Final Report
June 2007

RAPID ASSESSMENT OF RESTORATION PERFORMANCE MEASURES AT MULTIPLE SCALES IN THE GREATER EVERGLADES USING NEAR INFRARED REFLECTANCE SPECTROSCOPY (NIRS)

Matthew Cohen[1], Sabine Grunwald[2], Mark Clark[2] and K. Ramesh Reddy[2]
1 – School of Forest Resources and Conservation
2 – Soil and Water Science Department
Institute of Food and Agricultural Sciences
University of Florida

EXECUTIVE SUMMARY

Ecosystem change detection is an essential part of restoration. As the Comprehensive Everglades Restoration Plan reformulates the Greater Everglades ecosystem in an attempt to restore more historical processes, metrics of recovery are paramount, both for status reporting and adaptive management. In many respects, soils are a useful ecosystem performance measure: soils change at time scales well suited to large area diagnostic surveillance (i.e., not too rapidly and not too slowly), are sensitive and specific to numerous stressors, are ubiquitous and simple to sample, and integrate ecological process. However, soil properties cannot be reliably inferred remotely where vegetation cover is high; consequently, to understand ecosystem changes over large areas requires massive sampling campaigns, with enormous sample analysis costs to provide spatially informative estimates of system status, and high temporal density to infer system trends. The recent emergence of optical methods for sample analysis offer promise for alleviating some of the sample cost constraints on monitoring. Specifically, high-resolution diffuse reflectance spectroscopy (DRS) has been shown in numerous settings to offer tremendous accuracy for sample analysis at greatly reduced costs. This work explores the feasibility of using this technique for ecosystem monitoring of soils from the Greater Everglades.

We collected high resolution (1-nm band) spectra in the visible and near infrared regions of the electromagnetic spectrum from over 4000 samples from throughout the Greater Everglades. The samples had previously been analyzed for a suite of informative biogeochemical properties, and we used these as a library from which we statistically trained the spectra to predict the soil properties. This approach, referred to as chemometric modeling, lays the foundation for future application in the region by establishing which parameters can be reliably inferred in this manner, and deducing the algorithms for that prediction from samples for which only spectral information is collected. The library not only includes continuous metrics of soil quality (n = 10; e.g., total P, total Ca), but also categorical properties such as vegetative community type and soil-floc distinctions; we evaluated spectral prediction of this categorical properties also. Moreover, we expanded the library of soil properties by additionally measuring soil C quality (using a functional fractionation scheme) and total mercury (THg) concentrations.

Overall, the results strongly support the utility of this method for future ecosystem monitoring efforts. Using a hold-out validation data set to maximally represent the operational accuracy that we might expect applying the predictive models to new data, we observe prediction

error rates that compare favorably with laboratory accuracy, and extremely high precision (one of the hallmarks of this method). Further, we show high prediction accuracy for routine analytes, categorical targets (particularly between soil and floc) and, most importantly, for both THg and C quality. Our primary diagnostic for model efficiency for continuous parameters is the relative performance determinant (RPD), which is a unitless measure of the scale of mean prediction errors finding significant favor in the spectroscopy literature. A summary of RPD is provided below (Fig. I-A); validation values greater than 2.0 are considered excellent, while values greater than 1.5 are generally considered useful for mapping applications. Similarly, our metric of categorical model efficiency is the model odds ratio (OR). A summary of the model OR value for all categorical models is presented in Fig. I-B; typically, validation values > 10.0 indicate effective diagnostic models. As shown, almost all parameters selected show ample promise of routine spectral prediction.

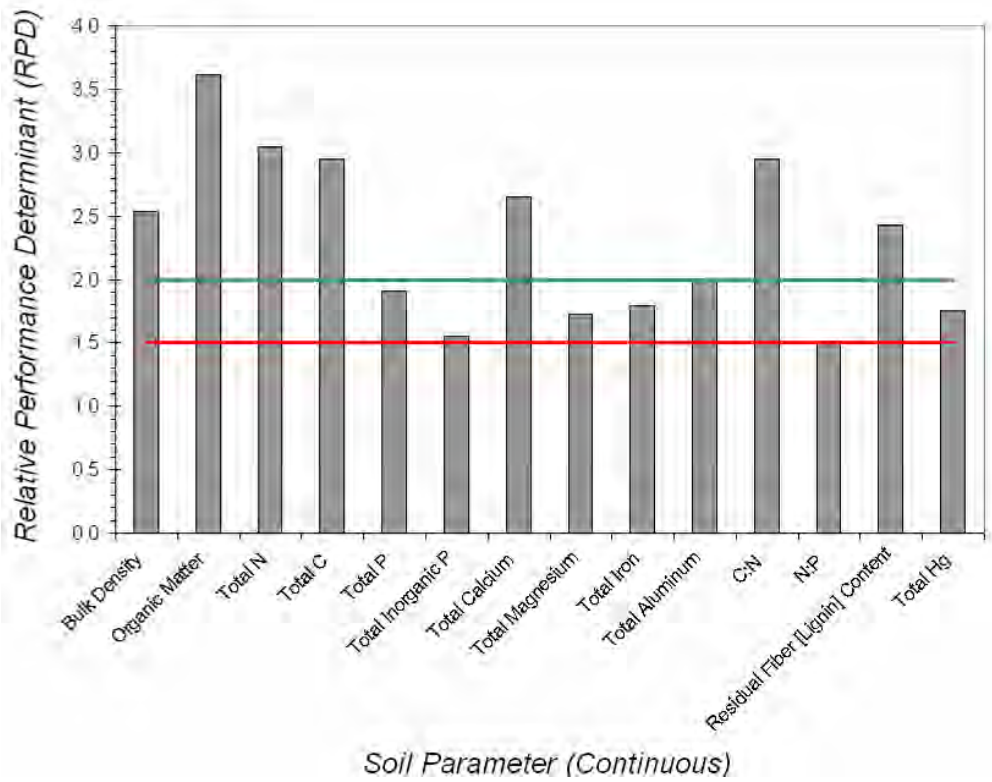

Fig. I-A. Relative performance determinant (RPD) for all continuous spectral models developed in this work. Threshold values of 2.0 (excellent) and 1.5 (useful for mapping) are depicted.

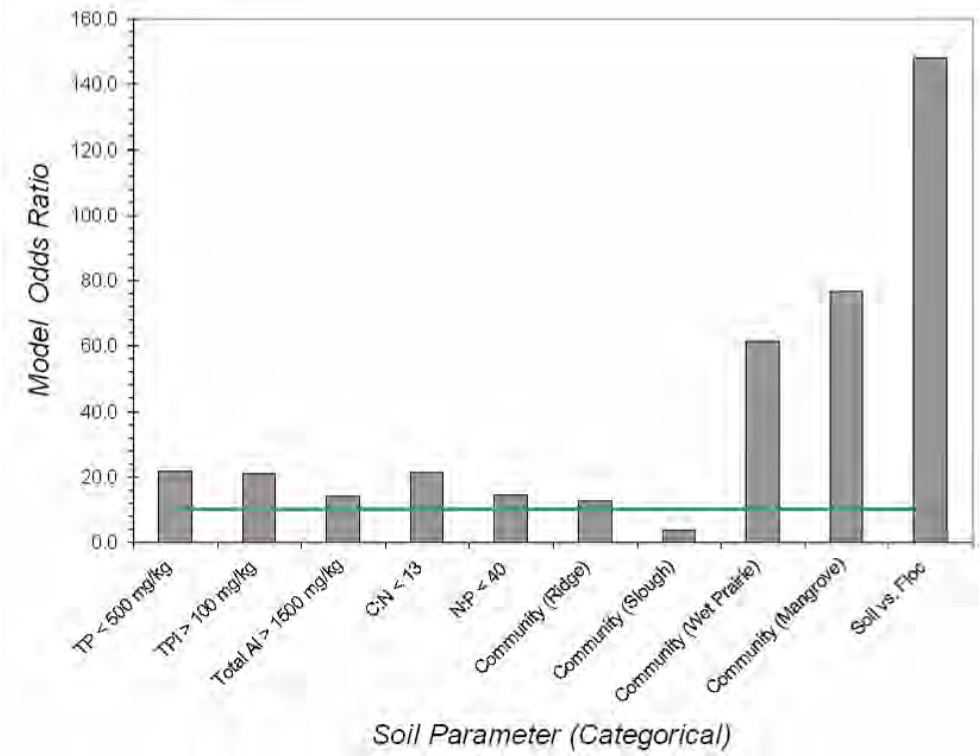

Fig. I-B. Model Odds Ratio (OR) for all categorical spectral models developed in this work. Threshold value of 10.0 is shown, above which model is generally considered highly useful.

Because spectra are a useful integrator of soil quality, there are compelling reasons to want to use them more overtly for ecosystem change detection; that is, rather than simply use them to predict soil properties, which themselves have uncertain association with ecological processes and condition, make spectral prediction of ecological condition directly. We use new methods for unsupervised classification of soil spectra to cluster samples; these clusters are shown to be ecological robust, integrating information about soil chemistry, vegetative community composition and hydrologic forcing in ways that support future exploration.

In all aspects, this report is a first effort to link this new method to the ongoing challenges of ecosystem change detection in the Greater Everglades. There are areas of substantial promise, and areas of remaining uncertainty. Areas in need of further work are:

1) *Characterization of ecological change gradients using spectra.* There is some preliminary evidence presented in this report that spectra may be useful for mapping change fronts (e.g., TP enrichment). These methods require further and more detailed explication.

2) *Local spectral library development.* There is strong evidence from this work to suggest that prediction errors are not geographically random. That is, soil processes at the local scale confound prediction to a moderate extent, suggesting that locally developed calibrations may be more accurate and robust. Further research is required to substantiate this.

3) *Operational accuracies.* As with all models of this type, the observed validation accuracies may be an overestimate of operational accuracies because of between instrument and between time errors in spectral reflectance. Efforts to quantify and control for these errors are paramount for routine implementation. Newer sensors are capable of much greater between instrument precision, which may be one fruitful area of exploration.

4) *Calibration free methods.* Ecosystem surveillance requires performance measures that are direct. While soils in the Everglades represent the total integration of ecosystem processes, the particular indicators used as performance measures remains uncertain. A direct approach, whereby spectra are calibrated to particular condition classes and/or stressor gradients is likely to prove extremely useful based on preliminary evidence presented here.

5) *Additional soil properties and extension to other substrates*. The success of the method for soils in the Everglades can measured in units of accuracy, but also in units of cost. We estimate that the comparative costs of the routine analyses and spectral methods differ by an order of magnitude ($77 vs. $7/sample). Moreover, the cost implications become more pronounced as more analytes are predicted. Given success with indicators of C quality, metal content, and ecological stoichiometry, we recommend the inclusion of additional informative analytes to the spectral library. Further, we recommend an extension of these methods, applied only to soils in this work, to plant tissues as well. Existing libraries of samples have been identified.

TABLE OF CONTENTS

EXECUTIVE SUMMARY ii
TABLE OF CONTENTS vi
1. INTRODUCTION 1
1.1 Everglades Assessment 1
1.2 Soil as a Performance Metric 1
1.3 VNIR Soil Measurement Technology 2
1.4 Objectives and Hypotheses 6
2. METHODS 8
2.1 Soil Library Development 8
2.1.1 Routine Analytes 8
2.1.2 Nutrient Ratios 11
2.1.3 Carbon Quality Analytes 12
2.1.3.1 Methane Production 14
2.1.3.2 Sequential Fiber Analysis 15
2.1.4 Total Mercury 16
2.1.5 Categorical Thresholds 17
2.1.6 Nominal Variables 18
2.1.6.1 Community Analysis 18
2.1.6.2 Soil-Floc Analysis 19
2.1.6.3 Soil Depth Analysis 19
2.1.7 Gradient Measurements 20
2.2 Soil Spectral Scanning 21
2.2.1 Precision Control 21
2.2.2 Data Management 22
2.2.3 Data Processing 22
2.2.4 Data Visualization 23
2.2.5 Developing Spectral Subsets 24
2.3 Chemometrics: Continuous Prediction 25
2.3.1 Statistical Tools 25
2.3.2 Model Validation 27
2.3.3 Model Diagnostics 27
2.3.4 Analysis of Model Residuals 28
2.4 Chemometrics: Categorical Diagnostic Prediction 28
2.4.1 Statistical Tools 28
2.4.2 Model Validation 28
2.4.3 Model Diagnostics 29
2.4.4 Analysis of Model Error Rates 29
2.5 Calibration Free (Unsupervised) Methods 29
2.5.1 Rationale 29
2.5.2 Statistical Tools 30
2.5.3 Metrics of Utility 31
2.6 Mapping 31
2.6.1 Mapping Model Error Rates 33
2.6.2 Mapping Soil Total Mercury 34
2.6.3 Mapping Soil Carbon Quality 36

3. RESULTS 37
3.1 Routine Analyte Summary 37
3.1.1 Summary Statistics 37
3.1.2 Correlations 37
3.1.3 Sub-Sampling 39
3.2 Spectral Library Visualization 40
3.2.1 Spectral Precision 40
3.2.2 Spectral Variance 43
3.2.3 New Sample Projections 45
3.3 Continuous Predictions 47
3.3.1 Comparison of Methods 47
3.3.2 Model Validation Results 49
3.3.2.1 Routine Analytes 49
3.3.2.2 Soil Nutrient Ratios 55
3.3.3 Spectral Information Content 58
3.3.4 Analysis of Model Residuals 59
3.3.5 Maps of Prediction Errors 62
3.4 Categorical Predictions 71
3.4.1 Soil Functional Thresholds 71
3.4.2 Error Rate Analysis 72
3.4.3 Nominal Analysis 73
3.4.3.1 Soil Depth 73
3.4.3.2 Community 74
3.4.3.3 Soil vs. Floc 75
3.4.3.4 Floc Probability with Depth 75
3.5 Analysis of Carbon Quality 77
3.5.1 Methane Production Summary 77
3.5.2 Fiber Analysis Summary 79
3.5.3 Spectral Prediction of Soil C Quality 82
3.5.4 Soil C Quality by Area and Community 85
3.5.5 Maps of Soil C Quality 86
3.6 Analysis of Total Mercury 88
3.6.1 Summary Statistics (Total Hg per mass and Total Hg per area) 89
3.6.2 Spectral Prediction of Total Hg (per mass) 91
3.6.3 Analysis of Semi-Variance 92
3.6.4 Analysis of Cross-Covariance 94
3.6.5 Analysis of Total Hg by Community Type 96
3.6.6 Analysis of Total Hg Gradients 98
3.6.7 Maps of Total Hg (per mass and per area) 98
3.7 Gradient Analysis – WCA2A 100
3.7.1 Spectral Variance with Distance 100
3.7.2 Classification Probability 100
3.8 Unsupervised Classification Results 102
3.8.1 EM Cost Sequence 103
3.8.2 EM Cluster Geography 104
3.8.3 EM Cluster Soil Properties 108

4. DISCUSSION 109
4.1 Spectral Feasibility for Sensing Soil Quality in the Everglades 109
4.2 Comparative Cost Analysis.....112
4.3 Integrating Soil Sensing into Large Area Assessment.....113
5. CONCLUSIONS AND RECOMMENDATIONS 114
6. REFERENCES CITED.....117
7. DATA AND APPENDICES.....120

1. INTRODUCTION

1.1 Everglades Assessment

As the restoration of the Greater Everglades continues, efforts to characterize baseline conditions and then monitor conditions in response to restoration are paramount. Restoration goals can only be assessed in a framework of robust performance measures, and management can only be adapted given timely feedback from the system on the state of those performance measures. Assessment of a system of the scope and complexity of the Everglades is a multi-metric challenge; that is, no single performance measure will suffice to provide useful feedback to managers and scientists about system status and trends. However, all performance measures that are selected need to possess three critical attributes: 1) the must be sensitive and specific to ecosystem change at reasonable time scales, 2) they must be useful across spatial domains of the system expected to respond to restoration, and 3) they must be sufficiently rapid and inexpensive to allow adequate temporal and spatial sampling density. Selection of performance metrics is therefore subject to important scrutiny. This report describes the potential of emerging soil assessment technology (spectral analysis) to permit inclusion of soils as performance measures by addressing the 3rd criteria listed above (cost vs. sampling density in space and time).

1.2 Soil as a Performance Metric

Changes in ecological processes are manifest in hydrologic, vegetative, faunal and soil dynamics, and key attributes of each of these are potential performance measures to assess Everglades ecosystem recovery. While hydrologic, vegetation and faunal studies abound in the Greater Everglades, substantially less emphasis has been placed on soil-based indicators of system condition. We suggest that ongoing research to quantify soil-based performance metrics of ecosystem condition offers several key advantages: first, soil development effectively integrates ecosystem dynamics over time and space, offering measures for which specific numeric benchmarks and targets can be set in spite of extant environmental variability; second, soil provides a ubiquitous indicator that can be related to changes in all aspects of ecosystem performance (hydrology, nutrient enrichment, landscape dynamics); finally, soil responds to anthropogenic influence generally, and nutrient enrichment/hydrologic alteration in particular, in relatively well understood ways in the Everglades system (Inglett et al. 2004, White and Reddy

2003, DeBusk and Reddy 2003, Reddy et al. 1993). Moreover, methods for measuring the properties of soils that may be useful as restoration benchmarks are well developed.

One reason soil remains relatively under-used as a performance measure for evaluating restoration progress in the Everglades is the significant cost associated with large area assessment. In particular, spatial and temporal variability necessitate large sample sizes for spatially heterogeneous systems, and the suite of effective non-redundant soil-based performance measures is large, ranging from basic indicators of soil biogeochemistry (e.g. total phosphorus, pH, cation conc., organic content) to measures of ecological toxicity (e.g., soil total mercury content) to complex integrative measures of soil quality (e.g. soil-floc boundaries, soil carbon quality). Comprehensive characterization of samples is necessary for effective integration and indication of ecosystem change, but expensive given high temporal and spatial resolution requirements. Sample processing costs increase geometrically with the complexity of analytes, spatial variability and temporal variability. Moreover, while vegetation response metrics in time and space can be resolved using airborne and satellite-borne sensing devices, no such remote measurement of soils is yet reliable enough to be useful for subtle ecosystem change detection.

As the diversity and cost of soil analysis grows along with demands on spatial and temporal change detection, the constraints on budgets demands examination of alternative soil analysis methods. Emerging methods for sensing soil quality using high resolution diffuse reflectance spectroscopy (DRS) have been validated across a wide range of analytes and environments. The multiple benefits of this technique are germane for large area ecosystem assessment: accuracy, high precision, greatly reduced cost and rapid sample throughput.

1.3 VNIR Soil Measurement Technology

Near infrared (NIR) diffuse reflectance spectroscopy (hereafter referred to as NIRS) is a proven technology for the rapid non-destructive assessment of materials, including plastics, industrial reagents, minerals and agricultural products. Spectral reflectance signature libraries of numerous material samples and composites have been cataloged (e.g. Clark 1999); from these libraries, unknown samples can be interpreted for functional and qualitative properties. The basic application involves collecting high-resolution reflectance signatures (e.g. 1-nm bandwidths) from a sample illuminated by a high intensity full spectrum light source; samples pre-processing typically involves only removal of water and physical macro-structure. Light

from the visible (350-750 nm) and near-to-mid infrared (750-2500 nm) regions of the electromagnetic spectrum is used for NIRS, but longer wavelength sensors (mid infrared – MIR) are also available. Under controlled conditions, reflectance signatures arise due to electronic excitation of atoms and vibrational stretching and bending of structural groups of atoms that form molecules and crystals. For example, fundamental vibrational features for organic matter (C-O, C-N, C-H, O-H functional groups) are observed in the mid- to thermal-infrared (2.5-25 μm) portion of the spectrum; however, overtones of these fundamental features occur at fractions of the fundamental frequency, which fall within the range typical of NIRS. Minerals (calcite, apatite, clays along weathering sequences) along with common cations also exhibit distinctive spectral reflectance characteristics due to light interference, facilitating rudimentary mineralogical description. Indeed, NIRS is among the principal tools for field mapping of geologic materials for purposes of mining.

Most applications depend on spectral matching; that is, a set of diagnostic absorbance features for a particular pure sample are noted and an unknown sample compared to that library. Quantitative inference of soil quality from spectral reflectance has consequently been severely limited by the heterogeneous character of the soil system and the absence of appropriate analytical tools for inference from complex, strongly co-linear reflectance spectra. However, with application of powerful statistical data mining tools (e.g. Partial Least Squares [PLS] regression, Classification and Regression Trees [CART], neural networks [NN], Multivariate Adaptive Regression Splines [MARS]) has allowed researchers to develop efficient correlations between spectral response characteristics for heterogeneous soil media and a wide array of standard soil and plant functional parameters (Ben- Dor and Banin 1995, Kooistra et al. 1997, Foley et al. 1998, Gillon et al. 1999, Chen et al. 2002). The method continues to be tested in a wide array of ecological regions; while no published data have been collected for low ash content soils like those typical of the Everglades, recent work (Cohen et al. 2005 – Fig. 1) suggests strong functional associations between numerous soil indicators and spectra for hydric soils with a wide range of organic content and nutrient concentrations. Critical reviews of the myriad applications of this technology for natural and agro-ecosystems now exist (Foley et al. 1998, Shepherd and Walsh 2007), and the recent literature points increasingly towards global calibrations (Brown et al. 2006) and operationalizing the methods for replacing or augmenting routine environmental sample analysis (Cohen et al. *in press*, Janik et al. 1997).

Formal methods for analyzing soil archives for which laboratory evaluations have been performed have been developed (Shepherd and Walsh 2002). This approach is founded on the development of Spectral Reflectance Libraries (SRLs). SRLs consist of soils for which spectral response curves and controlled laboratory analyses of functional indicators have been collected. From the SRLs, calibrations are developed to correlate reflectance to functional measurements; validated models can then be used to infer soil properties in incoming sample soils for the suite of laboratory indicators without direct analysis. Indicator analytes that have exhibited association with spectra in other studies include cation concentrations (Ca, Mg, Fe, Al), soil organic matter and ash content, soil texture, soil nitrogen and phosphorus. Moreover, preliminary evidence suggests that reliable models can be developed to infer organic carbon quality (e.g. lignin content – Shepherd et al. 2003), organic mineralization rates (Bouchard et al. 2003, Fystro 2002), soil hydraulic properties (Cohen et al. *in press*), and various measures of metal contamination (Kooistra et al. 1997). Cohen et al. (2005) present evidence that various microbial activity indicators (extracellular enzymes - acid phosphatase, β-glucosidase, dehydrogenase, peptidase) can be inferred with moderate accuracy from reflectance spectra. While for many of these parameters, it is strong co-linearity with readily observable soil

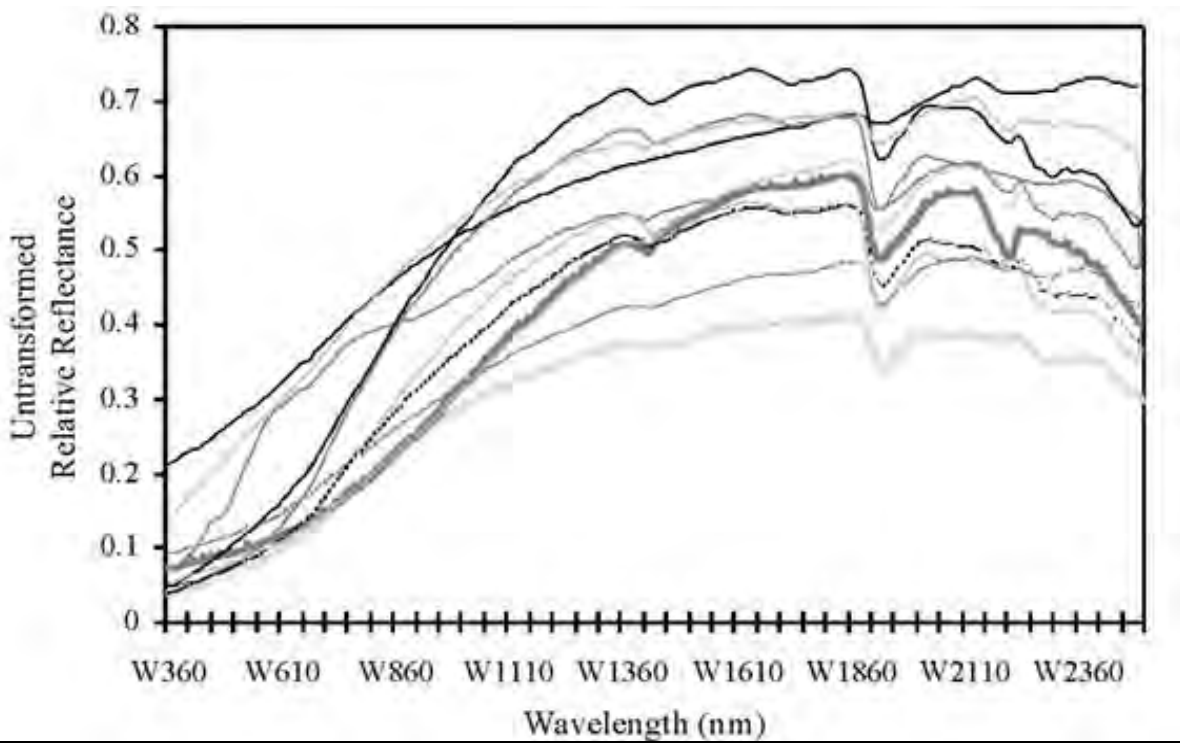

Fig. 1. Raw relative reflectance across the NIR spectrum versus wavelength for 11 selected end-member spectra from wetland soils (from Cohen et al. 2005).

attributes (e.g. soil carbon, soil clay content) and not direct reflectance effects that are being observed, that research showed that spectral information provides more predictive power than can explained by co-variance with other primary soil attributes. Other studies (Cohen et al. 2006, Vagen et al 2006) suggest that spectral measurements can be used to establish repeatable diagnosis of ecosystem condition across a suite of stressors, a result with direct relevance to change detection in the Greater Everglades.

The various cited advantages of NIRS are relevant to the question of ecosystem change detection and performance benchmarking within CERP. High analytical accuracy (i.e., efficient prediction of target parameters) is the central metric of utility, and the wealth of applications in the literature for which high accuracy has been demonstrated are compelling that the Greater Everglades system will be no different. Other advantages include high precision, which has important implications for maintaining between laboratory comparability, as well as comparability of results over time. Indeed, spectrometer precision is so much higher than standard analytical techniques that some portion of prediction error (i.e. spectral model accuracy) must logically be attributed to laboratory uncertainties. Yet another advantage of NIRS is the sample throughput. Our experience suggests that, given effective data and sample management protocols, more than 200 samples can be processed by a single technician in one day, and soil attribute inference can be automated so that results emerge for all parameters almost immediately. Given extremely basic sample pre-processing, this offers tremendous throughput potential. Finally, the advantage of cost is the most widely cited in the current literature. While the capital costs of the spectrometer are large (~$50,000), the per sample costs are minimal labor (see above) and electricity. We formally explore the comparative costs of conventional and spectral sample characterization in this work. With strategic spatial and temporal sampling of the environment, high-resolution surveys of soil condition and statistically powerful inference of the effects of human activities on soil function can be developed with relative ease. Most importantly, monitoring protocols that quantify restoration responses can be implemented with substantially reduced personnel and laboratory requirements.

Shepherd and Walsh (2002) pose a scheme for large-area diagnostic surveillance of ecosystems that are changing in response to anthropogenic forcing. That scheme (Fig. 2) forces formal definitions of “cases”, that is areas that are ecologically distal from their desired

condition, from which risk factors and management levers can be extracted, and ecosystem performance over time can be monitored. Spectral methods figure centrally in the screening test for the case definition (see Cohen et al. 2005 for a demonstration of the utility of this approach in heterogeneous landscapes) and for the direct measurement of incidence (changes through time). While spectral methods need not be part of this logical scheme, our intention with this work is to demonstrate their enormous capacity for improving monitoring and inference.

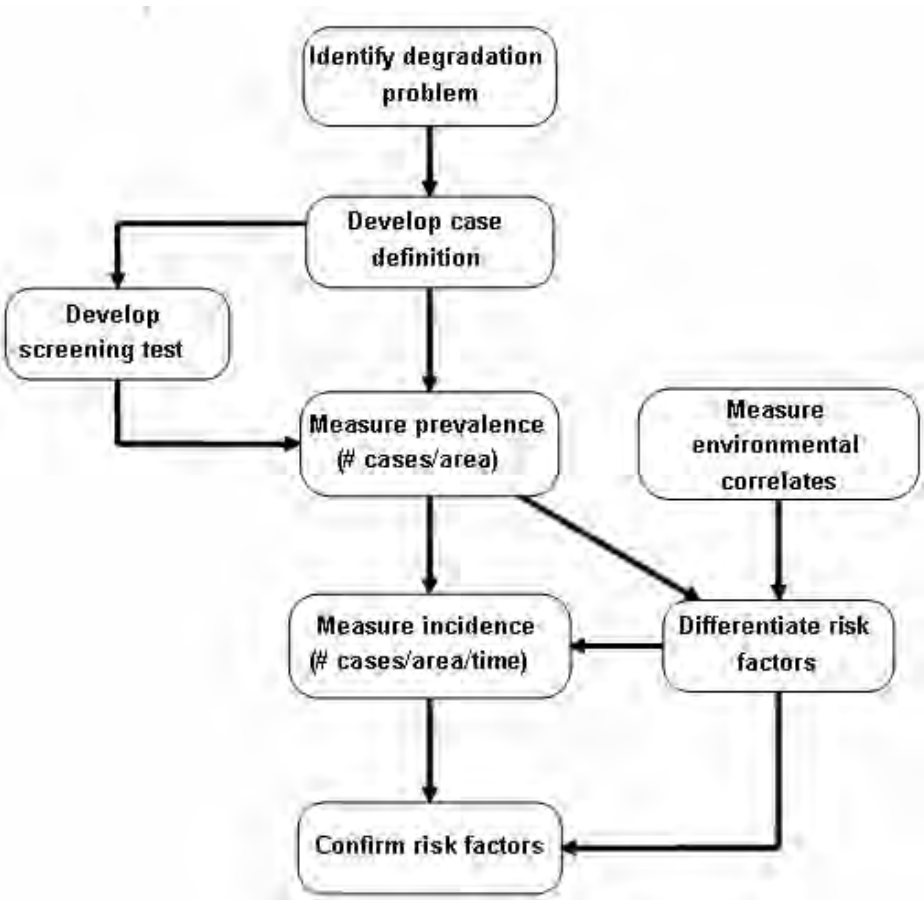

Fig. 2 – Logical scheme for large area ecosystem change detection and conditional assessment. Spectral methods appear to provide utility in the development of screening tests (impairment status) and in the estimation through space and time of environmental correlates.

1.4 Objectives

Our overall objective was to test the feasibility of a new soil sensing technology for streamlining the monitoring and assessment of ecosystem change in the Greater Everglades. The scope of work for this project, implemented over two years, had five objectives:

1) *Develop a spectral reflectance library for soils from the Greater Everglades for the purpose of spectral prediction and enhanced sample throughput.*

2) *Develop spectral library protocols for screening incoming samples and strategically enhancing library representation of landscape variability.*

3) *Develop and validate statistical models to relate reflectance spectra to measured soil properties.*

4) *Sub-sampling of the spectral library population for additional soil measurements, which will permit spectral prediction of those measurements across the entire population.*

5) *Mapping relevant variables and model residuals to examine spatial patterns at the scale of the Greater Everglades ecosystem and/or relevant subsets.*

This report synthesizes our findings across a suite of analytes collected on samples previously collected (Everglades Soil Mapping – K.R. Reddy, principal investigator), and a set of analytes measured and predicted specifically as part of this work. In particular, we demonstrate development of an analysis framework that starts with streamlined integration of spectral data collection, quality control and data management. Once reliable data on spectral and soil properties have been assembled, relationships between the data sets can be learned and validated. Incorporation of new measurements into the prediction data set is both useful from the perspective of the utility of the spectra (i.e., ability to predict a wider suite of potential performance metrics), but also because in so measuring the training data, we obtain data that are useful in their own right for monitoring. For example, our work explores the predictability of soil C quality and mercury concentrations using NIRS, but provides the ancillary benefit of direct measurements and analysis thereof even if the prediction efficiencies are insufficient for monitoring applications. Analysis of these primary variables (mercury, carbon quality, direct spectral inference) is embedded in the analysis of the NIRS prediction of them.

2. METHODS

The methods for this work are numerous, and frequently applied in specific combinations for specific analytes or objectives. We start by outlining the data set that forms the basis of the spectral reflectance library and then the derivatives and additions to that made as part of this project. In particular, the laboratory methods for determination of carbon quality (two parallel methods) and soil mercury are outlined here. Following explication of where the continuous target data came from, we discuss the origin of functional thresholds to define performance categories and those data that are nominal (community type, soil-vs.-floc). Then we describe the techniques for scanning each of the soils, how the data were quality controlled, processed and stored. Finally we discuss the statistical tools used for this work: this includes statistical methods for continuous and categorical predictions (chemometrics), methods for unsupervised learning from the spectral data, and methods for mapping (model residuals, new analytes).

2.1 Soil Library Development

2.1.1 Routine Analytes

This work builds on a comprehensive library of soils collected from throughout the Greater Everglades between April and December 2003 (Everglades Soil Mapping, ESM – K.R. Reddy, Principal Investigator – sites shown in Fig. 3). Soil samples from a total of 1405 sites from throughout the system were collected utilizing a 10 cm diameter medical grade stainless steel coring device. This device was a stainless steel section of tubing (1 mm thickness) with stainless steel handles (15 cm long by 2.5 cm diameter) welded 10 cm from the top of the core tube. The bottom end of the 70 cm tube was sharpened to facilitate coring in organic soils. At 132 sites, field triplicates were collected to examine analytical replication errors and zero-separation distance variability.

Cores were sectioned into floc, 0-10cm, and 10-20 cm increments in the field and sections placed into sealed in polyethylene bags and stored in coolers on ice until returned to the laboratory. The soil corer was washed in site water before and after each site visit. In many places, shallow to bedrock soils did not contain a 10-20cm profile. Likewise, a floc sample was not present at every site due to environmental or vegetative characteristics. When returned to the lab, soil samples were dried at 70 ºC for 3 days in plastic weigh boats and then ground in 20-mL HDPE scintillation vials with ceramic grinding balls to powderize the sample for analysis.

Samples were analyzed for total carbon (TC), total nitrogen (TN) total phosphorus (TP), total calcium (TCa), total magnesium (TMg), total iron (TFe), total aluminum (Tal), organic matter content (LOI – loss on ignition), HCl-extractable P (TPi – inorganic P), and bulk density (BD), using standard analytical methods at the NELAP-certified Wetland Biogeochemistry Laboratory at the University of Florida. A detailed description of site selection, field methods, analytical methods and spatial interpretation can be found in Corstanje et al. (2006), Bruland et al. (2006), Rivera et al. (2006) and Osborne et al. (*in press*).

A table of the locations of all sites by sampling area is provided (Table 1) to demonstrate the coverage of this effort. One important caveat is essential for assessing the global applicability of the maps and spectral models derived from this dataset: all sampling was done in areas where a helicopter could be safely set down, which systematically biases against areas where trees are a part of the biological assemblage. This is problematic most clearly in the Big Cypress Preserve landscape, where treeless areas are the exception rather than the rule, but also in tree islands. As such, while we refer to all models extracted from this dataset as applicable to the Greater Everglades, both mapping and spectral predictions should be treated as statistical extrapolations for tree-dominated sites. Further work to fill in these gaps may be of particular future interest.

Table 1. Summary of sites by geographic area and profile depth.

Number of Samples	Profile				
Geographic Area	1 - Floc	2	3	4	Total
Big Cypress (Central)	103	121	23		247
Big Cypress (North)	12	20	4		36
Big Cypress (South)	99	110	16		225
East Coast Buffer	13	33	30	3	79
Everglades National Park	169	373	273		815
Holey Land	21	68	60		149
Model Land	35	46	16		97
Rottenberger	10	54	54		118
Water Conservation Area 1	142	157	154	91	544
Water Conservation Area 2B	129	129	128		386
Water Conservation Area 2A	23	23	23		69
Water Conservation Area 3A north	40	177	170	29	416
Water Conservation Area 3A south	83	225	220	18	546
Water Conservation Area 3B	43	58	58	25	184
Total	**922**	**1594**	**1229**	**166**	**3911**

Note: 1225 sites with one core, 127 sites with triplicate cores, and 53 sites (incl. 5 with multiple cores) from a follow-up study.

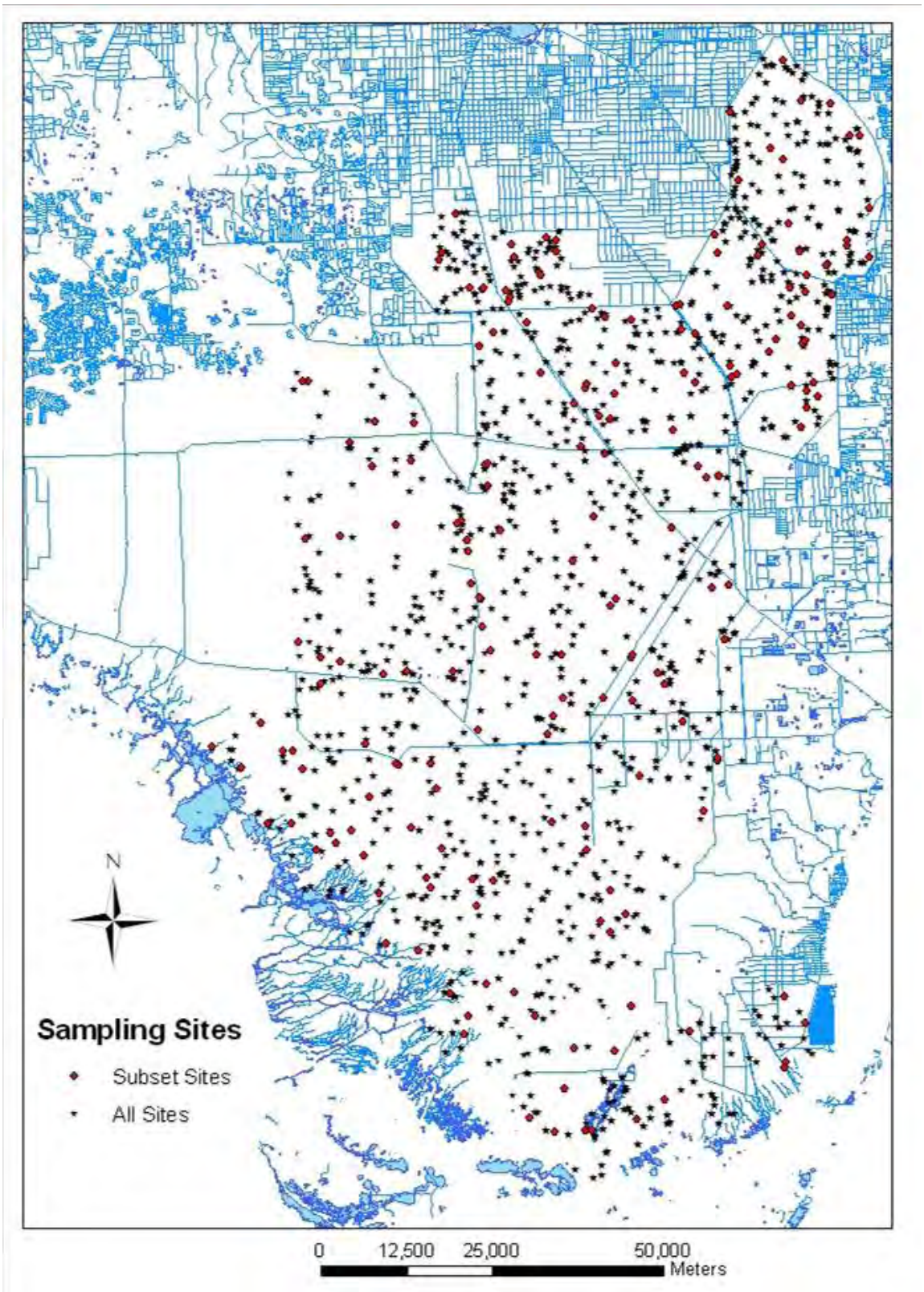

Fig. 3. Location of original ESM soil sampling sites and selected subset of sites for further analysis (n = 348; selection described in end-member characterization subsection).

Table 1 also shows the number of sites for which different soil depth profiles were included. Deeper profiles include 0-10 (Depth 2), 10-20 (Depth 3) and 20-30 (Depth 4) sections; only those sites where periphyton floc material was present (based on best professional judgement) had a Depth 1 sample. Figure 3 shows the location of all sites throughout the study region; a subset of the sites (n = 348) were selected strategically for soil carbon quality characterization (see below). A different subset (n = 600) were selected for total mercury (THg) analysis; both methods for selection are described below.

2.1.2 Nutrient Ratios

Based on measurements of C, N and P, we computed nutrient ratios that are ecologically values for interpretation (C:N, N:P). These were done on a molar basis to conform with existing stoichiometric inference. Calibrations to these nutrient ratios directly, rather than via predictions of the nutrient concentrations independently, was done because it eliminated one source of error. Indeed, this is emblematic of the potential utility of spectral methods (as described in Fig. 4). Much of the emphasis in this work is on the inferred relationship between soil properties and ecological condition (Pathway 1 – Fig. 4) which is itself an uncertain association. Spectral prediction of those soil properties (termed “environmental covariates” to illustrate that plant or animal tissue properties are also candidates for which literature suggests successful prediction) along Pathway 2 assumes first that the measurement of the soil properties is correct (a topic that will be addressed in greater detail later) and second that the relationship of the soil property to the ultimate target of ecological condition is adequate. Where a consensus “case-control” definition, similar to disease diagnostics in the medical field, can be established, an alternative mode of spectral inference is possible. That is, direct calibration of the spectra to the categories of ecosystem condition, obviating the middle step of prediction of environmental covariates (Pathway 3A). (Note: discussion of pathway 3B, which involves the use of unsupervised learning from spectral information is discussed later). This approach, used successfully in Cohen et al. (2006) allows a hardened case definition to be defined in a manner that is simple, rapid and inexpensive to monitor. Provided the covariance of Pathway 3A is as strong or stronger as that with the more conventional environmental covariates (Pathways 1 – 2), and this is generally expected, then there is significant benefit to this approach for surveillance of ecosystem change.

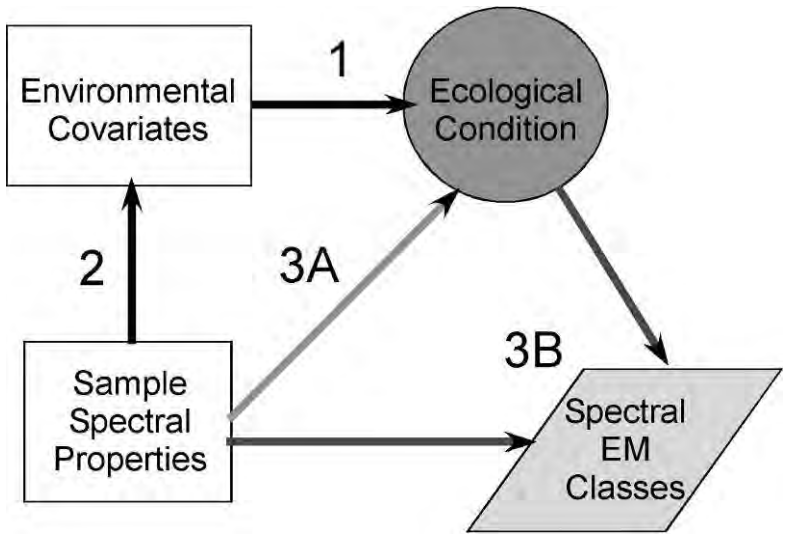

Fig. 4 – Conditional assessment using 1) existing environmental covariates, 2) spectral prediction of these environmental covariates, 3) direct spectral prediction via a) supervised calibration and b) unsupervised methods. Spectral EM classes refer to a priori *clusters defined from the spectra only using unsupervised machine learning techniques.*

2.1.3 Carbon Quality Analytes

Peat substrate quality is a crucial metric of peat systems because it encapsulates the sensitivity of the soil to changes in elevation (slow variable) and the availability of electron donors for microbial metabolism (fast variable). While numerous studies of Everglades soils exist, only recently has there been a more concerted effort to link soil quality measurements to ecosystem processes. One candidate metric, nutrient ratios such as C:N or N:P, are an important component of vegetative substrate quality, and, the amount and type of organic compounds and their recalcitrance (resistance to decomposition) are also important substrate quality components regulating decomposition. Clark and Reddy (2006) showed, for example, that soil development rates in ridges were nearly an order of magnitude higher than in sloughs because of both differences in production rates, but also in litter lignin content. They determined lignin content using sequential fiber analysis, a technique used in this work as well.

Characterization of soil organic matter quality (i.e., degree of recalcitrance, with more ligning rich, recalcitrant material termed higher quality) is an important potential indicator of ecosystem change. If, due to changes in hydrology, soil carbon is trending towards either more labile or more recalcitrant form, the implications for ecosystem C budgets are expected to be long-term and significant. In particular, if rates of peat accretion or the stability of accreted peat

change over time, these changes are expected to be extremely durable (long term changes). The mechanisms via which changes in soil organic quality might feedback to affect long term ecosystem dynamics are numerous, and include changes in soil pH dynamics (with implications for calcite formation/stability), changes in microbial community processing, changes in DOC production, and, in the long term, changes in soil elevation which directly effects the composition of peat based ecosystems. For the purposes of this work, there is strong precedent that NIRS can adequately approximate soil C quality (e.g., Shepherd et al. 2003, Skjemstad et al. 2006); given an adequate prediction and the geographic scope and density of the sample library, we can, for the first time, provide a spatial estimate of soil lability.

For measures of soil C quality, we use a nested analysis. First, we collected 16 sample cores from 4 locations (Fig. 5) in the Southern Everglades for high resolution analysis of soil C quality with depth. The objective of this part of the study was to compare different methods for indicating C quality (see methods below). At each site, 2 cores were obtained from the nearest ridge, and 2 from the nearest slough (note: all sites were within the ridge-slough mosaic). Cores were collected to 30 cm depth using 10-cm polycarbonate core tubes. Cores were capped in the field and kept below 4 C prior to sectioning. Cores were sectioned into 9 sections (0-2 cm, 2-4 cm, 4-6 cm, 6-8 cm, 8-10 cm, 10-15 cm, 15-20 cm, 20-25 cm, and 25-30 cm). No effort was made to distinguish periphyton floc from soil during core sectioning.

