USING SATELLITE HYPERSPECTRAL IMAGERY TO MAP SOIL ORGANIC MATTER, TOTAL NITROGEN AND TOTAL PHOSPHORUS

Baojuan Zheng

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Earth Sciences,
Indiana University

August 2008

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Lin Li, Ph.D., Chair

Pierre-Andre Jacinthe, Ph.D.

Master's Thesis
Committee

Gabriel M. Filippelli, Ph.D.

ACKNOWLEDGEMENTS

I wish to thank my thesis committee members, Lin Li, Gabriel Filippelli and Pierre-Andre Jacinthe for their endless guidance, support and inspiration. Especially, I wish to thank my advisor, Lin Li for providing invaluable insight into my work.

I would like to thank all the graduate students, staff and faculty in the Department of Earth Sciences at IUPUI for encouragement and assistance. Special appreciation is expressed to Xingtang Hu for participating in the field campaign and assisting in the laboratory, and to Kenny Brown for helping me with my coursework.

I wish to thank my family and friends for their unconditional support whenever I encounter difficulties. I am grateful to my parents for everything they have done for me.

Finally, thanks to IUPUI Research Support Funds Grant (RSFG) for funding support.

ABSTRACT

Baojuan Zheng

USING SATELLITE HYPERSPECTRAL IMAGERY TO MAP SOIL ORGANIC MATTER, TOTAL NITROGEN AND TOTAL PHOSPHORUS

Up-to-date and accurate information on soil properties is important for precision farming and environmental management. The spatial information of soil properties allows adjustments of fertilizer applications to be made based on knowledge of local field conditions, thereby maximizing agricultural productivity and minimizing the risk of environmental pollution. While conventional soil sampling procedures are labor-intensive, time-consuming and expensive, remote sensing techniques provide a rapid and efficient tool for mapping soil properties. This study aimed at examining the capacity of hyperspectral reflectance data for mapping soil organic matter (SOM), total nitrogen (N) and total phosphorus (P). Soil samples collected from Eagle Creek Watershed, Cicero Creek Watershed, and Fall Creek Watershed were analyzed for organic matter content, total N and total P; their corresponding spectral reflectance was measured in the laboratory before and after oven drying and in the field using Analytical Spectral Devices spectrometer. Hyperion images for each of the watersheds were acquired, calibrated and corrected and Hyperion image spectra for individual sampled sites were extracted. These hyperspectral reflectance data were related to SOM, total N and total P concentration through partial least squares (PLS) regressions.

The samples were split into two datasets: one for calibration, and the other for validation. High PLS performance was observed during the calibration for SOM and total N regardless of the type of the reflectance spectra, and for total P with Hyperion

image spectra. The validation of PLS models was carried out with each type of reflectance to assess their predictive power. For laboratory reflectance spectra, PLS models of SOM and total N resulted in higher R^2 values and lower RMSEP with oven-dried than those with field-moist soils. The results demonstrate that soil moisture degrades the performance of PLS in estimating soil constituents with spectral reflectance. For *in-situ* field spectra, PLS estimated SOM with an R^2 of 0.74, N with an R^2 of 0.79, and P with an R^2 of 0.60. For Hyperion image spectra, PLS predictive models yielded an R^2 of 0.74 between measured and predicted SOM, an R^2 of 0.72 between measured and predicted total N, and an R^2 of 0.67 between measured and predicted total P. These results reveal slightly decreased model performance when shifting from laboratory-measured spectra to satellite image spectra. Regardless of the spectral data, the models for estimating SOM and total N consistently outperformed those for estimating total P. These results also indicate that PLS is an effective tool for remotely estimating SOM, total N and P in agricultural soils, but more research is needed to improve the predictive power of the model when applied to satellite hyperspectral imagery.

