impact of seasonality on leaf spectral response suggests that an understanding of the time domain is an essential step in their development.

4.1 VARIATION IN CLASSIFICATION STRUCTURE AND ARRANGEMENT

Variation in leaf spectra between seasons indicates a potentially fundamental difference in plant function between wet and dry forest environments, also noted in Sánchez-Azofeifa *et al.* (2009). Rather than the actual reflectance or spectral signature of a given species, we are concerned with the extent and location of the greatest variation in the signature between seasons. At Fort Sherman, our wet forest site, this variation was most pronounced along the red edge, between 550 and 750 nm. This spectral region is notable in vegetation studies as containing the wavelengths typically used in the estimation of chlorophyll, indicating that the primary driver to spectral variation is likely variation in the relative abundances of leaf chemical components. Conversely, the dry site, Parque Natural Metropolitano, showed substantially less variation in this spectral region, with the greatest sources of variation found at longer wavelengths in the near- and shortwave-infrared regions, where reflectance is governed largely by absorbance by water. This is perhaps unsurprising in a drier forest environment where species are particularly sensitive to moisture variation (Murphy and Lugo 1986), but it indicates that spectral variation cannot be simply tied to a common source, and should be considered in the context of local-scale ecology.

Our results indicate strongly that the variation imposed by seasonality on leaf optical properties is more than sufficient to affect the results of clustering and classification processes applied to the spectral data. If seasonality were unimportant, the same type of spectral data, clustered in the same manner, would yield clusters where the same species were found in close association to each other in both seasons. Our results indicate that this is seldom the case, consistency between the seasons not exceeding 25%, and as low as 6.6 % at the dry forest site. That consistency should be lower at the drier of our study sites is generally consistent with Asner *et al.* (2009), who found stronger associations between biological and spectral properties in wetter, cooler forest environments. However, this conclusion stands in contradiction to other studies, which found that classification at the structural group level using leaf level spectra was more successful in dry forest environments (Castro-Esau *et al.* 2004; Kalacska *et al.* 2007).

Horton's R_b has been used to quantify not only the complexity of river systems, but also variation in the branching of vegetative shoots as a means of evaluating genotypic plasticity (Oohata and Shidei 1971; Whitney 1976). In comparing R_b between seasons at our two sites, we find that not only are the final clusters affected by interseasonal variation, but also the internal structure of the dendrograms. The absolute difference in R_b between seasons also shows a relationship to the sample size taxonomic complexity of the data being processed. While differences are found both at all orders (including the overall mean) when the entire datasets (all tree + liana species) are considered, these differences are greatly muted when only the liana species were considered. Trees only were not considered independently as the number of species in this group did not provide an adequate sample.

4.2 EFFECTS OF SEASONALITY ON CLASSIFICATION ACCURACY

Classification was successful at each site, with accuracies ranging from 80.43% to 93.48%, provided that both training and testing data were drawn from the same season. This is consistent with the accuracies reported by Clark *et al.* (2005), with an accuracy of 92.0% in classifying a set of 7 tree species in Costa Rica, and Castro *et al* (2006), who reported accuracy of better than 80% at each of six sites in Costa Rica, Panama, and Mexico. Certainly sample size has an influence on overall accuracy. Castro *et al.* (2006) report a decreasing trend in classification test accuracy from approximately 85% for 20 species to approximately 80%, projecting a linear decrease to 69% with 100 species, with accuracy eroding beyond that. Our results show a similar relationship between sample and accuracy, though our sample size is smaller overall, with our greatest accuracies, both within-season and between-season, found where sample size is smallest. Where only trees are considered, dropping the overall sample to 8 species at both sites, we achieve an average within season accuracy of approximately 90%, in line with Clark *et al.* (2005). Where the sample is largest, accuracy drops to values similar to Castro *et al.* (2006), with an average of 82% found for a sample of 30 species of mixed trees and lianas.

5. CONCLUSIONS

Our results confirm that inter-seasonal variation in leaf optical properties is measurable and sufficient to preclude automated classification of species at the leaf level using a "database" approach. Analysis of dendrograms derived from an agglomerative clustering of full-range spectra show strong differences in the arrangement of species when spectral data is collected in the wet versus the dry season, and that neither clustering seems to follow any pattern consistent with species taxonomy or structural group. Further, even using the moderate sample size of the current study, we found differences in the internal structure of the dendrograms, quantified by comparison of Horton's bifurcation ratio. The effect that these seasonal differences in spectral properties has on the potential for automated species classification is reflected in the dramatic decrease in accuracy found when comparing within- and between-season classification accuracy using a non-parametric classifier. Our results here indicate a general ten-fold decrease in overall accuracy when a classifier trained using data from the wet season is applied to data from the dry season, or vice versa. This result reinforces previous work, which found a strong influence of environment on classification accuracy (Castro-Esau *et al.* 2004; Kalacska *et al.* 2007).