After selecting one method (fiber analysis) we chose a sub-sample (n = 348) from the entire dataset (n = 3911) to analyze. This subsample was obtained by doing a principal components analysis (PCA) of the 10 measured soil properties for the ESM dataset, and partitioning the first 5 axes into zones (4 zones for axes 1 & 2; 3 zones for axes 3, 4 and 5). One sample was selected from each hypercube zone. These should, theoretically, most efficiently express the diversity of the ESM population; we note here that we did not stratify by geography or depth in this effort (though we summarize how well this sampling captured that variability). Only 348 of the 432 zones contained samples; one was randomly selected from each.

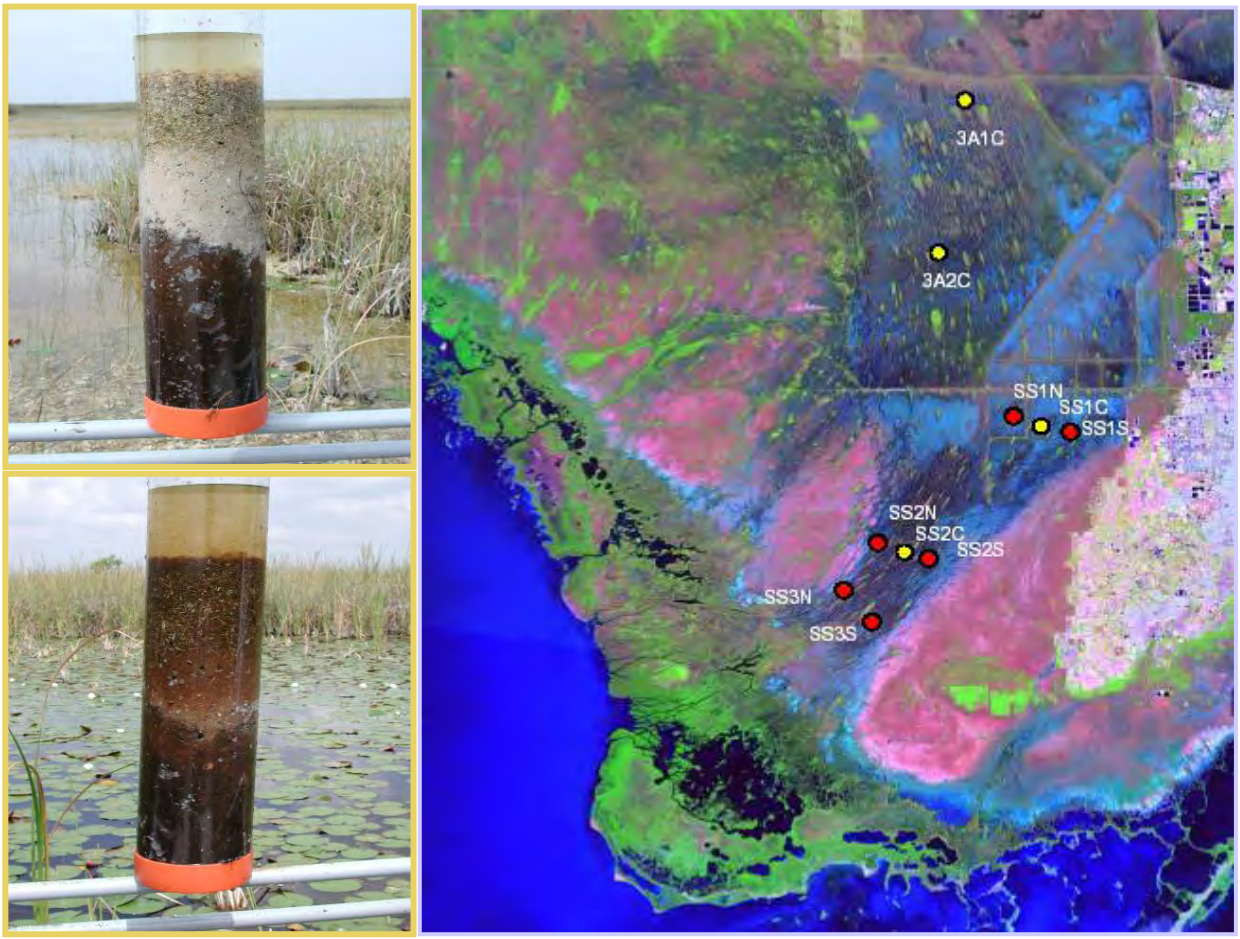

Fig. 5 – Locations (yellow circles) and photographs of cores (top – SS1C, bottom, SS2C).

2.1.3.1 Methane Production

There is no standard method to determine the lability of soil carbon; instead we used to complementary methods. The first examines the rate of methanogenesis under laboratory controlled conditions (temperature, moisture, head space gas composition). Rates of methane production, measured using a gas chromatograph (Shimadzu) during 14-day incubations at constant temperature was our first method. Specifically, 3 replicate incubation vials were seeded with 2 grams of field-moist soil from each of the core sections; deionized water was added to ensure well hydrated conditions. The head space of the incubation vials were purged with N_2 to ensure anoxia, and then incubated at 25 C in a dark environment. Head space gas samples were obtained at 0, 3, 5, 7, 10 and 12 days after initiation. After 7 days, the head space was purged to avoid negative feedback on methanogens as methane increases in concentrations. Cumulative

methane production over 12 days was used to estimate production rates in μg/g soil/day. Methane production was compared between cores and across depths using two-way ANOVA.

2.1.3.2 Sequential Fiber Analysis

Lignin is the most abundant aromatic polymer present in soils and is the principal regulator of organic matter recalcitrance, particularly under anaerobic conditions (Criquet et al., 2000). Lignin is centrally responsible for providing rigidity and structural support to vascular plants; lignin is also strongly resistant to bacterial decay because of its complex aromatic structure. The mosaic of vegetation, ranging from ridges (comprised of monotypic sawgrass) and tree islands to sloughs (with submerged aquatics) and periphyton provides highly variable quality input materials for the formation of soils. More labile plant tissue components such as sugars and starches are likely to be more abundant in vegetation with lower structural requirements. As such, measuring the most recalcitrant organic structural compounds, contained within the residual fiber fraction, yields insight into primary regulators of substrate quality.

While there are standard methods for the determination of lignin in samples (Rowland and Roberts, 1999), they are time consuming and labor intensive. As an alternative, functional fractionation schemes (Fig. 6) have high routine precision and strong covariance with more pure chemical classifications. Functional fractionation of carbon quality is straightforward operationally, and provides nearly synonymous information about recalcitrance to decay. In this scheme, samples are exposed to increasingly intensive detergents intended to remove starches (neutral detergent fraction – NDF), then hemi-cellulose (acid detergent fraction – ADF). The remaining material is partitioned into cellulose and lignin by digestion in strong sulfuric acid; the mass of sample remaining is the residual fiber, functionally defined herein as lignin. Sequential extraction is recommended Mould and Robbins 1981) because it improves dissolution of cell wall proteins and minimizes influence of condensed tannins on detergent fiber residuals (Terrill et al. 1994), which may alter recoveries of residual fiber and lignin (Rowland and Roberts 1999). Sequential fiber fractionation was done on triplicates of 2-g samples from each core section. The intent was to examine covariance between a functional method (fiber fractionation) and a process level measurement (methane production). Ultimately, we chose one method for application to the ESM data. Precision and simplicity of fiber fractionation led us to conclude that those data were more informative. We selected to use that method as a proxy for C quality.

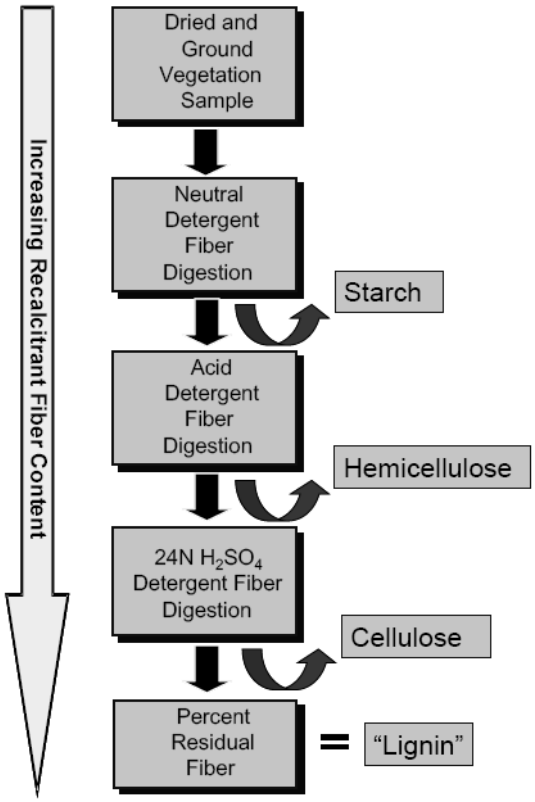

Fig. 6 – Schematic of sequential fiber analysis (after Lewis 2005).

After making that determination, we expanded the scope of our fiber fractionation efforts to include the 348 samples selected using Latin hypercube sub-sampling (see above). Each sample was run in triplicate, and the relative fraction of each used as a target for spectral prediction. Our principal interest, however, was not in the spectral characterization of the more labile fractions, but in our capacity to estimate the recalcitrant fraction (residual fiber ~ lignin) because it is that fraction that holds the most functional significance to long term ecosystem dynamics, and might reasonably be expected to act as a useful performance measure.

2.1.4 Total Mercury

Previous evidence (Kooistra et al. 1997) indicates that some level of metal contamination can be resolved using spectral methods. In the Greater Everglades, the principal metal contaminant of concern is mercury, though primarily not in elemental form. Our intent with the following measurements was to determine the total mercury concentrations of a subset of soils, and estimate spectral covariance.

A subset of 600 samples from the total data set were analyzed for total mercury per unit mass (THg_M – mg Hg/kg soil). These samples were selected from the population using a Latin hypercube subsampling protocol using principal components axes derived from the basic soil biogeochemical factors to define the sampling space. THg_M concentrations in these 600 soil samples were determined following an acid digestion of a pre-weighed dry sample (~1 g) with a mixture of concentrated $HCl/HNO_3/HF$ mixture, in acid-cleaned and marble capped volumetric flasks (Hossner 1996, Donkor et al. 2005). Samples were heated overnight to a refluxing boil on a hot plate, and diluted after cooling to a known final volume with Nanopure® water. Hg concentrations in the resulting digestate were analyzed by the $SnCl_2$ reduction technique, dual gold amalgamation and detection by cold vapor atomic fluorescence spectrometry (CV-AFS) (Bloom and Crecilius 1983). QA/QC criteria were met by the use of reagent blanks, standard solutions, and a certified reference material (IAEA-405).

2.1.5 Categorical Thresholds

From the perspective of ecosystem monitoring, conditional classes are often more widely used than continuous variables; for example, agronomic testing of soils typically results in classes of soil performance rather than continuous predictions (Cohen et al. 2007). From these categorical predictions, management recommendations are made. As such, spectral prediction of performance categories may ultimately be more meaningful than continuous prediction, particularly within a probabilistic surveillance framework (i.e., what is the probability of observing a degraded condition given risk and recovery factors). Moreover, where continuous spectral prediction of soil chemical properties is of limited efficiency, spectral screening models that delineate categories of soil condition are frequently possible.

There are numerous potential categories to examine; some categories are natural while others are assumed. For example, vegetation community or soil-vs.-floc are naturally nominal variables that by necessity need to be evaluated in a categorical modeling framework. Other categories (e.g., TP < 500 mg/kg vs. TP > 500 mg/kg) are functionally defined thresholds in otherwise continuous responses. They are, however, meaningful from an ecological and regulatory perspective, and are possibly useful benchmarks for ecosystem change detection. The particular soil properties for which we have selected to explore categorical modeling are total P,

total inorganic P, soil-floc (previously developed), ridge-slough, total Al, and the C:N and N:P ratios. The thresholds selected for categorical modeling are defined in Table 2.

Table 2. Summary of categorical model thresholds.

Model Target	Category Threshold†
Total P	Case > 500 mg/kg
Total Inorganic P	Case > 100 mg/kg
Total Al	Case < 1500 mg/kg
C:N ratio	Case < 13
N:P ratio	Case < 40
Soil vs.Floc	-
Community Type	-

† - Case (response = 1) definitions. No definition is necessary for existing categorical data (e.g., soil vs. floc).

2.1.6 Nominal Variables

In addition to continuous properties segmented into functional classes for diagnostic assessment, there are several ecosystem properties that may well prove to be useful performance measures (or at least measures of process) that might benefit from spectral prediction for monitoring and mapping purposes. While chemometric methods (described below) for nominal data are less well developed, several pattern recognition tools allow robust interpretation of complex spectral data. Our effort here was simply to develop first-order inference about feasibility of nominal variable prediction. Further work would be required in each case to refine the models for routine implementation.

2.1.6.1 Community Analysis

We developed spectral prediction models for the community types that were observed during ESM sampling. We note here that this is a strongly unbalanced library because many of the communities that are important in the Everglades mosaic were not sampled because of logistical and safety constraints (principally related to the ability to land a helicopter proximate to areas with trees). As such, the spectral differences between community types focus almost entirely on herbaceous communities (sawgrass ridges, emergent sloughs, wet prairies). The few samples from tree islands and mangrove forests are insufficient to permit robust categorical model development.

2.1.6.2 Soil-Floc Analysis

One important performance metric in the Everglades is the prevalence of calcitic flocculent material in space, and the extent to which that material is integrated in soil profiles. The prevalence question can, and has been addressed, using best professional judgement of what constitutes floc; a robust case definition may well be helpful to avoid between analyst errors, and also to diagnose areas where the probability of categorical designation to a fluc class is changing through time. As such, a spectral diagnostic model of soil vs. floc is an important component of demonstration of methodological feasibility in this work. Moreover, given a reliable diagnostic prediction of soil vs. floc, vertical profiles of class assignment probabilities can be computed. Our hypothesis is that, if floc material is being integrated into the soil in areas where floc is present at the surface, the class assignment probability to floc will decline gradually with depth. If, instead, floc is a surface phenomenon only and is not integrated into soils, the transition of class assignment probabilities will be abrupt. To test this hypothesis, we collected 4 cores from WCA2A in areas where floc material was abundant at the surface. Cores were 30 cm long and were sectioned finely (2 cm sections); each section was dried, milled and scanned, and assignment probabilities for each were determined from the categorical models developed from the spectra. Plots of assignment probabilities to the floc class vs. depth were used to draw inference about incorporation of floc into soil. The location of the 4 cores is shown in Fig. XX.

2.1.6.3 Soil Depth Analysis

The ESM sample population is partitioned into soils collected from different depths. Because of the dynamic nature of the Everglades ecosystem, particularly with respect to soil subsidence and accretion processes, monitoring of soil depth profiles may be an important metric. We used the extant soil depth data to determine if there are significant and systematic spectral differences between soils of different depths when evaluated across the ecosystem. If significant variability does exist, then diagnostic spectral models may be helpful in mapping areas where significant oxidation has occurred, or, alternatively, where peat soils are accreting quickly. As with many of the arenas explored in this work, positive calibration to soil depth is primarily suggestive of avenues of more detailed work; the current study (a feasibility study) was of insufficient scope to follow up where strong calibrations were observed.

2.1.7 Gradient Measurements

One of the most innovative developments in the NIR spectroscopy literature is the direct covariance between spectra and environmental gradients (see Shepherd and Walsh 2007 for a discussion of applications). In short, spectra integrate environmental conditions and can be ordinated to match up to environmental gradients. One technique (discussed below under calibration free methods) is to extract single-metric signals from the spectral data and analyze covariance with known gradients. To examine this technique further, we selected the TP gradient well documented in northern WCA2A (Debusk et al 2001, Reddy et al. 1991); we sampled every 200 m from areas known to be profoundly enriched with TP down to areas designated as background TP concentrations (Fig. 7). Samples were collected to 10-cm deep using a stainless steel coring device, and stored on ice until they could be dried, ground and scanned. No chemical analysis was done on these samples.

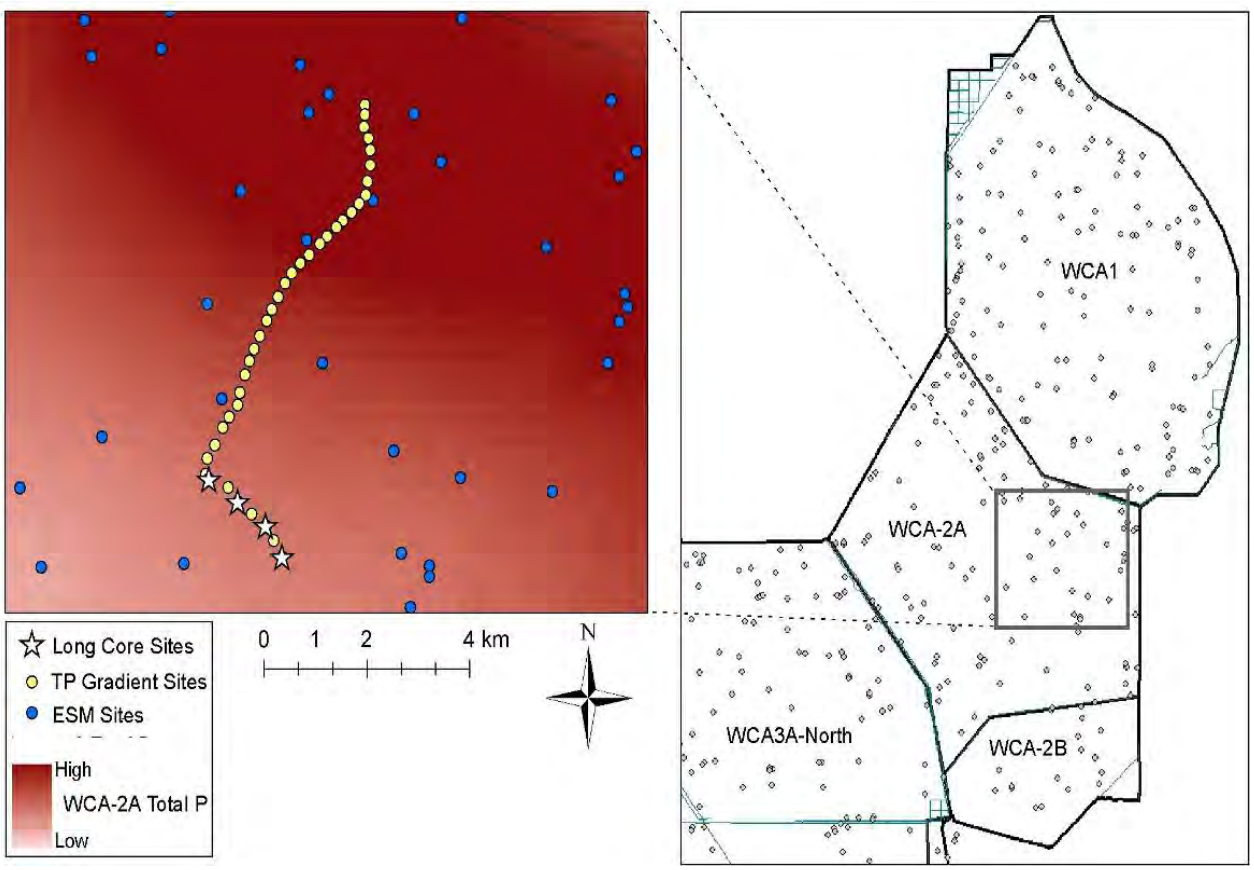

Fig. 7 – Locations of TP gradient sampling. Samples were obtained every 200 m along the known TP gradient. Four long cores (collected to examine spectral diagnosis of floc vs. soil with depth were also collected in the unimpacted zone.

2.2 Soil Spectral Scanning

All soils were scanned using a post-dispersive spectroradiometer; the instrument, a FieldSpec Pro FR manufactured by Analytical Spectral Devices (Boulder CO), collects diffuse reflectance information in the visible (350-750 nm) and near infrared (750-2500 nm) regions with a spot size of 11 mm (field-of-view). Integrated fiber optic cables transmit diffuse reflected light to three internal spectrometers. Junctions between the spectrometers as well and the optical extremes where the tungsten light source provides low signal are areas of low signal:noise, and are omitted from our analyses. The spectral resolution of the FieldSpec Pro FR instrument (as measured using the Full Width at Half Maximum – FWHM) is 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. The internal reflectance data are sampled to 1 nm bands for reporting.

Our optical set-up consists of a high intensity (50 W) halogen land (3000 K) directed upwards through borosilicate glass dishes containing soil samples. The light incidence angle was set to 25 degrees to maximize diffuse reflected signal while minimizing specular reflectance, with the viewing angle set to nadir. Borosilicate glass dishes, which have minimal absorbance characteristics in the visible and near infrared regions, were filled with 5 to 10 grams of soil (sufficient soil to cover the entire 6 cm diameter). This set-up, described in Shepherd et al. (2004) minimizes stray light interference, maximizes sample throughput and also reduces the optical effects of differential sample grinding by providing a smooth surface (glass dish) against which the sample rests; all of these advantages minimize variability between sample replicates. Diffuse reflectance is measured in relative terms, compared with a Spectralon (LabSphere, Hutton, NH) acting as white-reference. Spectralon is a highly lambertian (between 257 and 10,600 nm) thermoplastic material that provides very high (~99%) and nearly uniform relative reflectance properties in the range of our selected sensor. As with the samples, Spectralon is scanned through borosilicate glass, which has nearly perfect optical transmission in the VNIR.

2.2.1 Precision Control

Each sample was scanned 4 times, with a 90-degree rotation between each replicate, and each replicate itself consists of 25 integrated spectrometer observations. Multiple observations permit examination of analytical precision, which is one of the widely-cited advantages of spectroscopy vis a vis conventional laboratory techniques. Our data storage and retrieval system

includes data management algorithms that allow sample precision among replicates to be evaluated in real time. A precision index (PI) was defined as follows:

$$PI = 100 * \frac{\sum_{N} RMSE_i / \mu_i}{N} \qquad [1]$$

where N is the number of wavebands (N = 2150), RMSE is the root mean squared error across the four replicates for band i, and μ is the mean reflectance in that band. We set the threshold for data acceptability to a PI below 3%, meaning that samples where the PI exceeds 3% are rejected by the database and rescanned.

2.2.2 Data Management

Accepted scans are entered into a Microsoft SQL Server database hosted by IFAS. Middleware software (written in Java) permits on-line data handling, advanced query functionality and multiple options for data retrieval (resampling, various transformations, data formats). The database developed for this work can be viewed at http://swsd.ifas.ufl.edu/soil/mainwindow.swf); this interface software was developed to link to a SQLServer back-end, and permit web-based queries of both spectral and biogeochemical properties. This database set-up was selected for two reasons: first, it permits web accessibility to a centralized database, making sharing spectral data easier, and collaboration among researchers with spectral data more efficient. Second, the IFAS servers are routinely backed-up and checked for data integrity.

2.2.3 Data Processing

Each scan with the FieldSpec Pro yields a spectrograph of raw relative reflectance consisting of 2150 points in the region between 350 and 2500 nm. This enormous dimensionality and auto-covariance between proximate bands makes statistical analysis of these data somewhat problematic. Therefore, perhaps the most important reason for central database storage is that almost all of the analyses of spectral data that are done in the literature are performed after data pre-processing is performed. While the algorithms to do much of this pre-processing are well developed, there are cases where subtle differences led to poor between-instrument comparability, rendering models derived from the spectra un-usable by other machines. Among the common pre-treatments for the data include resampling, to reduce the

dimensionality of the data sets to more realistically reflect the spectrometer capabilities; instrument resolution is between 3 and 10 nm while the reported data have a resolution of 1 nm. A typical pre-treatment is to resample the raw data to 10-nm windows, a process that results in the loss of no predictive information (Shepherd and Walsh 2002). Further data processing includes derivative transformation, which can be done in numerous ways. Derivative transformation with 2nd order smoothing, which is done to reduce the effects of optical set-up, grinding, ambient light and power source variability, is an option in the database. In general, much more stable calibrations are obtained when overall albedo (average reflectance) is controlled for, which is the result of derivative transformation. Additional pre-processing steps (mean centering, convex hull removal) are either done as a routine part of the statistical modeling, or were shown to have no significant effect on the model stability.

2.2.4 Data Visualization

Visualizing the spectra is of particular value for two reasons. First, it is often instructive to see the spectra from population end-members (selected using statistical subsampling routines described below) to appreciate the subtle differences that frequently differentiate samples. This is particularly true when discussing the potential of using a diagnostic feature approach to spectral inference rather than the statistical learning approach advocated here. While it is well known that spectra of pure materials can be precisely identified using spectra (particularly from the mid-infrared), the very obvious similarities in spectrograph shape between soils with quite different chemical properties underscores the need to examine relationships within the spectra rather than just identifying spectral diagnostic features. Fig. 1 shows end-member spectra from a data set with much greater mineralogical diversity than would be expected in the Greater Everglades, and there is still strong concordance among the spectrographs.

Second, the operational use of spectroscopy for sample prediction when using statistical inference (chemometric models) depends on the complete representation by the reference library. Samples with spectra that fall outside the multi-parameter space bounded by the spectral library should be considered outside prediction range; subsequent inclusion of these samples in the reference library after laboratory analytical evaluation will ensure that the library will grow to be more inclusive of extant edaphic diversity. One way to visualize the properties of the spectral reference library is to extract principal components (PCs) from the high dimension spectral data

sets. Plotting the PC scores for the reference library vs. the samples to be predicted from any chemometric models will reveal samples that lie outside prediction bounds. We develop this principal components analysis, extracting 5-6 latent variables that can be used to visualize characteristics of the population and subsets thereof. For this particular effort, because all samples have both spectral and chemical data, we partition the population into a training set and a validation set. The PC scores permit estimation of whether the partitioning approach used yields a sufficiently representative subset of the data to be used for model training. Similarly, where we select a subset of the samples for additional chemical analyses, we can use the PC scores to ensure that the selected samples maximally represent the population. By performing this data visualization step at the outset, we maximize the generality of any chemometric models that are developed.

2.2.5 Developing Spectral Subsets

One of the important objectives of visualizing the sample spectra (and indeed summarizing the statistical variability in the soil chemical observations) is that it is cost prohibitive to run additional laboratory analyses on all 3911 soils that were obtained during field sampling. The costs of developing the existing library, where all samples were analyzed for 10 analytes were enormous; in order to expand the list of analytes, we required a subset of the samples that was optimally representative of the variability in the full population, but strategically selected to minimize the requisite sample size. We used Latin hypercube subsampling, a protocol for sampling that ensures representation from each area of a multi-dimensional sample space, to identify this subsample. Our target was to select approximately 350 samples for analysis of organic matter lability (four functional pools of organic C) and mercury content.

We used two approaches for this task. We first sampled from the multi-dimensional sample space defined by the soil chemical observations only; our rationale for this was that the soil chemistry are the more accepted analytical techniques. Operationally this consisted of the following 3 steps:

1) Develop a principal components analysis on all 3911 samples using the 10 soil analytes. Extract only the first 5 axes.

2) Divide the scores for each axis into 4 segments (for axes 1 and 2) or 3 segments (axes 3, 4 and 5).

3) Randomly select 1 samples from each multi-variate segment (e.g., zone 1-1-1-1-1 represents the first segment for each axis, while zone 1-2-3-3-3 represents the first zone for axis 1, the second zone for axis 2 and the third zones of the remaining axes). This results in a total of 432 possible zones, but since only 348 zones are represented in the data set, our sample size from this approach was 348 samples.

After selected the subset in this way, we evaluated the extent to which the full variability of the data set was represented with respect to spectra. In this way we can confirm that any predictive models developed between the spectra and the soil chemistry are selected to most efficiently represent the full range extant in the system (as sampled); the application of these models to the remaining ~3500 samples is consequently considered much more reliable.

2.3 Chemometrics: Continuous Prediction

2.3.1 Statistical Tools

Chemometrics refers to the statistical tools for relating spectral observations to potential sample covariates. Since the spectra are profoundly complex, with overlapping absorbance features and harmonic overtones, the so-called soft modeling approach of statistical inference is required. Modern chemometrics are based on advances in machine learning and pattern recognition; however, despite the complexity of the model technique, the assessment of model efficiency continues to be done using the familiar regression (for continuous data) and discriminant (for categorical data) diagnostics. Here, we summarize the methods that we selected for predictive modeling, and then the diagnostics that we used to evaluate their performance. It is not an exhaustive list. We evaluated several simpler techniques first (in particular principal components regression [PCR] which has been used by authors for similar purposes – Chang et al. 2001) before concluding that the two methods presented here are by far the most reliable. The primary activity to date has been development of chemometric models relating sample spectral responses to measured chemical and physical properties. This process was done using two methods that are gaining primary favor in the developing literature on soil prediction from spectra. These are partial least squares regression (PLS) and stochastic gradient

boosted tree regression (SGBT). While our primary objective was not methodological comparison, recent literature (Brown et al. in press) documents the increased efficacy of SGBT vs. PLS for prediction.

Spectral complexity of soils necessitates soft-modeling wherein spectra are related to chemical and physical properties via statistical models. A chemometric approach allows entry of numerous statistical techniques. We compared two popular techniques, outlined below; both were implemented in Statistica 7.1 (StatSoft, Inc. 2005).

The first and most common is partial least squares regression (PLS). This approach, like PCA extracts canonical variables from the data. However, in contrast to PCA, in which linear combinations are defined based only on total variance reduction among the co-linear predictors, PLS linear combinations are defined based on covariance with a target variable. Decomposition of the predictor-by-sample matrix into principal components is conditioned to maximize covariance between the PCs and target parameters. PLS is highly regarded for its simplicity, replicability and comparative avoidance of statistical over-fitting.

Another method gaining favor (Brown et al. 2006) for its flexibility and accuracy is a non-parametric data mining tool called stochastic gradient boosted tree regression (SGBT). SGBT is analogous to well-known tree-based analytical methods, which are based on recursive binary splitting (Breiman et al. 1984). In basic tree-based regression, the algorithm identifies a partitioning variable at each binary split (or parent node), that maximizes the purity or minimizes within-node deviance of the resulting two nodes (daughter nodes). This algorithm, applied recursively, allows partitioning of the original data into increasingly pure subsets based on simple decision rules. In a regression setting, each sample allocated to a particular terminal node is given the mean value of that node, and this value is compared to the observed value. Problems with standard tree-based regression include optimality concerns and over-fitting. These are addressed in SGBT models, wherein large numbers of smaller trees ($n_{nodes} \sim 3$; selecting single splits from among predictors) are grown, each building on the last by incorporating the previous residuals in an additive weighted expansion; observations with larger residuals are preferentially weighted in subsequent iterations. This generic algorithm is rapid to implement, allows inclusion of predictors for which pair-wise associations are relatively weak, and can be used with non-normal data and for non-linear responses (Friedman 2002).

2.3.2 Model Validation

For soft-modeling of spectra, or chemometric modeling, one critical requirement is a data set large enough to permit separation of the sample data into training and verification sets. Many of the statistical tools have a propensity to over-fit to the training data. To provide a reasonable measure of predictive accuracy under implementation conditions, a hold-out validation set is retained. For spectral prediction of properties measured on all 3911 samples, model training was performed using 67% of the data (~ 2600 samples); models were evaluated using the remaining ~ 1300 samples. For target parameters available for only a fraction of the data set (e.g., total mercury, lignin content) we simply maintained the same relative proportion of test samples.

Method utility is inferred based on model efficiency for these validation data in all cases.

2.3.3 Model Diagnostics

Model efficacy for continuous prediction is based, in all cases, on standard prediction errors, and goodness-of-fit criteria. In all cases, we rely on model diagnotics from hold-out validation for model assessment because of the tendency of the modeling tools to overfit to the data when presented with the large number of predictors (n ~ 200) that spectra provide. We use a trio of indicators to permit model comparison:

1) *Coefficient of determination* – this familiar model diagnostic (r^2) summarizes the proportion of the total variance explained by the model.
2) *Standard error of validation* – this diagnostic is, in some ways, a more meaningful indicator of model efficiency because it evaluates the magnitude of residuals. The SEV is the root mean squared error (RMSE) between predicted and observed for the hold-out validation data set. It is reported in the physical units used in model development.
3) *Relative performance determinant* –RPD is the ratio of population standard deviation for a particular analyte to SEV for that analyte. The primary purpose of computing this statistic is to permit model comparison between analytes (where SEV is unit specific, RPD is unitless). There is also relatively well-established literature precedent for evaluating model performance on the basis of RPD. Dunn et al. (2002) and Chang et al. (2001) suggest that models with RPD values greater than 2 are immediately useful for prediction, while models

with RPD values less than 1.5, though potentially useful for some very rudimentary information, are of limited prediction utility. Intermediate values are useful for certain applications, particularly those where mapping is required. We use these thresholds to evaluate our model performance.

2.3.4 Analysis of Model Residuals

Frequently, where a predictive model fails is as informative as where it is successful. We examine model residuals in two ways; first, we examine covariance between model residuals and various plausible confounders (other soil analytes, distance from canals). Second, we can map the spatial distribution of errors to determine if there are consistent geographic patterns to the error that might suggest some local unmeasured confounder. Generic methods for mapping are described below; in this case, we map the residual error for each of the modeled soil properties.

2.4 Chemometrics: Categorical Diagnostic Prediction

2.4.1 Statistical Tools

Categorical chemometrics use the same spectral information, but rather than targeting a continuous response, the spectra are used to quantify to probability of class membership. For this task, we make use of gradient boosted tree classification (GBT), a method well suited to complex data sets like the spectral libraries employed here, and with the distinct advantage of being non-parametric and insensitive to predictor co-linearity. The method works along the same lines as classification trees (CT - Breiman et al. 1986) – which employ binary recursive partitioning rules to sequentially allocate samples to target classes based on predictor levels – but assembles a large number of small trees in an additive weighted expansion that solves some of the fundamental problems of the CT approach. The GBT method has been used successfully for spectral prediction (Brown et al. 2006), and several low-cost statistical packages now provide a tool-set for this kind of analysis. Other methods for categorical discrimination (discriminant analysis, logistic regression) have been tried as part of this research as well, and in each case, the GBT approach outperforms them when measuring validation model efficiencies, and those results are omitted from this report.

2.4.2 Model Validation

As with continuous spectral prediction, all models were developed (trained) using 67% of the data and evaluated using the remaining 33% of the samples. Method utility is inferred based on model efficiency for these validation data only; as with continuous models there is a strong tendency for over-parameterized models to perform well below calibration expectations when applied to a hold-out data set.

2.4.3 Model Diagnostics

For each, we evaluate model efficiency using:

1) *Overall accuracy* –number of correctly classified samples over total number.
2) *Specificity* – percentage of negatives correctly predicted.
3) *Sensitivity* – percentage of positives correctly predicted
4) *Odds ratio* –ratio of the odds of correct prediction to odds of incorrect prediction. (Agresti 1990).

The odds ratio (OR) in particular provides insight into model efficiency because it balances the sensitivity and specificity metrics; values greater than 10.0 are generally considered to be indicative of a useful diagnostic model provided that the prevalance of cases (i.e. proportion of total samples meeting one of the threshold criteria in Table XX) is a substantial fraction of the total population; where prevalence is low, corrections to the OR are required.

2.4.4 Analysis of Model Error Rates

Each model provides a classification prediction and a probability of class assignment, allowing us to quantify how errors arise. In particular, we can condition the error on various confounding covariates (e.g., soil properties, location, distance from canal) to see if there are systematic reasons for misclassification. Given significant confounders, we can identify necessary ancillary information that might improve future classification accuracy.

2.5 Calibration Free (Unsupervised) Methods

2.5.1 Rationale

All discussion of spectral technique to date has focused on prediction of existing environmental indicators. In particular, we focus on attributes of soil nutrition (total P) or quality (organic content and lability, Ca/Fe/Al/Mg content). Despite observed success at

predicting these environmental covariates that are used as proxies for multi-attribute environmental quality, we contend that spectra may be useful in a more fundamental way. In previous work (Cohen et al. 2006) we showed that spectra may actually out-perform biogeochemical measurements in diagnosing environmental condition with respect to accuracy and repeatability. While spectra alone won't provide a degradation mechanism, which is clearly a critical part of environmental diagnosis, the ability to reliably map and monitor environmental condition using an indicator that, in principle, integrates across the soil environment is appealing for adaptive management. Figure 4 underscores the need for effective diagnosis in order to reliably identify management sensitive covariates; here we explore the feasibility of using spectra directly (i.e., not calibrated to proxy variables for assessing environmental condition).

2.5.2 Statistical Tools

Techniques for this are two-fold. First, there is the possibility of direct supervised classifications. Figure 4 summarizes the difference between indirect spectral models and direct spectral models. The direct calibration to ecological condition is not trivial, primarily because it relies of some scientifically acceptable benchmark that can provide a calibration target. In the past, total P has been used as one proxy for environmental quality (particularly along the WCA2A nutrient gradient); other proxies include plant community composition, extracellular alkaline phosphatase activity (which is moderately predictable using spectral models), invertebrate community composition, and uptake pathways of nitrogen that result in differential isotopic fractionation. Our work here is a preliminary investigation on the utility of spectra as a direct indicator for conditional assessment (Pathway 3A in Fig. 4). We focus on categorical prediction of phosphorus because no reliable benchmark for ecological condition exists. Specifically, we develop categorical models using spectral reflectance as the predictor set and TP > 500 mg/kg as the case (response = "1") and TP < 500 mg/kg as the control (response = "0"); we used standard binomial model diagnostics to evaluate effectiveness, including the validation odds ratio, model sensitivity and specificity, in addition to the overall prediction accuracy. We use spectrally extracted distance measures between samples based on principal components analysis as a measure of similarity between cores at the center of WCA2A in areas designated as minimally ecologically impaired and cores at the ecologically degraded edges. Our goal is to understand where, in the sequence of ecological indicators that includes P enrichment, macro-

invertebrate community changes, periphyton floc changes and vegetative community changes we observe concomitant changes in soil spectra. We use spectral similarity analysis in WCA2A as a tool to screen spectral features; these results are presented here, while the more detailed analysis of the WCA2A gradient using calibration-free methods remains an ongoing effort.

In Figure 4 there is an additional under-studied relationship. Pathway 3B links unsupervised categories of spectra to ecological condition. Again, because there is no agreed upon definition of an ecological case (i.e., degradation) vs. control (i.e., intact), concordance between *a priori* classes and estimated condition are not immediately possible. However, we pursue this technique because of its appeal in taking advantage of extremely high analytical precision for spectral measurements. Various unsupervised classification techniques exist, including the well known k-mean clustering algorithm. However, nearly all standard tools force a user-specified number of clusters, and class assignment is based on some distance metric from self-organizing cluster centroids. This both limits the ecological specificity of the technique because the number of clusters is generally unknown, and provides a deterministic answer to what is more realistically a probabilistic question (i.e., allocating sample to class Y rather than evaluating the probability of various class memberships). A modern alternative to the k-means approach uses a maximum likelihood approach to determine the probability of class membership; samples are assigned to the class with the highest probability, but these probabilities are retained for more in-depth examination. This process, called expectation maximization (Witten and Frank 2000), has several important advantages for this application. First, the number of classes does not need to be specified a priori as with k-means clustering. Rather, a v-fold cross validation procedure helps determine the degree of evidential support for different numbers of classes, selecting the categorical resolution that is best supported. As such, EM is a much more powerful data mining/machine learning tool, similar in pattern recognition potential to self-organizing feature maps (Kohonen maps) but more scalable. Second, the EM approach can be scaled easily to large numbers of co-linear predictors, which is the situation faced with spectral data. Third, EM permits entry of categorical predictors which maximizes the generality of the application for situations where there are known ordinal and nominal confounders. Finally, as mentioned above, the algorithm provides classification probabilities instead of just class assignments, which can be useful where the physical basis of assignments is unknown, or where there is observed discordance between a priori spectral classes and environmental categories.

This approach was implemented using the data-mining tools in Statistica v. 7.0 (StatSoft, Inc. Tulsa, OK).

2.5.3 Metrics of Utility

Absent an agreed upon case definition, assessment of pure spectral classes is limited to the degree to which soils cluster in meaningful ways according to biogeochemical properties. After identifying the most statistically defensible number of EM classes from spectra, we revisit associations with soil measurements. We use ANOVA to evaluate soil differences between inferred EM classes as a means of understanding the functional implications.

Moreover, it is possible to map locations of spectral cluster membership to see if clusters group spatially, or by depth. To the extent that they do, changes in class membership probabilities through time may be useful for quantitative ecosystem change detection.

2.6 Mapping

One of the objectives of this work is to explore the spatial patterns that emerge from our consideration of point-level data. Interpolation-based mapping using the tools of geostatistics is applied to only a subset of the analytes that have been spectrally analyzed. Specifically, only those analytes measured under this contract are mapped; maps of the original measured data are not. However, we do provide maps of chemometric model residuals (error term between predicted and observed values) as a means of visualizing prediction errors in space and determining the universality of spectral predictions for the Greater Everglades system. In addition, we provide maps of residual fiber content of soils, expectation maximization classes, and total mercury concentrations. While it would be advantageous to examine spatial pattern in each ecosystem partition area (i.e., WCA1, ENP) separately because of the strong delimiting effect of canals, for the purposes of this visualization effort, we selected to examine the data as a whole. Further, all maps developed herein are for the top layer of soil (0-10 cm); we omit mapping the floc layer because it is discontinuous, and also deeper soil layers because we are most interested in soils of recent origin nearer the surface.

In all mapping efforts, we compared two methods: ordinary kriging and local spline interpolation. Kriging, which is a well-known interpolation technique based on development of a semi-variogram, honors the observed data and develops a statistical prediction of unsampled

points based on nearby samples; the influence of nearby samples is quantified using semi-variance analysis. Kriging follows three steps. First, we develop a semi-variogram at the scale of the entire dataset that determines the extent and magnitude of spatial autocorrelation. The output of each semi-variance analysis is a measure of the range (spatial length of significant autocorrelation), nugget variance (semi-variance at zero separation distance), sill variance (semi-variance beyond the range) and the relative structure (nugget-to-sill ratio). Second, we use the geometric properties of the model semi-variogram (fit to the experimental semi-variance data at different separation distances) to predict values at unsampled locations; the model semi-variogram can take several forms, be isotropic or directional, and can be a composite of multiple model structures. The parameters of the best-fit semi-variogram model describe the distance and degree of influence that measured points have on unmeasured points; the interpolation process is then simply the application of those parameters to unsampled locations.

Finally, we examine cross-validation efficiency of the kriging model to observational data. This cross-validation procedure iteratively predicts the level of sampled points based only on the points in their vicinity, and compares that prediction with the observed level. In this way, all samples are used to evaluate the efficiency of the interpolation algorithm so that some estimate of prediction uncertainty can accompany the map. As part of the prediction error evaluation, we report the root-mean-squared error (RMSE) as a quantitative measure of residual variability, and the r^2 between predicted and observed as a measure of variance reduction.

In general, we follow the described protocol for all maps that produced. However, for mapping soil mercury, we explored options for more sophisticated mapping in some detail to ensure that resulting map products take fuller stock of the spatial patterns observed.

2.6.1 Mapping Model Error Rates

Maps of model residuals were made using ordinary kriging; spatial statistics (nugget variance, sill:nugget, range) were reported for all parameters as well as metrics of spatial autocovariance of the residuals. Where significant autocovariance exists, we infer that local conditions (geochemical, biological) lead to low-incidence spectral differences that are missed by a global chemometric model. As such, areas where there are locally elevated residuals may be candidates for more local calibration models, as data permit. Development of these local calibrations was not possible in the scope of this work.

2.6.2 Mapping Soil Total Mercury

Geostatistical methods were used to characterize spatial variation and map total mercury per unit mass (THg_M) and per unit area (THg_A) across the Everglades. Mercury variation was analyzed for spatial pattern across the Greater Everglades, among hydrologic partitions, with distance from canals (as evidence for surface hydrologic Hg delivery), and by vegetative community type (as evidence for differential enrichment). Because P enrichment is a primary anthropogenic influence impacting ecosystem structure and function in the Everglades (Noe et al. 2001) and because P distribution is relatively well understood (Bruland et al. 2006, Reddy et al. 1991), we were particularly interested in associations between THg_M and TP concentrations. In particular, we examine global covariance between the two, and also compared concentration profiles with increasing distance from the canals. We expected TP to decrease with distance from canals; if THg_M follows the same pattern, it would suggest that surface water delivery was active. Finally, THg_M, TP and TC concentrations were contrasted among sub-regions of the Everglades, and between impacted and un-impacted areas of the Everglades. All exploratory data analyses were done using Statistica 6.0 (Statsoft, Inc. Tulsa, OK). All interpolations, correlations, semi-variances and cross semi-variances were computed per unit mass (THg_M) and per area (THg_A).

Semivariance analysis quantifies spatial dependence among observations of a particular variable, measured as the degree of similarity of paired observations as a function of sample spacing or lag (h). Spatial dependence was analyzed using semivariograms, which plot semivariance against lag spacing (Webster and Oliver 2001):

$$\hat{\gamma}\ (\mathrm{h}) = \frac{1}{2n} \sum_{i=1}^{n} \{z\ (x_i) - z(x_i + h)\}^2 \qquad \text{Eq 1.}$$

where, n is the number of sample point pairs separated by a distance h, and $z(x_i)$, $z(x_i + h)$ are the sample measurements at points within that lag distance. Because semivariograms are sensitive to the presence of spatial outliers, observations identified as such using Anselin's Local Moran's Index (z-score < -1.96; Anselin 1995) were omitted from semi-variance analyses (though not from exploratory analyses). In order to identify spatial autocorrelation structure at different scales, semi-variograms were computed for regular lag spacings ranging from 100 to 1000 m. The optimal lag spacing (500 m) was selected based on the stability of the semi-variances, and the minimum number of sample pairs within each lag (all were > 30).

Spatial interpolation was performed using Ordinary Kriging (OK), the most common kriging method. OK provides a Best Linear Unbiased Estimator (BLUE) because error variance is minimized, predictions are linear combinations of available data, and mean error is reduced to zero (Isaaks and Srivastava, 1989). Detailed methodology is described in Isaaks and Srivastava (1989), Goovaerts (1997) and Webster and Oliver (2001). After exploratory spatial data analysis (trend surface detection) in ArcGIS 9.0 (ESRI, Redlands, CA), semivariograms were modeled in Variowin (Pannatier 1996). OK was done in Geostatistical Environment Modeling Software – GEMS (Remy 2004), and final OK predictions were developed on a 200 x 200m grid.

To estimate model error at unsampled points a validation procedure comparing model results with cross-validation observations was performed. Given the sample size (n = 600) and spatial extent (8220 km^2), cross-validation was employed, wherein estimation of prediction error is made by interpolating using all the samples except one, and comparing predicted and observed; iterative application of this process until all sites have been "held-out" and predicted permits representation of prediction errors without sacrificing data density (Goovaerts 1997). Prediction quality was assessed by the fit between observed and predicted values at validation sites using the mean error (ME), root mean squared error (RMSE) and Pearson correlation coefficient (r). ME measures systematic bias in predictions and indicates whether the model is, on average, over- or under-estimating. RMSE measures prediction accuracy, while r quantifies covariance between predicted and observed.