Lin Li, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Non-point Source Pollution

Nonpoint source (NPS) pollution is caused by rainfall or snowmelt moving over and through the ground. Sources of nonpoint pollution include runoff from farms, pasture, range, abandoned mines, construction sites and so forth. Pollutants can be carried away and finally deposited into rivers and lakes as water moves from upstream to downstream. Agricultural NPS was identified as the leading source of water quality impairment in assessed streams, rivers, lakes and reservoirs in the United States, and NPS is one of the top probable sources of impairments for rivers, streams, lakes and reservoirs in the State of Indiana (2004 National Assessment Data).

Eagle Creek, Morse and Geist Reservoirs are important part of Indianapolis drinking water system. However, these reservoirs have algal bloom problems almost every year from late summer to early fall. The nutrient status of these reservoirs ranges from mesotrophic to eutrophic due to large amounts of nutrient input from their corresponding watersheds: Eagle Creek Watershed, Cicero Creek Watershed and Fall Creek Watershed (Indiana Department of Environmental Management, 2002, 2004, 2006). Since the dominant land use in these three watersheds is row-crop agriculture (> 50%), fertilizer and pesticides are the main source of pollutants. If farmers over applied fertilizer, excessive nutrients, such as nitrogen (N) and phosphorus (P) can leach out from the soil and be carried into these reservoirs. This nutrient discharge promotes the occurrence of algal blooms that finally degrade drinking water quality. Eagle Creek Watershed Plan (2005) concluded that nutrient concentrations in all streams in Eagle Creek Watersheds frequently exceed the national average for watersheds with 50 - 75%

agricultural use. Total P load entering the reservoir is estimated to exceed 40 metric tons/year and total N load 550 metric tons/yr in 2004 (Eagle Creek Watershed Plan, 2005). These excessive amounts of nutrients affect water quality in Eagle Creek Reservoirs. Reducing nutrient loads into these water bodies relies on effective watershed and farm management strategies.

In conventional agriculture, the spatial variability of soils and crop is not taken into account, and fertilizer, herbicide, insecticide, fungicides and irrigation are uniformly applied across the field, causing over or under usage of these chemicals (Mandal et al., 2000). Because oven application of these agrichemicals could exacerbate NPS pollution, conventional agricultural practices need to be evaluated. Precision farming (PF) has emerged as a management option that maximizes agricultural productivity and minimizes the risk of environmental pollution. PF employs a site specific management approach to crop production; hence the term "site-specific farming" (Grisso et al., 2002). Forty farmers practicing PF in Northern Ohio achieved a 70 - 80% reduction in nutrient loading to a nearby stream and saved about 25 - 35% of their chemical application cost (Horsley, 1995). The implementation of PF requires knowledge of the spatial distribution of soil properties and crop conditions in order to properly tailor application of agrichemicals to crop needs.

Traditionally, the spatial variability of soil properties is obtained via field grid sampling: farmers or agricultural specialists collect soil samples that are later analyzed for individual constituents such as nitrogen (N), phosphorus (P), etc. This conventional procedure is labor-intensive, time-consuming and expensive. The advances in remote sensing (RS), geographic information system (GIS) and global positioning system (GPS)

allow for rapid mapping of soil properties. These technologies, especially remote sensing, not only provide synoptic view of a watershed, but also allow for multi-temporal assessment of the same watershed. They provide a rapid soil mapping tool for precision farming and watershed management.

1.2 Spectral Response of Soil Constituents

Soil is a complex system with highly variable in physical and chemical properties (Ben-Dor et al., 1999). Both chemical and physical chromophores have spectral responses and affect the shape and nature of a soil spectrum. This comprises the physical basis for the application of remote sensing to map soil properties. Soil chemical chromophores are materials in a soil system that absorb incident radiation in discreet energy levels. Soil has three major chemical chromophores: mineral (mostly clay and iron oxides), organic matter and water (Ben-Dor, 2002).

The spectral features of clay minerals are associated with oventones and combination modes of fundamental vibrations of functional groups in the infrared region. In general, three major spectral regions are active for clay: 1.3 - 1.4, 1.8 - 1.9, and 2.2 - 2.5 μm (Ben-Dor, 2002). Iron oxides can change soil color significantly due to selective light absorption in the visible range. Soil iron content is highly correlated with the absorbance in the spectral region 0.6 to 1.0 μm.