We recognize, however, that our results here are based on analysis of purely optical data, with spectral indices standing in for measured biochemical data. The promising results of Asner and Martin (2008) and Asner *et al.* (2009) build on relationships between leaf optical properties and leaf chemistry, and exploration of the strength and consistency of these relationships across temporal and geographic gradients must be a priority of research to come. Our findings using purely spectral data do point to the value of further study, expanding the seasonal data collection to include a biochemical survey of the leaves collected to complement the spectral analysis. Such an expanded analysis would allow for the exploration of not just the extent of seasonal spectral variation, but also the sources and drivers to this change.

REFERENCES

Adams, J. B., D. E. Sabol, et al. "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon." Remote Sensing of Environment 52(2): 137-154.1995

Andrade, J. L., F. C. Meinzer, et al. "Water uptake and transport in lianas and co-occurring trees of a seasonally dry tropical forest." Trees-Structure and Function 19(3): 282-289.2005

Asner, G. "Biophysical and biochemical sources of variability in canopy reflectance." Remote sensing of environment (USA).1998

Asner, G. (2008). Hyperspectral Remote Sensing of Canopy Chemistry, Physiology, and Biodiversity in Tropical Rainforests. Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. M. Kalacska and G. A. Sanchez-Azofeifa, CRC**:** 261-296.

Asner, G. and R. Martin "Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests." Frontiers in Ecology and the Environment 7(9): 269-276.2008

Asner, G., R. Martin, et al. "Leaf chemical and spectral diversity in Australian tropical forests." Ecological Applications 19(1): 236-253.2009

Asner, G. P. "Biophysical and biochemical sources of variability in canopy reflectance." Remote Sensing of Environment 64(3): 234-253.1998

Castro-Esau, K., G. Sanchez-Azofeifa, et al. "Monitoring secondary tropical forests using space-borne data: implications for Central America." International Journal of Remote Sensing 24(9): 1853-1894.2003

Castro-Esau, K., G. Sanchez-Azofeifa, et al. "Variability in leaf optical properties of Mesoamerican trees and the potential for species classification 1." American Journal of Botany 93(4): 517-530.2006

Castro-Esau, K. L., G. A. Sánchez-Azofeifa, et al. "Discrimination of lianas and trees with leaf-level hyperspectral data." Remote Sensing of Environment 90: 353-372.2004

Clark, M., D. Roberts, et al. "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales." Remote Sensing of Environment 96(3-4): 375-398.2005

Cochrane, M. A. "Using vegetation reflectance variability for species level classification of hyperspectral data." International Journal of Remote Sensing 21(10): 2075-2087.2000

Coppin, P. R. and M. E. Bauer "Change detection in forest ecosystems with remote sensing imagery." Remote Sensing Reviews 13(3): 207-234.1996

Fuentes, D., J. Gamon, et al. "Mapping Canadian boreal forest vegetation using pigment and water absorption features derived from the AVIRIS sensor." Journal of Geophysical Research-Atmospheres 106(D24).2001

Huemmrich, K. F., T. A. Black, et al. "High temporal resolution NDVI phenology from micrometeorological radiation sensors." Journal of Geophysical Research.1999

Kalacska, M., S. Bohlman, et al. "Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels." Remote Sensing of Environment 109(4): 406-415.2007

Kalácska, M., G. A. Sanchez-Azofeifa, et al. "Estimating leaf area index from satellite imagery using Bayesian networks." IEEE Transactions on Geoscience and Remote Sensing 43(8): 1866-1873.2005

Murphy, P. G. and A. E. Lugo "Ecology of Tropical Dry Forest." Annual Reviews in Ecology and Systematics.1986

Oohata, S. and T. Shidei "Studies on the branching structure of trees. I. Bifurcation ratio of trees in Horton's law." Jap. J. Ecol 21: 7-14.1971

Sanchez-Azofeifa, G. A., K. L. Castro-Esau, et al. "Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lanias and trees: Implications for remote sensing in tropical environments." Remote Sensing of Environment 113: 2076-2088.2009

Schnitzer, S. A. "A mechanistic explanation for global patterns of liana abundance and distribution." Am Nat 166: 262-276.2005

Schowengerdt, R. "On the estimation of spatial-spectral mixing with classifier likelihood functions." Pattern Recognition Letters 17(13): 1379-1387.1996

Sims, D. A. and J. A. Gamon "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures, and developmental stages." Remote Sensing of Environment 81: 337-354.2002

Strahler, A. N. "Quantitative analysis of watershed geomorphology." American Geophysical Union Transactions 38: 913-920.1957

Ustin, S. L., D. A. Roberts, et al. "Using Imaging Spectroscopy to Study Ecosystem Processes and Properties." BioScience 54(6): 523-534.2004

Whitney, G. G. "The bifurcation ratio as an indicator of adaptive strategy in woody plant species." Bulletin of the Torrey Botanical Club 103(2): 67-72.1976

Xiao, X., Q. Zhang, et al. "Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest." Remote Sensing of Environment 94(1): 105-122.2005

Zhang, X., M. Friedl, et al. "Monitoring vegetation phenology using MODIS." Remote Sensing of Environment 84: 471-475.2003