Because of the spatial extent of the study area and range in THg_M concentrations observed, we contrasted spatial patterns among the sub-regions of the Everglades. However, sample sizes within each of the sub-regions fall short of recommended minimum samples size (n = 100) for estimating semivariograms (Webster and Oliver 1992). Based on geographic proximity and ecological similarities, the landscape was clustered into five agglomerated regions (Fig. 1): Big Cypress National Preserve (BCNP – lumping three sub-regions); Everglades National Park (ENP – lumped with the Model Lands); Water Conservation Areas 1, 2A and 2B (WCA1/2); Water Conservation Area 3A North, Holeyland, and Rotenberger ($WCA3A_N$); and Water Conservation Areas 3A South and 3B ($WCA3A_S$/B).

Spatial structure of THg (per mass and area) within each sub-region was identified using semivariograms computed using Eq. (1). We contrasted spatial correlation of THg_M with other soil properties (TP and TC) using cross-variograms between sub-regions that represent impacted

(WCA3AN) and unimpacted (ENP) parts of the Everglades; site impact status was based on a TP threshold of 500 mg/kg (DeBusk et al. 2001). Cross-semivariogram analysis quantifies spatial correlation between two variables such that using information about one can enhance interpolation accuracy for the other (Webster and Oliver 2001):

$$\hat{\gamma}_{uv}\ (\mathrm{h}) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z_u\ (x_i) - z_u\ (x_i + \mathrm{h})\}\{z_v\ (x_i) - z_v\ (x_i + \mathrm{h})\} \qquad \text{Eq 2}$$

where, z_u and z_v are variable measurements, x_i and x_i + h are sampling locations separated by a distance h, and *m* is the number of sample point pairs separated by a distance h. The structural variance (Q), computed as the ratio of partial sill to total sill (= nugget variance + partial sill) was used to evaluate the spatial variance explained by the semivariogram (Morris 1999). Values approaching 1.0 indicate strong spatial structure while values near 0.0 indicate either low spatial structure or structure at spatial scales larger or smaller than those observed.

2.6.3 Mapping Soil Carbon Quality

Ordinary kriging was applied to spectrally estimated soil carbon quality. The measured data were spatially sparse; only 348 samples were analyzed, and these were obtained at different depths to ensure full representation of the biogeochemical diversity of the ESM population. As such, interpolation of the direct observations was not considered feasible for this work. Contingent on observing strong spectral prediction, maps can be developed of spectrally inferred soil C quality since predictions can be made at all locations and depths for which spectra were available. We examined standard geostatistical metrics (sill:nugget, range) of autocorrelation as well spatial pattern and estimated error from the interpolation.

3. RESULTS

3.1 Routine Analyte Summary

3.1.1 Summary Statistics

Soil analyses were performed during 2003-2004. For each site, between 3 and 4 depths were sectioned, and each depth section was analyzed for bulk density, organic matter content (LOI), total N, total C, total P, total inorganic P, total Ca, total Mg, total Fe, and total Al. Mean and standard deviation data for each of the measured parameters are provided in Table 3. What is clear from the table is the extreme range of soil properties present in this work; organic matter contents range from less than 1% to over 99%, total P concentrations range from 18 to nearly 2000 mg/kg, and iron and calcium concentrations vary over 3 orders of magnitude. This variation is essential for spectral modeling; datasets where samples are homogeneous with respect to functional attributes tend to be more sensitive to over-fitting and poor generality.

Table 3. Summary data for ESM reference library.

	Average	Std. Dev.	Max	Min
Bulk Density	0.23	0.23	1.88	0.00
Organic Matter (%)	64.01	30.88	99.51	0.91
Total C (g/kg)	337.27	139.44	544.48	3.88
Total N (g/kg)	23.01	10.92	46.25	0.26
Total Aluminum (mg/kg)	6328	9453	125455	150
Total Calcium (mg/kg)	89136	107989	499776	414
Total Iron (mg/kg)	6559	6750	137045	19
Total Magnesium (mg/kg)	2261	1743	21902	0
Total Phosphorus (mg/kg)	346.16	249.36	1953.40	17.90
Total Inorganic Phosphorus (mg/kg)	84.28	102.98	1428.03	2.29

3.1.2 Correlations

There were strong correlations between the parameters; Table 4 summarizes these correlations. Several of the variables (bulk density, total P, total inorganic P, total Ca, total Mg, total Fe and total Al) required log transformation prior to analysis; this transformed data was used for correlations and for subsequent chemometric development. Also shown in Table 4 (along the main diagonal) are the r^2 values for multiple regression prediction of each parameter. That is, a model was developed to predict each parameter based on all the others; this value provides a useful benchmark for later analysis of spectral prediction models because it

conceptually quantifies the degree to which the spectra are acquiring information about a particular parameter that is independent of the suite of chemical measurements.

Table 4. Summary of correlations between ESM soil analytes. Along the main diagonal (shaded grey) is the r^2 for predicting each analyte from other analytes. This is a useful benchmark to assess spectral models. Note that moment calculations are for untransformed data; correlations use transformed data.

	Organic Matter (%)	*Total Nitrogen*	*Total Carbon*	*Bulk Density*	*Total P*	*Total Inorganic P*	*Total Ca*	*Total Mg*	*Total Fe*	*Total Al*
Organic Matter (%)	0.87‡									
Total N (g/kg)	0.93	0.90								
Total C (g/kg)	0.98	0.92	0.86‡							
Bulk Density†	-0.71	-0.70	-0.70	0.66						
Total P (mg/kg)†	0.63	0.60	0.56	-0.52	0.58‡					
Total $P_{inorg.}$ (mg/kg)†	0.41	0.39	0.36	-0.46	0.86	0.33‡				
Total Ca (mg/kg)†	-0.60	-0.60	-0.50	0.31	-0.34	-0.09	0.69			
Total Mg (mg/kg)†	0.21	0.08	0.22	-0.17	0.30	0.31	0.23	0.37		
Total Fe (mg/kg)†	-0.13	-0.06	-0.16	0.22	0.10	0.12	0.09	-0.02	0.54	
Total Al (mg/kg)†	-0.24	-0.18	-0.32	0.44	0.09	0.03	-0.02	0.01	0.69	0.69

† - Parameters natural log transformed to meet normality assumptions.
‡ - These r^2 values are for prediction without using the most obvious correlate (for TC = LOI, for TP = TPi).

Figure 8 shows the soil attribute correlations graphically. For many of the relationships, the low correlation coefficients in Table 4 may be due to more complex inter-relationships. For example, the relationship between organic matter and calcium appears to be non-linear, with both low and high calcium values covariant with low organic matter; this is presumably a function of calcite (high Ca) and sand (low Ca) end-members, both of which have low carbon contents. Similarly, the association between P and Ca appears to be non-linear; this result is surprising given well known associations between Ca and P sorption. We surmise that, generally, TP increases with Ca until extreme levels, at which point insufficient P is available for sequestration. A spatial analysis of the locations where Ca and P are inversely correlated would illustrate this potential mechanism; such analyses are beyond the scope of this work.

There is evidence of some outliers. None were removed from this preliminary spectral modeling effort, but subsequent iterations of the predictions might warrant their removal.

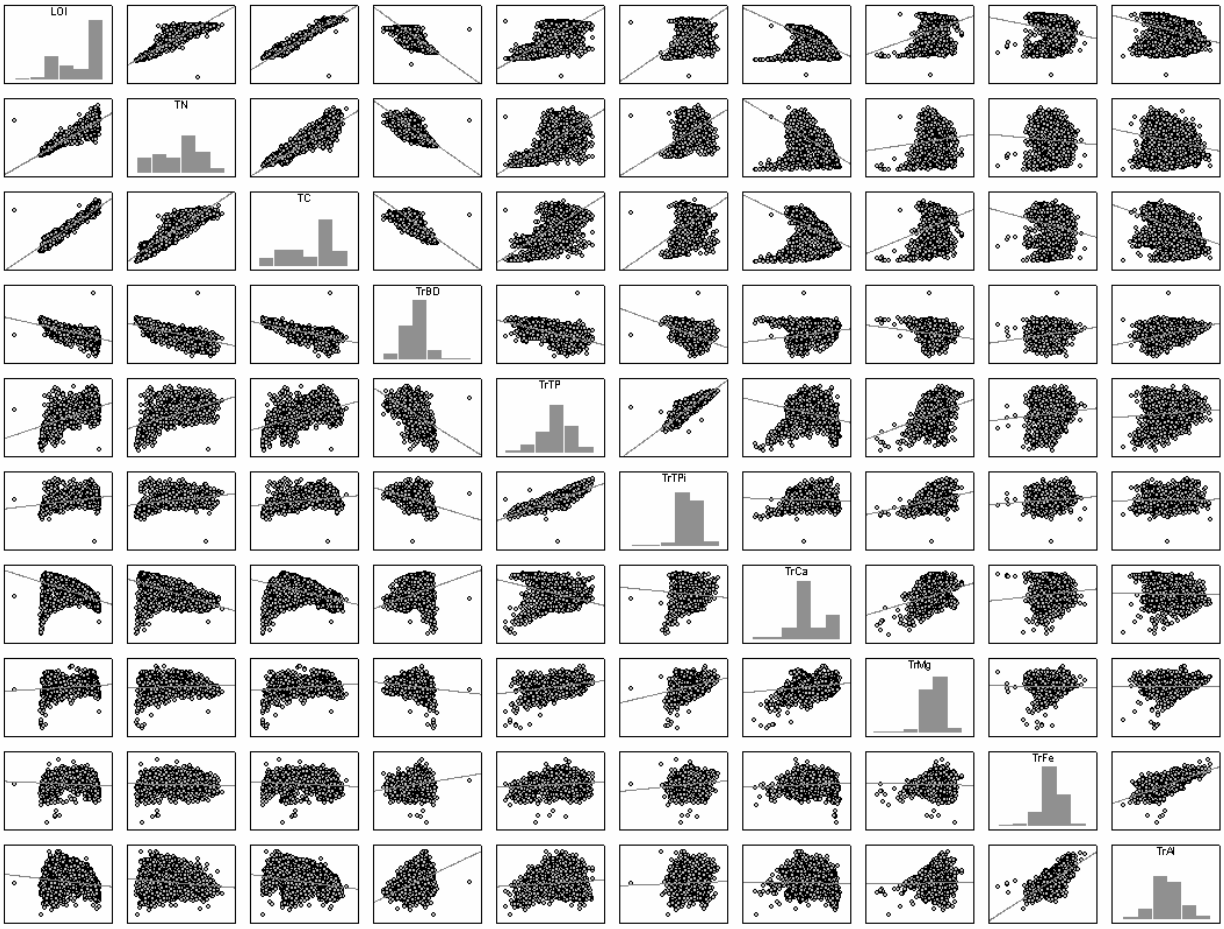

Fig. 8. Scatter plots for each of the analytes. Along the main diagonal is a histogram. Units are excluded from this plot for clarity. Refer to Table XX for units and magnitude.

3.1.3 Sub-Sampling

We used the measured soil attributes to extract latent factors from a principal components analysis. Our rationale is that many of the observed attributes are correlated, and a smaller number of variables formed by linear combinations of the original data can explain the variance in the data set. This reduced dimensionality allows us to sub-sample the population of soils into representative subsets; this is useful for two reasons. First, the development of chemometric models requires separation of the training data into calibration and validation subsets. While random sampling is typically effective at ensuring that the calibration data set is representative of the whole population when the calibration set is large, we anticipate requiring much smaller calibration sets for certain portions of this work. To ensure that data subset selected to train the model (which relates spectra to soil properties) is representative of the population, we used Latin Hypercube subsampling (LHS); this technique draws samples in a stratified random manner,

with the strata defined by the principal components axes. Operationally, we divided each of the first 2 PC axes into 4 zones, and an additional 3 PC axes in 3 zones. This creates 4^2x3^3 (n=432) zones in PC space; one sample is selected at random from each zone with at least 1 sample to develop a subset with maximal probability of adequately representing the population. Overall, 348 sites were selected (84 zones had no samples). The selected samples (drawn from the entire population over space and depth) are shown in Fig. 9; Table 5 summarizes the soil properties if the population and the subset. Clearly the subset spans the variance contained in the first 6 principal components axes (PCA was run on soil chemical properties, not soil spectra), which cumulatively account for more than 85% of the original dataset variance.

Table 5. Summary of soil properties for sub-sample

	Mean Value				
	Population	Sample	Population SD	Δ/SD†	p-value
Bulk Density	0.20	0.23	0.22	12.2%	0.03
Organic Matter (% LOI)	64.74	61.11	30.87	11.8%	0.04
Total Aluminum (mg/kg)	6.21E+03	6.06E+03	9.12E+03	1.7%	0.76
Total Calcium (mg/kg)	8.75E+04	9.35E+04	1.08E+05	5.5%	0.33
Total Carbon (mg/kg)	342.4	325.6	139.0	12.1%	0.03
Inorg. Phosphorus (mg/kg)	86.4	79.5	110.0	6.2%	0.27
Total Iron (mg/kg)	6.52E+03	6.12E+03	6.66E+03	6.0%	0.28
Total Magnesium (mg/kg)	2.25E+03	2.21E+03	1.71E+03	2.5%	0.63
Total Nitrogen (mg/kg)	23.25	21.86	10.92	12.7%	0.02
Total Phosphorus (mg/kg)	351.4	331.9	253.4	7.7%	0.17

† - Δ = difference between population and sample mean; SD = population standard deviation

3.2 Spectral Library Visualization

3.2.1 Spectral Precision

Instrument precision for post-dispersive spectrometers like the one used for this work are extremely high, particularly in comparison to typical laboratory methods. As such, we defined an unacceptable replicate (n = 4) variance to be 3%, averaged across the entire spectrum. This quantity was termed the precision index; Fig. 10 summarizes precision indices of accepted scans for the entire ESM database (n = 3911). We conclude that, overall, the instrument selected performs at a precision level sufficient to support subsequent modeling. We note, however, that roughly 1 in 20 samples were rejected and rescanned, underscoring the need for real-time precision evaluation. Fig. 11 shows spectra from two samples, one which was written to the database (PI = 0.95) and one that was not (PI = 3.45).

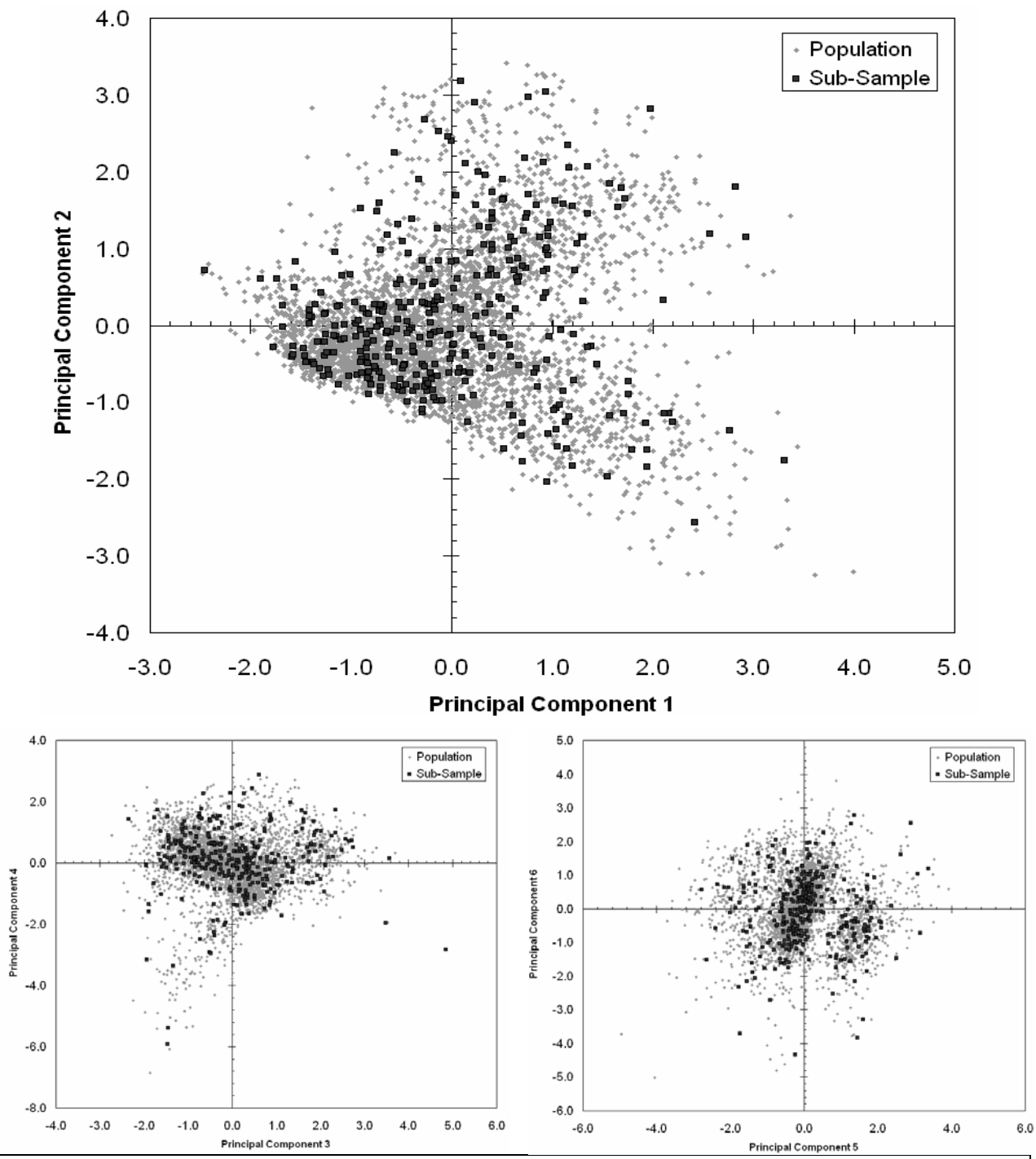

Fig. 9 – Principal components space for ESM population (grey circles) and a Latin hypercube sub-sample thereof (n = 348; black squares). These 6 PC axes account for over 85% of the total population variance.

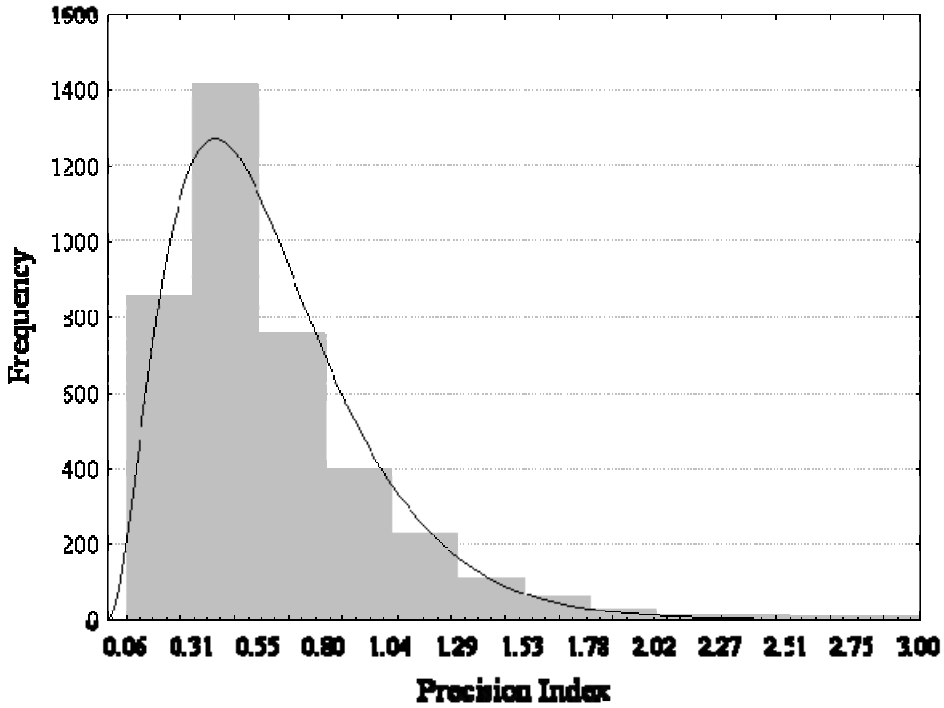

Fig. 10 – Spectral precision indices (PI - %) for all samples from the ESM spectral library.

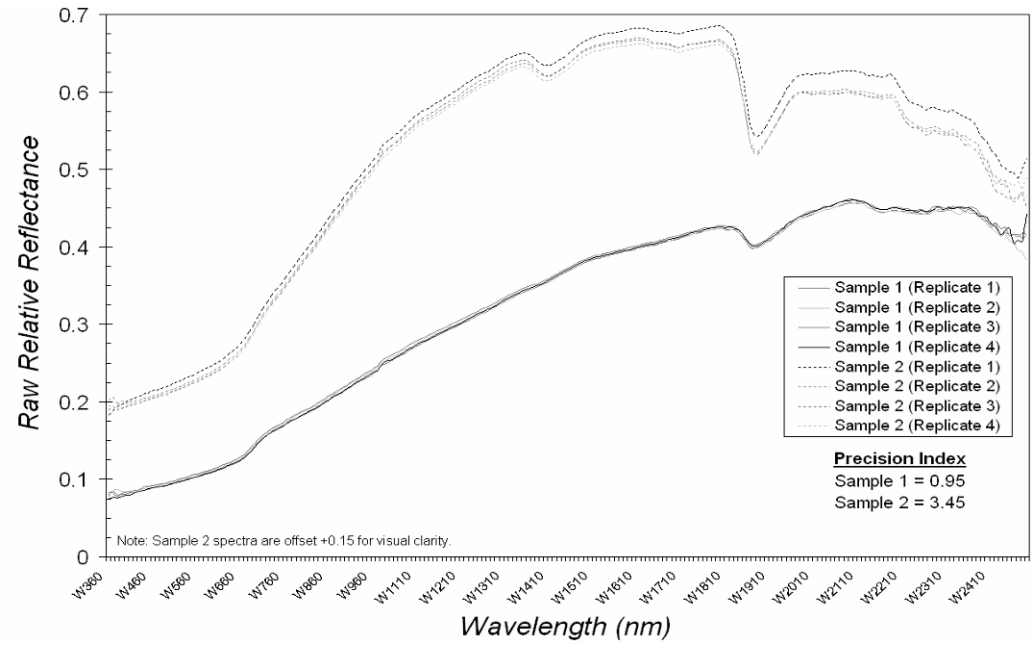

Fig. 11 – Spectrographs from two samples (n = 4 replicates) with associated precision index values. As shown, Sample 1(lower albedo spectra) would pass our precision criterion and be written to the database; Sample 2 (higher albedo spectra) would be rejected and rescanned.

3.2.2 Spectral Variance

End-member samples from the entire ESM library were selected based on the Latin hypercube sampling described above. Their respective spectra (1st derivative transformed for analysis) are displayed in Fig. 12. Notably, there are numerous regions with high spectral variability. Diagnostic regions for prediction of soil functional attributes are, however, frequently found in areas with comparatively low but systematic variability. As such, mechanistic interpretation of complex spectrographs has not proven a particularly effective approach. That is, the location and intensity of particular absorbance features is not directly mapped to prediction of functional attributes. Indeed, many of the statistically robust models developed in this work (described below) are comprised of pairwise correlations between a particular waveband and a particular attribute that are fairly weak; it is in a multi-variate setting where associations are conditional that prediction efficiency is obtained.

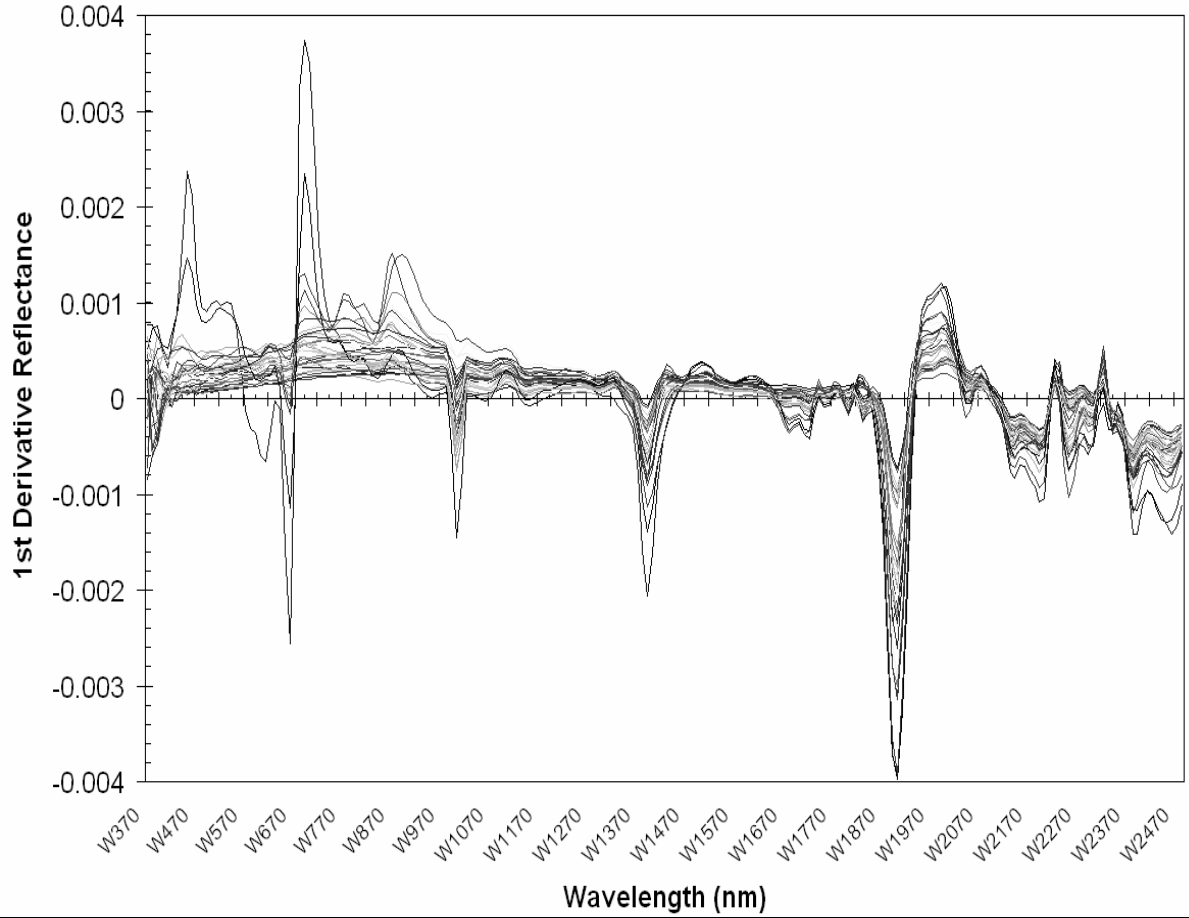

Fig. 12 – End-member spectra from the ESM data set. Spectra are derivative transformed with 2nd order smoothing.

A key feature of chemometric modeling is that statistics for processing high dimensional datasets like spectra are sensitive to over-fitting. Consequently, model efficiency for the training data may be an overestimate of actual performance when applied to new data. As such, a hold-out validation data set is selected to assess expected performance for new samples more realistically. To do this we partitioned the ESM population into training (67%) and validation (33%) based on the Latin hypercube sampling regimen previously described (choosing 33% of the data in each zone for validation). A visualization of the calibration and validation data sets is given in Fig. 13, demonstrating that they both capture the variability of the total population.

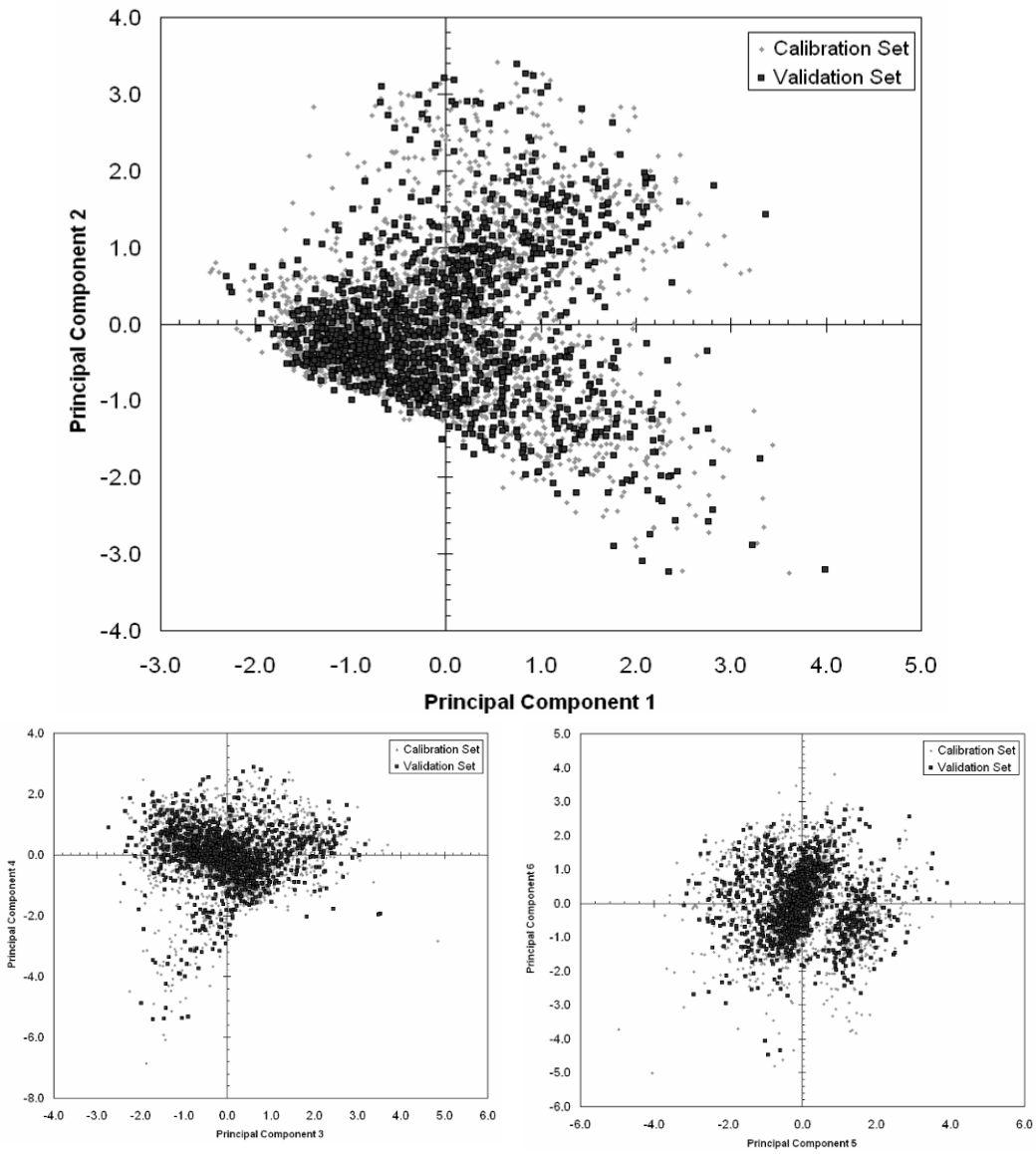

Fig. 13 – Calibration and validation data sets in spectral principal components space. This figure illustrates how well both sets represent to total population.

3.2.3 New Sample Projections

One stated advantage of NIRS approach to soil analysis for large-area ecosystem monitoring is that the quality and representation of the spectral reference library can grow with additional sampling efforts. In our case, there were over 200 additional samples collected as part of project activities beyond the original ESM library, and these provide a useful benchmark for both the representation of the system provided by the ESM dataset (n = 3911 samples from throughout the Greater Everglades used as our reference library), and the protocols that we've put in place to screen new samples for their concordance with that reference library. Figure 14 summarizes the results of this examination, which applied the loadings from a principal components analysis of the reference library spectra to the new samples. What we show there is that all samples except 1 collected for project activities (black squares) fall within the bounds of the spectral library (grey circles). The sample that falls outside of the reference bounds is a very deep core section (25-30 cm) which was not well represented in the reference library. The observation of reference library adequacy holds true for all of the principal components (1 through 6, which explain over 85% of the original variance). Further, we are able to apply the linear combinations that comprise the PCA extremely rapidly to new samples, which, in principle, allows us to identify samples falling outside the library bounds that require full laboratory analysis and subsequent inclusion in the library. Ordinarily, project would not start with as comprehensive a library as we have here; we fully expected that very few new samples will fall outside the calibration bounds. However, if this technique becomes more widely used, we foresee samples from tree islands, STAs, lake sediments and Florida Bay sediments being outside the current bounds. As spectral prediction of these kinds of samples becomes necessary, it will be increasingly important for the library to grow in a systematic, efficient way; protocols put in place in this work permit that future enhancement.

Another observation from the bi-plots in Figure 14 are obvious clusters of the new samples into two clouds in each plot. In Fig. 14A, the clouds delineate perfectly (accuracy = 100%) between marl soils and peat soils (an obvious spectral distinction). In Fig. 14B and 14C, the cloud corresponds dramatically to ridge and slough (separation accuracy is 94.5% on this basis alone), suggesting strongly spectrally distinctive peats between the two. The cloud in Fig. 14C corresponds almost perfectly (accuracy = 98.4%) to an aspect of sample preparation that we have not discussed: the process of grinding. At each site, three cores were collected in each

ridge and each slough, and two cores (Core 1 and Core 2) were hand ground using a mortar-and-pestle, followed by ball-milling to pulverize the samples. The third core was only hand ground to preserve texture for the methane digestion work. Spectral separation between these two methods underscores the need for standardized protocols if spectral methods are to be widely employed.

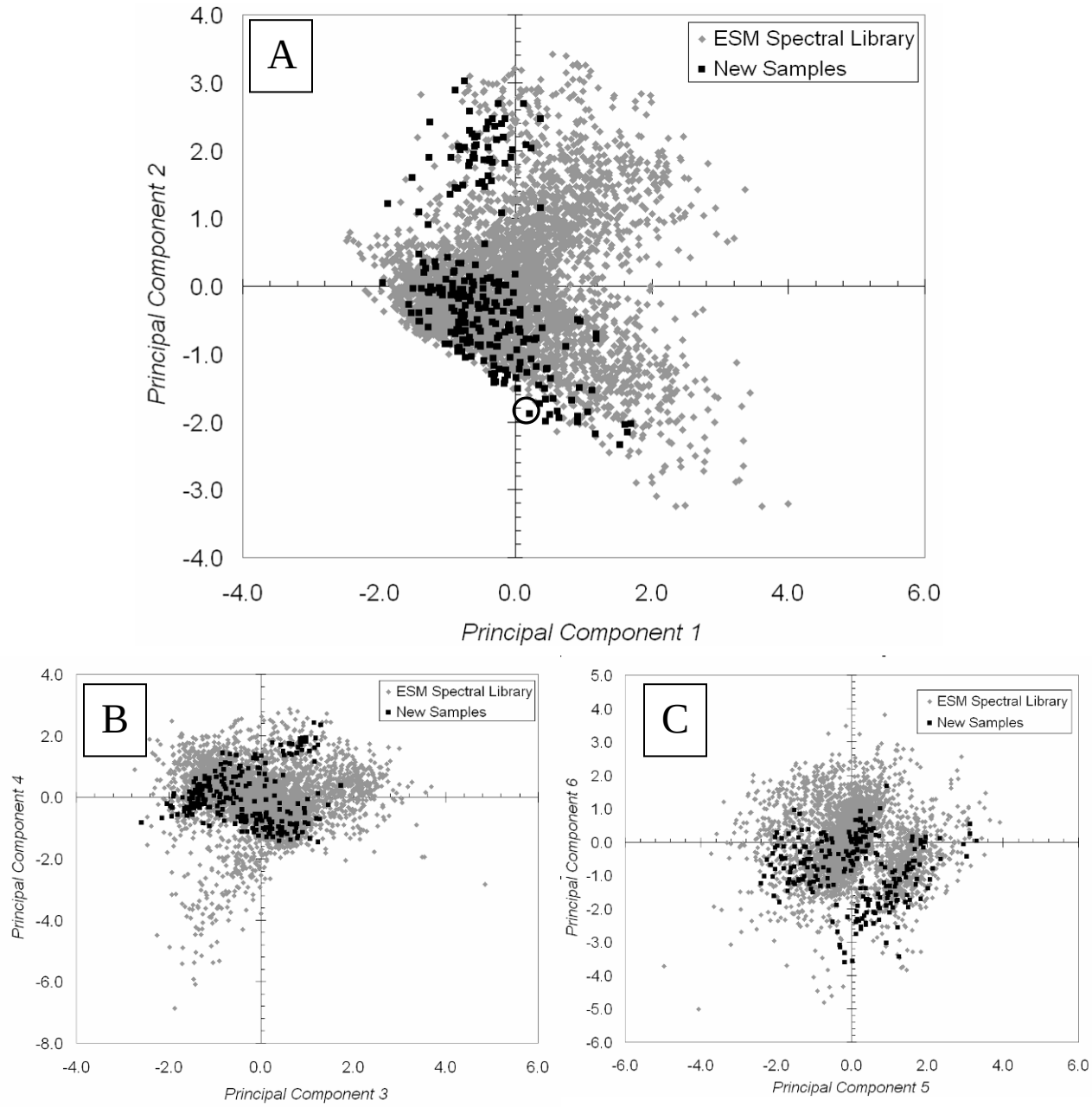

Fig. 14 – Results of introducing new spectra to reference library. Only one sample (circled) falls outside the multi-dimensional hull defined by the PC spectral space and would be incorporated into the library after full laboratory analysis. The distinct clouds evident in each bi-plot discriminate between a) peat and marl, b) ridge and slough, and c) sample grinding protocol.

3.3 Continuous Predictions

3.3.1 Comparison of Methods

Fig. 15 summarizes the comparison of two measures of model efficiency between two chemometric prediction methods (PLS and SGBT – see methods) across all of the parameters modeled in this study; Fig. 15A summarizes the coefficient of determination (r^2) between predicted and observed, and Fig. 15B summarizes the RPD, which is a measure of error (population standard deviation divided by the standard error of prediction). Note that the r^2 values can be compared with the values along the main diagonal in Table 4 for a proximate measure of the efficacy of spectra vis-à-vis inter-correlations between measured parameters. Both the r^2 and RPD values are for the holdout validation data set (33% of the data retained from algorithm training for more realistic estimation of predictive accuracy). What is clear from both comparisons is that there is limited difference in predictive power. In general, the SGBT appears to outperform PLS, but not for all parameters (total C and total Al are more effectively predicted using PLS). Overall, the degree of difference in predicting each parameter is so minor as to be functionally irrelevant. In this situation, the ability to extract information efficiently and apply the algorithms to new data easily are perhaps more important for deciding which method to use. Because of the computational intensity of SGBT and the much greater need for customized programming to implement the resulting algorithms (methods for PLS implementation are already very well developed in the commercial market), there are reasons to select PLS as the preferred method. Despite these considerations, we have selected to report the result of SGBT for all further analyses because of the moderate improvement in RPD that is observed for some soil properties (LOI, TN, Ca, bulk density).

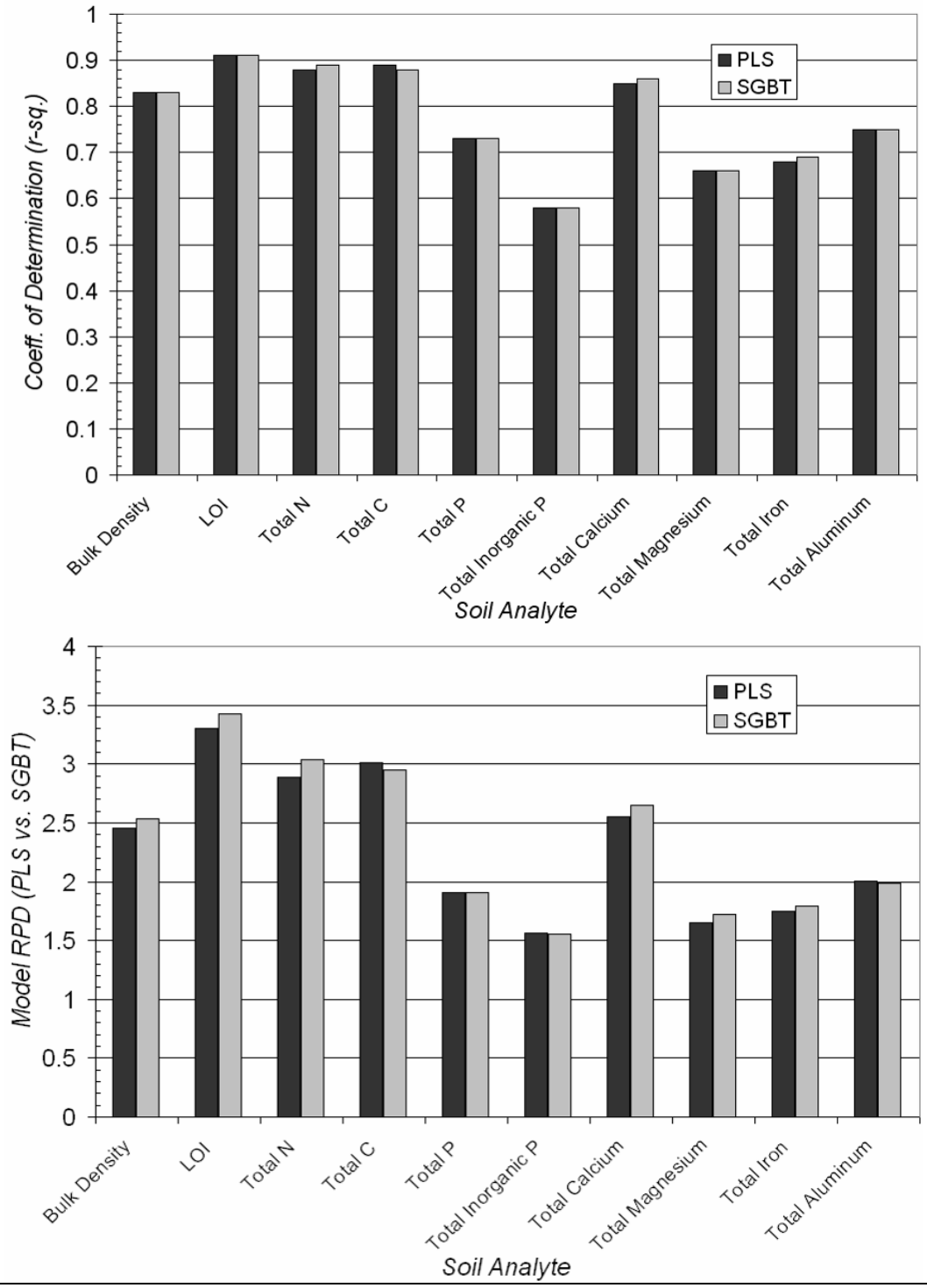

Fig. 15. Methodological comparison of partial least squares (PLS) regression and stochastic gradient boosted tree (SGBT) regression for chemometric prediction of soil properties. A) coefficient of determination (r^2) between predicted and observed, B) RPD measure of model efficacy, computed as the ratio of the population standard deviation to the standard error of prediction (values above 2 are considered directly applicable to routine analysis, values above 1.5 are considered to be of utility in certain situations).

3.3.2 Model Validation Results

3.3.2.1 Routine Analytes

The figures that follow depict the predicted vs. observed values for each of the 10 soil analytes measured in the ESM study (Figs. 16 – 25). In each case, the left figure is the calibration result (67% of the full data set) while the right figure is the validation result. For each, 4 measures of model efficacy are reported. The first is the slope of the fitted regression line; values significantly different from 1 illustrate bias in the prediction. Second is the coefficient of determination (r^2), which describes the fraction of the total variance in the observations described by the predictions. Third is the standard error (either of calibration or validation); this is the root-mean square error (RMSE) between predicted and observed, and can be interpreted in the specific units of the attribute being analyzed (note that many of the analytes have been natural log transformed prior to model development so the units are not the familiar mg/kg or % values in which results are typically reported). Finally, we report the RPD for each model, which is a uniform measure of model fit that can be compared across analytes. Values less than 1.5 are considered to be of limited utility, while values greater than 2 are usually indicative of models that have broad utility for a wide array of applications. In previous work (Cohen et al. 2007), we showed that the RPD for replicate laboratory measurements of the same parameter was between 1.5 and 2.5; that is, where spectral models exceed this selected thresholds, they are possibly as effective as laboratory analytical accuracy. More detailed examination of error sources (spectral vs. laboratory) are critically needed to establish the functional accuracy of spectra vis-à-vis labs.

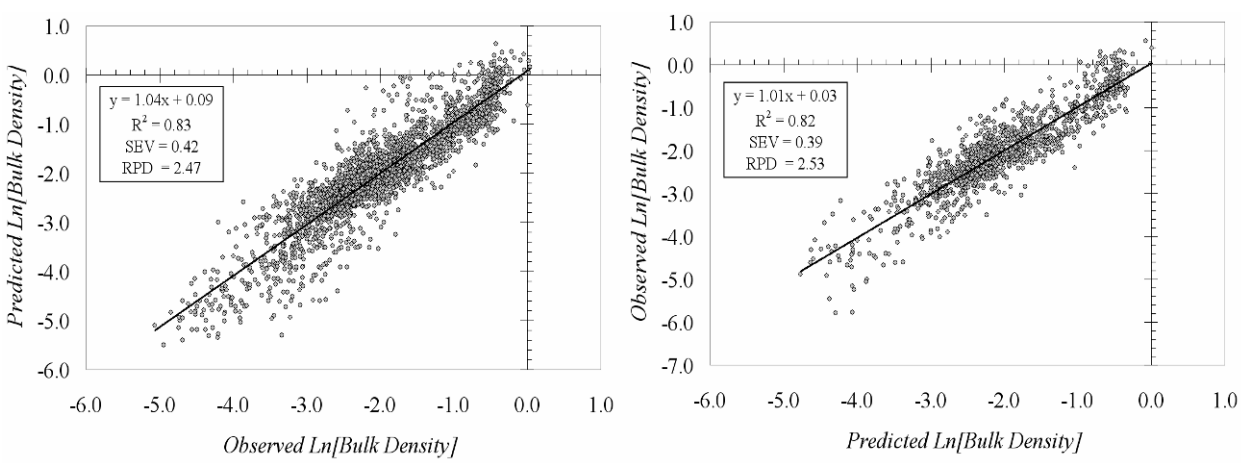

Fig. 16. Scatter plot of spectrally predicted vs. observed for bulk density observations; A) calibration and B) validation.