Soil organic matter (SOM) consists of plant and animal residues in different stages of decomposition (NRCS, 2008a). SOM has strong influence on soil reflectance, absorbing light through the VIS-NIR-SWIR (visible-near infrared-shortwave infrared) region. Ben-Dor et al. (1997) observed decreased soil albedo across the VIS-NIR-SWIR

region as the organic matter aged. They suggested that the reflectance slope between 450 and 638 nm may be useful for identifying both the degradation condition and the parent material status of the organic matter and the reflectance slope between 680 and 800 nm may be used as a general parameter for monitoring organic matter maturity. Moreover, organic matter was found to be correlated with the spectral region 0.5 to 1.2 μm (Mathews et al., 1973) and that the region 0.9 to 1.22 μm is good for mapping soil organic matter (Beck et al., 1976).

Three major forms of water exist in soil: hydration water incorporated into the lattice of the minerals, hygroscopic water adsorbed on soil surface areas as a thin layer and free water filling soil pores (Ben-Dor, 2002). Like SOM, moisture decreases soil reflectance in the VIS-NIR-SWIR region. Spectra of a mixed system of water and minerals, exhibit OH absorption features at around 0.95 μm (very weak), 1.2 μm (weak), 1.4 μm (strong) and 1.9 μm (very strong).

1.3 Objective

The main objective of this study was to test the capability of partial least square regression for mapping soil organic matter, total N and total P using satellite hyperspectral imagery. To the author's best knowledge, this is the first attempt to map SOM, total N and total P using satellite hyperspectral imagery. The secondary objective was to generate soil organic matter, total N and P distribution maps for the three Central Indiana watersheds. Specifically, the study was aimed at addressing how PLS performs when applied to spectral data acquired by sensors on different platforms and how factors (i.e. moisture effect, signal to noise ratio of spectral data) affect the prediction accuracy.

The specific research tasks were completed by comparing PLS modeling results for field-moist soils with those for oven-dried soils, and by assessing PLS results for three levels of spectral data: laboratory, field measured, and Hyperion image spectra. These comparisons are important because they provide effective strategies for improving prediction accuracy in the future.

1.4 Previous Research

1.4.1 Overview

A number of investigations have been done with multi- and hyper-spectral reflectance data to characterize soil properties. A general review on remote sensing of soil properties was published by Ge et al. (2006). As early as the late 1960s to early 1970s, soil scientists began to investigate the possibility of using multispectral remote sensing data for differentiating surface soils. The legacy of Landsat satellites began in 1972 with the launch of Landsat 1. Since then, millions of multispectral satellite images have been available for soil survey and mapping. Several studies have used multispectral remote sensing imageries to estimate soil properties (Chen et al., 2000; Ray et al., 2004; Dematte et al., 2007). However, many early attempts were deemed not appropriate for precision agriculture because multispectral images cannot yield enough prediction accuracy for estimating SOM, particle size and clay minerals with the limited spectral resolution. A new approach, namely, hyperspectral remote sensing emergent in the early 1980s, provides detailed spectral information in dozens or hundreds of narrow, adjacent spectral bands. With the progress and maturity of this technology, hyperspectral remote sensing has found a wide range of applications in mapping soil types and quantifying soil

constituents. In the laboratory, the method of using hyperspectral data to predict major soil properties has been well established (Dalal et al., 1986; Morra et al., 1991; Masserschmidt et al., 1999; Chang et al., 2001; Udelhoven et al., 2003; He et al., 2007). The application of airborne hyperspectral imagery for mapping soil properties has recently emerged (Ben-Dor et al., 2002; Uno et al., 2005; Bajwa et al., 2005; Stevens et al., 2006; Farifteh et al., 2007). While satellite hyperspectral imagery has become available since 2000, few attempts have been made to use satellite hyperspectral imagery for mapping soil properties.