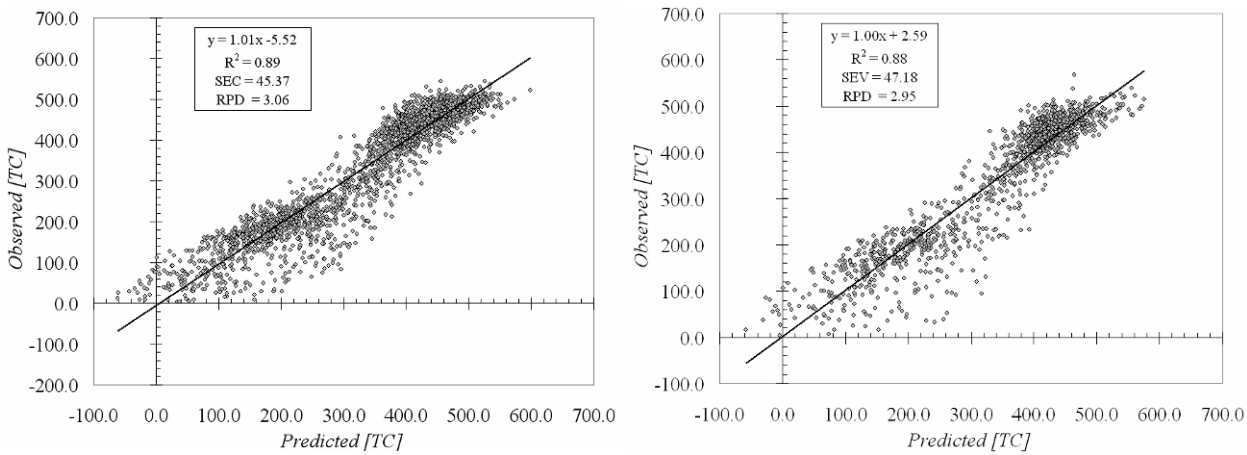

Fig. 17. Scatter plot of spectrally predicted vs. observed for total carbon observations; A) calibration and B) validation.

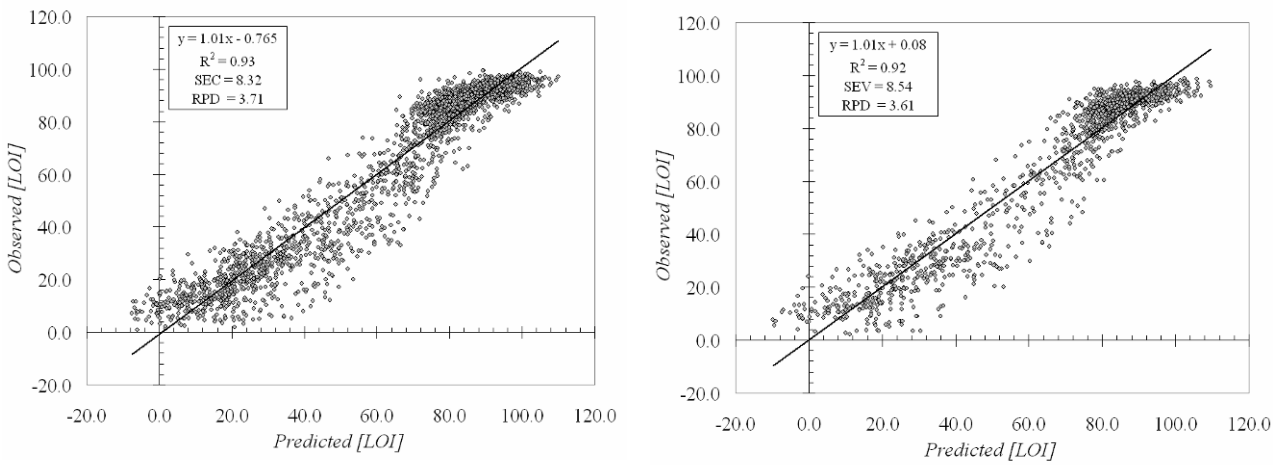

Fig. 18. Scatter plot of spectrally predicted vs. observed for organic matter observations (via loss-of-ignition); A) calibration and B) validation.

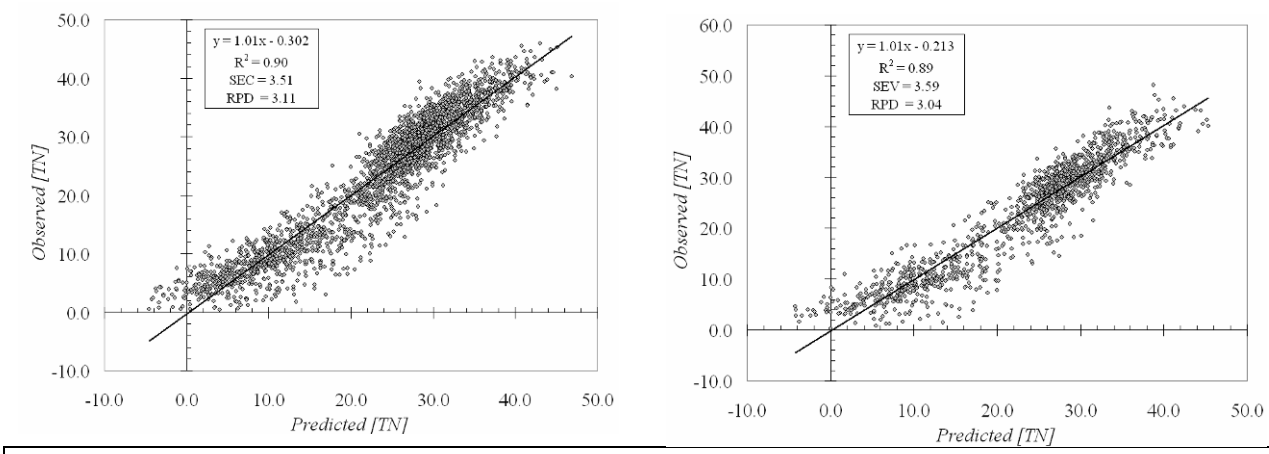

Fig. 19. Scatter plot of spectrally predicted vs. observed for total nitrogen observations; A) calibration and B) validation.

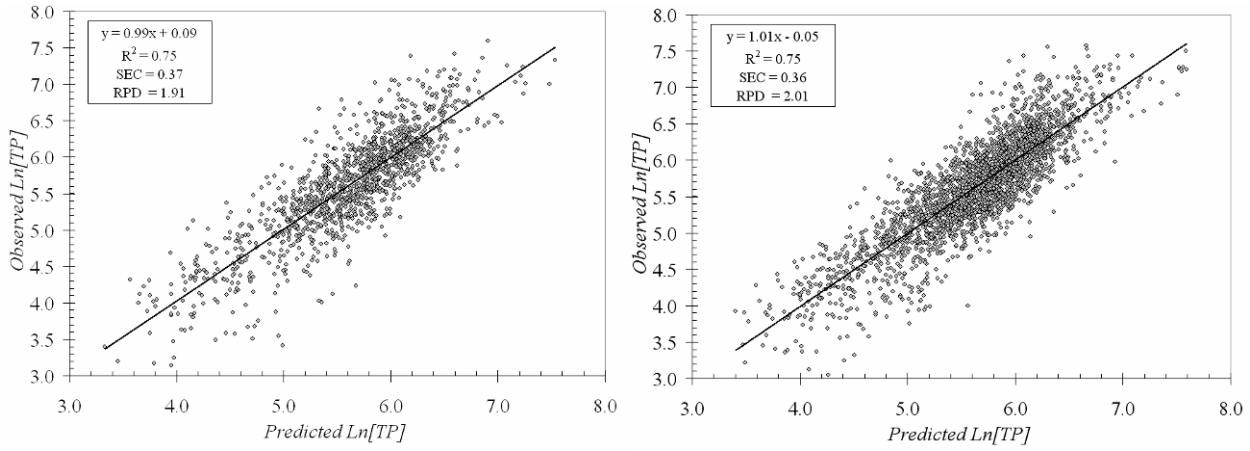

Fig. 20. Scatter plot of spectrally predicted vs. observed for total phosphorus observations; A) calibration and B) validation.

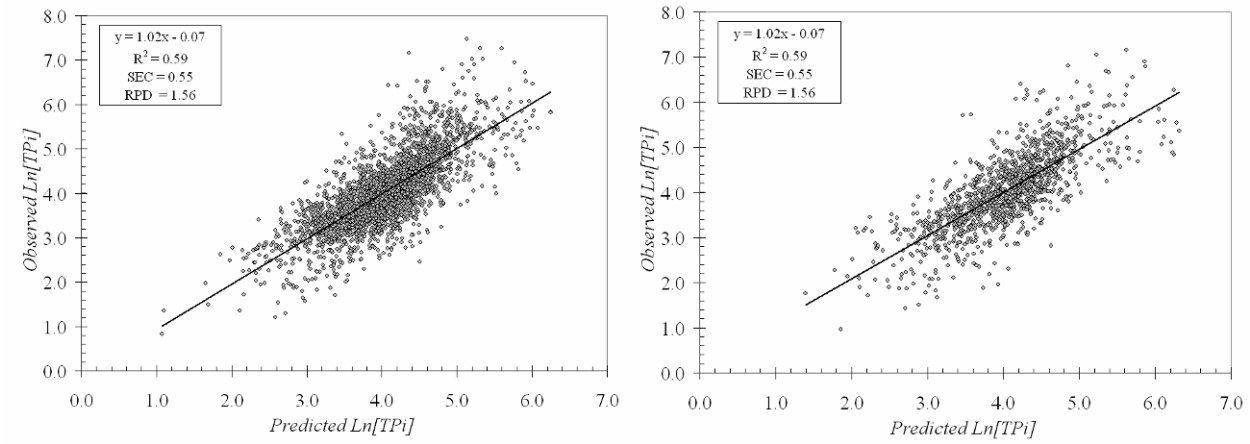

Fig. 21. Scatter plot of spectrally predicted vs. observed for total inorganic phosphorus (HCl-extractable) observations; A) calibration and B) validation.

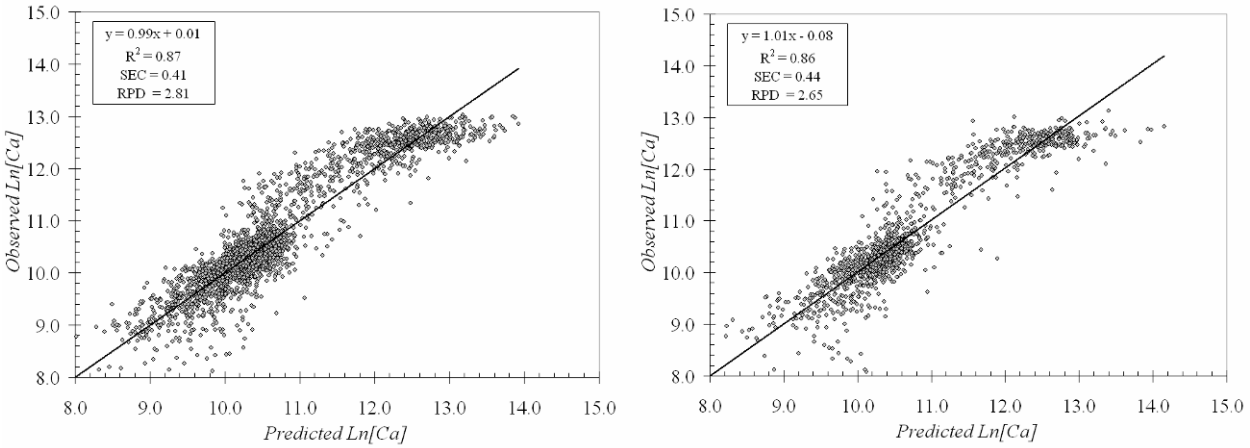

Fig. 22. Scatter plot of spectrally predicted vs. observed for total calcium observations; A) calibration and B) validation.

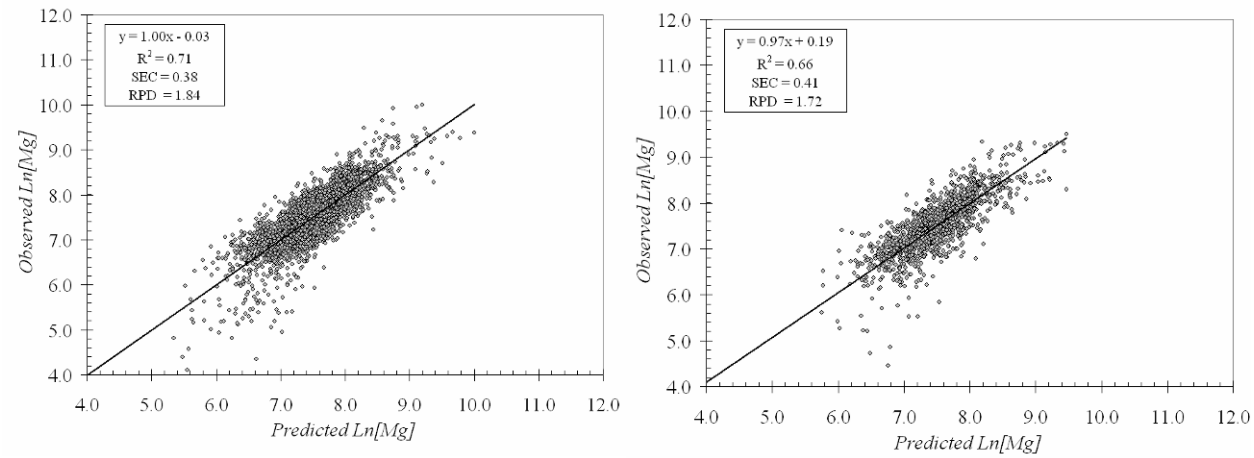

Fig. 23. Scatter plot of spectrally predicted vs. observed for total magnesium observations; A) calibration and B) validation.

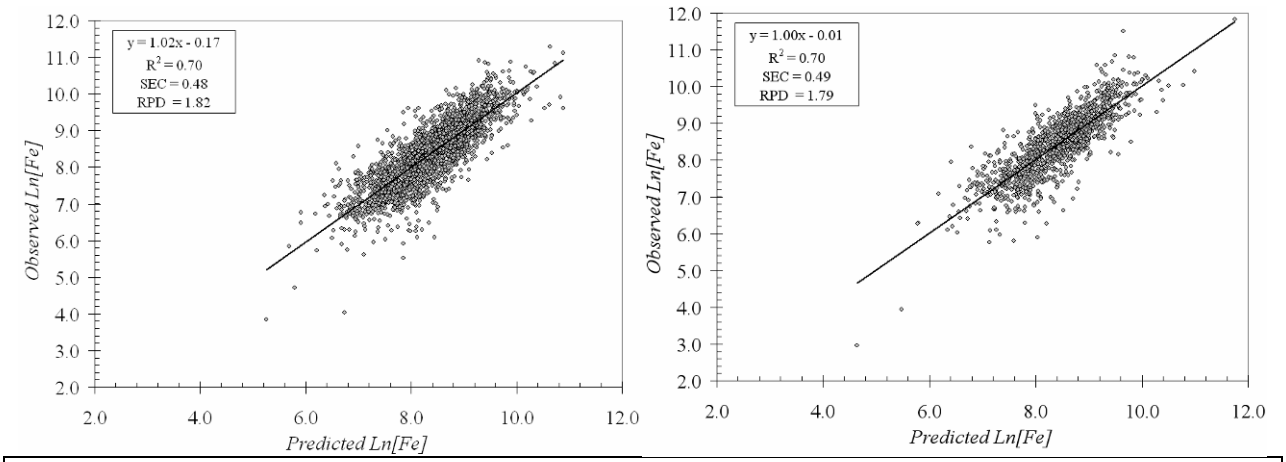

Fig. 24. Scatter plot of spectrally predicted vs. observed for total iron observations; A) calibration and B) validation.

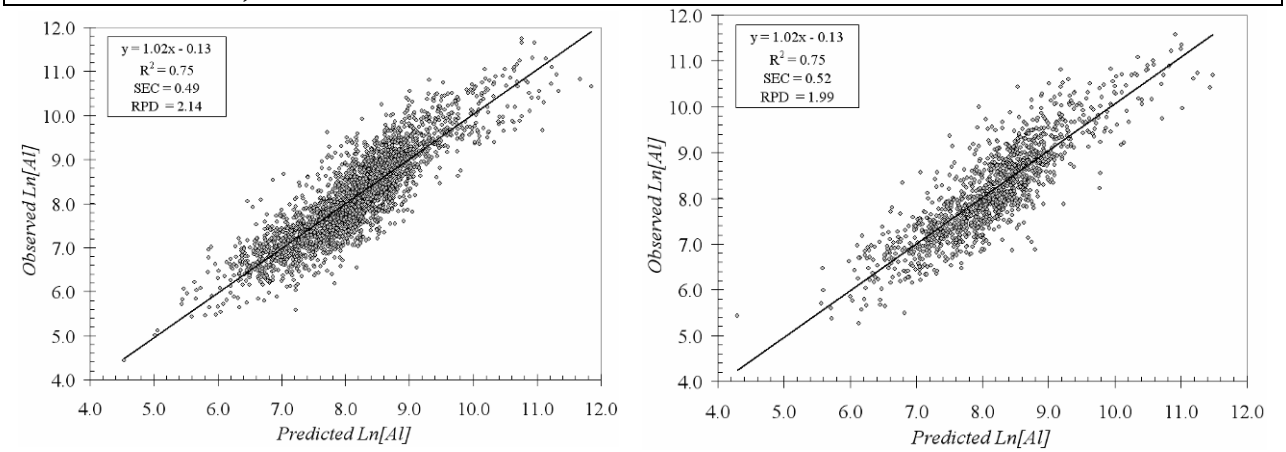

Fig. 25. Scatter plot of spectrally predicted vs. observed for total aluminum observations; A) calibration and B) validation.

One of the primary concerns with prediction algorithms developed from high dimension data sets (e.g., spectral data sets) is that they will be over-fit to the data; that is, models will fail to realize the same level of accuracy for a new set of data because they are predicated on circumstantial relationships in the training data set. This is the motivation for partitioning the SRL into training and verification subsets, and reporting spectral prediction potential using the results obtained for the verification data subset only. Any over-fitting is assumed to be removed when predictions are applied to a new data set. To determine the extent of overfitting in this data set and using these modeling techniques, we compare the RPD (which scales the error rate to the parameter standard deviation) for calibration and validation. We expect a small drop-off in accuracy; where the drop-off is large, we infer that the modeling technique is excessively sensitive to non-systemmatic variability in the spectral data. From Fig. 26, we infer that there is no evidence of serious overfitting in these models (reported for the SGBT method only). In all cases except for bulk density (where model fit improves in validation because of several outlier samples in the calibration set) model accuracy is reduced by less than 7% (max. = Total Al).

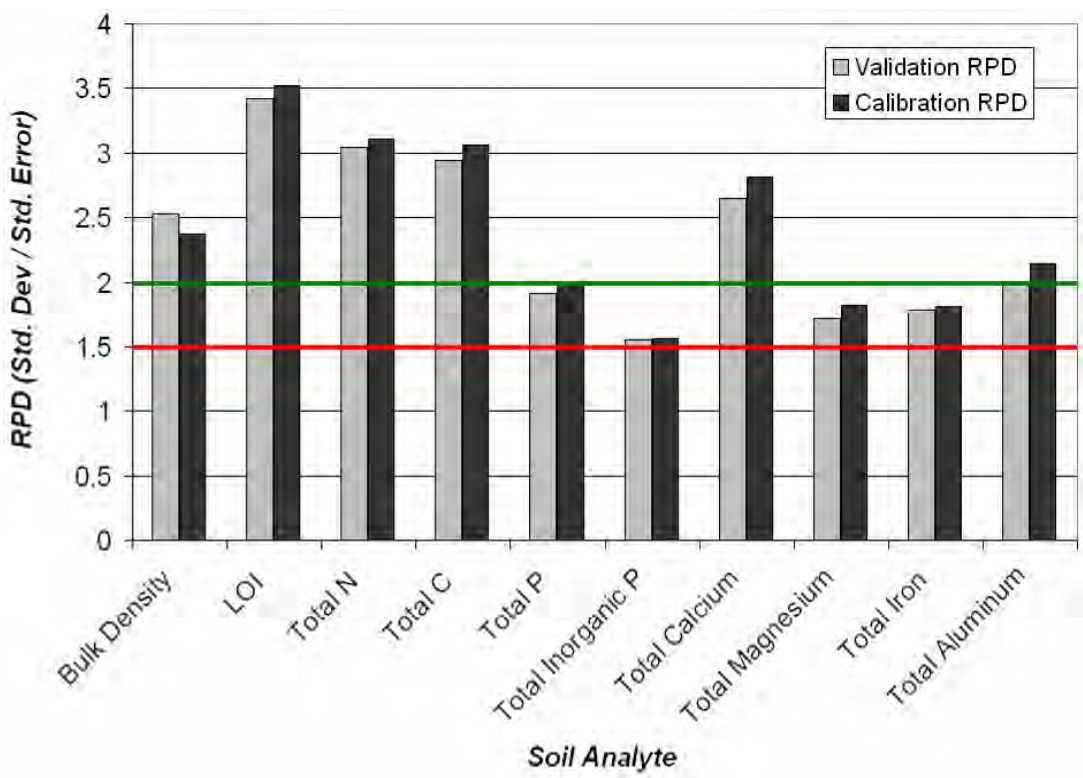

Fig. 26. Comparison of model accuracy between training data (calibration) and verification (holdout validation). RPD summarizes model error rates in a manner that is comparable across parameters.

Several features of spectral predictability of soil properties can be observed in Figs. 16 – 25. First, spectral models do relatively well at predicting all of the measured properties; analytical accuracy compares favorably with expected accuracy levels expected for conventional laboratory measurements (Fig. 27); typically we expect laboratory analyses (as measured using the relative % of standard deviation [RSD]) to be within +/- 15% of the true value 95% of the time. The median value for all analytes is less than 7% RSD; 75% quartiles are less than 15% in all cases except bulk density. Note that the error rates for P, inorganic P, Ca, Mg, Fe and Al are for log transformed data.

Combining this RSD information with the high levels of RPD (Fig. 26) compared with typical literature values (e.g., Dunn et al. 2002, Chang et al. 2001) suggest that this application of NIR spectroscopy offers significant and immediate application in service of improved restoration assessment. In as much as soil can provide a useful performance measure for ecosystem recovery, NIR spectroscopy can provide a viable alternative to conventional chemical analyses, and a significant opportunity to increase the spatial and temporal density of sampling.

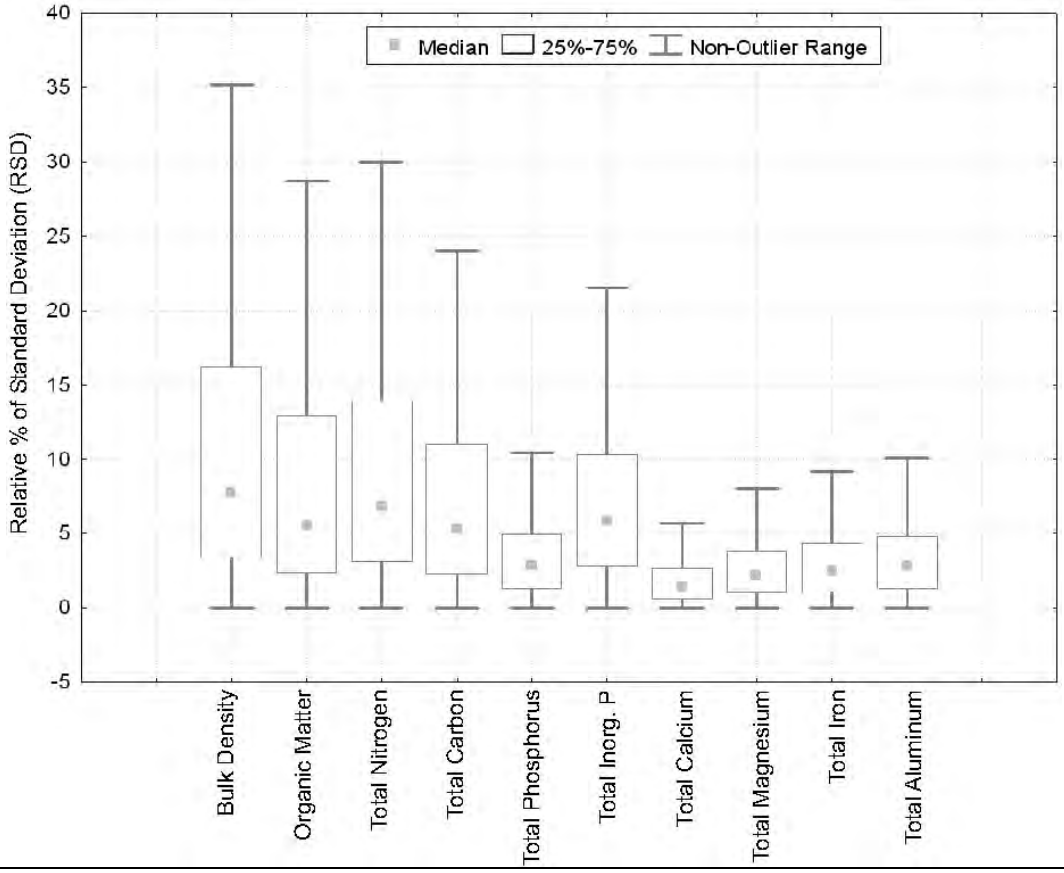

Fig. 27. Summary of relative % of standard deviation (RSD) values across analytes.

3.3.2.2 Soil Nutrient Ratios

The ESM dataset permits computation and prediction of key ecological metrics of soil quality based on the relative abundance (on a molar basis) of nutrients. Here we present results of prediction of C:N (a metric of soil carbon lability) and N:P (a metric of potential nutrient limitation and eutrophication sensitivity) ratios. For both, we present the overall distribution of the ratios across the entire population (Fig. 28 for C:N and Fig. 29 for N:P), and then a plot of the spectral prediction applied to validation data via which we assess whether direct prediction (that is, vs. prediction of C, N and P separately followed by computation of ratios) is a worthwhile endeavor.

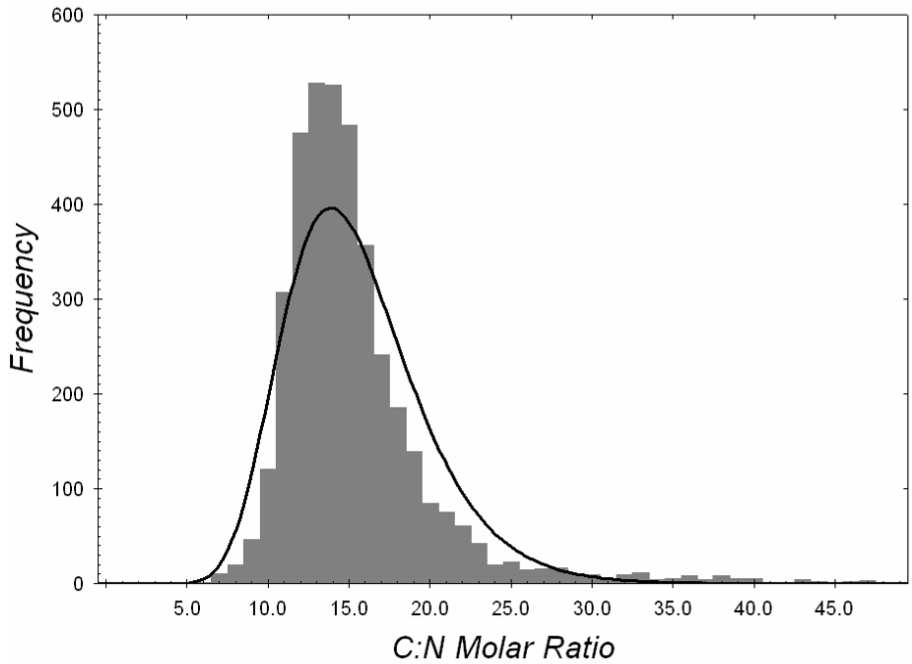

Fig. 28 – Distribution of C:N ratios across the Greater Everglades.

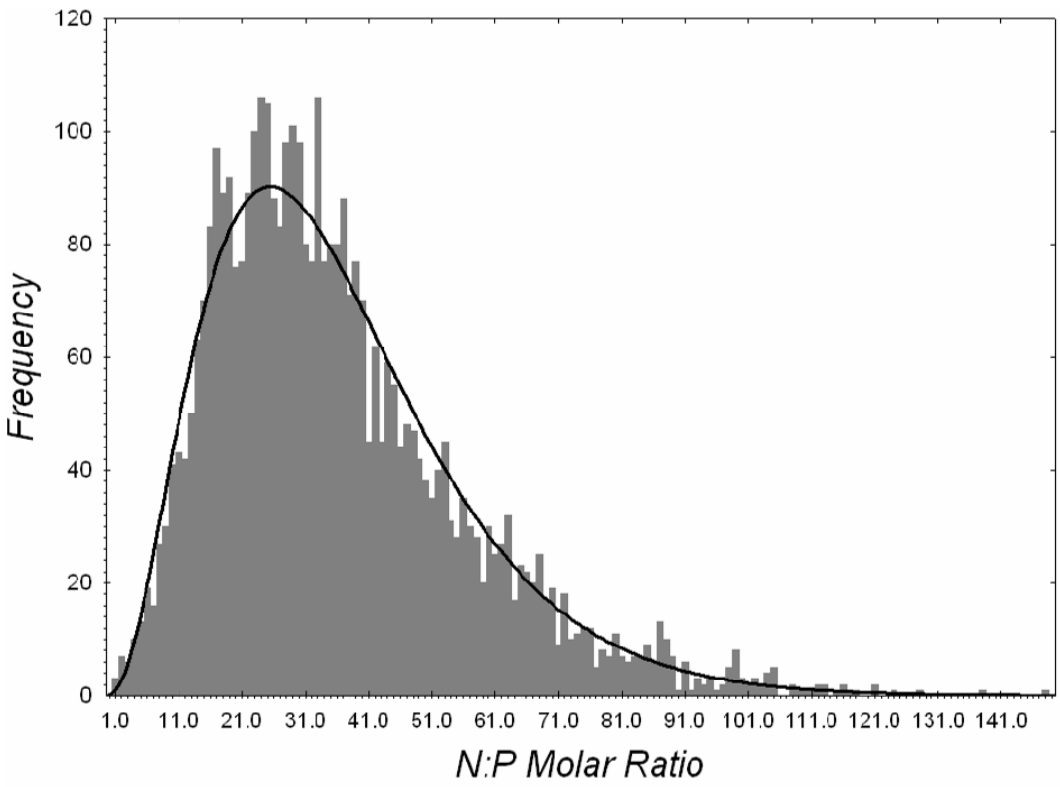

Fig. 29 – Distribution of N:P ratios across the Greater Everglades.

Spectral prediction of C:N (Fig. 30) shows an extremely strong association, with an RPD value of 2.95. As such, we conclude that this performance metric can be robustly be predicted using spectral methods. Because C:N is informative about the potential lability of soil carbon (see soil quality section below), this is extremely encouraging for ecosystem-level monitoring of C processes. We observe a weaker but still useful prediction of N:P (Fig. 31); the RPD value of 1.51 suggests that spectral prediction may be most useful for mapping broad scale patterns, and not for local high-accuracy applications. However, regional mapping of N:P and inference of global trends through time and space are informative, and can be simplified using spectral prediction.

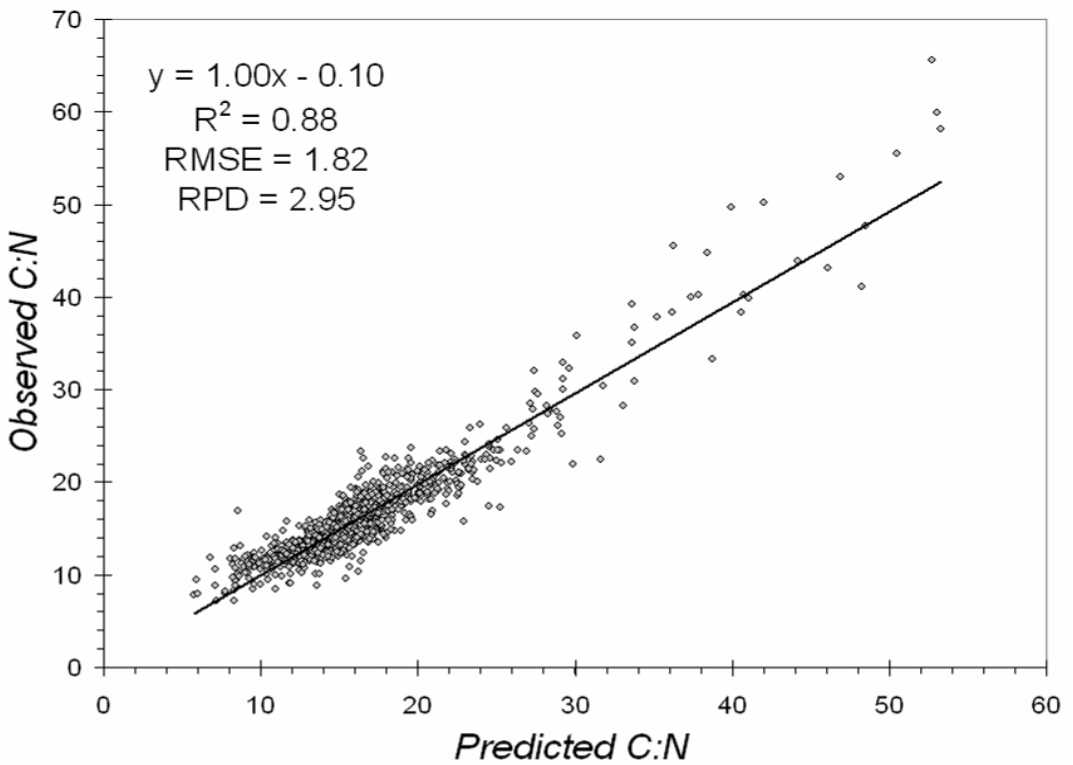

Fig. 30 – Validation data spectral prediction efficiency for C:N.

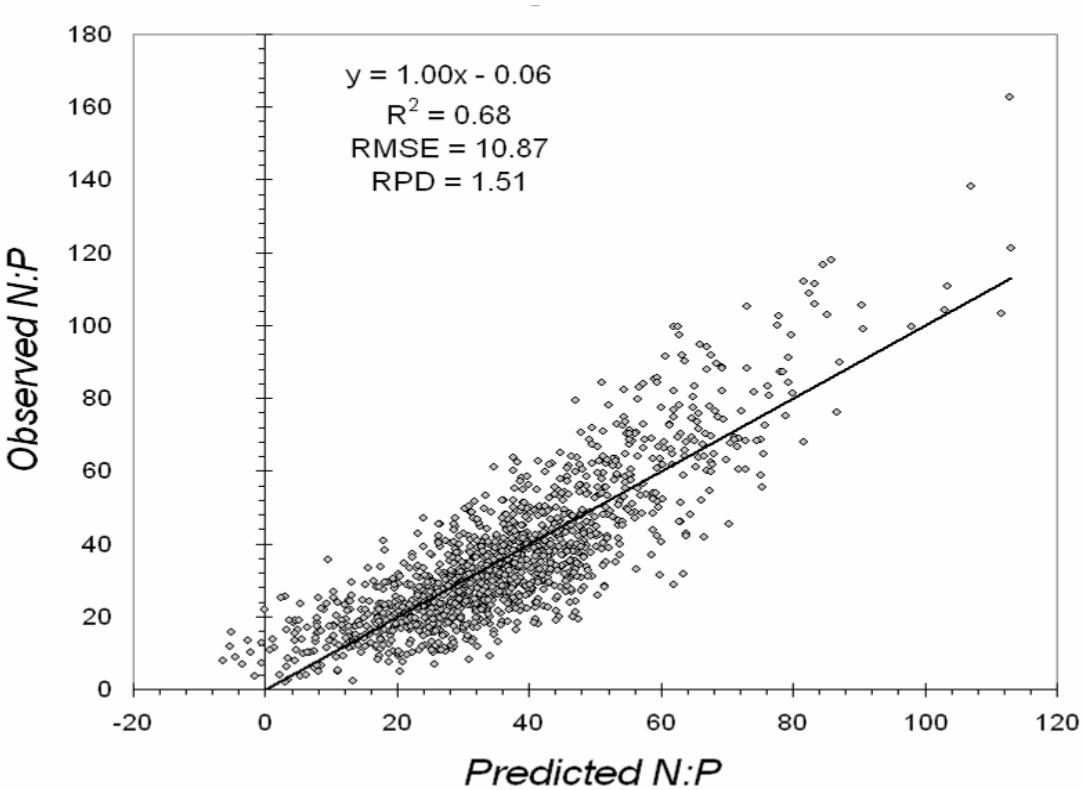

Fig. 31 – Validation data spectral prediction efficiency for N:P.

3.3.3 Spectral Information Content

One of the primary critiques of soft modeling (i.e., statistical prediction) as opposed to direct mechanistic inference (i.e., peak matching) is that elucidation of how prediction occurs is left undefined. In many applications of analytical spectroscopy, mechanistic inference can be achieved by comparing observed spectrograph peaks and absorbance features to know specimens; simple mixtures can be evaluated in this way as well using simple end-member unmixing models. However, when the media from which reflectance data are being obtained is highly heterogeneous, as is the case with soils, plant tissues, animal tissues, and many other potential environmental samples, the ability to match particular peaks with particular constituents is intractably complex. Add to this complexity the fact that reflectance information in the NIR region is not fundamental frequencies, but overtones and it becomes clear that soft-modeling is the only practical manner in which to extract information from the spectra. The complexity of the associations between spectra and the various soil analytes is presented in Fig. 32, which shows the relative importance of each waveband in the chemometrics developed using the SGBT method. This illustrates the large number of wavebands that contribute information to the statistical prediction, and also the degree to which there is significant variability in waveband importance across different analytes. Notably, the pairwise correlation between particular bands and the numerous target variables already examined are generally low. Indeed, many regions of the spectrum that appear to have strong predictive potential based on their importance to the SBGT prediction routine have extremely weak marginal correlation with the target variables. What is critical to remember is that the conditional correlations (that is, the association when controlling for other predictors) are what lead to improved prediction potential, and that these are not immediately evident from pairwise correlation. As such, the ability to make spectral predictions is embedded in the entire spectrum, not a particular waveband, because the predictive capacity of a particular spectral region is conditioned on spectral responses in other regions.

Relative importance graphs of this type are also useful for underscoring the utility of wide range spectrometers; cheaper instruments often sacrifice spectral range of observation, and Fig. 32 illustrates why that may reduce accuracy.

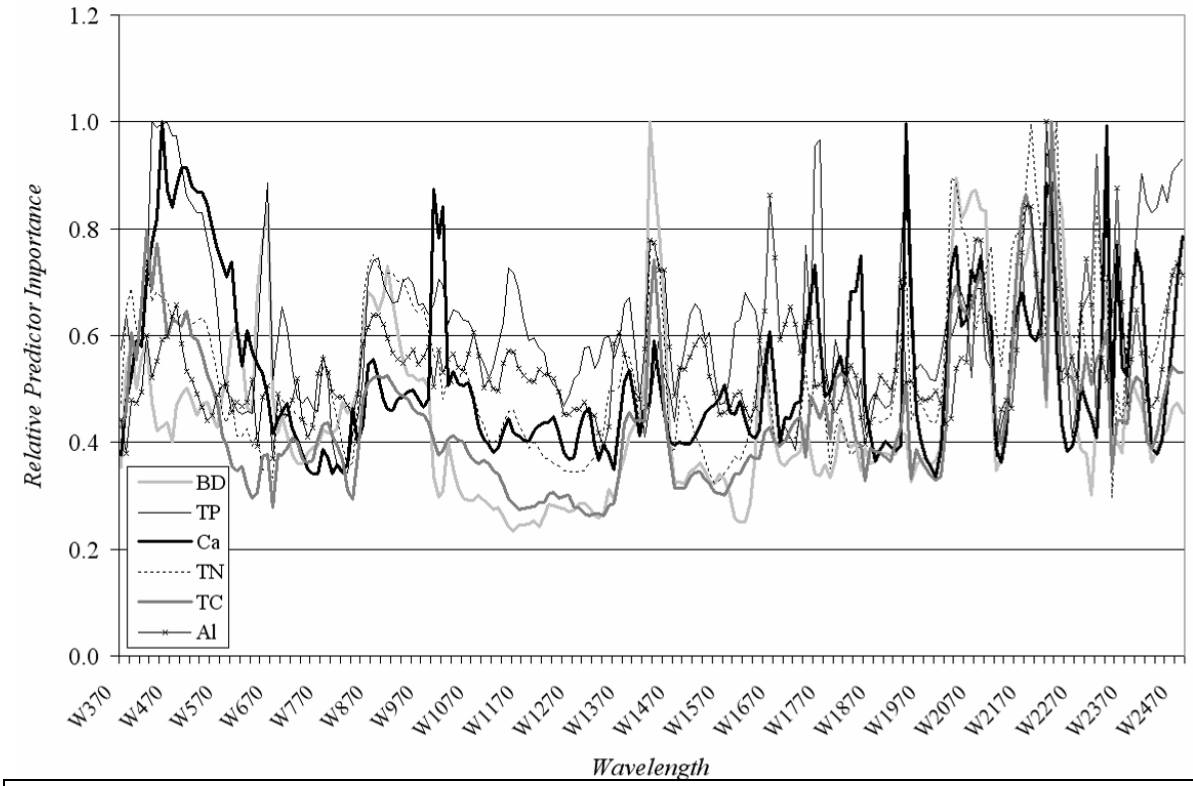

Fig. 32. Relative importance of wavebands for prediction of various soil properties.

3.3.4 Analysis of Model Residuals

Model predictions suggest that the spectral method may be useful and effective in Everglades monitoring and process assessment. However, there remains in all of the models some unexplained variance in the target variable that may or may not be due to errors that can be controlled. Since the objective of spectral modeling is not necessarily to use only spectra, but any ancillary information that might be available, we examine here covariance between the error and other attributes of each sample. Table 6 shows covariance between the model errors (prediction residuals) and 1) observed soil properties, 2) other prediction errors, and 3) distance from canals. We observe significant correlation between the observed errors and the original parameter values (e.g., correlation between bulk density model residuals and bulk density observations); this suggests some systematic bias in the predictions. Since the correlation is, in each case, positive, we infer that the spectral model generally truncates the range and spread of the observed data. This can be adjusted in the predictions if the degree of truncation is uniform across the observed range, which given the evidence for homoscedasticity in Figs. 16 – 25, we can safely assume.

We observe some modest correlation between the error rates (columns in Table 6) and other parameters as well. In all, however, we conclude that these are latent correlations due to, for example, correlation between TP and TP_i (Table 4). Finally, we observe some correlation between the observed errors for one parameter and errors observed in other parameters. This error covariance, particularly between parameters whose values are strongly correlated suggests that the spectra are keying in on similar properties for each prediction, resulting in similar errors. One critique of the NIR method for soils analysis is that it principally sees organic matter, and predictions fall from covariance between OM and other properties. The positive correlation of LOI errors with other parameter errors illustrates that this critique holds to a moderate extent. However, we also see covariance between TP error and metal concentration errors, suggesting that the spectral features mistakenly used for TP prediction are the same as those for metals, which are, notably, different from the features used for OM prediction.

Table 6 - Residual error correlation with observational covariates.

		Residual Error									
		Bulk Density	Organic Matter	Total N	Total C	Total P	Total Inorg. P	Total Ca	Total Mg	Total Fe	Total Al
Observational Covariates	Bulk Density	0.38			-0.21				-0.21		
	Organic Matter		0.31	0.26	0.29						
	Total N		0.28	0.35	0.29						
	Total C		0.31	0.29	0.35						
	Total P	-0.21				0.50	0.44	0.21			
	Total Inorg. P	-0.22				0.52	0.65	0.26			
	Total Ca							0.36	0.20		
	Total Mg	-0.26			0.23	0.20		0.26	0.55		
	Total Fe									0.57	0.30
	Total Al									0.28	0.50
	Canal Distance										
	Fiber Content										
	BD Error		-0.25	-0.23	-0.28	-0.20		-0.20	-0.25		
	LOI Error			0.74	0.87	0.25		0.27	0.34		
	TN Error				0.77	0.27		0.22	0.25		
	TC Error					0.22		0.39	0.37		
	TP Error						0.80	0.43	0.36	0.27	0.35
	TPi Error							0.41	0.28	0.21	0.24
	TCa Error								0.49	0.25	
	TMg Error										0.25
	TFe Error										0.50
	TAl Error										

We also explore prediction errors as a function of categorical factors. For example, are errors similar across hydrologic partitions, or community types, or depth?

We observe evidence for significant error differences among hydrologic partitions. In general, the magnitude of that covariance is small compared with the total error, suggesting that the operational significance of the observed differences is small. However, many of the mean differences are significantly different than 0, revealing systematic spatial bias in the predictions. This bias is quantified in Figs. 33 and 34, and can be visualized in interpolated error maps.

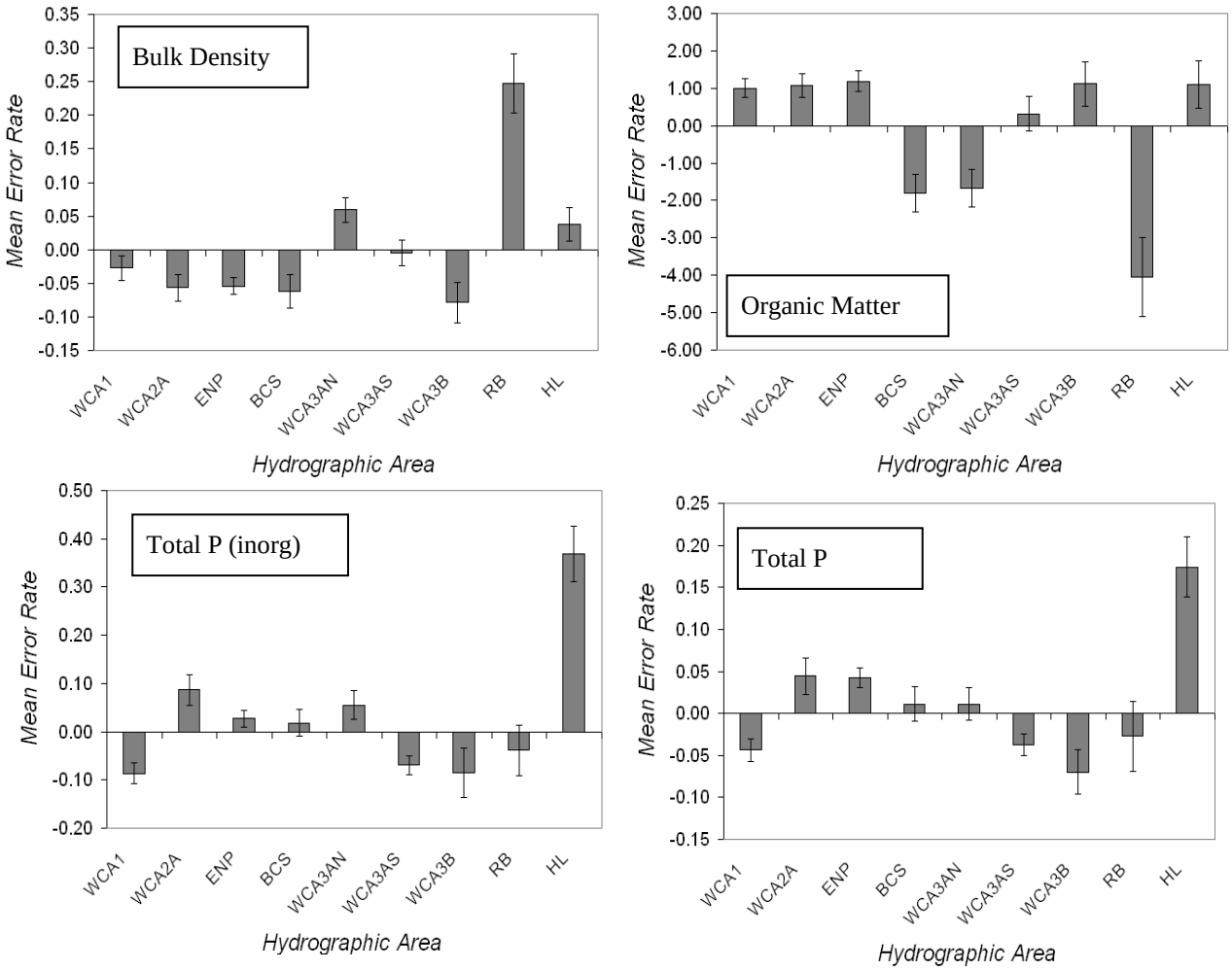

Fig. 33 – Chemometric prediction errors by hydrographic area.

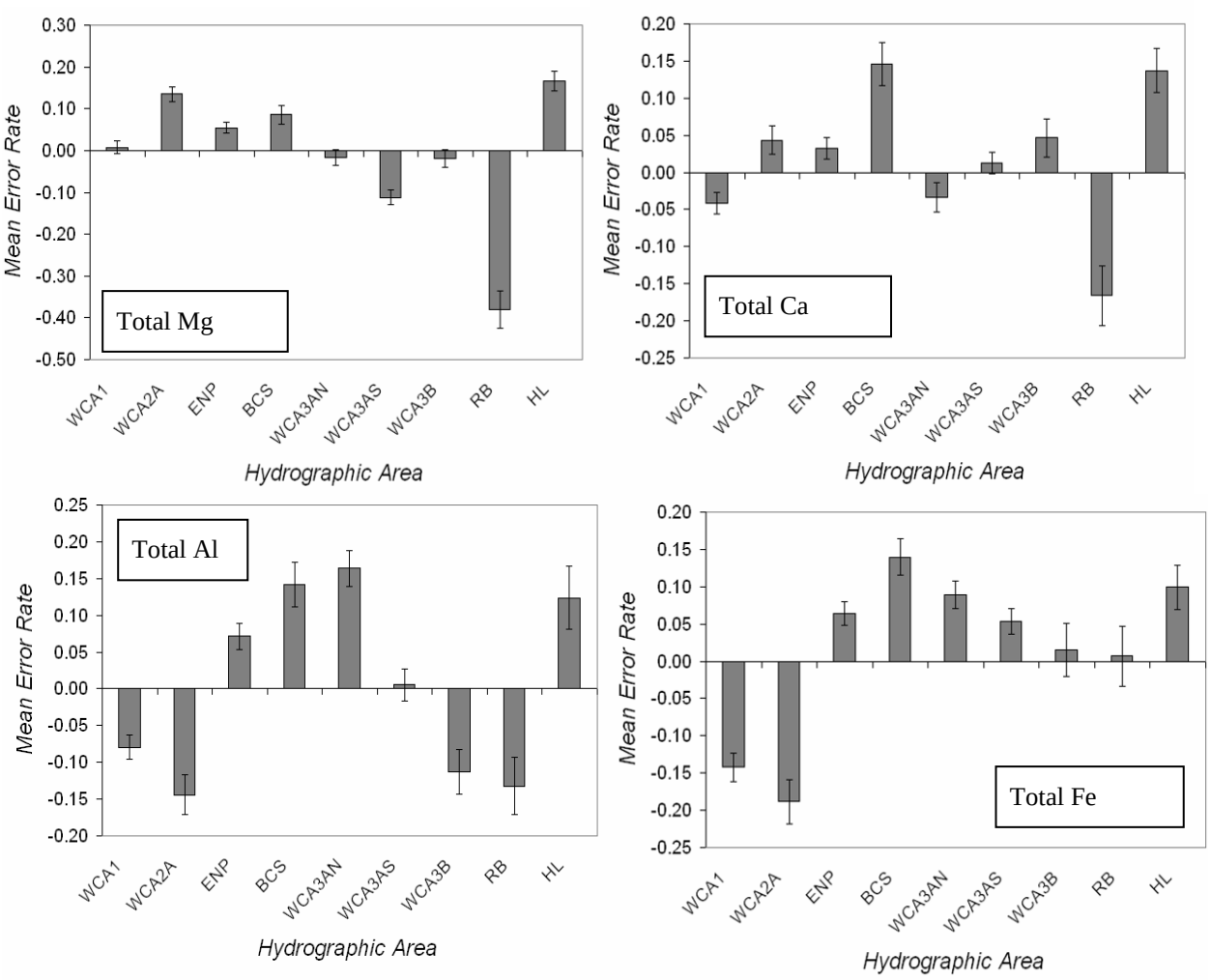

Fig. 34 – Chemometric prediction errors for cations by hydrographic area.

Most notably, we observe consistent overestimation for all parameters in Holeyland (HL) and nearly universal underestimation in WCA1. Moreover, nearly all hydrographic areas were had error rates significantly different from 0, illustrating that the global calibration is geographically confounded. We hasten to point out that while the error trends are significant, they are of small magnitude. One implication is that local calibrations (i.e., a spectral library consisting only of samples from WCA1 or ENP) might be more stable. This is a research avenue warranting further consideration.

Comparison of error rates across community types indicates no significant variation from a mean error. Large overestimates of total C were observed for wet prairie habitats, but these errors were not significant at $p < 0.05$.

Error rates by depth were significant principally for the floc and upper soil profiles; prediction errors in the deepest profile were generally smaller (Fig. 35). The most important errors were a general underestimate of bulk density in the flow layer, and an overestimate of the total Al content in the upper 10 cm of soil. As before, despite significant error differences at different depths, these differences were small, and likely to be of little operational significance.

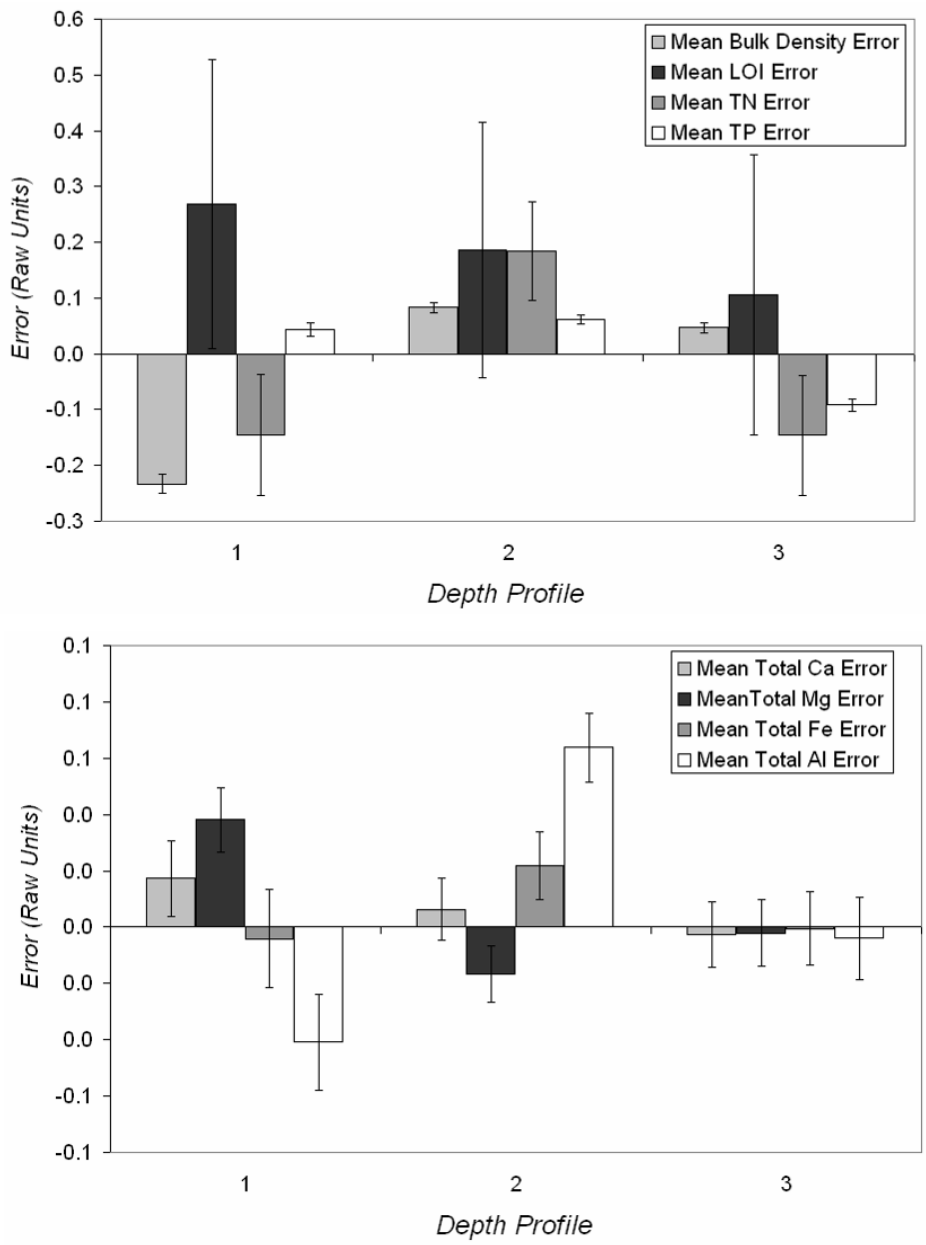

Fig. 35 – Error rates by soil depth.

3.3.5 Maps of Prediction Errors

One of the most important avenues of further inquiry from Year 1 chemometric development activities is a deeper understanding of prediction errors. We are currently developing a detailed understanding of the role of confounders on prediction errors; a summary of that analysis will be included in the final report. For this report, we have examined the prediction errors in space to better understand any systematic spatial pattern to the observed disagreement between observations and associated spectral predictions.

The first analyte for which we map the prediction errors is total P (Fig. 36). This map reveals that there are areas where the spectral model appears to systematically underestimate the observed TP values, and similarly, areas where it appears to overestimate observed values. In this cases, the error rates are quite small, and the structure of the spatial pattern is practically non-existent (see Table 4 – relative structure and cross-validation r^2).

In a similar vein, the spatial pattern of error rates for organic matter content prediction (Fig. 37) and total calcium prediction (Fig. 38) identify areas where there appears to be systematic bias in the predictions. Notably, northern WCA3A is under predicted for all three parameters, and the bulk of southern Shark River Slough is over predicted. This suggests avenues of additional analytical work that are beyond the scope of this project, but might include entry of easy-to-measure covariates (like water depth based on the EDEN data product) into the chemometric model as a means of decreasing sensitivity to these regional biases.

We present the error spatial patterns for the other parameters in Figs. 39 and 40. As with TP, TCa and LOI, we note that there is significant evidence for strong spatially structured errors. This result has never been demonstrated in the literature before, and is the source of some important inference: specifically, we interpret the significant geostatistical error patterns to the potential utility of local calibration libraries versus global libraries. Given the propensity in the NIRS literature for global models, this finding is crucially important, suggesting that errors in global models are expected to be autocovariant (due to local scale differences in the way that soils express particular properties). We note that error maps for TC, TMg and TAl are the most strongly patterned, while maps for bulk density and total N are weakly structured.

Quantification of the degree of spatial structure for both the uppermost soil layer (Profile 2) and the floc layer (Profile 1 – maps not shown) is provided in Table 7. Of particular note is the relative structure parameter which quantifies to degree to which the data are autocovariant.

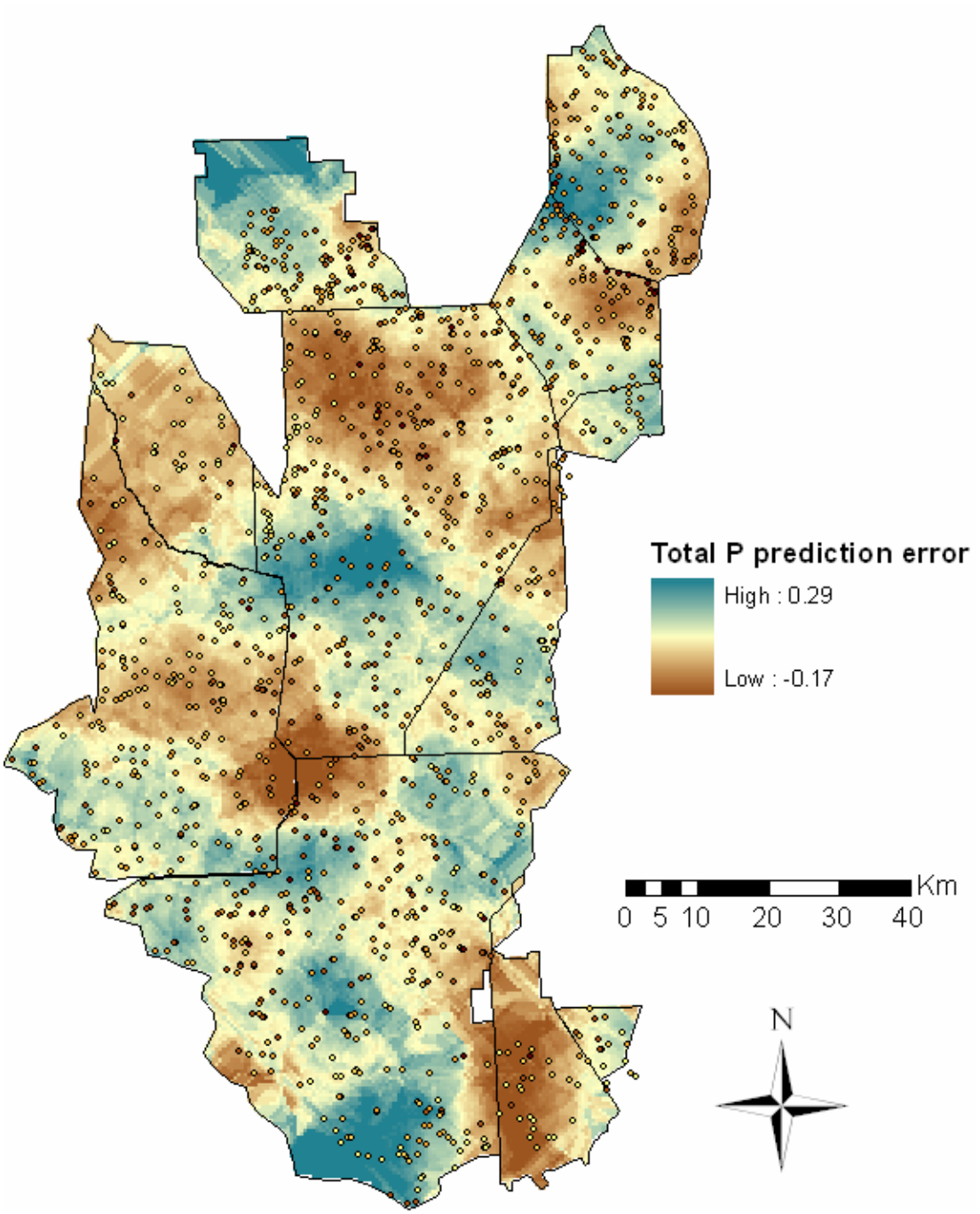

Fig. 36 – Interpolated error rates for TP chemometric model. Note the small error magnitude; error units are Ln[mg/kg].

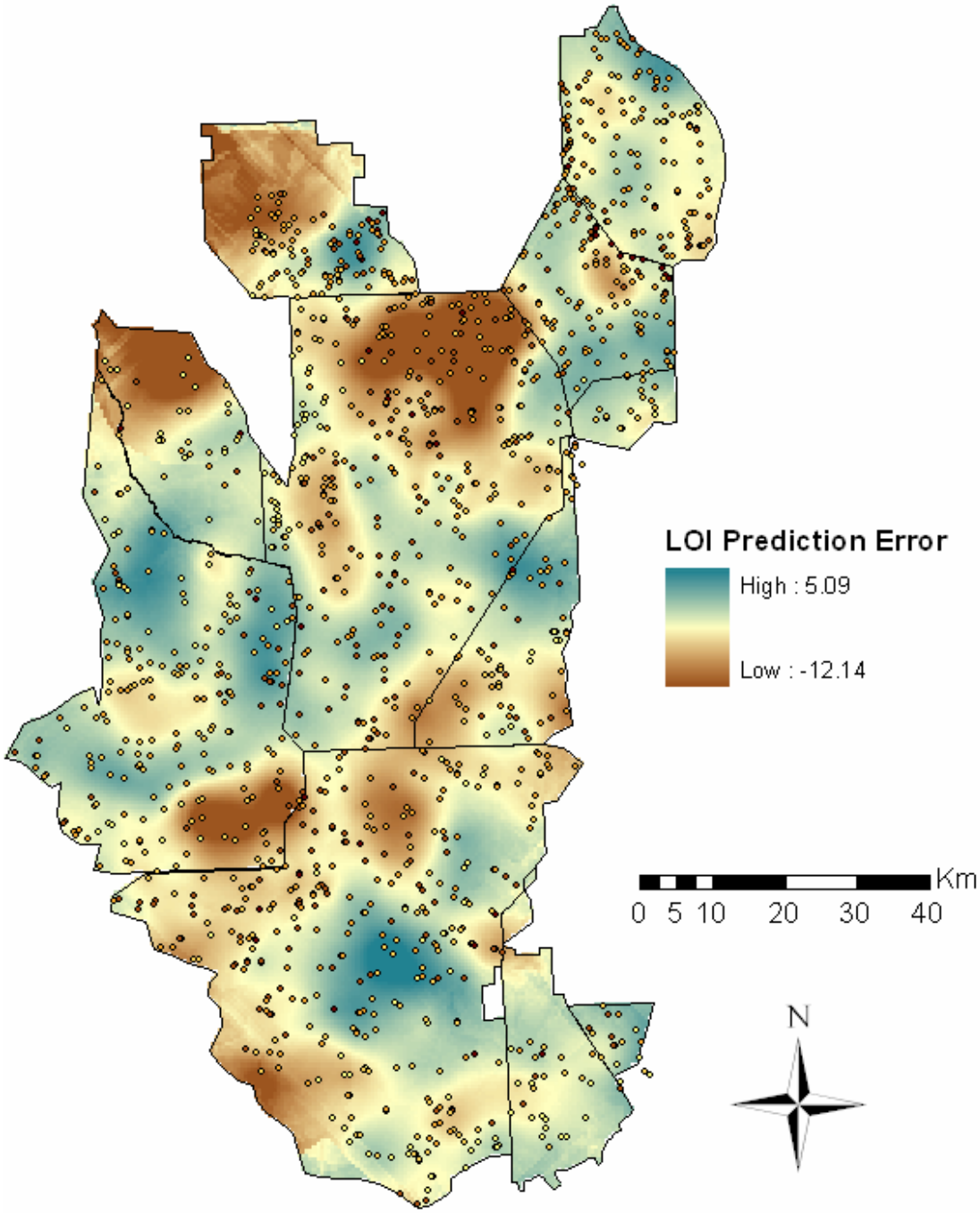

Fig. 37 – Interpolated error rates for LOI chemometric model. Error units are % loss on ignition.

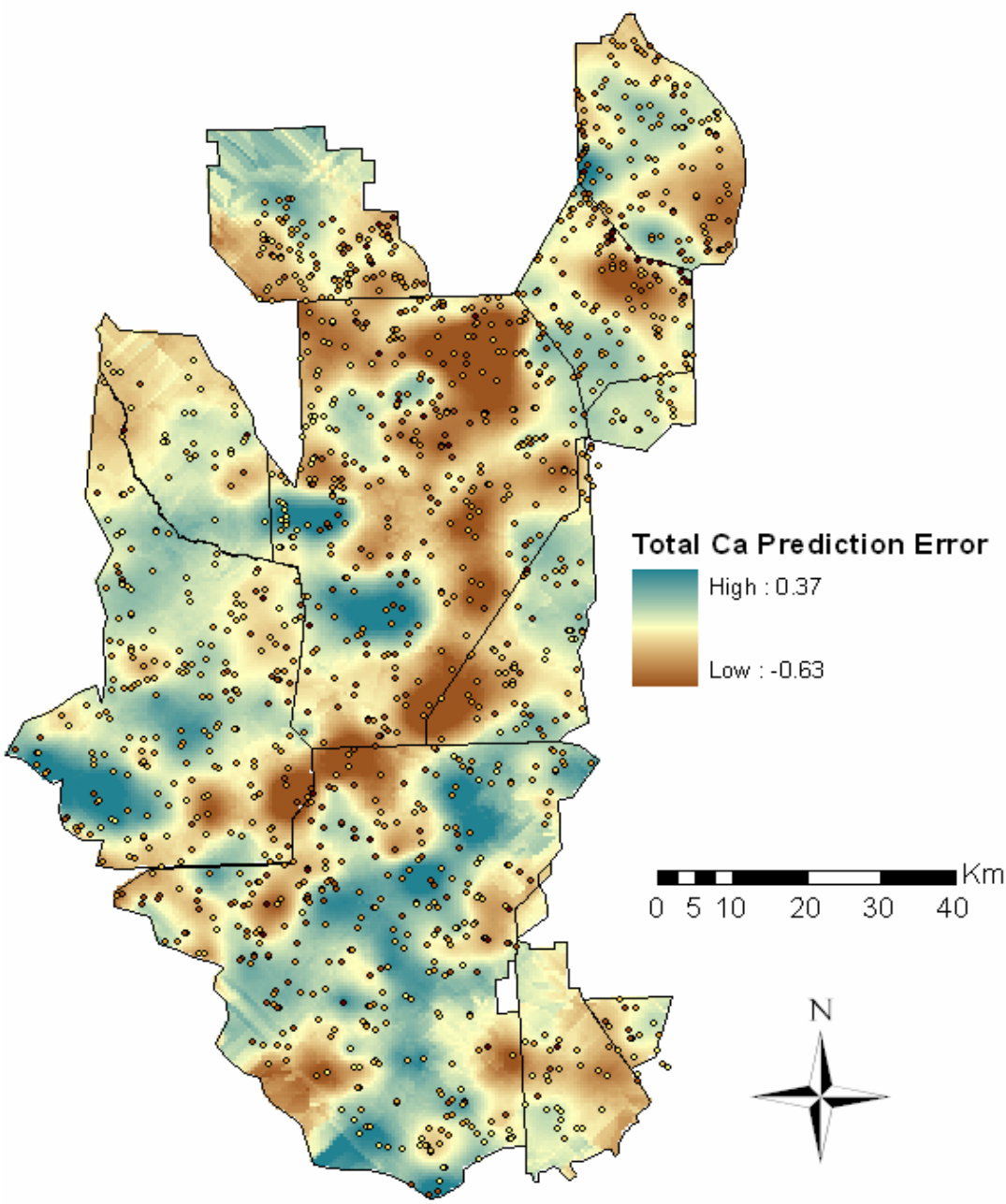

Fig. 38 – Interpolated error rates for Total Ca chemometric model. Error units are Ln[mg/kg].

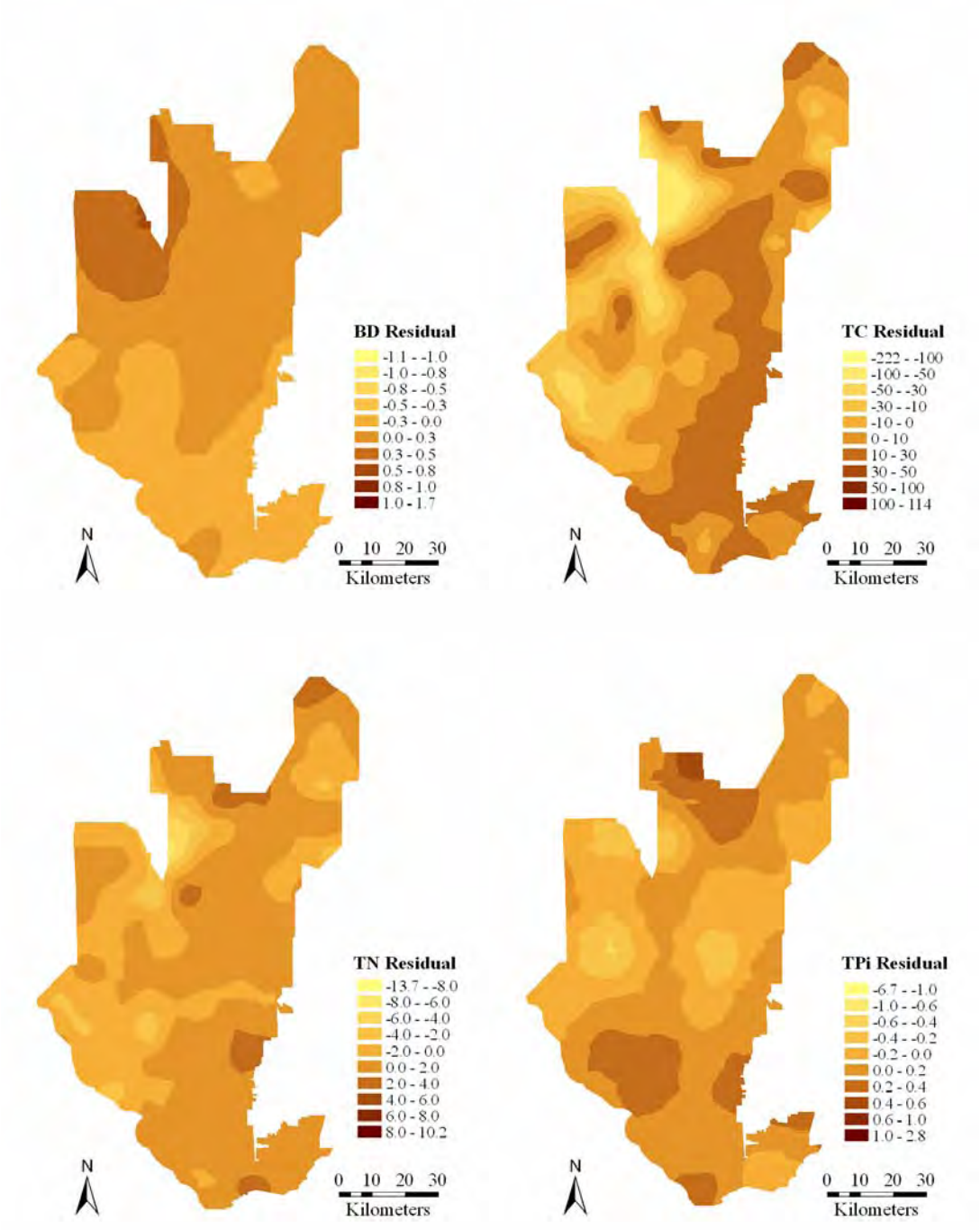

Fig. 39 – Maps of chemometric prediction residuals for ESM soil properties (bulk density – BD, total C – TC, total N – TN and total inorganic P – TP_i).

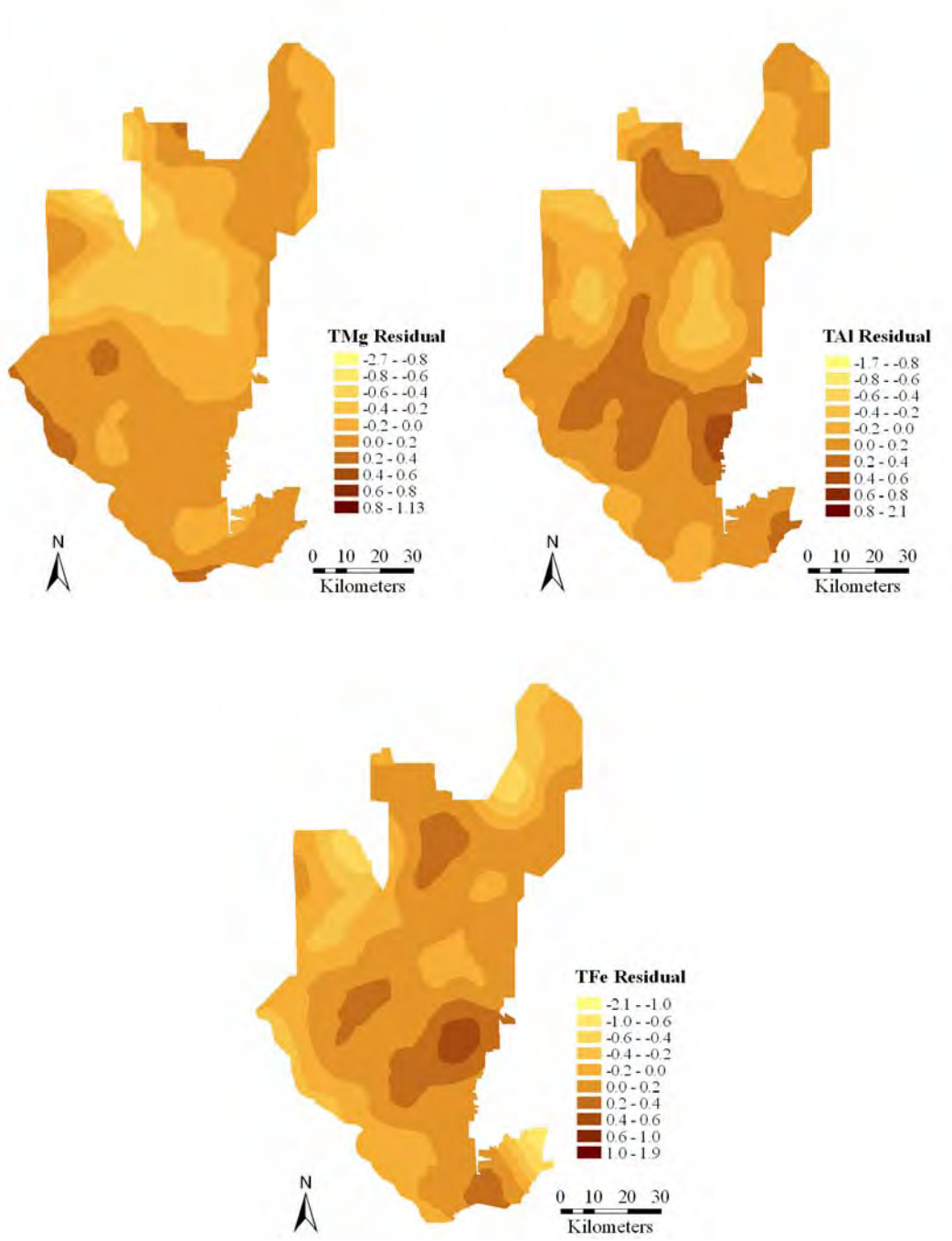

Fig. 40 – Maps of chemometric prediction residuals for ESM soil properties (total Mg – TMg, total Al – TAl and total Fe – TFe_i).

From Table 7 we infer that all of the properties have substantial spatial structure; none exhibits the expected "pure nugget" condition wherein errors are spatially independent. Moreover, there is relatively little difference in the relative structure between the parameters

(from 19% for TP to 36% for total Mg), suggesting that all could possibly benefit for local calibration libraries. We also infer that the range of error autocorrelation is large (23,000 – 59,000 m) suggesting that what principally structures the error is not fine scale variability. Error geostatistics for the floc layer are slightly different, with both shorter ranges (indicative of more local level process variability) and higher relative structure (indicative of more spatially structured errors). As with soil, there are no parameters for which errors are spatially independent. Interestingly, the parameter with the lowest relative structure in the upper soil (TP_i) has the highest relative structure in the floc.

Table 7 – Summary of geostatistical parameters for chemometric prediction errors. Nugget (short range semi-variance), sill (maximum semi-variance) and range (distance over which semi-variance increases to sill) are reported along with the relative structure (defined as the ratio of the partial sill to the total semi-variance, which equals the partial sill plus nugget semi-variance). High relative structure indicates parameters with strong spatial pattern in their prediction errors.

		Profile 2 (0 – 10 cm soil)			
Parameter	Units	Nugget Semi-Variance	Partial Sill Semi-Variance	Semi-Variance Range (m)	Relative Structure
Bulk Density	g/cm3	0.09	0.03	5.93E+04	23.8%
Organic Matter	%	62.13	30.32	2.74E+04	32.8%
Total N	g/kg	9.69	3.08	1.63E+04	24.1%
Total C	g/kg	1535.30	673.89	2.49E+04	30.5%
Total P	Ln[mg/kg]	0.10	0.03	5.93E+04	22.3%
Total P (inorg)	Ln[mg/kg]	0.27	0.06	5.36E+04	19.0%
Total Ca	Ln[mg/kg]	0.13	0.05	2.57E+04	29.3%
Total Mg	Ln[mg/kg]	0.11	0.06	5.93E+04	36.2%
Total Fe	Ln[mg/kg]	0.17	0.06	4.44E+04	26.2%
Total Al	Ln[mg/kg]	0.19	0.05	2.30E+04	20.1%

		Profile 1 (Floc)			
Parameter	Units	Nugget Semi-Variance	Partial Sill Semi-Variance	Semi-Variance Range (m)	Relative Structure
Bulk Density	g/cm3	0.17	0.11	6.03E+04	39.4%
Organic Matter	%	45.48	23.61	4.38E+04	34.2%
Total N	g/kg	8.67	2.64	3.52E+04	23.3%
Total C	g/kg	1067.20	611.35	5.11E+04	36.4%
Total P	Ln[mg/kg]	0.09	0.03	1.08E+04	24.7%
Total P (inorg)	Ln[mg/kg]	0.15	0.12	6.01E+03	43.4%
Total Ca	Ln[mg/kg]	0.13	0.04	9.74E+03	21.3%
Total Mg	Ln[mg/kg]	0.09	0.03	2.13E+04	24.9%
Total Fe	Ln[mg/kg]	0.21	0.08	1.56E+04	26.8%
Total Al	Ln[mg/kg]	0.21	0.05	1.90E+04	19.5%

3.4 Categorical Predictions

Chemometric modeling is clearly of significant value for this application, but several parameters have weaker than desired accuracy. Further, some aspects of the soil that might be useful if spectral predictions were developed are not continuous variables. As such, a second analytical effort was directed at the development of categorical models, where the target variable is either functionally defined into classes from continuous observations (e.g., soil TP > 500 mg/kg is a broadly useful indicator of enrichment) of is in fact categorical (e.g., soil vs. floc, community type, depth in the current data set).

We developed categorical prediction models using the same SGBT method as previously described, but in a mode that permits categorical targets (in fact, the tree-based algorithms that underlie the SGBT method are more intuitively suited for categorical targets). Other methods have been explored previously (e.g. canonical discriminant analysis, logistic regression for binary or ordinal variables), but the combination of high predictor dimensionality, severe co-linearity and moderately weak pairwise associations between individual bands and the target variables lead to the condition where these methods are significantly less effective at inference. No comparative method results are shown here, but typically SGBT outperformed the other methods by 10-15% in validation accuracy.

The first variable that we developed a categorical prediction of is the depth of each soil sample. While this is unlikely to be of immediate value (ascertaining the depth of a sample is a field observation not a laboratory one), it may prove to be of significant utility when assessing the degree to which soils at the surface deviate from soils at depth. This may be useful for crude community history mapping, surface oxidation studies, or variability in historic peat depositional dynamics. Table 9 summarizes the results. Overall accuracy levels of 72% in validation are of relatively limited utility, but given the variability of soil profiles across the entire system, this particular modeling question is worthy of additional attention. In particular, it will be useful to determine if source region (ENP, WCA, Big Cypress) affects the predictive accuracy, and where confounded prediction is prevalent (i.e., in areas with significant historic enrichment).

3.4.1 Soil Functional Thresholds

One of the original objectives of this work was to explore categorical screening models where continuous prediction of soil analytes was insufficient. The results to date (both this

report and our previous report) suggest, in general, that continuous spectral prediction of soil analytes is sufficiently successful that categorical models may not be needed. However, despite the relative success of continuous prediction, there are still numerous opportunities to explore categorical modeling, both by designating functional thresholds in otherwise continuous observations (e.g., TP performance threshold of 500 mg/kg) or by examining the spectral separability of ecosystem properties that are naturally categorical (e.g., soil vs. floc, dominant vegetation). This section summarizes the success of spectral prediction of categories.

Table 8 provides a simple comparative summary of all binary spectral models developed for this work. For each model we report the functional threshold that delineates case from control (in the case of continuous variables only), the overall calibration and validation accuracy, the sensitivity and specificity, and finally the odds ratio for correct classification. Overall, we conclude that spectral screening models have strong potential application. In some sense, this is a method looking for an application; in particular, were there to be an effort to delineate ecological condition, perhaps based only on best professional judgment, then the target variable for spectral screening would be substantially more interesting. It is not our objective to provide this effort, so we focus on other classification problems to illustrate the method and how the efficacy of a predictive model might be evaluated.

Table 8. Summary of categorical model efficiency.

Model Target	Category Threshold†	Validation Prediction Accuracy	Validation Sensitivity‡	Validation Specificity‡	Validation Odds Ratio§
Total P	Case > 500 mg/kg	85.5%	87.9%	74.9%	21.7
Total Inorganic P	Case > 100 mg/kg	84.2%	86.6%	76.4%	21.0
Total Al	Case < 1500 mg/kg	80.4%	81.9%	75.6%	14.1
C:N ratio	Case < 13	83.2%	84.2%	80.0%	21.4
N:P ratio	Case < 40	81.3%	82.2%	75.9%	14.6
Soil vs.Floc	n/a	92.6%	92.1%	92.7%	148.1

† - Case (response = 1) definitions. No definition is needed for some categorical data (e.g., soil vs. floc).
‡ - Sensitivity = correct positives; Specificity = correct negatives
§ - Typically, OR values greater than 10 indicate effective diagnostic screening models.

3.4.2 Error Rate Analysis

As with continuous predictions, the incidence of errors as a function of known sample attributes (depth, geography, vegetation community) can be informative about how to use these models in practice. We examined error probabilities conditioned on both continuous and

categorical variables that might be considered covariates. In general, error rate differences were small; for example, no significant differences in error probabilities were observed for soil depth or community type. As with continuous chemometrics, there were significant differences among hydrographic regions, but as before these were of small magnitude.

3.4.3 Nominal Analysis

3.4.3.1 Soil Depth

Categorical prediction of soil depth is principally of value if it can aid in peat depositional studies. The existing library is coarsely sectioned, so the calibration is primarily suggestive of potential rather that a conclusive demonstration. The results (Table 9) are fairly compelling; overall, nearly 72% of samples are correctly classified from 4 soil depths. Error occurs mostly with proximate depths, suggesting that the model can readily discriminate deep from shallow samples. Moreover, the most confounded class is the deepest (profile 4) which had few samples and would not necessarily be expected to be much different from profile 3 samples. What is most encouraging about this first-order estimate of spectral prediction of soil horizons is that the soil depth prediction is not conditioned on geographic area. This global model is both

Table 9. Spectral prediction of soil depth (1 – floc, 2 – 0 to 10 cm, 3 – 10 to 20 cm, 4 – 20 to 30 cm)

Calibration		Predicted			
Observed		1	2	3	4
	1	513	65	2	17
	2	88	752	117	121
	3	14	93	616	101
	4	1	7	33	80
Overall Accuracy			**74.8%**		

Validation		Predicted			
Observed		1	2	3	4
	1	270	33	7	15
	2	40	367	81	28
	3	3	71	261	70
	4	1	5	14	25
Overall Accuracy			**71.5%**		

stable and, given the expected uncertainty, remarkably accurate. As such, it is expected that localized models (i.e., those from one hydrographic region) would perform better. Our study of finely sectioned cores (see below) is also suggestive of systematic spectral variance with depth. Where truncated soil profiles exist (e.g., due to peat subsidence or fire), spectra may be a rapid and robust tool for diagnosing and mapping. Recent evidence from the literature (Skjemstad et al. 2006) suggests that NIR spectra can readily discriminate charcoal C from peat C, allowing relative inexpensive and spatially explicit reconstructions of fire history in peat systems.

3.4.3.2 Community

Another application of categorical modeling that may have some immediate utility is the prediction of community based on spectra. The clearest application of such a model would be backcasting through cores; if reliable discrimination models could be developed between ridge and slough (as an example), soil profiles could be examined to determine whether a given site had changed between the two community types during the period represented by that soil profile. Table 10 summarizes the spectral separability of 4 major vegetation community assemblages; as shown, there is considerable support for spectral screening of ridge and wet-prairie (presumably due to the spectral influence of marl sediments), but poorer discrimination between ridge and

Table 10 – Summary of spectral prediction of community type. Odds ratios > 10.0 are typically considered useful for diagnostics testing.

Calibration	Predicted					
Observed	Wet	Ridge	Slough	Mangrove	Accuracy	Odds
Wet Prairie	395	40	16	57	61.4%	54.0
Ridge	26	949	151	14	70.8%	14.7
Slough	26	148	261	11	45.8%	5.3
Mangrove	2	1		64	83.1%	152.9
Overall Accuracy	64.0%					

Validation	Predicted					
Observed	Wet	Ridge	Slough	Mangrove	Accuracy	Odds
Wet Prairie	171	21	11	18	60.9%	61.5
Ridge	8	385	81	8	70.0%	12.7
Slough	11	71	110	10	40.7%	3.5
Mangrove		2		25	71.4%	76.6
Overall Accuracy	61.0%					

slough. A closer examination of ridge-slough soils reveals stronger discrimination when considered as a binary classification problem; the validation accuracy and odds ratio for this binary model are 79.5% and 12.6, respectively. This suggests moderate accuracy; the utility of this level of accuracy for ecological back-casting is beyond the scope of this work, but would be an important component of any follow-up activities.

3.4.3.3 Soil vs. Floc

Categorical prediction that may have more immediate ramifications is discrimination of soil from flocculent material. Currently there is no repeatable way to define the difference other than best professional judgment (which represents the "truth" condition for this model). Table 11 summarizes the results of delineating between floc and soil. We observe high overall validation accuracy levels (93%) with a classification odds ratio over nearly 150 (10 is typically used for identifying a useful diagnostic model). We consider this to be of substantial utility for objective mapping of the presence and perhaps status of floc development across the area.

Table 11. Summary of spectral prediction of soil vs. floc.

Calibration	Observed	
Predicted	Floc	Soil
Floc	498	29
Soil	113	1980
Overall Accuracy	94.6%	
Sensitivity	94.5%	
Specificity	94.6%	
Odds Ratio	300.9	
Validation	Observed	
Predicted	Floc	Soil
Floc	257	22
Soil	74	938
Overall Accuracy	92.6%	
Sensitivity	92.1%	
Specificity	92.7%	
Odds Ratio	148.1	

3.4.3.4 Floc Probability with Depth

Given the obvious success at discriminating between soil and floc, we expanded this segment of the analysis to test the hypothesis that floc material is not being incorporated in the

soil. Floc is spectrally distinct from soil (at least in the coarse sectioning within the ESM sample set); therefore, if floc is being integrated into the upper soil profile, there should be a residual spectral signal thereof. If floc is not being incorporated – by extension it is a transient mass with no durable accretion products – then the floc-soil boundary should be sharp. The methods used to discriminate soil from floc provide both a binary classification and a class assignment probability. Using finely sectioned long cores collected in floc-forming sloughs in central WCA2A, we examined the vertical changes in floc assignment probability. Fig. 41 summarizes the results for the 4 cores. In general, we observe moderately high floc class assignment probabilities for at least the first 6 cm of the profile; in two of the cores there are high probabilities as deep as 18 cm. This suggests that floc is more than simply a transient surface phenomenon, and some of floc materials are being incorporated in soils to some depth. However, after 18 cm, the probability of floc class assignment drops effectively to zero. Moreover, in two of the cores (Core 1 and Core 3) the decline in probabilities is abrupt, suggesting stark transitional boundaries, which would indicate limited integration of floc material into the soil. Cores 2 and 4 both have high floc probabilities in the surface sections, and then another band of apparent floc at 8-12 cm below the surface. Marl accumulation layers, perhaps resulting from historical floc accumulation or from some vertical profile in Ca availability or pH, are the likely explanation.