The success of mapping soil properties with hyperspectral imagery relies on robust spectral-compositional models linking soil compositional parameters to hyperspectral reflectance. A number of spectral-compositional models have been proposed in previously published studies (Ge et al., 2006). Among these models, multiple linear regression analysis (MLR), principal component regression (PCR) or partial least squares regression (PLS) are commonly used to quantify hyperspectral soil data.

MLR uses a linear equation to correlate a response variable (i.e., chemical concentration) with two or more explanatory variables (i.e., spectral wavelength). The number of spectral wavelengths that could be used in MLR is limited because a larger number of spectral bands than the number of samples can result in rank deficiency problems.

Unlike MLR, PCR and PLS are full-spectrum methods. PCR is simply principal component analysis (PCA) of spectra followed by a regression against chemical compositions, while PLS is a rotated PCA applied to both spectra and chemical compositions and then finds the best relationship between them. Studies have shown that

PLS model achieves the optimal prediction model more rapidly than PCR, and is computationally more efficient (Haaland and Thomas, 1988a; Li, 2006). Recently, several new statistical tools have been utilized for soil mapping, such as artificial neural networks (ANN) and boosted regression trees (BRT). Fariftch et al. (2007) showed that soil salinity can be mapped by either PLS or ANN with similar model performance. Brown et al. (2006) built BRT and PLS models with soils collected from all around the world for determining soil organic carbon, inorganic carbon, clay, cation exchange capacity (CEC) and Fe. They found that BRT was far superior to PLS and contributed this result to BRT's ability to incorporate complex, non-linear relationships and interactions. Ge et al. (2007) made a comparison between wavelet analysis and PLS for clay content estimation and found that both methods have similar prediction ability. However, they claimed that the wavelet model was more suitable for sensor development because it had fewer regressor variables and the capability of physical interpretation.

1.4.2 Close-range hyperspectral remote sensing of soil properties

Visible/near-infrared spectroscopy (VNIRS) has been successfully used for quantifying organic matter or organic carbon (C) and total N in the laboratory. Ingleby et al. (2000) applied MLR to reflectance of selected wavelengths from 390 to 2500 nm for organic carbon prediction. Five MLR models were built for five fields and resulted in coefficients of determination (R^2) greater than 0.73. Chang et al. (2001) used PCR to estimate total C and total N from NIR spectra (R^2 = 0.87 and 0.85, respectively). Van Waes et al. (2005) derived an optimum partial least squares (PLS) model to estimate organic C and obtained an R^2 of 0.88 for validation. Other studies using PLS yielded high

correlations between measured and predicted SOM and total N (Ludwig et al., 2002; Chang et al., 2005; He et al., 2007), and between measured and predicted C:N ratios and the percentages of O-alkyl C (Terhoeven-Urselmans et al., 2006). With one exception, all these studies used spectra of dried and sieved soils so that the interference of particle size and moisture was minimized and the prediction models performed well.

Several researchers applied multivariate statistical analysis to soil spectra measured before soil drying and/or sieving. Hummel et al. (2001) used undisturbed soil spectra (1603 - 2598 nm) to estimate via stepwise multiple linear regression (SMLR), producing R^2 values greater than 0.72. Udelhoven et al. (2003) obtained relatively low R^2 values between measured and predicted organic C ($R^2 = 0.60$) and total N ($R^2 = 0.62$) from PLS modeling of natural soil spectra measured without removing soil moisture. Chang et al. (2005) compared PLS modeling results for air-dried soils with those for the moist soils and concluded that several properties (organic C, total C, inorganic C, total N, cation exchange capacity, %clay and moisture) of moist soil can be estimated using the NIRS-PLS technique. But the model was not transferable to a new study site. Sudduth et al. (1996) proved that the prediction capability of a PLS model decreases as the geographic range extends. Cozzolino et al. (2006) found that prediction accuracy of soil organic C varies for different particle size fractions, and the highest correlation, R^2 of 0.90, between measured and estimated soil organic C was yielded for the smallest particle size, indicating decreased variability in particle size might improve organic prediction with NIR spectra. Field measured spectra were also used for assessing organic C and N, but reported predictions were poor. Coleman et al. (1991) investigated the capability of field measured spectra for mapping SOM, and obtained an R^2 of 0.69 via MLR