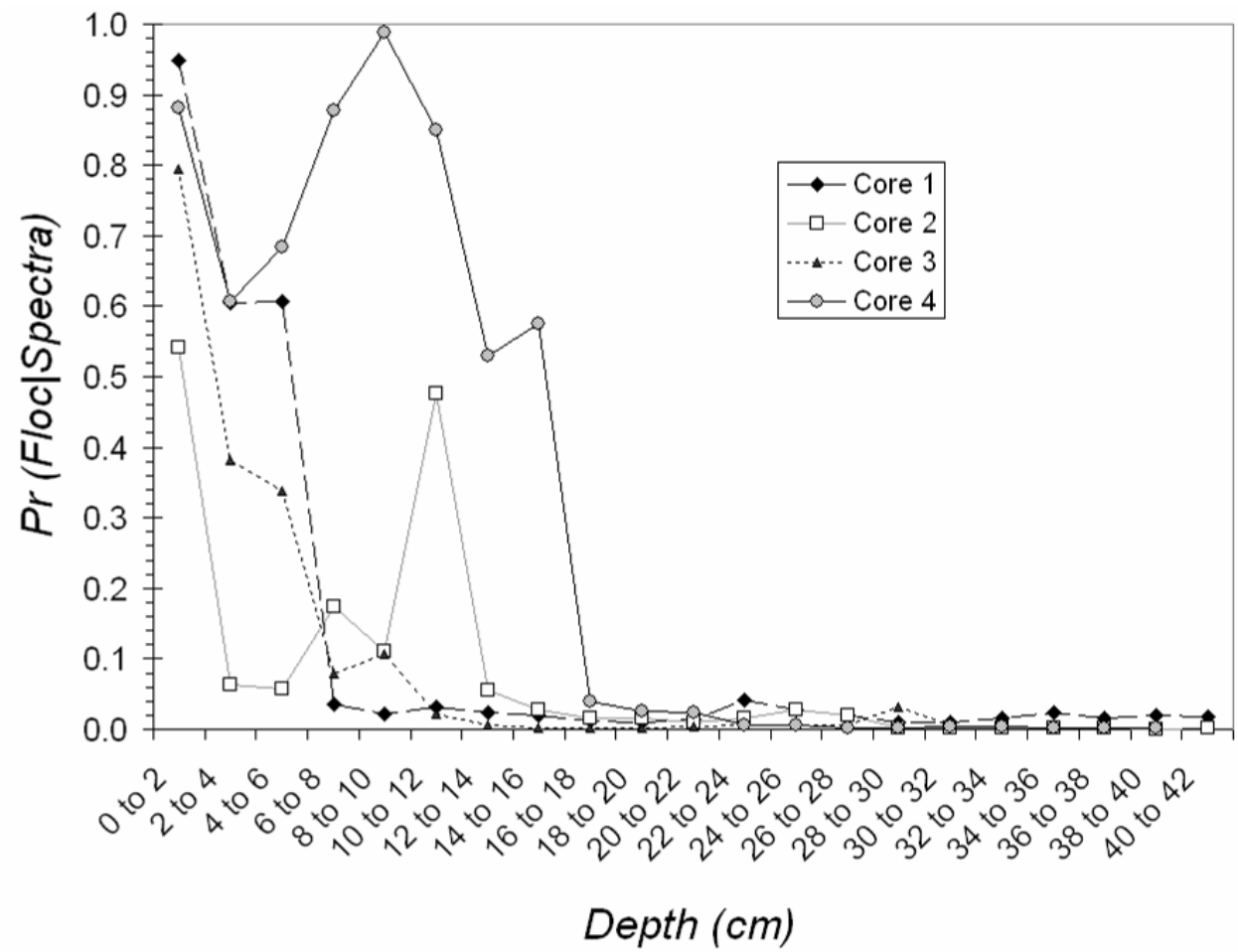

Fig. 41 – Probability of sample classification as floc vs. depth for 4 cores (Fig. XX) obtained from sloughs in WCA-2A. The graph shows variable but dramatic decline in floc classification probabilities given the sample spectra suggesting only moderate incorporation of the floc material into the soil profile. Note that the 0 depth is the top of the floc layer.

3.5 Analysis of Carbon Quality

3.5.1 Methane Production Summary

Methane production was measured in controlled incubations over 12 days; average rates of methanogenesis are plotted vs. the depth of the soil sample in Fig. 42 and 43. Error bars refer to standard errors derived from triplicate measurements from each sample depth.

As expected, the trend with depth is a decrease in methane production (corresponding to more recalcitrant material) and an increase in residual fiber (or lignin) content, which is relatively unavailable for microbial metabolism. We note that site SS1C (see Fig. 5) acts as an outlier in each case; this site is a marl prairie, with shallow calcitic soils and very little peat

accretion. Indeed, the vegetative community that persists at this site is typical of drier conditions than the communities present at the other three sites.

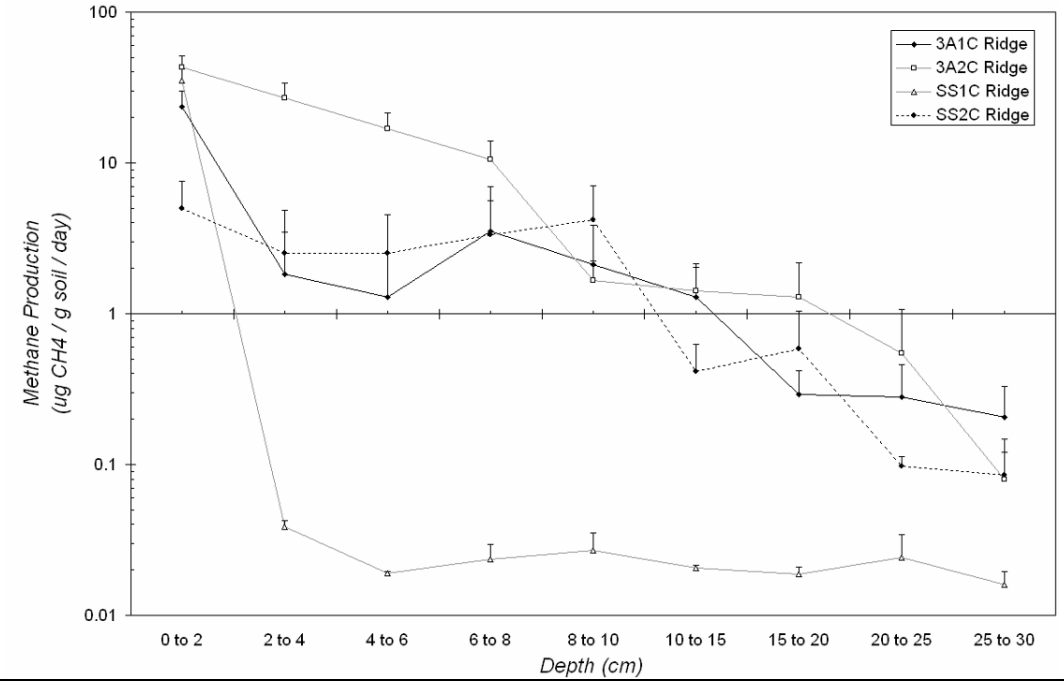

Fig. 42 – Methane production (µg/g soil/ day) for ridge cores.

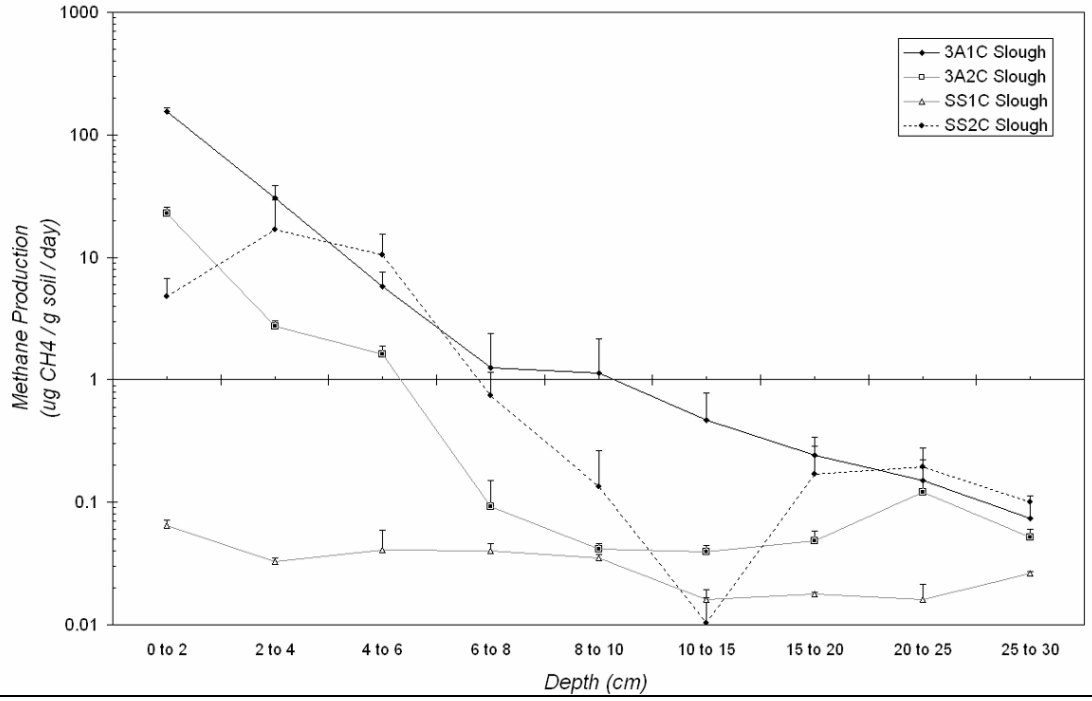

Fig. 43 – Methane production (µg/g soil/ day) for slough cores.

Several notable observations are evident from these graphs. First, the trend of increasing C recalcitrance with depth is relatively uniform between ridge and slough, which was surprising because the elevated lignin content of sawgrass litter vis a vis litter from slough co-dominants (*Utricularia* spp., *Nymphaea odorata*, *Eleocharis cellulosa*, *Rhynchospora inundata*). Indeed, the rates of methane production and variance with depth are remarkably similar given known differences in the input litter quality from the different communities.

A second relevant observation is that the analytical precision of the methane digestion protocol is low (particularly vis-à-vis the fiber analysis – see below), as demonstrated by the magnitude of the error bars between triplicate observations. The coefficient of variation (std. dev./mean) for the methane protocol averaged over 81%, while for the fiber analysis protocol, the same value averaged less than 20% over all samples. This suggests that, from a spectral modeling perspective, the use of fiber analysis is a more robust target because there is less potential for the calibration to be confounded by laboratory error.

3.5.2 Fiber Analysis Summary

The fiber analysis of the same set of sectioned cores shows a trend in the fraction of residual fiber in each sample that is consistent with the methane incubations (Figs. 44 and 45). In general, with the exception of the marl soil sample (SS1C), the residual fiber content increases monotonically with depth. The error bars are small (they report the standard deviation of three analytical replicates run in separate batches) suggesting that whatever signal we observe here is highly repeatable.

An equally important observation from a process perspective is that the fraction of peat residual fiber at depth is actually higher in slough (~50-60%) than in the ridge (~40%), omitting the data from SS1C. This is reinforced by weak evidence of increased methane production at depth in ridges (~ 0.3 μg/g/day) compared with sloughs (0.1 μg/g/day). We infer that despite strong differences in litter quality, the material that comprises the peat is of similar quality.

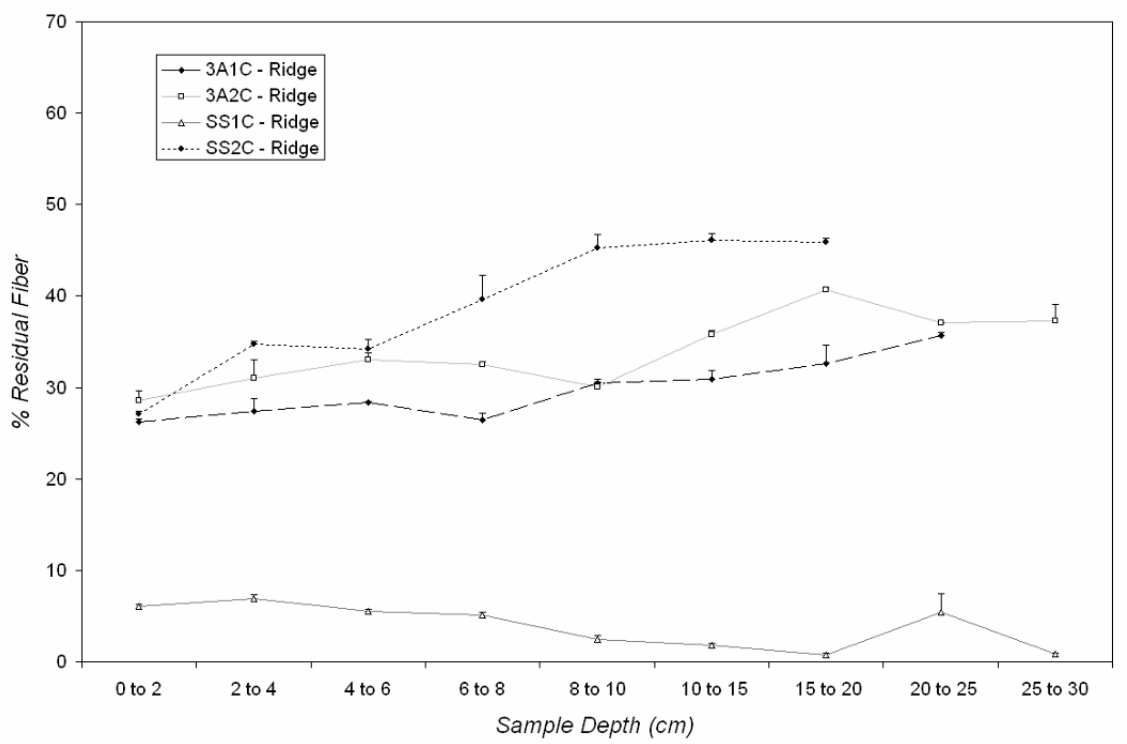

Fig. 44 – Residual fiber fraction (%) for ridge cores.

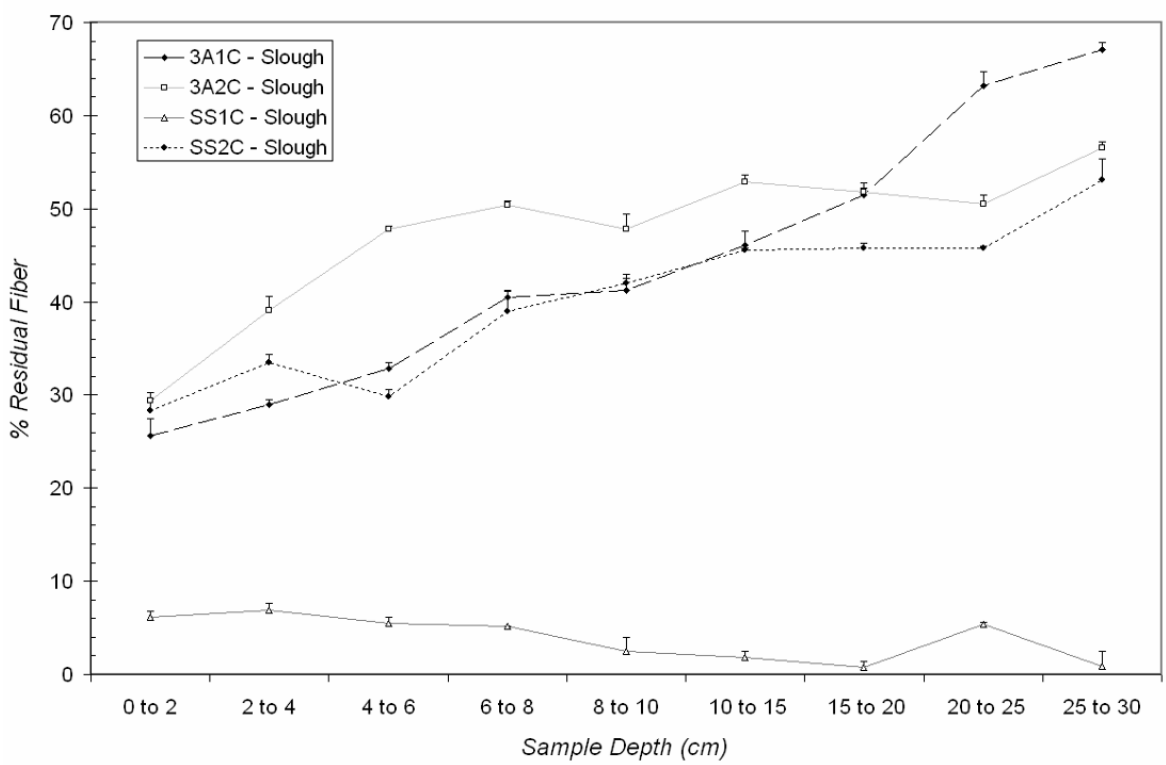

Fig. 45 – Residual fiber fraction (%) for slough cores.

We compared the two metrics of C quality using ordinary least squares regression; to linearize the methane production data, they were reported on a $\log_{10}$ scale. We observe moderate cross-correlation between them (Fig. 10 – $r^2 = 0.30$; $p < 0.001$) in the expected direction (as fiber content increases, methane production potential decreases); note that this analysis omits the data from site SS1C. We attribute the weaker than expected association primarily to reduced analytical precision for the methane analysis. We further note that the leverage of four points (Fig. 46) suggests that they may constitute analytical outliers. The r^2 value increases to 0.51 with those points omitted; notably all four points are from the deepest core sections (25-30 cm).

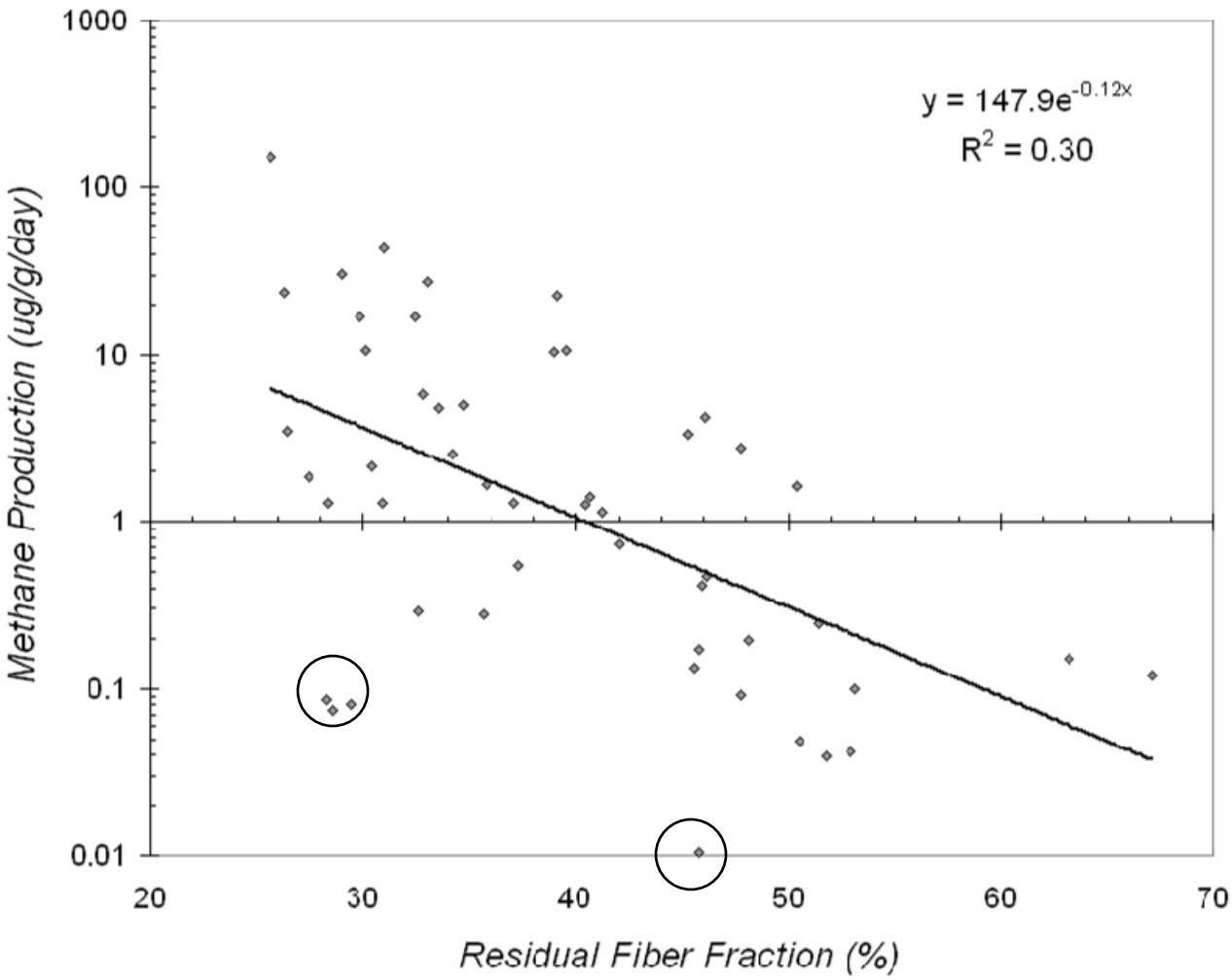

Fig. 46 – Cross-correlation between residual fiber content (%) and methanogenesis (μg/g/day) for all samples. Circled data points exhibit high leverage and large residuals, suggesting that they may be analytical outliers.

3.5.3 Spectral Prediction of Soil C Quality

In addition to running C-quality protocols on these additional cores, we selected 350 samples from the original data set on which to run similar analyses. After observing the poor analytical precision of the methane digestion approach and the concordance between that method and the fiber analysis method, we selected to run only fiber analysis on those 350 samples.

The selection of samples was done to maximize representation of the overall population. We used a Latin-hypercube sub-sampling design, predicated on soil spectra. This protocol partitions multi-variate space defined by spectral principal components into segments (n = 432) from which we selected, at random, one sample; 84 segments had no samples, resulting in 348 samples that maximally cover observed variability in the data set (n = 3911). Fig. 47 shows spectral locations of the sub-sample with respect to the population. We make this effort to represent the full data set to the maximum extent possible so that any subsequent spectral prediction made based only on the sub-sample will not encounter samples that are spectrally different.

After identifying the sub-sample set and performing the fiber analysis procedure, we summarized the results according to various attributes of interest, including community type and sample depth (floc, 0 – 10 cm, 10 – 20 cm). As expected, both exhibited significant differences ($p < 0.01$). The effect of depth is as expected (Fig. 49), with significant increases in residual fiber content between floc and soil ($p < 0.001$); differences between the two soil profiles (0 to 10 cm and 10 to 20 cm) were non-significant ($p = 0.14$), but trended as expected. Differences in residual fiber content between community types yielded several surprising results (Fig. 50). First, soils in mangrove ecosystems were significantly lower in residual fiber than other ecosystems. Tree-islands had higher residual fiber content than any other system; both these inferences are suspect because of small sample sizes for each. We note here that all comparisons made in this section are for actual fiber observations, not for spectral inference of fiber content.

Among the community types with sufficient replication to be reliable, wet prairies had significantly lower residual fiber than ridge and slough; sloughs had higher fiber content than ridges, but the p-value for that post-hoc contrast was only marginally significant ($p = 0.02$). We reiterate that our objective here is not to exhaustively describe patterns of C quality across the Greater Everglades, but to demonstrate that these data are of interest, and to consider the feasibility of their prediction using spectra sample spectral reflectance.

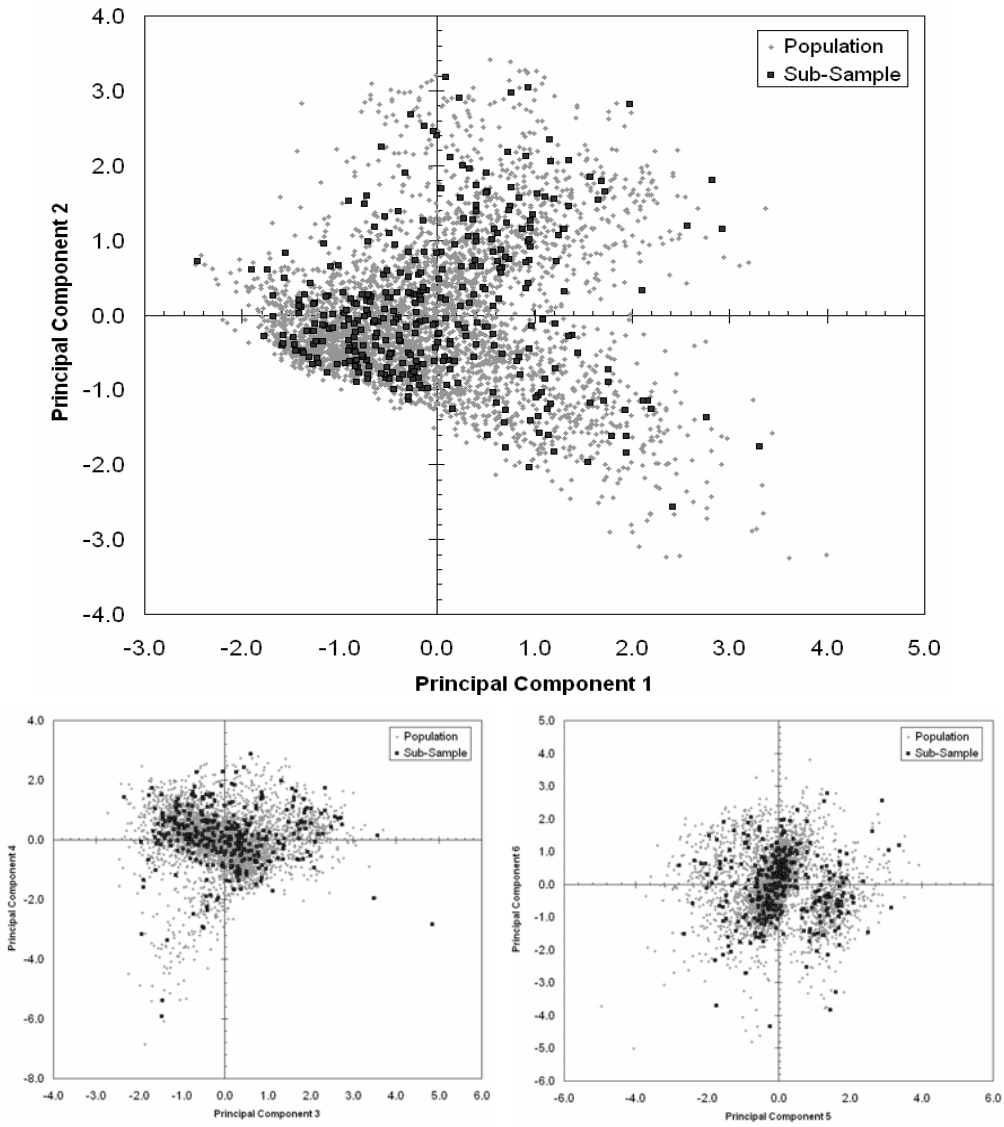

Fig. 47 – Representation of the overall spectral population (reference library – grey diamonds; n = 3911) by the Latin hyper-cube sub-sample (black squares; n = 348).

Predicting litter and soil C quality using spectra is supported by recent research in this field. Shepherd et al. (2003) and Bouchard et al. (2001) demonstrate the central concept in manure and litter, respectively. Here we only summarize the results of chemometric development between spectra and residual fiber predicted for the 350 samples from throughout the Greater Everglades. While we developed spectral models predicting methane production and fiber analysis for the core study, those results are suspect because of auto-covariance issues created by multiple samples being from the same core. However, we did observe stronger

spectral prediction efficiency for the fiber analysis than for methane (validation r^2 values of 0.93 and 0.67, respectively) underscoring methodological uncertainties in the latter protocol.

Fig. 48 shows the hold-out validation comparison of predicted and observed residual fiber content. As shown, the validation r^2 is 0.83 between predicted and observed; as with the results presented previously, we place more emphasis on the RPD and SEV values. In this case, the RPD is 2.47, suggesting comparatively strong prediction efficiency that the literature suggests is of immediate and broad utility.

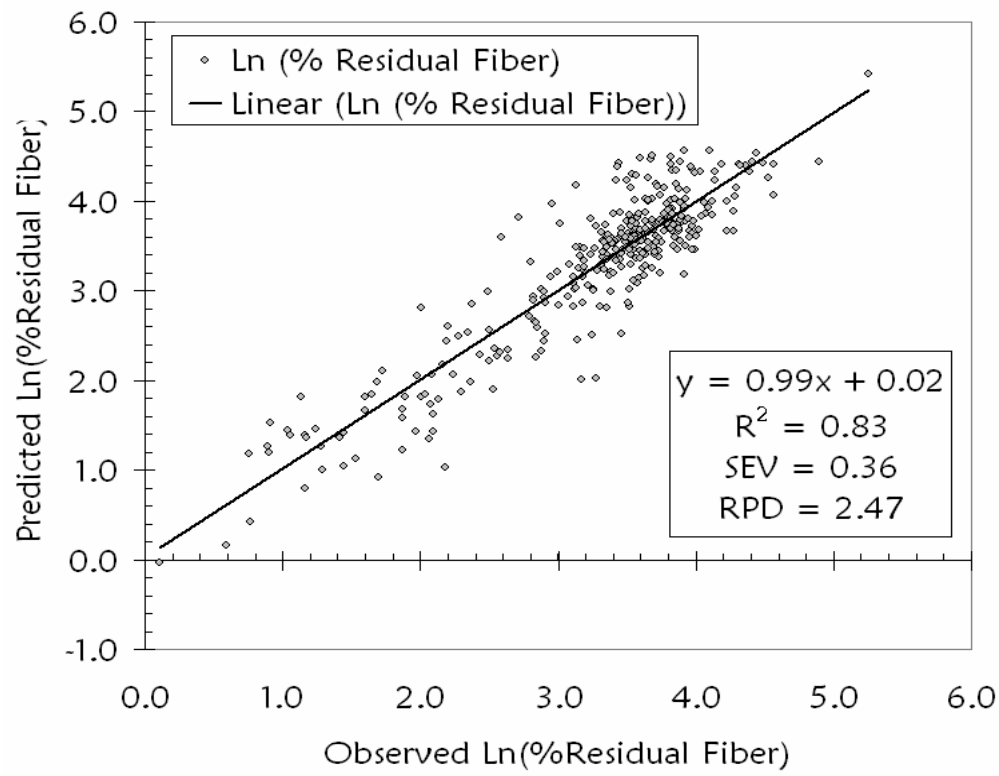

Fig. 48 – Predicted vs. observed residual fiber content for the hold-out validation data set. Values were natural log transformed prior to analysis to meet assumptions of normality.

In addition to predictions of the residual fiber pool, we develop calibrations to other pools (Neutral Detergent Soluble – Waxes/Fats/Carbohydrates, Acid-Detergent Soluble - Hemi-Cellulose, Strong Acid Soluble - Cellulose) for comparative purposes, and to further determine the resolution of the spectral prediction. The results are somewhat less encouraging for these other pools, as summarized in Table 12. Clearly, the spectral method is most sensitive to the bulk differences between recalcitrant organic material and bioavailable material (residual content), and our ability to discriminate categories of bioavailability is limited. In particular,

spectral prediction of the neutral detergent soluble fraction (%NDF) is poor. The other two fractions approach the literature threshold (RPD > 1.5) for useful chemometrics.

Several modifications to our selected analytical procedure might yield better results in future work. In particular, our ability to control for the solubility of the mineral fraction of the sample is quite limited, and sample pretreatments to exclude organic material followed by the sequential fiber protocol could yield more experimental control over this unknown.

Table 12. Summary of chemometric performance for fiber fractions.

Fiber Fraction	Validation r^2	Validation SEV	Validation RPD
%NDF	0.39	12.78	1.26
Ln[%ADF]	0.59	0.62	1.48
Ln[%H_2SO_4]	0.54	0.63	1.38
Ln[Residual]	0.83	0.36	2.43

%NDF – neutral detergent fraction; %ADF – acid detergent fraction, %H_2SO_4 – strong acid soluble fraction, Residual – lignin/residual fiber fraction

3.5.4 Soil C Quality by Area and Community

Given the spectral prediction of residual fiber content, we can estimate patterns of C quality as a function of depth (Fig. 49) and community type (Fig. 50). Both show significant differences. The relationship with depth is as expected, with low fiber content in floc.

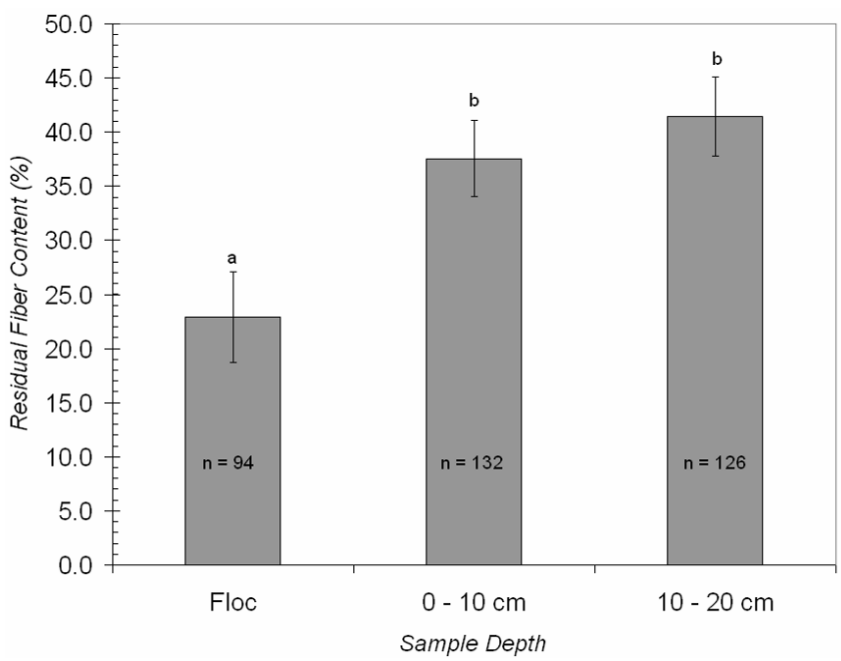

Fig. 49 – Residual fiber fraction compared across sample depth. Different letters denote communities with significant differences (p < 0.05); sample sizes for each are given.

Differences among communities are more difficult to interpret. Tree islands (n = 4) had the highest residual fiber content; while this is consistent with what would be expected of organic soils that form in comparatively oxic environments, more samples are required to substantiate the trend. Interestingly, sloughs generally had significantly higher residual fiber than ridges, challenging the notion that soils in sloughs, which are made from labile plant tissues, should themselves be more labile.

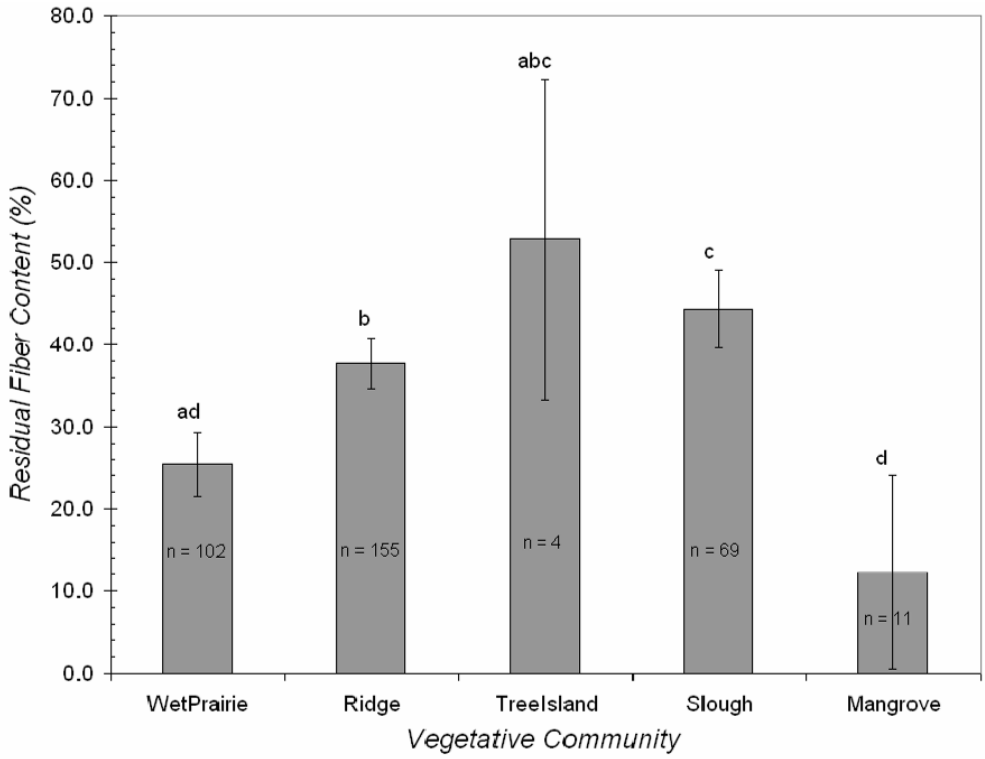

Fig. 50 – Residual fiber fraction compared among dominant vegetation type. Different letters denote communities with significant differences ($p < 0.05$); sample sizes for each are given.

3.5.5 Maps of Soil C Quality

One objective of developing the spectral model predicting soil residual fiber content was visualization of that quantity across the entire ecosystem. Figure 51 shows that product; Table 2 summarizes the relevant mapping parameters (nugget and sill variance, range, cross-validation accuracy) for this and all other maps presented in this work. Fig. 51 indicates significant spatial pattern to the soil carbon quality with variability over 3 orders of magnitude. The highest quality C is found in soils in northern Big Cypress Preserve and, to a lesser extent, WCA1 (Loxahatchee). The fringes of Shark River Slough in Everglades National Park have soils with

low residual fiber content. Note that this map was produced using the data inferred for the top soil layer (i.e., not surface floc or deeper soil profiles).

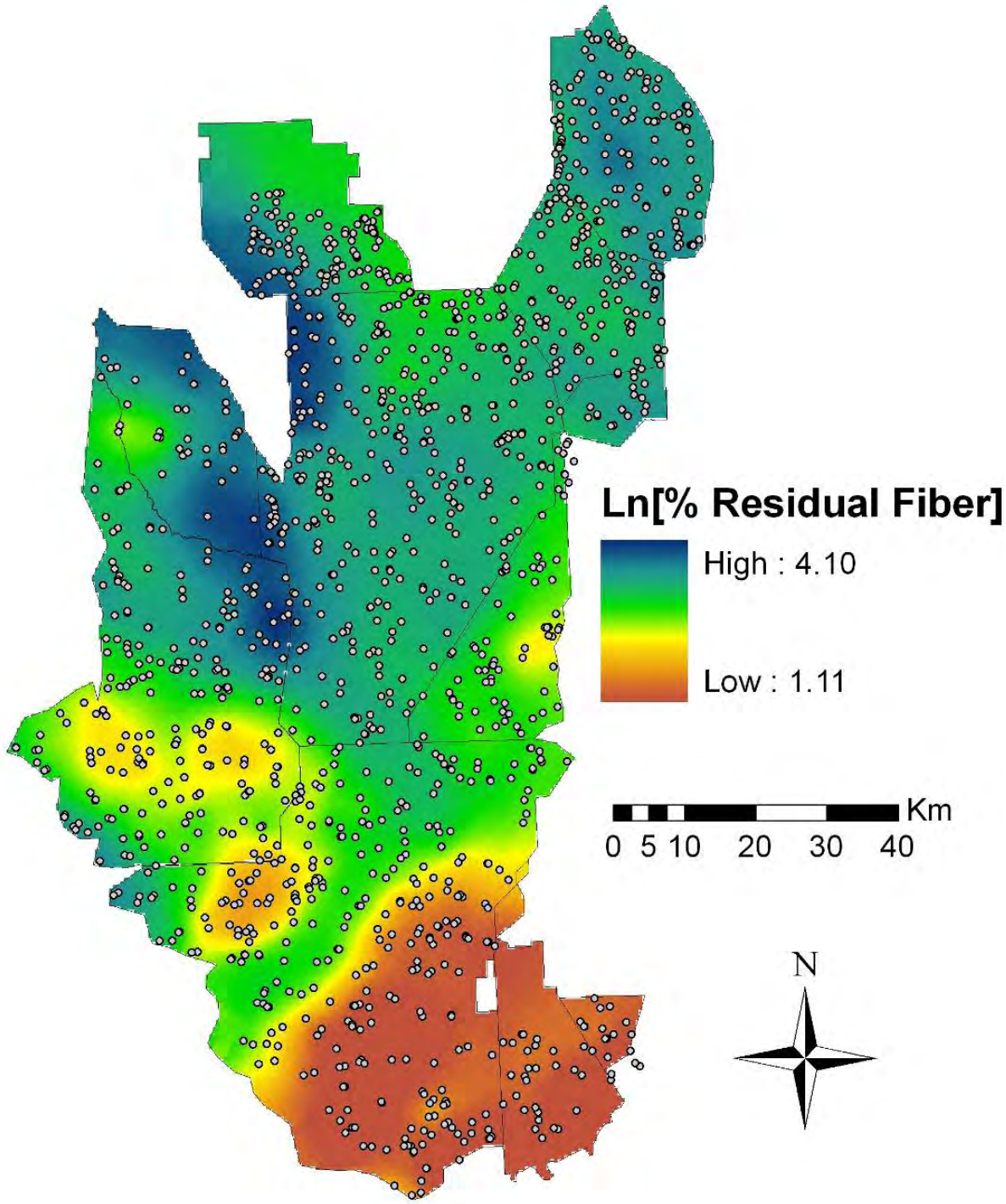

Fig. 51 – Map of residual fiber content across the Greater Everglades produced by applying the chemometric model for that parameter, constructed using a subset of 348 samples, to the entire reference library (n = 1594 in profile 2 – 0-to-10 cm).

3.6 Analysis of Total Mercury

Samples were distributed throughout the Everglades, with no evidence for sampling bias to a particular area (Fig. 52). The average minimum distance between samples was 1902 m. In general, samples were sparser in the rocky southeastern region of ENP and WCA 3B (mean distance = 2077 m and 2611 m, respectively) and in Big Cypress National Preserve (2330 m) than in WCAs (WCA1 – 1834 m, WCA2A – 1446 m, WCA3AN – 1767 m, WCA3AS – 1905 m). Sample densities in smaller hydrologic zones were higher (HL – 1175 m, MOD – 1210 m, RB – 1283 m). Overall, the sample density, with over 25% of samples within 1000 m of another site, supports exploration of lag spacings for semi-variogram analysis between 100 and 1000 m.

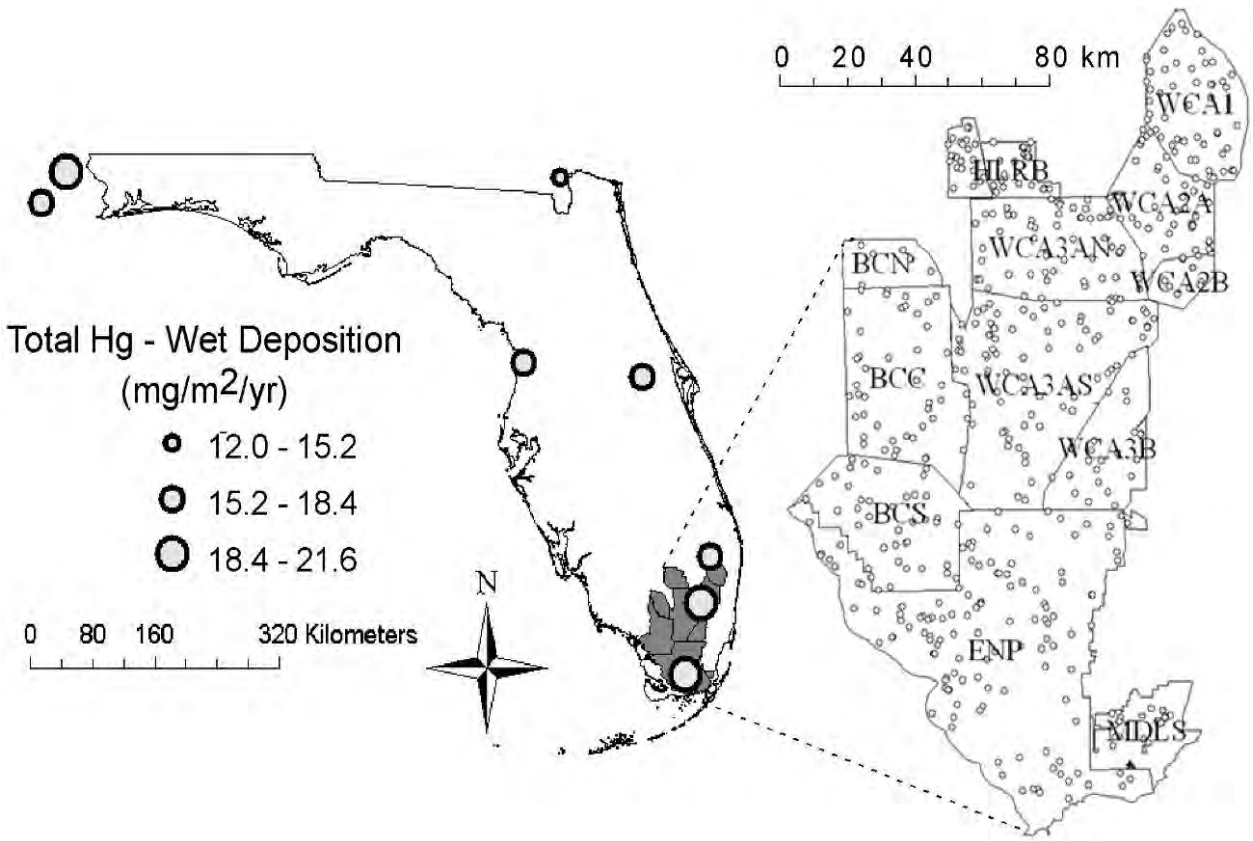

Figure 52. Soil sampling locations across the Everglades. Shown on the Florida map are National Atmospheric Deposition Program Hg monitoring stations and recent (since 1997) annual average wet deposition rates ($\mu g/m^2/yr$) (NADP 2007). Hydrologic partitions are labeled on the Everglades map (BC = Big Cypress Preserve, WCA = Water Conservation Area, ENP = Everglades National Park, HLRB = Holeyland/Rotenberger tracts; MDLS = Model Lands)

3.6.1 Summary Statistics (Total Hg per mass and Total Hg per area)

The mean mercury concentration (THg_M) across the Everglades was 0.156 mg/kg (Table 13) with concentration above 0.2 mg/kg at 168 sites. The mean THg_A was 2.3 mg/m^2, approximately 100 times the estimated annual loading rate of 25 μg/m^2/yr given a sampling depth of 10 cm. Variances around both means are large with the coefficient of variation (CV) roughly 80% for both, and distributions for both are moderately skewed right. The maximum concentration (THg_M = 0.917 mg/kg) occurred in western WCA3A; 32 sites (5%) had concentrations > 0.4 mg/kg, 15 of which were in WCA3A and 12 of which were in WCA2A and WCA1. Further, while no sites in the BCNP had THg_M > 0.2 mg/kg, 58% of sites in WCA1 (34 out of 59 sites) and 60% of sites WCA3AS (51 out of 85 sites) were above the threshold.

Bulk density values, which were used to compute THg_A, varied dramatically across the Everglades with mean values in ascending order: WCA1 (BD = 0.09 g cm^{-3}) < WCA2A (0.11 g cm^{-3}), WCA3AS (0.12 g cm^{-3}) < WCA3B (0.13 g cm^{-3}) < WCA2B (0.18 g cm^{-3}), WCA3AN (0.19 g cm^{-3}) < ENP (0.24 g cm^{-3}) < BCNP (0.67 g cm^{-3}).

Table 13. Statistical properties of THg_M and THg_A across the Everglades

Variable	THg_M (mg kg^{-1})	THg_A (mg m^{-2})
N	600	600
Mean	0.156	2.267
SE mean	0.006	0.066
Min	0.002	0.075
Max	0.917	12.038
Med	0.126	1.888
StDev	0.138	1.622
CV †	88.461	71.548

Correlations of THg_M with other measured biogeochemical properties (TP, TC, LOI and BD) were significant (Table 14), but correlations between those same variables and THg_A were non-significant. The absence of correlation between THg_A and principle indicators of soil type and condition (except TP in enriched sites) offers evidence for depositional uniformity in space. Covariance between THg_M and soil properties was strongest with TC, and weakest with TP. Average TP, TC, and THg_M observations by region, ranked by mean TP from left to right (Fig 4), suggest positive covariance between TP and both THg_M and TC in areas with low TP (BCNP, ENP WCA3AS). In areas with elevated TP (WCA1, WCA2A, WCA3AN), the association with

THg and TC reverses. This non-linear association is reinforced by a scatter plot of THg_M and TP (Fig 53), which shows strong positive correlation below TP concentrations of 500 mg/kg, but moderate negative correlation above that level. Notably, soil with TP > 500 mg kg^{-1} are considered P enriched (DeBusk et al., 2001). Using this threshold, we evaluated correlations between variables in unimpacted (TP < 500 mg/kg) and impacted (TP > 500 mg/kg) sites (Table 14). Correlations between THg_M and TC were consistently positive, but stronger in unimpacted sites. In contrast, correlations between THg_M and TP were negative in impacted soils and positive in un-impacted soils; all correlations were statistically significant at $p < 0.05$.

Table 14. Pearson correlation coefficient of THg_M and THg_A with biogeochemical properties, and with TP and TC in impacted and un-impacted areas of the Eveglades

	THg_M	THg_A	TP	TC	TN	LOI	BD
Overall (n = 600)							
THg_M	1.00	0.63*	0.22*	0.58*	0.60*	0.58*	-0.48*
THg_A	0.63*	1.00	-0.09*	0.01	0.07	0.03	0.04
Un-impacted (n = 461)							
THg_M	1.00	0.62*	0.51*	0.61*	0.63*	0.61*	-0.51*
THg_A	0.62*	1.00	0.08	0.04	0.09	0.06	0.01
Impacted (n = 139)							
THg_M	1.00	0.76	-0.35*	0.39*	0.39*	0.40*	-0.32*
Thg_A	0.76*	1.00	-0.30*	0.07	0.13	0.06	0.11

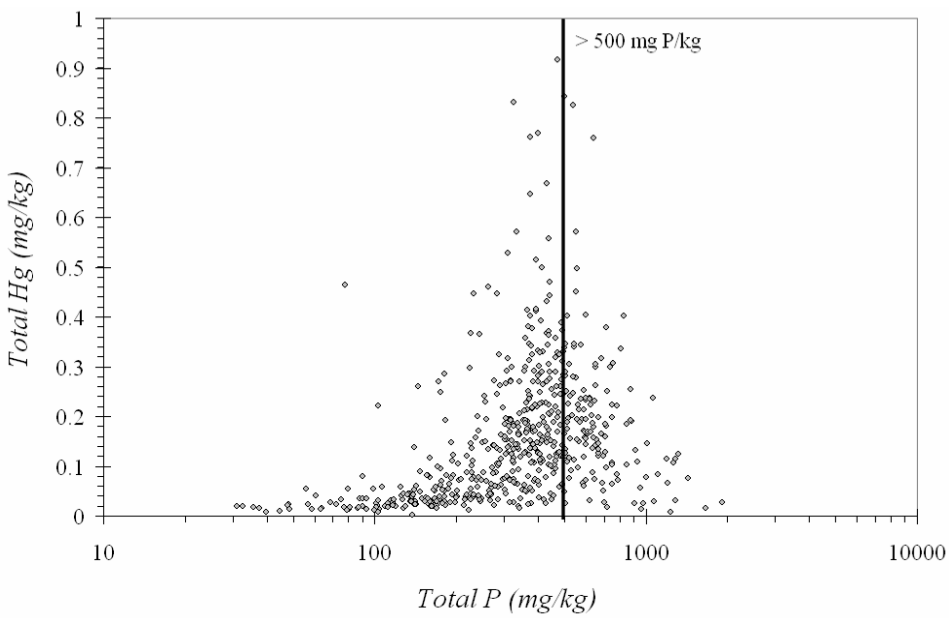

Fig. 53 – Relationship between sample total P and total Hg. TP enriched areas show a negative correlation with THg, while below 500 mg/kg the correlation is positive.

3.6.2 Spectral Prediction of Total Hg (per mass)

As with all other parameters in this work, the principal objective of our NIRS feasibility study was to determine if methodological accuracies are sufficient for routine monitoring and assessment. We applied that same set of chemometric tools to the prediction of Total Hg from spectra, and observed relatively encouraging results (validation accuracy; Fig. 54). While the error rates are somewhat high and there is substantial residual error, the model suggests that spectral prediction is a plausible method for increasing the spatial or temporal density of sampling. Given the need for high accuracy measurements, and given the low overall concentrations, this method is most likely to be useful in an integrated sampling effort, wherein routine analyses are performed simultaneously, and spectral prediction is used principally to increase sample densities. Given this constraint, intrinsic uncertainty about the manner in which total Hg can be spectrally predicted, and the size of the data set for which total Hg measurements were obtained, all subsequent analyses are performed on the raw data only.

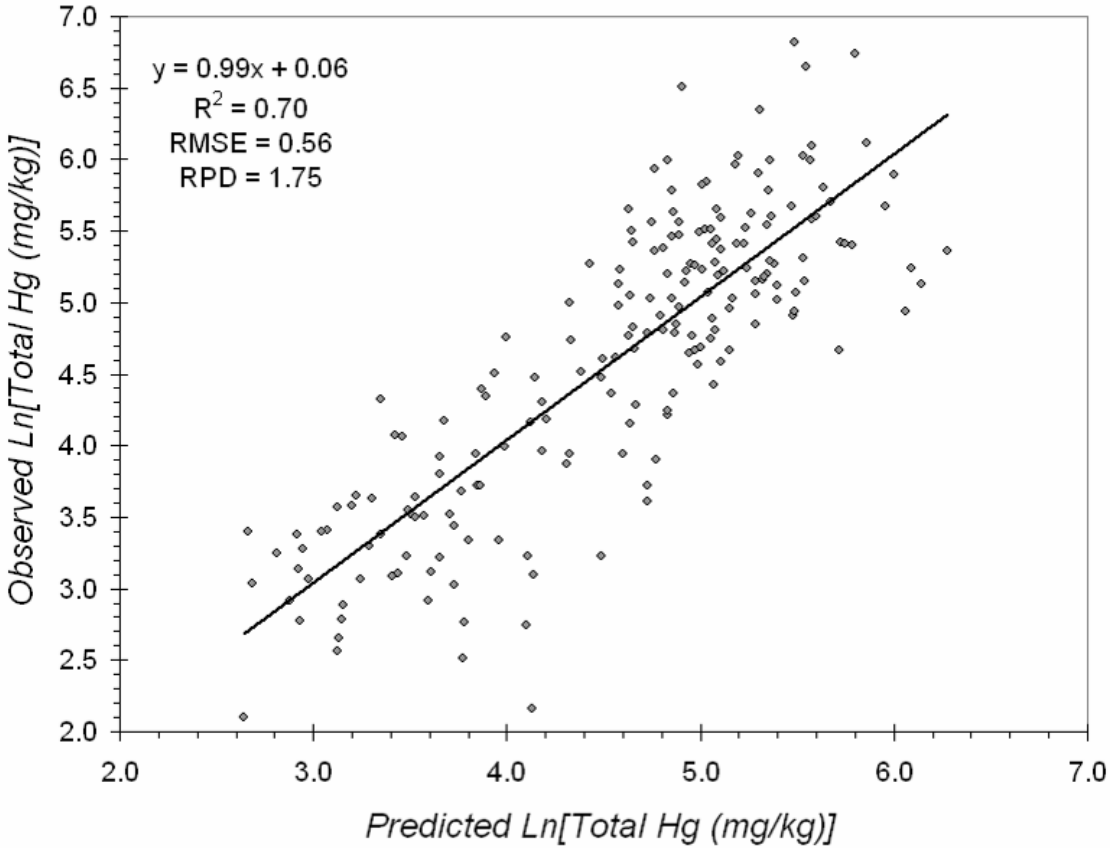

Fig. 54 – Validation efficiency of spectral prediction of Total Hg across the Greater Everglades. Some of the observed correlation is with Total C content, but correlations are much higher for spectra (r = 0.84 for spectra vs. 0.58 for TC; Table XX)

3.6.3 Analysis of Semi-Variance

Semivariograms for THg_M and THg_A showed high semivariance at short lags, indicating potential spatial outliers. Based on local Anselin's Moran I, 17 observations were identified as local spatial outliers, and removed for semivariance analysis. Spatial autocorrelation of both variables was modeled using a double spherical structure (Fig 55); modeled parameters are given in Table 15. We observed a clear sill in semi-variance at a range of 30 km, suggesting that a linear semi-variogram model (Stober et al. 2001) is inappropriate for interpolation.

Overall, semivariance analysis indicates strong spatial structure for THg_M with a relative structure parameter (Q) indicating that over 80% of the spatial variance is explained by the model semivariogram. That is, only 19% of spatial variance (nugget variance) remains unexplained by the semivariogram model. The semivariogram for THg_A explained less of the total variance (56%) suggesting spatial structuring was absent or occurring at different scales than our observations. There was no evidence of significant anisotropy in either variable.

The spatial extent of THg_M and THg_A in quantized ranges indicate that over 77% of the area had THg_M values below 0.2 mg/kg, and less than 2% had concentrations more than double that threshold. This result, combined with our exploratory analyses that suggested 5% of observations exceeded 0.4 mg/kg illustrates the localized nature of severe mercury enrichment. That is, large areas had low THg_M levels, and hot spots were geographically constrained. While there is no regulatory threshold for THg_A, we observed that nearly 30% of the landscape had levels in the upper 10 cm exceeding 100 years of deposition at current rates.

Table 15. Semivariogram parameters for THg_M and THg_A and OK cross-validation errors.

Variable	Semivariogram parameters †						Error statistics‡		
	Nugget	Partial sill §	Range (m) §	Partial sill ¶	Range (m) ¶	Q#	ME	RMSE	r
THg_M	0.004	0.00459	4,498	0.012	37,398	80.6	0.0006	0.098	0.70
THg_A	1.009	0.426	3,998	0.858	42,998	56.0	0.0024	1.461	0.45

† All parameters are for spherical semivariogram model
‡ ME, Mean Error; RMSE, Root mean squared error; (Schloeder et al., 2001)
§ Parameters for first spherical model
¶ Parameters for second spherical model

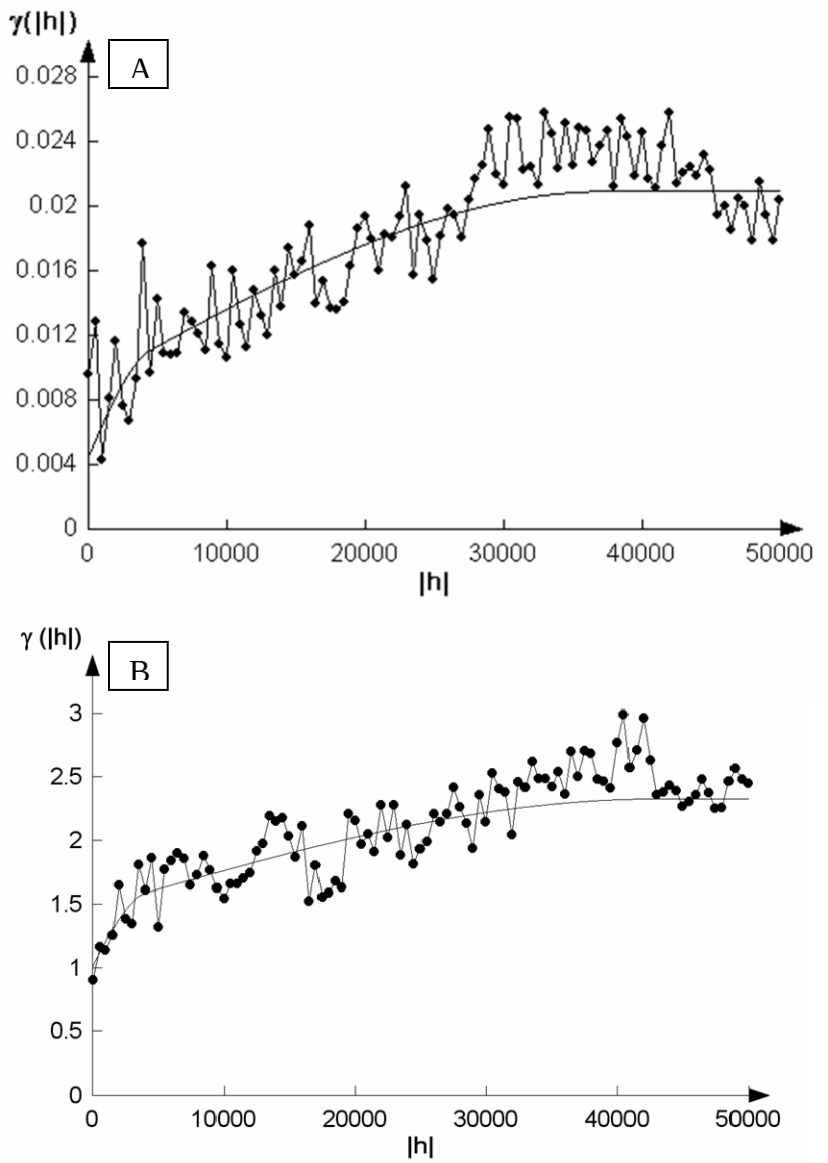

Fig. 55 Modeled global semivariograms for a) THg_M and b) THg_A with 500 meter lags.

Cross-validation efficiency for both THg_M and THg_A (plotted in Fig. 56 and summarized in Table 15) showed strong spatial agreement between predicted and observed to THg_M, but reduced agreement for THg_A. The mean error for both is close to 0 substantiating that kriging predictions were unbiased. The RMSE for THg_M is 0.098 mg/kg; the RSME was as high as it is principally because predictions for 6 validation sites, two of which were under-predicted and four that were highly over-predicted; the high correlation value (r = 0.70) underscores the relative prediction efficiency for most validation sites. RMSE for THg_A was high (1.46 mg/m^2), with an overall fit that was fair (r = 0.45) suggesting less consistent spatial structuring.

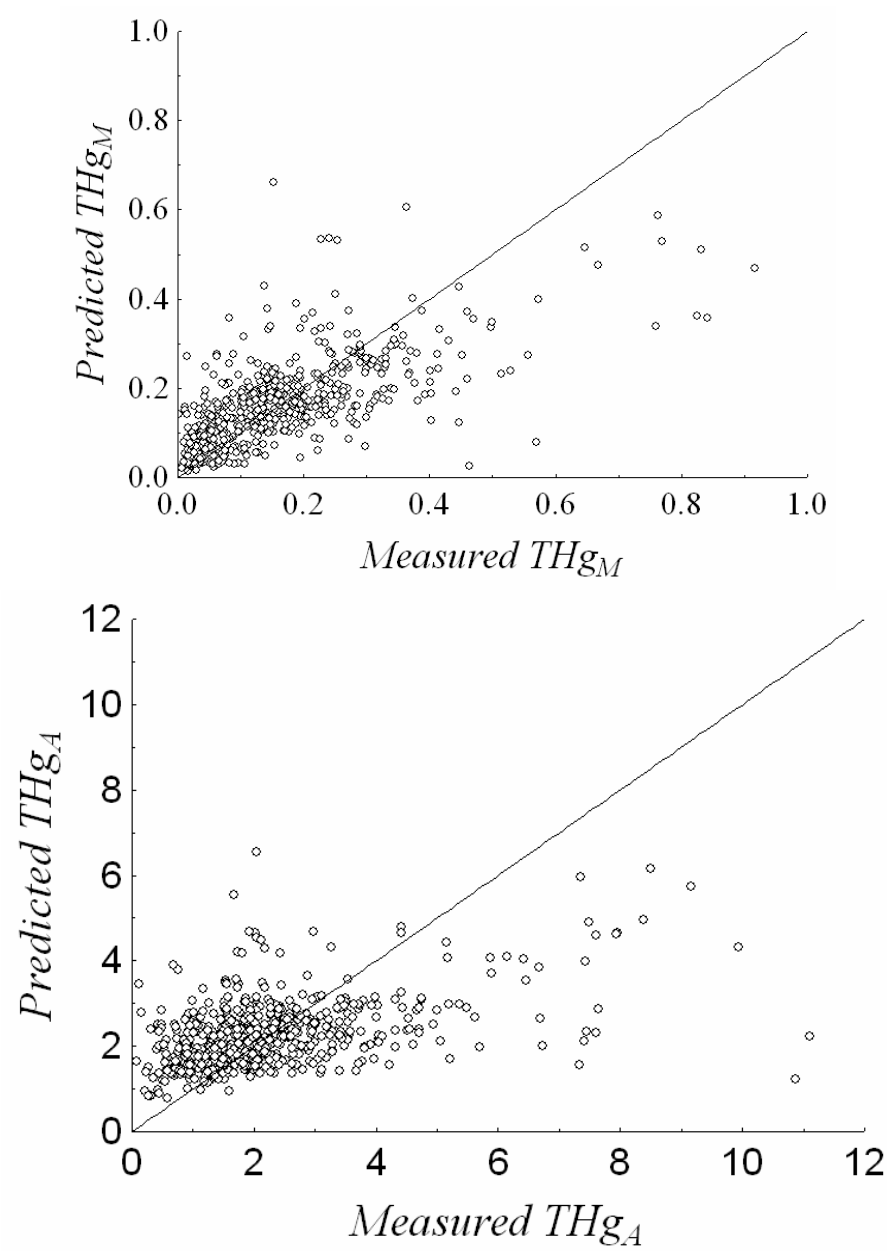

Fig. 56 – Cross-validation efficiency for kriging models for total Hg per mass and per area.

3.6.4 Analysis of Cross-Covariance

Sample variances for THg_M differed markedly across sub-regions, with low variances observed for BCNP (0.0007), WCA3AN (0.004), ENP (0.008), and high values in WCA1 (0.014), WCA2A (0.015) and WCA3AS (0.039); overall variance was 0.019. These differences suggest that spatial structure could also differ from the global spatial structure observed above. Spatial structures within individual regions were analyzed using semivariograms, and modeled

parameters (Table 16) suggest smaller ranges for the sub-regional models compared to the global model; this is expected because semivariograms are scale dependent (Qi and Wu 2005). Among sub-regions, ENP exhibited the largest range (11,539 m), which could be attributed to larger spatial extent. Sill values, which also varied among sub-regions primarily respond to mean concentrations, so lower values were observed in BCNP, ENP and WCA3AN, with larger sills in WCA3AS&B and WCA1&2. The relative structure quantifies the degree to which the model semivariogram captures spatial variance, and all values for sub-regional models are less than the global model (Q = 80.6 – Table 15).

Using sub-regional models to make predictions did not improve interpolation accuracy; the overall RMSE between predicted and observed for cross-validation using the sub-regional models was 0.101 mg/kg compared with 0.098 mg/kg for the global model.

We examined cross-semivariances in an effort to identify ancillary soil measurements (TP or TC) that could be used to improve interpolation. Cross-semivariograms at the regional scale were significant for TC only (Table 16), but did not improve interpolation accuracy despite the availability of 744 additional sample sites at which TC was measured. Cross-validation RMSE (0.099 mg/kg) and predicted vs. observed correlation (+0.71) were the same as univariate interpolation (Table 15). We also selected 2 sub-regions based on mean total P concentrations, one impacted and the other unimpacted for more geographically confined investigation; ENP

Table 16. Parameters for semivariograms for THg_M in stratified regions, and cross-semivariogram of THg_M with TP and TC in selected impacted and unimpacted areas

Areas	N	Var.	Model	Nugget	Partial Sill	Range	Q †	Mean Error	RMSE	r
			Semivariograms							
BCNP	95	THg	Spherical	0.00015	0.0005	5400	76.1	-0.001	0.026	0.30
ENP	155	THg	Spherical	0.00249	0.0044	11539	63.9	0.001	0.092	0.42
WCA1/2	120	THg	Spherical	0.00710	0.0070	8320	49.6	0.003	0.104	0.49
$WCA3A_N$	113	THg	Spherical	0.00035	0.0006	7474	64.4	-0.002	0.064	0.54
$WCA3A_S$/B	117	THg	Spherical	0.01080	0.0224	3549	67.5	-0.001	0.156	0.53
			Cross-semivariograms							
Global§	600	THg-TC	Double Spherical	0.48	1) 2.40 2) 8.88	1) 7800 2) 47000	95.9	0.000	0.099	0.71
ENP	155	THg-TC	Spherical	0.8	6.42	10749	87.6	0.0018	0.084	0.48
ENP	155	THg-TP	Double Spherical	2.4	1) 6.55 2) 0.88	1) 1199 2) 10000	67.7	-0.0019	0.075	0.62
$WCA3A_N$	113	THg-TC	Spherical	0.5	5.1	10750	90.2	-0.0063	0.064	0.67

† Q, Relative structure expressed as %

§ Greater Everglades, integrating across compartments, including TC measurements for spatially denser prediction.

(mean TP = 341 mg/kg) was selected as un-impacted, and WCA3AN (mean TP = 539 mg/kg) was selected as impacted. Cross-semivariograms parameters (Table 16) indicate that Thg_M was spatially cross-correlated with both TP and TC in ENP, but was cross-correlated with TC only in WCA3AN. The relative structure of the modeled semivariograms shows that spatial variation was explained equally well in both sub-regions using TC as a covariate, and substantially better than the relative structure of THg_M alone. Using the stronger TC cross-semivariogram for co-kriging interpolation yielded results that were markedly improved for both sub-regions.

3.6.5 Analysis of Total Hg by Community Type

To examine evidence for differential enrichment by biological processes, we compared the THg_M levels among dominant community types (Fig 57A); there were clear global differences with mean THg_M spanning a gradient from 0.06 mg/kg in depressional marshes (DM) and wet prairies, to a high of ~ 0.20 mg/kg in ridges, tree islands and sloughs. Analysis of variance (ANOVA) of THg_M by community type was significant; post-hoc comparisons showed mean THg concentration differed significantly between wet prairie and both ridge and slough.

Despite significant global differences, variability in mercury deposition and peat accretion rates makes comparison of geographically proximate sites more meaningful. Using only ridge, slough and wet prarie communities because of sample size constraints, we selected the closest site of a different community type to make a sample pair; sites where the nearest pair was > 5,000 m away were removed from this analysis. Pairwise analysis (paired t-test) of differences between community types was performed. The histogram of paired differences (Fig. 57B; paired concentration diff. = slough minus ridge or wet prairie minus ridge) shows weak evidence of systematic differences in enrichment. Paired t-tests were equally significant for both comparisons ($p = 0.08$), suggesting that ridges have higher concentrations than proximate wet priaries, but lower concentrations than proximate sloughs. We surmise that there are no strong community effects, and that any global effects are more likely a function of deposition or peat accretion conditions than indicators of differential risk of mercury enrichment among ecotypes.

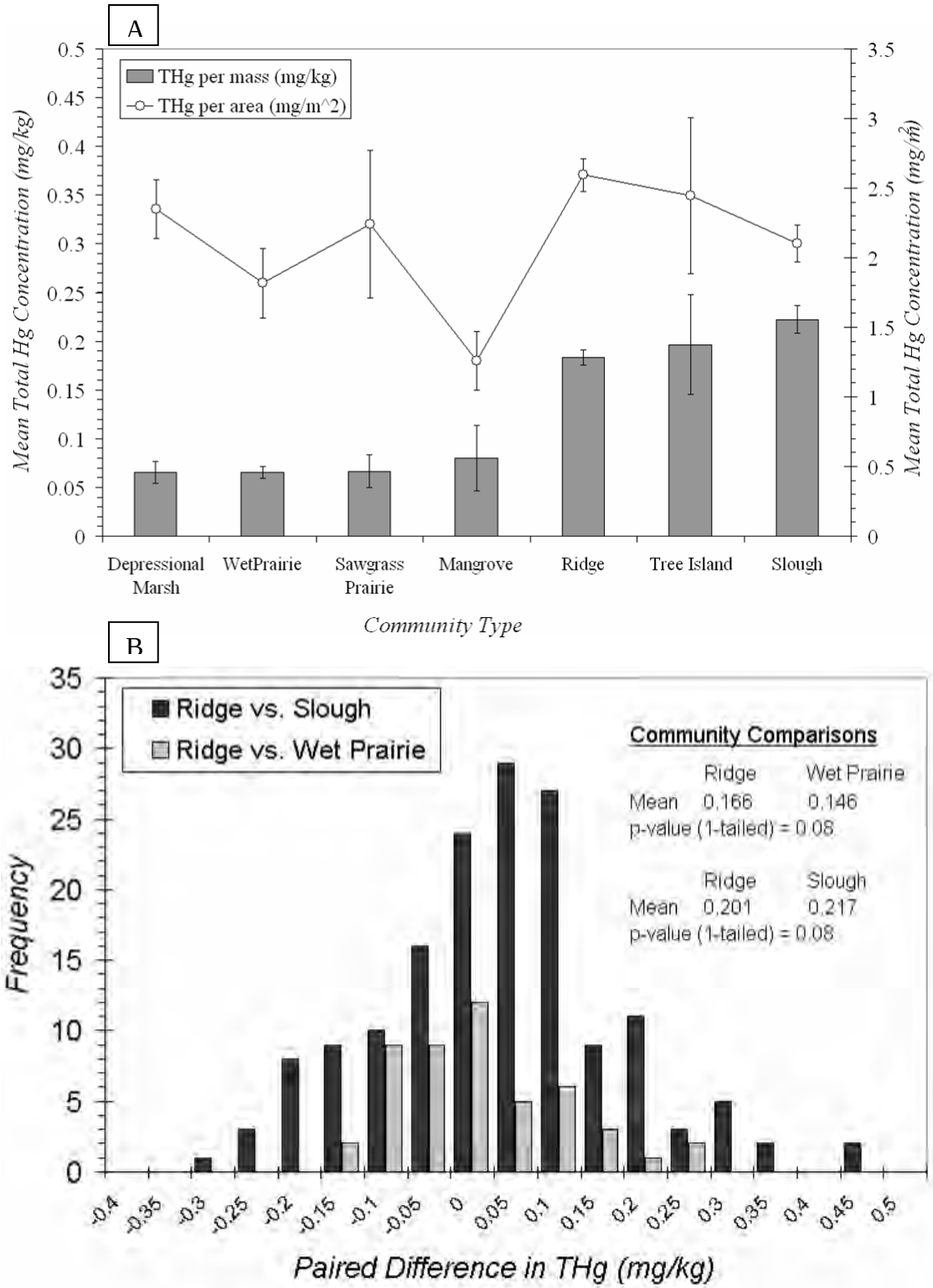

Fig. 57 (a) Mean (and standard error) THg_M and THg_A levels by community type regionally. (b) Histogram of paired differences between communities (max. pair separation = 5,000 m) with paired t-test results comparing means.

3.6.6 Analysis of Total Hg Gradients

To further explore the relationship between mercury and phosphorus, THg_M and TP concentrations were evaluated at multiple buffer distances from canals (0-500 m , 500-1000 m, 1000-1500 m, 1500-2000 m, 2000-2500 m, 2500 – 3000 m, 3000-4000 m, 4000-5000 m, 5000-6000 m) (Fig 58). Regression of concentrations vs. distance was significant ($p < 0.05$) for both TP and THg_M, but the direction of covariance was opposite. Specifically, TP was negatively correlated ($r = -0.73$) with distance from canals, while THg_M was positively correlated ($r = 0.77$). The association between distance from canals and THg_A was positive but non-significant ($r = 0.34$, $p = 0.37$), suggesting offsetting trends (decreasing bulk density and increased THg_M).

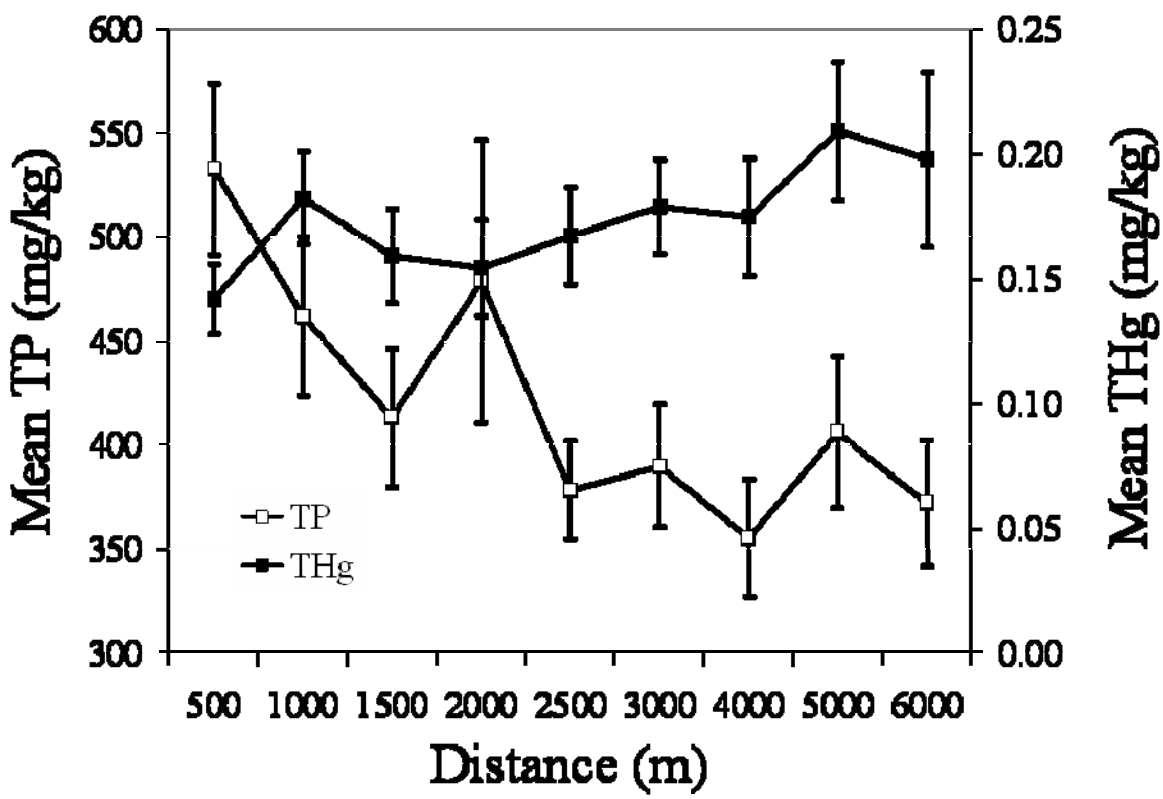

Fig. 58 – Relationship between TP and THg and distance from canals. Slopes of lines are both significantly different from 0 ($p < 0.001$); the slope is positive with distance for THg and negative for TP

3.6.7 Maps of Total Hg (per mass and per area)

Ordinary kriging prediction maps (Fig. 59) show significant hot-spots and strong differences in spatial heterogeneity among sub-regions. The THg_M prediction map shows hotspots in western WCA3AS, northern Holeyland (HL), southern WCA2A and WCA1. Enrichment zones in Shark River Slough and the coastal marshes in western ENP were also observed. The hot spot in western WCA3AS had eight soil samples with THg_M concentrations ranging from 0.646 to 0.917 mg/kg (average 0.774 mg/kg); notably, those same locations had relatively low TP concentrations (325.41-638.96 mg/kg; average 439.51 mg/kg), and THg_M and TP levels in these sites were weakly correlated ($r = 0.16$; $p = 0.58$). The map of THg_A hot spots also shows the regions in WCA3AS and HL, plus a moderate hot spot in WCA3B, but the other hot spots (WCA1, WCA2A) are absent.

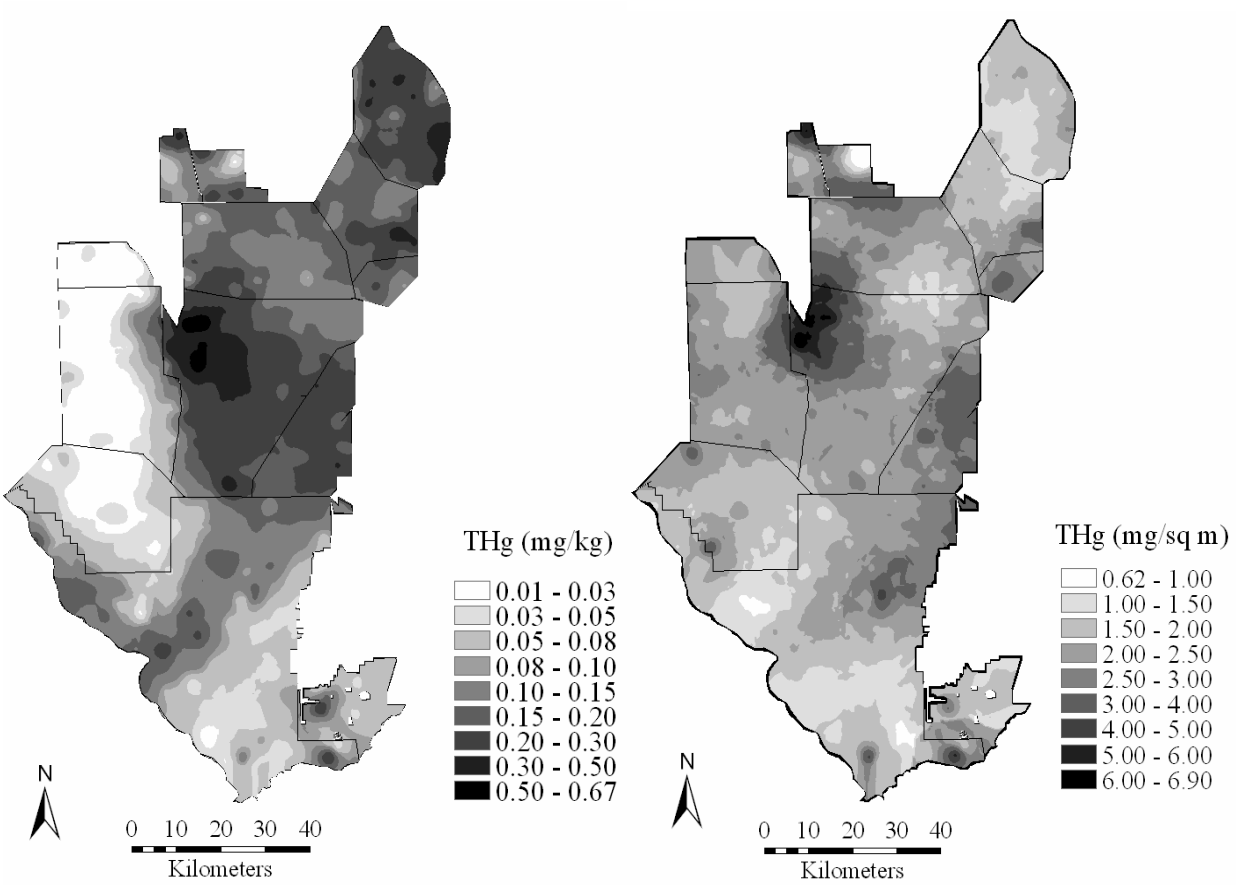

Fig. 59 – Interpolated maps of total Hg (per mass, left and per area, right). Regional hotspots and trends are clear from these maps.

3.7 Gradient Analysis – WCA2A

3.7.1 Spectral Variance with Distance

We plotted spectral variance as a function of distance from canal to generate Fig. 60; this illustrates that there are systematic changes in spectra across a known enrichment gradient. The degree to which that variance can be used for prediction of P enrichment is described below.

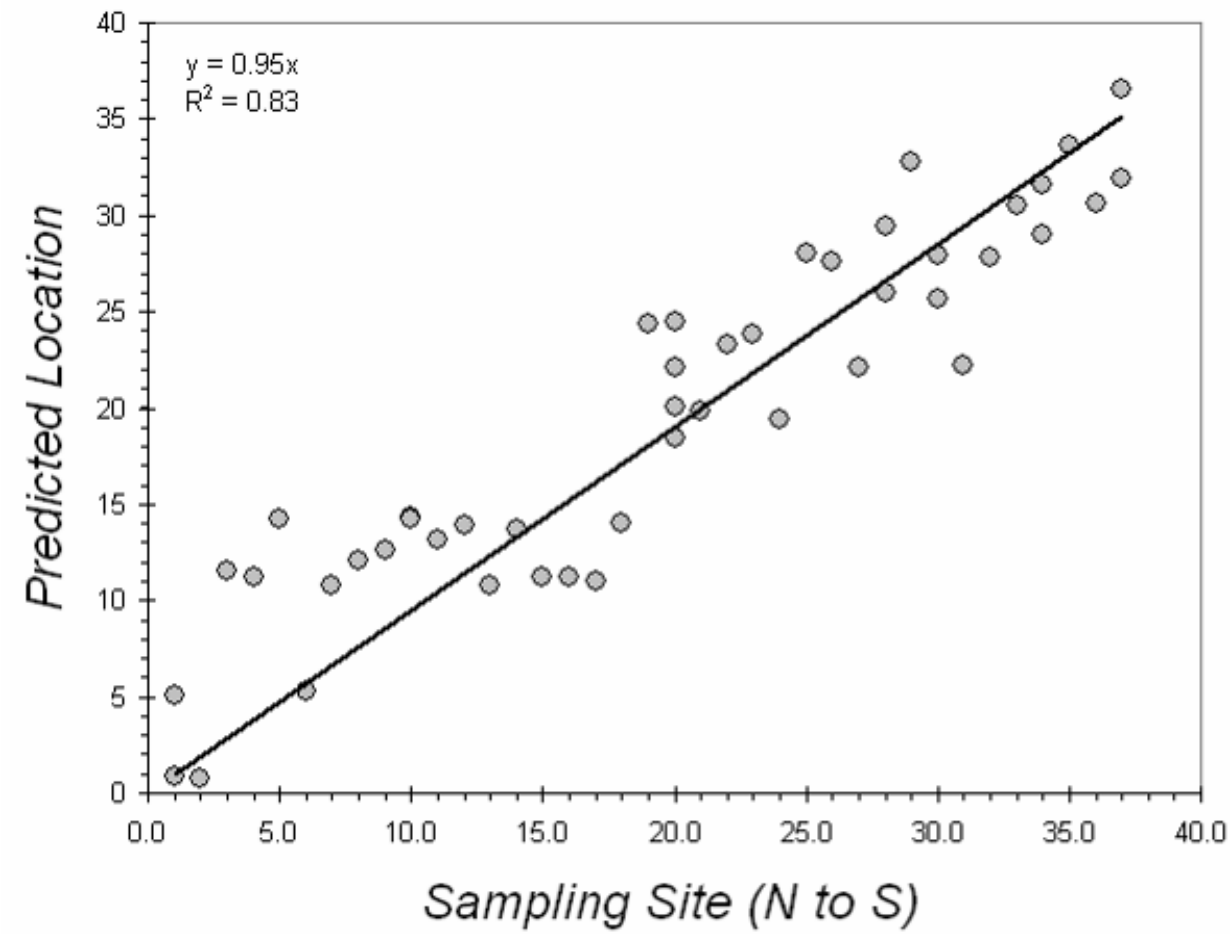

Fig. 60 – Distance from canal (x-axis) vs. variance inferred distance from canal. We infer that there are strong spectral variance gradients present along the sample transect.

3.7.2 Classification Probability

To examine the TP gradient in WCA2A using spectra, we developed a spectral diagnostic model specific to WCA2A. Based on several strongly diagnostic regions (Fig. 61), we developed a binary model that was 87% accurate, with a classification odds ratio of 22.6. This model was applied to the spectra from the WCA2A gradient and the probability of class assignment plotted as a function of distance (from known enriched to uneriched) (Fig. 62).

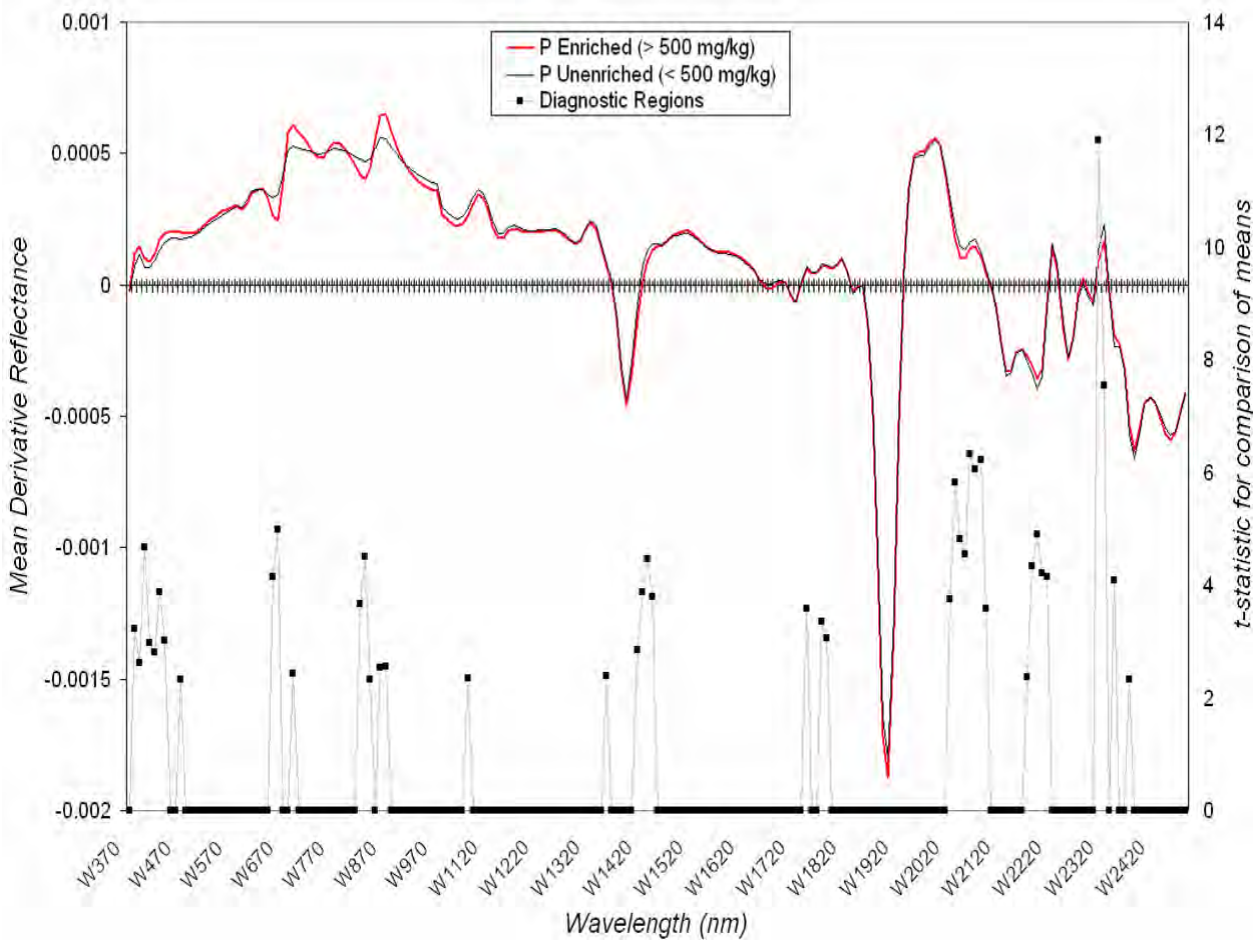

Fig. 61 – Spectral diagnostic regions for discrimination between high and low TP classes. T-statistics for contrast are only illustrative of spectral diagnostics; a multivariate model using all spectra was constructed for class assignment probabilities.

What is most striking about the probability function is that it declines so abruptly. Between 4800 and 5600 m south of the canal (see Fig. 7 for spatial reference), the probability of being in the TP > 500 mg/kg class goes from 0.94 to 0.08. This supports the contention that spectra are changing systematically in response to ecosystem stressor gradients in the Everglades. What should, of necessity, follow from this work is a more explicit spatial and temporal examination of where the spectral change front lies vis-à-vis other indicators of change along this gradient.

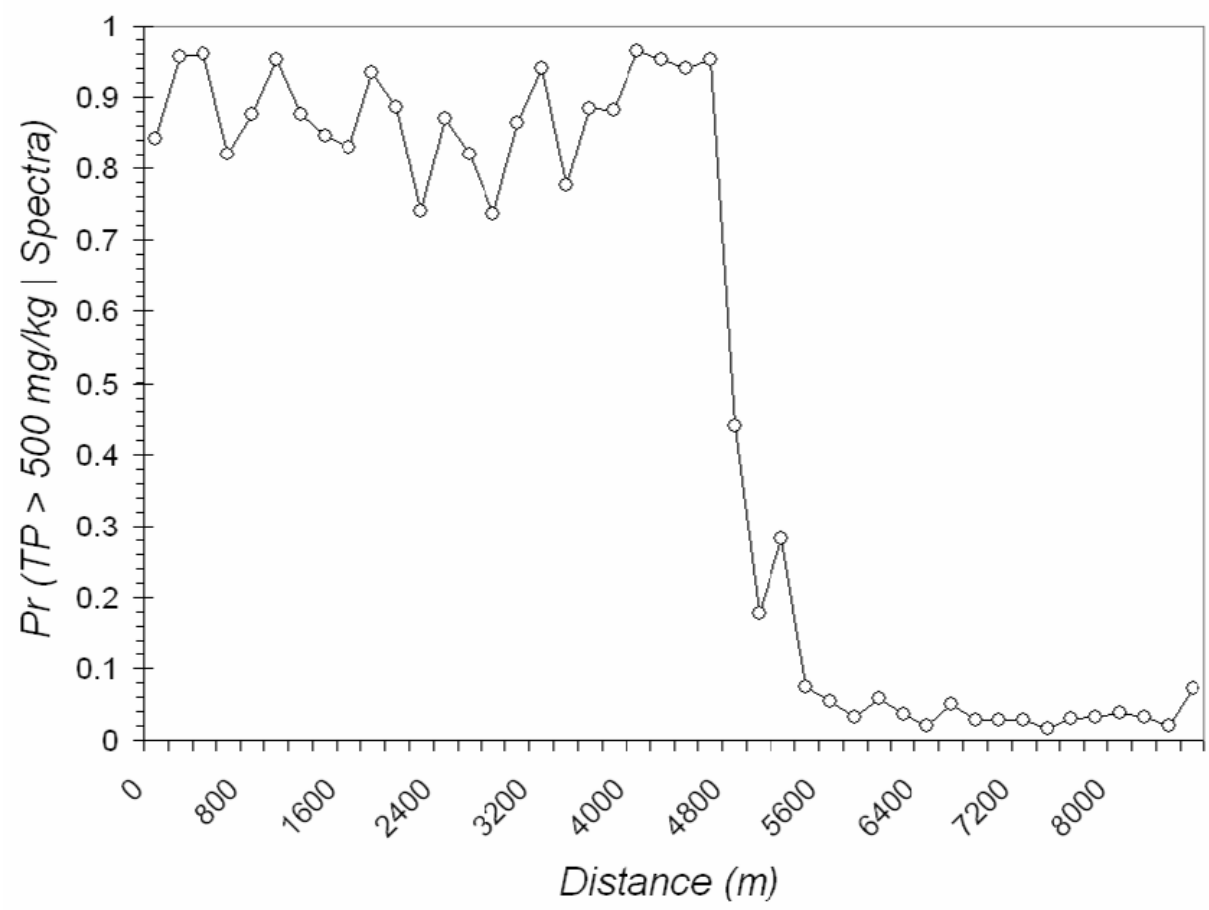

Fig. 62 – Class assignment probability to enriched TP class as a function of distance in WCA2A.

3.8 Unsupervised Classification Results

The results of supervised calibrations are clearly effective; there are numerous parameters for which spectra provide a reasonable estimate of functional attributes. However, there remains the conceptual problem that, while spectra can predict the biogeochemical properties of the soils, this dies not sufficiently inform monitoring in the absence of a case definition. Borrowing from the medical sciences where case definitions are specified and prevelance/incidence statistics are computed, we suggest that what is needed for all large area assessment of condition is a repeatable and adapatable definition of condition. While there are some relatively obvious candidates (e.g., TP concentrations), they alone are not sufficient for multi-stressor systems. Since spectra are integrated snapshots of current and historical soil processes, their intrinsic differences across the landscape may be informative about ecosystem performance. This section

examines techniques for extracting clusters from the spectral data, and then interprets the spectral clusters based on biogeochemistry and geography.

3.8.1 EM Cost Sequence

Unsupervised clustering applied to spectra allows inference of cluster memberships independent of a prior specified objectives; that is clusters self-organize rather than respond to assumed propertied. The drawback of this approach is that there are rarely situations in which the number of clusters is known; using a cross-validation technique, the EM algorithm allows use to select the optimal number of clusters based on the likelihood ratio test, which evaluates the probability of the data given various cluster sets. The cluster set that minimizes the LRT (or deviance) is the most defensible statistical agglomeration of the data. Fig. 63 shows the number of clusters vs. the deviance, and supports the selection of 6 clusters as statistically optimal.

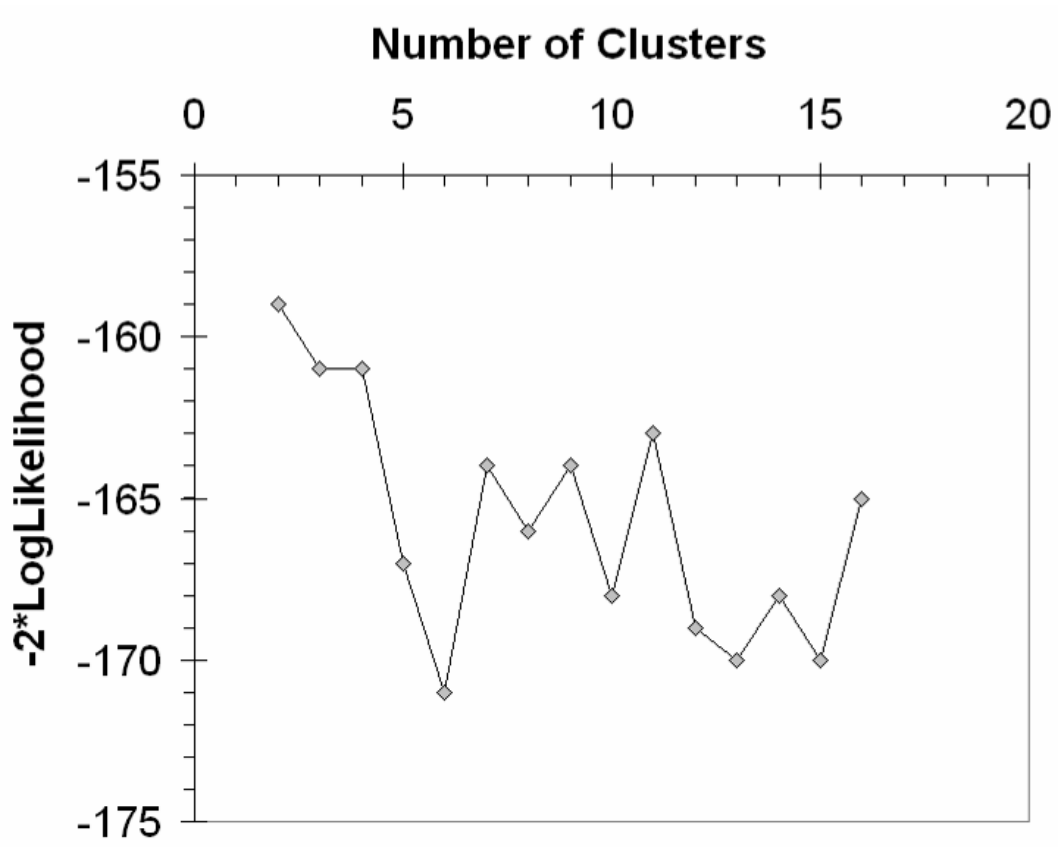

*Fig. 63 – Cost sequence of EM algorithm with different cluster numbers. The maximum likelihood is observed at 6 clusters (deviance, or -2*LogLikelihood, is the cost metric that the EM algorithm minimizes)*

To understand the EM clusters, we began by examining the manner in which known categorical properties sorted according to EM cluster. In general, the clusters were mixtures of community types and areas; clusters 2 and 6 also contained more floc samples than would be expected by chance (Fig. 64).

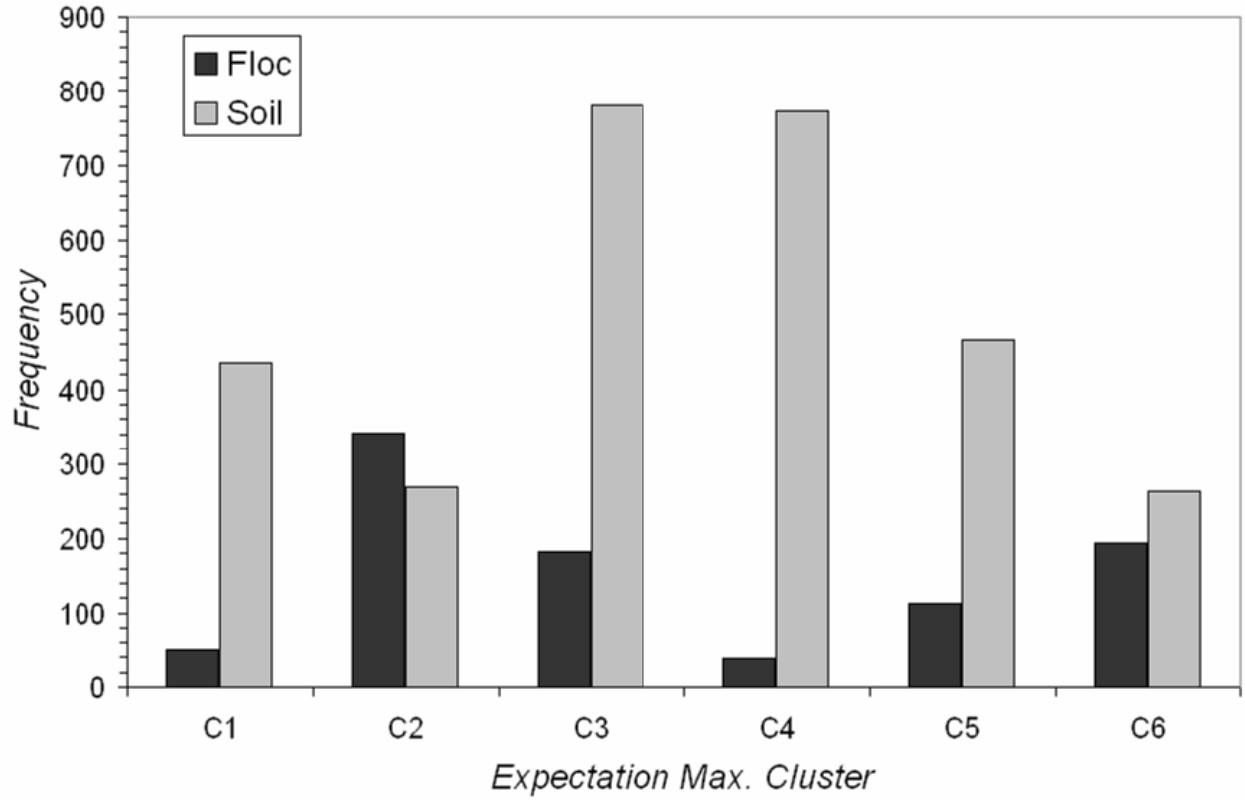

Fig. 64 – Frequency of floc and soil samples by EM cluster.

3.8.2 EM Cluster Geography

In the absence of some conditional assessment to which to compare, we explored the spatial and biogeochemical associations of the EM clusters. Figs. 65 – 67 show the assigned EM classes for floc (profile 1), upper soil (profile 2) and lower soil (profile 3) in space. There are clear agglomerations of cluster assignments, with clusters 1 and 2 corresponding to the calcitic soils of Big Cypress Preserve and the Rocky Glades east of Shark River Slough. Clusters 3 and 4 appear to be peat soils that form under hard-water conditions, while clusters 5 and 6 appear to be soft-water peats in WCA1, WCA2A and WCA3AS. Cluster 6 in particular appears to correspond with areas of high nutrient impact, and Cluster 3 appears to correspond with historically deep water peats along the primary Everglades flowpaths. In general, the concordance between

spectral clusters and ecosystem processes is strong; full exploration of this concordance will require vetting by the scientific community more attuned to the particular character of the Everglades system.

There is also strong concordance between clustering at deep and shallow samples within sites; that is, cluster assignments appear to be relatively consistent across depths, which lends further support to the contention that pure spectral categories may be both robust and useful.

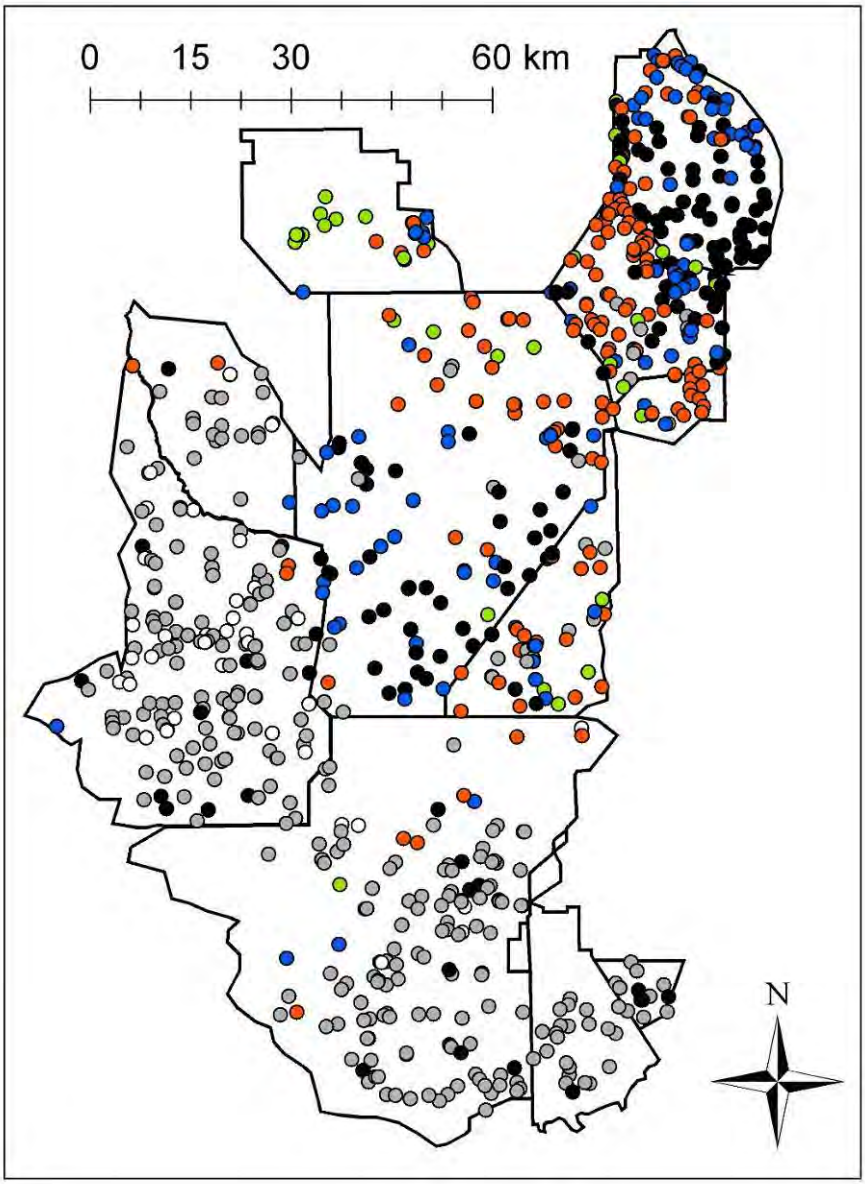

Fig. 65 – EM clusters for floc layer samples.

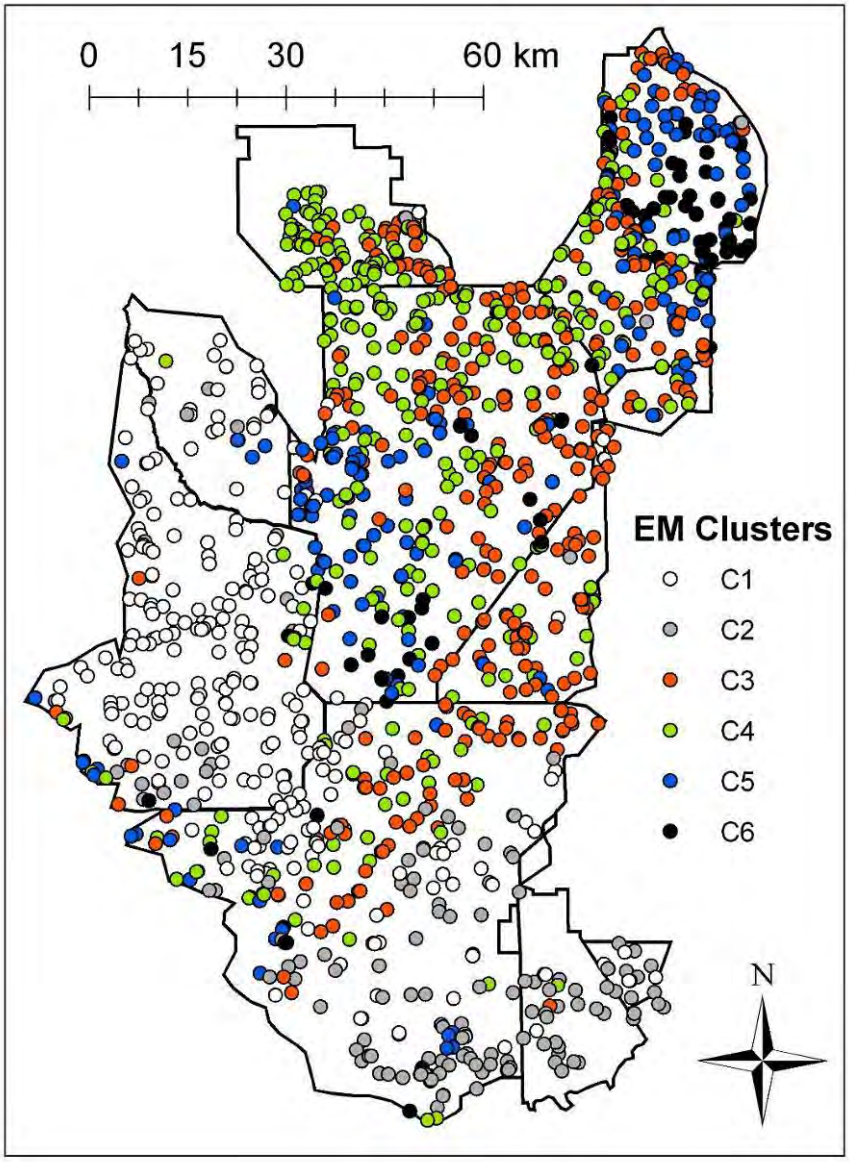

Fig. 66 – EM clusters for profile 2 data (0-10 cm).

Cluster assignments also appear to show some fidelity to particular hydrographic regions. For example, the boundary between WCA1 and WCA2A is obvious based on the abundance of clusters 5 and 6; similarly, the boundary between WCA3AS and WCA3AN is readily apparent based on the prevalence of cluster 5. Because the hydrologic and nutrient forcing varies significantly between these regions, the fact that the EM algorithm is highlighting edaphic variability offers some assurance that this approach could yield significant monitoring benefits.

It is also evident from the maps that there is significant interspersion among categories (e.g., 3 and 4, 5 and 6). We parsed cluster assignments by community type, a local scale confounder of large scale pattern and observed strongly significant differences between vegetative communities using Chi-square tests of independence. Evidence for statistical dependence between clusters and communities was strongly significant ($\chi^2 = 2813$, df = 30, $p << 0.001$). Specifically, we observed that wet prairies are almost exclusively in C1 and C3, ridge communities are dominant in C2 and C6, and sloughs dominant in C5. Mangrove fringe systems, of which there were only 112 observations, were exclusively (91%) assigned to C3.

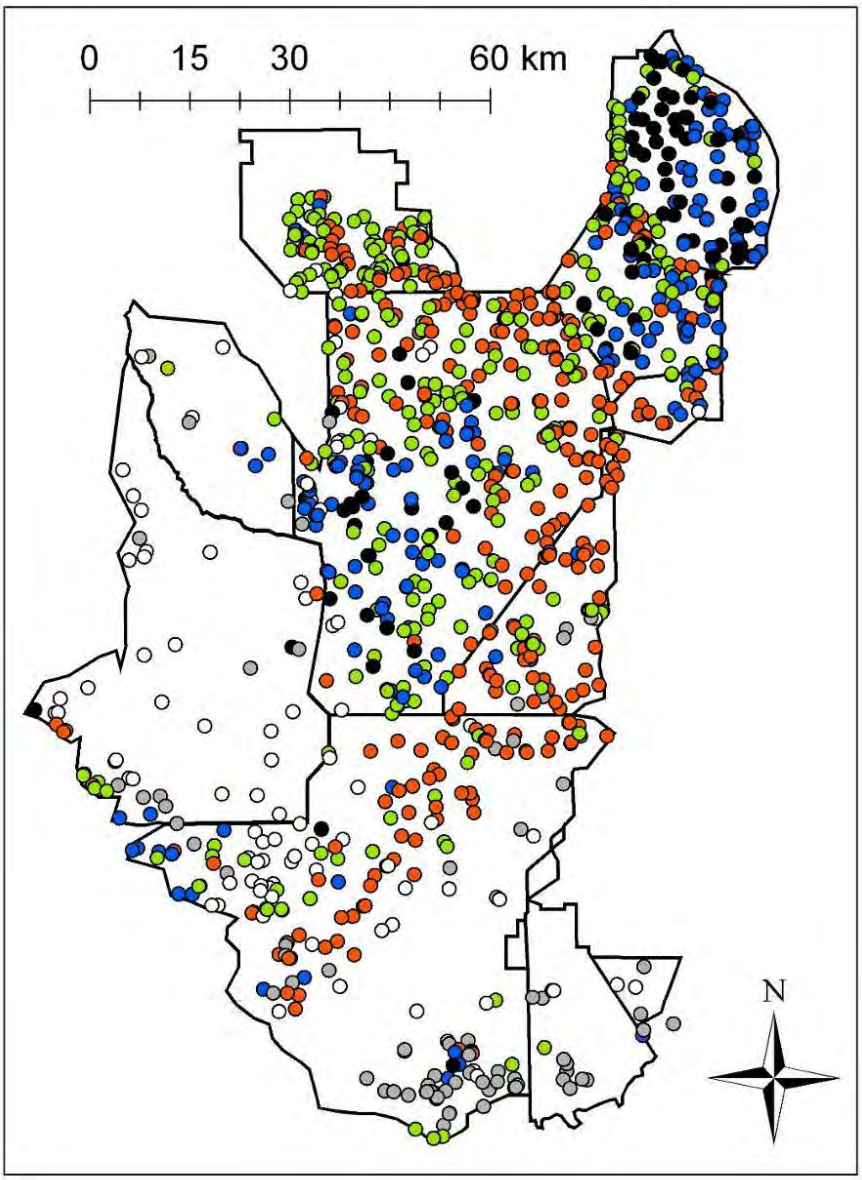

Fig. 67 – EM clusters for profile 3 samples (10 – 20 cm)

3.8.3 EM Cluster Soil Properties

EM clusters were also evaluated based on comparative chemical and physical properties. Fig. 68 summarizes the means and standard deviations of the 6 selected EM clusters for the 10 parameters measured in the ESM dataset. All clusters were significantly different ($p < 0.01$) from all other clusters except where noted.

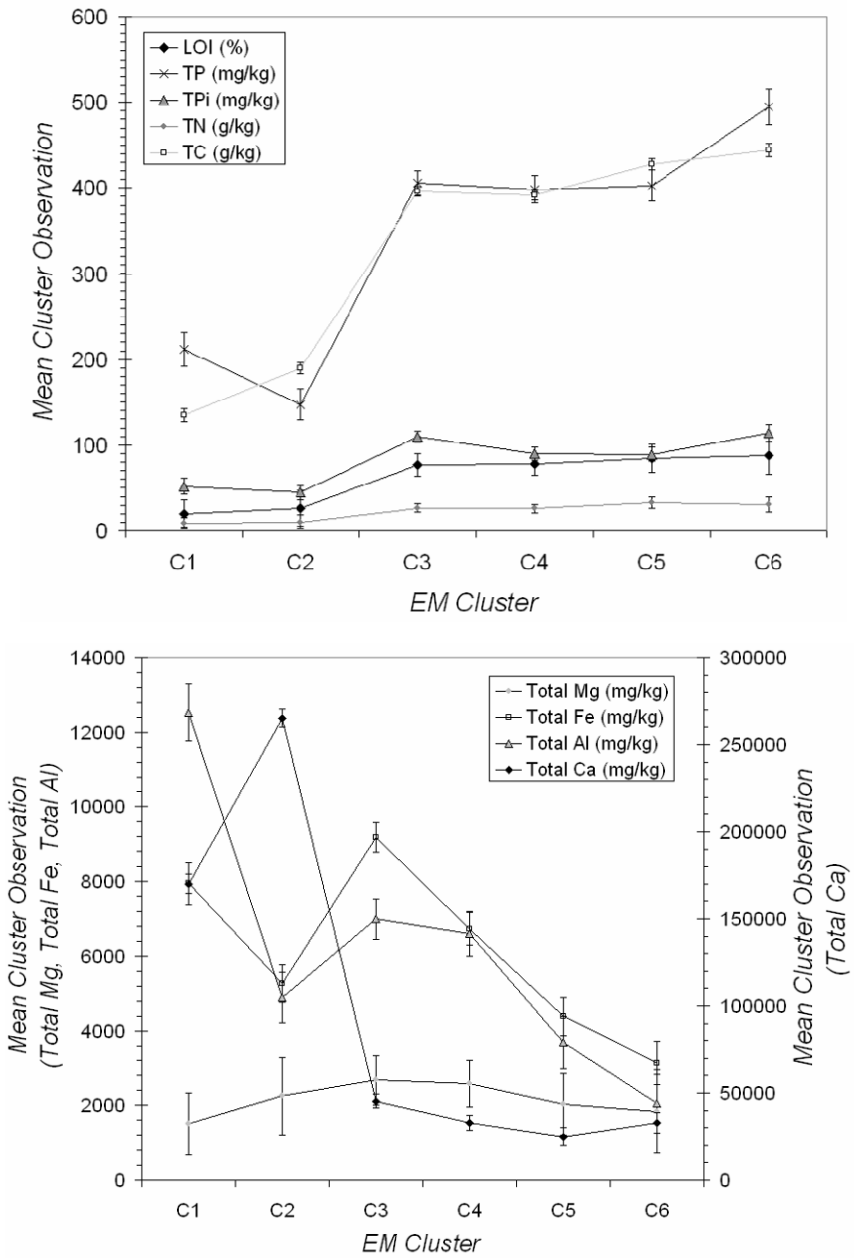

Fig. 68 – Summary of EM cluster chemical and physical properties. C1 and C2 were not significantly different in LOI or TN; C3 and C4 were not significantly different for TC, TP, TAl, or TN. None of the clusters differed in TMg. All other pairwise contrasts were significantly different at $p < 0.01$.

4. DISCUSSION

4.1 Feasibility of Sensing Soil Quality in the Everglades

This report has documented investigations of NIR spectroscopy for the analysis of soils across the Greater Everglades. Overall, the results suggest that the method is highly feasible, and worth pursuing as a component of ecosystem scale diagnostic surveillance. Among the advantages, clearly articulated in the literature, that we reinforce in this work are high accuracy (frequently rivaling the expected accuracy for laboratory measurements), low cost (discussed in more detail below), rapid throughput (~ 200 samples per day for 1 instrument) and high precision. Moreover, we note that there are no chemical wastes generated and little technician training required. The overall implication is that process and monitoring measurements can be collected at much greater spatial and temporal density than was previously possible, improving the statistical power of ecosystem change detection.

We note that the recent development of long-term study plots (T. Phillippi, *unpublished*) will require regular high-resolution monitoring of the myriad performance measures; as budgets constrain the sampling of soils and plant tissues, NIRS becomes a useful tool. Moreover, our work on calibration free methods for the interpretation of spectra are perhaps most germane for regular monitoring of local areas. Spectral category transitions, and the ecological implications thereof, have been shown in the literature (Vagen et al. 2006) to be a more effective covariate of environmental change than basic soil properties. Further work in this regard is clearly needed, but our work underscores the potential of this line of inquiry.

The results in this work also underscore the flexibility of the spectral method. While the concept has been widely criticized for environmental samples, with critics arguing that the only spectral response is due to organic matter, and all calibrations are responding to covariance with OM, we show more stable and accurate predictions than could be predicted by covariance with OM. As such, the method appears to be applicable to a wide range of analytes, including functional analytes not previously demonstrated in the literature (e.g., residual fiber content of peat soils). We reiterate that the list of analytes for which spectral methods have been demonstrated continues to expand, and with each analyte that can successfully be predicted, the cost and time advantages grow.

There are also significant advantages beyond cost: spectral methods offer a means to operationalize and standardize categorical definitions. For example, the designation of soil vs.

floc, or intact vs. degraded, or ridge vs. slough is based on observer judgments. While in most cases this is adequate, there is strong evidence from other systems that observers frequently disagree, and often with important consequences for management. Because spectra are both integrative and precise, definitions arising from their interpretation are likewise integrative and precise. Cohen et al. (2005) used this to significant advantage in the diagnosis of soil degradation across landscapes, observing that observer definitions of cases were neither uniform in time, nor among observers.

An additional observation from this work that is critically important for further exploration is the ability to use classification probabilities to examine ecological gradients. We observe an abrupt change in the classification probability to low TP (< 500 mg/kg) along the P enrichment gradient in WCA2A. This change is far more abrupt than the change in P, and the change occurs further along the gradient than many of the other metrics of ecosystem change (e.g., cattail dominance) suggesting that the spectra may be sensing edaphic changes early in the transition process. Further research on these kinds of gradients are well warranted given their potential for monitoring.

A key insight from the description of where and how spectral errors arise is also suggestive of additional research. In general, the NIR literature is moving increasingly towards global models; that is, spectral libraries have been assembled that are representative of huge geographic areas and applicable to a wide range of soil types. This is clearly attractive from the perspective of centralized laboratories. Following that trend, we developed a global library for the Everglades based on comprehensive sampling. While our chemometrics are highly effective, both vis-à-vis analytical labs and other applications of the spectral methods, they appear to have strong spatial structure to their error. We explored this using geostatistical methods and observed high levels of relative structure and particular hot-spots of over- and under-estimation that are clearly not spatially independent. As such, local level calibrations may be more effective because they allow the statistical algorithms to identify the locally optimal bands, not the globally optimal. This demands a larger set of library samples, or the general inclusion of geographic information into the statistical model. We hasten to point out here that while there are spatially dependent errors, the error rates are, in most cases, quite small. The marginal value of reducing these errors needs to be weighed against the additional costs and logistics of partitioned prediction models.

One of the most important considerations for applications of spectroscopy for routine analysis is the concordance of measured reflectance signatures between instruments. Castro-Esau et al. (2006) demonstrate significant and systematic errors in a between-instrument comparison, particularly where optical set-up varies. This significant error presents the significant logistical constraint that calibrations developed using date from one instrument are not useful for predicting properties of samples scanned using a second instrument until a conversion algorithm can be identified. If the variability between sensors is due to non-systematic variability, then the problem is compounded further. For the purposes of routine application of spectral soil predictions in the Greater Everglades, there are several immediate options. The first is to use the instrument used to develop these calibrations to scan all subsequent soils for which spectral prediction is required. This is not an insurmountable task; as discussed, this method permits greatly enhanced sample throughput, on the order of 200+ samples per day. Predictions can be done in real time, and data storage has been greatly facilitated by our efforts to develop a data ingestion and storage system that not only compiles the data in a scalable and web-accessible database, but also permits numerous user-selected computations for improving data quality and analysis. With this rate of sample throughput, a single spectrometer can reasonably be expected to process somewhere between 25,000 and 50,000 samples per year, which should be adequate for routine monitoring needs of CERP and beyond.

The second is to develop inter-instrument calibrations so that spectra from one instrument can be reliably mapped to the spectral that would have been obtained on another instrument. This is logistically problematic, but is of considerable value for the purpose of field sampling. While we have scarcely discussed this option in this report, the long-term potential for spectroscopy will be fully realized as a field deployment technique. Current impediments to collecting spectra and quantitatively inferring soil properties in the field (that is, without preprocessing) are the confounding spectral effects of water and stray light. Sample vestibules can reasonably be expected to prevent stray light effects, and our group has been developing algorithms to permit removal of water effects that may soon be field operable. The possibility of a field deployment of the spectrometers might necessitate two or more spectrometers because of the potential risks of field observations, and the need for multiple field crews. However, inter-

instrument variability is primarily an engineering hurdle, and immediate implementation of spectroscopy for routine analysis of soils (and indeed plant tissues and possibly water) is possible because our analysis concludes viable prediction for a wide array of soil analytes.

A third option, emerging from recent literature on MIR interferometric spectroscopy is a move from the NIR region to sensors with massive sample throughput (~600 per day) operating in the more stable MIR region. These sensors, which have been available for analytical chemistry for decades but only recently used for environmental sample analysis, are far more stable within and between instruments because rather that depending on a white reference for reflectance calibrations, they use interferometry to relate diffuse or total attenuated reflectance to the characteristics of the light source. As the scope of spectral methods grows, these instruments are likely to be the work-horses of future large area monitoring efforts. There cost is competitive (or lower) than NIR instruments. Principal among their drawbacks is that they are not field portable.

4.2 Comparative Cost Analysis

One of the objectives of this work was to determine if spectral methods provide a reasonable alternative to full chemical characterization when dealing with the large number of samples necessary for ecosystem characterization. This question has two parts: the first concerns the analytical accuracy of the methods. The discussion above demonstrates that the NIR method can provide highly reliable results for the suite of analytes selected, often providing sufficient accuracy to meet stringent laboratory performance requirements. Since instrument precision is typically below 1% error between sample replicates, these results underscore the strong potential for NIR to be part of the large area assessment toolkit.

The second part of the methodological evaluation involves the reduction in costs and sample processing time between spectral methods and routine laboratory analyses. Table 17 summarizes this information for comparison. We estimate, based on our in-house costs of sample analyses, that spectral analysis is an order of magnitude less expensive than conventional methods; this contrast is further underscored by pointing out two additional considerations. First, cost savings magnify with the number of analytes; where spectral prediction of more expensive analytes is demonstrated (e.g., extracellular enzymatic activity and carbon quality), then the cost comparison is further enhanced. Second, spectral prediction is made on soils that

are processed and handled in a standard manner. Where soils are archived in this standard manner (dried and in air-tight containers), it may be possible to predict soil properties on old samples after some significant period of storage, provided that an contemporary library exists.

Table 17. Comparison of analytical costs between spectral analysis and conventional methods.

Conventional Method		Spectral Method	
Analyte	$$ Per Sample	Analyte	$$ Per Sample
Soil Pre-Processing	5	Soil Pre-Processing	5
Organic Matter (LOI)	3	Scanning	1
Total Nitrogen	6	Data Processing	1
Total Carbon	6		
Bulk Density	1		
Total P	13		
Total Inorganic P	11		
Total Ca	8		
Total Mg	8		
Total Fe	8		
Total Al	8		
Total	77	Total	7
** Est. Processing Time = 90 days		** Est. processing time = 10 days	

Including the costs of fiber analysis ($30/sample) and mercury analysis ($15/sample) increases the relative benefit of spectral prediction. With each additional analyte, this cost difference grows, as do the labor and time requirements for measurement. Data from the literature supports application of NIRS to both soils and plant tissues, to indicators of function ranging from C3 vs. C4 soil OM inputs, metal concentrations, mineralogy, and soil aggregation capacity. While not all of these are relevant to the Everglades, some may be, and others, yet unidentified may prove to be effectively predicted using these techniques.

4.3 Integrating Soil Sensing into Large Area Assessment

Adoption of new technologies is frequently a slow process, for obvious reasons. The vetting of scientific data and the independent testing of emerging methods should be a tentative and cautious process. In the case of NIR spectroscopy, the body of literature on the topic has swelled in recent decades as more researchers are demonstrating feasibility in numerous environments and for numerous analytes. Our suggestion for operational use in the Everglades is to begin with parallel measurement systems; that is, do both laboratory standard methods and spectral prediction and validate (and update) the models. One way to make optimal use of the

technique is to oversample the landscape, collecting far more samples than can be reasonably be processed in the lab. Do routine lab analysis on a subset and spectral methods on all. The additional power that emerges from high density spatial sampling can aid in process or monitoring inference, and the method can simultaneously be validated and offer new insights.

A project currently starting through the MAP/Recover program on mapping soil nutrients will use this technique. We will densely oversample in all the study sites and use this oversampled data to both validate the method and expand the footprint and resolution of our study.

The described application is useful for demonstration; however, to really incorporate NIRS into Greater Everglades ecosystem assessment will require attention be paid to the issues that this report has raised. First, how will we deal with temporal and between instrument variability; while this issue is likely far more problematic for analytical laboratories than for this high-precision method, it is still an important component of uncertainty to quantify. Second, local level calibrations and the scope of reference libraries (both in space and in analyte diversity) need to be considered. Finally, the use of calibration-free and gradient-based methods should be explored in substantially more detail.

5. CONCLUSIONS AND RECOMMENDATIONS

This report has demonstrated the utility of NIR spectral methods for the prediction of numerous soil attributes. Principal among our conclusions is that NIR represents a potentially important component of the large area surveillance toolbox. As CERP continues to make changes in the local and regional environmental system, robust (i.e., spatially and temporally dense) sampling of ecosystem response is critical for providing useful management feedback. NIRS offers a mechanism to do this for relatively low cost. While it is important to emphasize the cost differences between the two methods since they are so pronounced, we reiterate that there are compelling reasons to adopt NIRS as a tool even without those savings. Among our arguments is that the sensors are highly precise, so errors (generally large and unquantified) between laboratories and over time can be minimized. More importantly, spectra appear to be a robust integrator of the soil environment, responding to multiple stressors is relatively predictable ways. As such, direct calibration techniques and unsupervised methods offer substantial promise for objective characterization of ecosystem status and trends. Finally, we suggest that while all of the efforts herein were for samples returned to the laboratory (making sampling effort the principal constraint), recent developments in our lab support the use of the spectrometer used here for field campaigns. That is, scans under variable moisture conditions may approach the same level of predictability, meaning that the sensor could be brought in the field, and nearly 80% of per sample costs (pre-processing) could be eliminated.

Other areas of recommended follow-up based on the results we obtained are listed below:

1) *Characterization of ecological change gradients using spectra.* There is some preliminary evidence presented in this report that spectra may be useful for mapping change fronts (e.g., TP enrichment). These methods require further and more detailed explication.
2) *Local spectral library development.* There is strong evidence from this work to suggest that prediction errors are not geographically random. That is, soil processes at the local scale confound prediction to a moderate extent, suggesting that locally developed calibrations may be more accurate and robust. Further research is required to substantiate this.
3) *Operational accuracies.* As with all models of this type, the observed validation accuracies may be an overestimate of operational accuracies because

of between instrument and between time errors in spectral reflectance. Efforts to quantify and control for these errors are paramount for routine implementation. Newer sensors are capable of much greater between instrument precision, which may be one fruitful area of exploration.

4) *Calibration free methods.* Ecosystem surveillance requires performance measures that are direct. While soils in the Everglades represent the total integration of ecosystem processes, the particular indicators used as performance measures remains uncertain. A direct approach, whereby spectra are calibrated to particular condition classes and/or stressor gradients is likely to prove extremely useful based on preliminary evidence presented here.

5) *Additional soil properties and extension to other substrates*. The success of the method for soils in the Everglades can measured in units of accuracy, but also in units of cost. We estimate that the comparative costs of the routine analyses and spectral methods differ by an order of magnitude ($77 vs. $7/sample). Moreover, the cost implications become more pronounced as more analytes are predicted. Given success with indicators of C quality, metal content, and ecological stoichiometry, we recommend the inclusion of additional informative analytes to the spectral library. Further, we recommend an extension of these methods, applied only to soils in this work, to plant tissues as well. Existing libraries of samples have been identified.

6. REFERENCES CITED

Agresti, A. 1990. *Categorical Data Analysis.* Wiley-Interscience, New York, NY

Anselin, L. 1995. *Local indicators of spatial association.* Geographical Analysis 27: 93-115.

Arfstrom, C., A. W. Macfarlane, and R. D. Jones. 2000. *Distributions of mercury and phosphorus in everglades soils from water conservation area 3A, Florida, USA.* Water, Air, and Soil Pollution 121: 133-159.

Beebe, K.R., Pell, R.J., & Seasholtz, R.B. 1998. *Chemometrics: A Practical Guide.* Wiley-Interscience, New York, NY

Ben-Dor, E. & Banin, A. 1995. *Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties*. Soil Sci. Soc. Am. J. **59**, 364–372.

Bouchard, V., Gillon, D., Joffre, R. & Lefeuvre, J.C. 2002. *Actual litter decomposition rates in salt marshes measured using near-infrared reflectance*. J. Exper. Mar. Biol. Ecol. **290**, 149-163

Breiman, L., J. Friedman, C.J. Stone, & R.A. Olshen. (1984). *Classification and regression trees*. Chapman & Hall, CRC Press, Boca Raton, FL.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D. & Reinsch, T.G. 2006. *Global soil characterization with VNIR diffuse reflectance spectroscopy*. Geoderma **132**, 274-290

Bruland, G. L., S. Grunwald, T. Z. Osborne, K. R. Reddy, and S. Newman. 2006. *Spatial distribution of soil properties in Water Conservation Area 3 of the Everglades.* Soil Sci. Soc. Am. J. 70:1662-1676.

Chang, C.W., Laird, D.A., Mausbach, M.J., & Hurburgh, C.R. 2001. *Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties.* Soil Sci. Soc. Am. J. **65**, 480-490

Cohen, M.J., K.D. Shepherd and M.G. Walsh. 2004. *Empirical Reformulation of the Universal Soil Loss Equation for Erosion Risk Assessment in a Tropical Watershed.* Geoderma 124:235-252

Cohen, M.J, Prenger, J.P. & DeBusk, W.F. 2005. *Visible-Near Infrared Spectroscopy for Rapid, Non-Destructive Assessment of Wetland Soil Quality*. Journal of Environmental Quality **34**, 1422-1434

Cohen, M.J., S. Dabral, W.D. Graham, J.P. Prenger and W.F. Debusk. 2006. *Evaluating Ecological Condition using Soil Biogeochemical Parameters and Near Infrared Reflectance Spectra.* Environmental Monitoring and Assessment 116:427-457

Cohen, M.J., R. Mylavarapu, I. Bogrecki and D. Lee. 2007. *Testing Near Infrared Spectroscopy for Routine Agronomic Soil Analyses.* Soil Science 172:469-485

Corstanje, R., S. Grunwald, K. R. Reddy, T. Z. Osborne, and S. Newman. 2006. *Assessment of the spatial distribution of soil properties in a northern everglades marsh.* J of Environ Qual. 35: 938-949.

Cozzolino D. & Moron A. 2003. *The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics*. J. Agric. Sci. (Cambridge) **14,** :65-71

Craft, C. B., and C. J. Richardson. 1993. *Peat accretion and N, P, and organic C accumulation in nutrient-enriched and unenriched everglades peatlands.*

DeBusk, W. F., S. Newman, and K. R. Reddy. 2001. *Spatio-temporal patterns of soil phosphorus enrichment in Everglades Water Conservation Area 2A.* J. Environ. Qual. 30: 1439-1446.

Dunn, B.W., Batten, G.D. & Ciavarella, S. 2002. *The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australia*. Aust. J. Exp. Agric. **42**, 607-614

Dvonch, J.T., et al. 1999. *Use of Elemental Tracers to Source Apportion Mercury in South Florida Precipitation.* Environ. Sci. Tech. 33:4522-4527

Fearn, T. 2000. *Savitzky-Golay filters*. NIR News **6**, 14-15

Friedman, J.H., 2002. *Stochastic gradient boosting*. Comput. Stat. Data Anal. **38**, 367–378.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation.* Oxford Univ. Press, New York, NY.

Guentzel, J. L., W. M. Landing, G. A. Gill, and C. D. Pollman. 1995. *Atmospheric deposition of mercury in Florida: the FAMS project.* Water, Air and Soil Pollution 80: 393-402.

Janik, L.J., Merry, R.H. & Skjemstad, J.O. 1998. *Can mid-infrared diffuse reflectance analysis replace soil extractions?* Australian Journal of Experimental Agriculture **38**, 681-696

Johnston, C.T. & Aiochi, Y.O. 1996. Fourier Transform Infrared and Raman Spectroscopy. *In Methods of Soil Analysis Part 3: Chemical Methods* (ed. J.M. Bigham). SSSA-ASA, Madison, Wisconsin

Kooistra, L., Wehrens, R., Leuven, R.S. & Buydnes, L.M.C. 1997. *Possibilities of visible-near infrared spectroscopy for the assessment of soil contamination in river floodplains*. Anal. Chem. **446**, 97-105.

Malley, D.F., Lockhart, L., Wilkinson, P., & Hauser, B. 2000. *Determination of carbon, carbonate, nitrogen and phosphorus in freshwater sediments by near-infrared reflectance spectroscopy: rapid analysis and a check on conventional analytical methods.* Journal of Paleolimnology **24**, 415-425

McBratney, A. B., and R. Webster. 1986. *Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates*. J. Soil Science 37: 617-639.

McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. *Mid-infrared and near infrared diffuse reflectance spectroscopy for soil carbon measurement*. Soil Sci. Soc. Am. J. **66**, 640–646.

National Atmospheric Deposition Program (NRSP-3). 2007. NADP Program Office, Illinois State Water Survey, 2204 Griffith Dr., Champaign, IL 61820.

Noe, G. B., D. L. Childers, and R. D. Jones. 2001. *Phosphorus biogeochemistry and the impact of phosphorus enrichment: Why is the Everglades so unique?* Ecosystems 4: 603-624.

Pannatier, Y. 1996. *Variowin: software for spatial data analysis in 2D.* Springer-Verlag, New York, NY.

Reddy, K.R., W. F. DeBusk, Y. Wang, R. DeLaune, and M. Koch. 1991. *Physico-chemical properties of soils in the Water Conservation Area 2 of the Everglades* – Final Report, Soil and Water Science Dept., University of Florida, Gainesville, 214 pp.

Reeves, J.B., McCarty, G.W. & Meisinger, J.J. 1999. *Near-infrared reflectance spectroscopy for the analysis of agricultural soils*. J. Near Inf. Spec. **7**, 179-193

Rood, B. E., J. F. Gottgens, J. J. Delfino, C. D. Earle, T. L. Crisman. 1995. *Mercury accumulation trends in Florida Everglades and Savannas Marsh Flooded Soils.* Water, Air, and Soil Pollution 80: 981-990.

Shepherd, K.D. & Walsh, M.G. 2002. *Development of reflectance spectral libraries for characterization of soil properties.* Soil Sci. Soc. Am. J. **66**, 988-998

Shepherd, K.D. and M.G. Walsh. 2007. *Review: Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries.* Journal of Near Infrared Spectroscopy 15:1-20

Stober, Q.J., K. Thornton, R. Jones, J. Richards, C. Ivey, R. Welch, M. Madden, J. Trexler, E. Gaiser, D. Schmidt and S. Rathbun. 2001. *South Florida Ecosystem Assessment: Phase I/II – Everglades Stressor Interactions: Hydropatterns, Eutrophication, Habitat Alteration and Mercury Contamination.* USEPA Report #904-R-01-003

StatSoft, Inc. 2005. STATISTICA (data analysis software system), version 7.1. www.statsoft.com.

Urban, N. H., S. M. Davis, N. G. Aumen. 1993. *Fluctuations in sawgrass and cattail densities in Everglades Water Conservation Area 2A under varying nutrient, hydrologic and fire regimes.* Aquatic Botany 46: 203-223.

Vagen, T., M.G. Walsh and K.D. Shepherd. 2006. *Stable isotopes for characterization of trends in soil carbon following deforestation and land use change in the highlands of Madagascar*. Geoderma 135:133-139

Vagen, T., M.G. Walsh and K.D. Shepherd. 2006. *Sensing landscape level change in soil fertility following deforestation and coversion in the highlands of Madagascar using Vis-NIR spectroscopy.* Geoderma 133:281-294

Vaithiyanathan, P, C.J. Richardson, R.G. Kavanaugh, C.B. Craft and T. Barkay. 1996. *Relationships of Eutrophication to the Distribution of Mercury and the Potential for Methylmercury Production in the Peat Soils of the Everglades.* Environ. Sci. Tech. 30:2591-2597

Webster, R., and M. A. Oliver. 1992. *Sample adequately to estimate variograms of soil properties.* J. of Soil Science 43: 177-192.

Webster, R. and M. A. Oliver. 2001. *Geostatistics for Environmental Scientists*. John Wiley and Sons, Ltd. Chichester.

Witten, I., H., & Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques.* New York: Morgan Kaufmann.

7. DATA AND APPENDICES

[Provided on CD-ROM]