regression. Udelhoven et al. (2003) used laboratory and field spectra to estimate organic C and N, but the predictions from field spectra ($R^2 < 0.69$) were poor compared to the PLS with laboratory. The authors attributed the poor prediction with field spectra to the strong interference of soil structure, and claimed that PLS prediction model is only applicable to geologically homogeneous areas.

Several researchers have employed mid-infrared (MIR) spectral region and PLS to determine SOM and total N in soil (Masserschmidt et al., 1999; Ehsani et al., 2001; Reeves et al., 2001; McCarty et al., 2002). Reeves et al. (2001) and McCarty et al. (2002) reported PLS modeling with MIR generated a better result than using NIR for estimating organic C, and attributed the low performance of PLS with the NIR spectra to the interference of vibration overtones and combination bands. However, modeling with MIR spectra requires to dilute soil samples with KBr and to remove particle size effect, making its "real world" application impractical.

Unlike SOM and total N, soil P is not easy to estimate from laboratory spectral data. Chang et al. (2001) used PCR with NIR reflectance (1300 to 2498 nm) to estimate Mehlich-3 P (an operationally-defined pool of plant available P), but reported a low coefficient of determination ($R^2 = 0.40$). Moron and Cozzolino (2007) applied PLS to both VIS-NIR and NIR reflectance for P prediction, and obtained a poor prediction with the highest R^2 of 0.61. Moron and Cozzolino (2007) also found that P concentration measurement methods affect the accuracy of the NIR calibration, and recommended to take into account soil P measurement methods when NIR calibrations are compared. Similarly, poor prediction of soil P concentration was achieved in studies by Ludwig et al. (2002), Udelhoven et al. (2003) and He et al. (2007). Only Bogrekci et al. (2005a)

demonstrated a strong relationship between total P and UV-VIS-NIR absorbance spectra of dried and sieved soil with a validation R^2 value of 0.92 resulting from the PLS model. Bogrekci et al. (2006) also demonstrated a significant improvement in the prediction of soil P after removing the moisture effect.

1.4.3 Airborne hyperspectral remote sensing of soil properties

Only a few studies have utilized hyperspectral imagery and multivariate statistical analysis to map soil properties. Ben-Dor et al. (2002) applied MLR to Digital Airborne Imaging Spectrometer (DAIS-7915) data. Bajwa et al. (2005) used PLS to estimate organic matter from RDACS/H-3 (Real Time Digital Airborne Camera System H3) spectral data. Using Compact Airborne Spectrographic Imager (CASI) data, Uno et al. (2005) developed a field-scale SOM estimation models through SMLR and artificial neural network (ANN). Both Bajwa et al. (2005) and Uno et al. (2005) showed relatively low R^2 values (< 0.56) and this was possibly due to the narrow range of spectra and coarse spectral resolution. Stevens et al. (2006) mapped SOC using CASI-2 data and PLS with an R^2 of 0.85 for validation. Hyperspectral Mapper (HyMap) airborne hyperspectral data were used to map organic C and total N (Selige et al., 2006) and soil salinity (Farifteh et al., 2007). These studies showed the potential of airborne hyperspectral imagery for mapping organic matter, total N and soil salinity. Detailed results from these studies are summarized in Table 1.

Table 1. Summary of information about remote sensing of soil properties

Soil properties	R^2	MR	PSR	Spectral data	Author and Year	Method
Organic matter	0.69			Field (multispectral data)	Coleman et al., 1991	MLR
Organic matter	0.84	X		Lab (MIR: 4000 - 400 cm^{-1})	Masserschmidt et al., 1999	PLS
Organic C	> 0.73	X	X	Lab (390 - 2500 nm)	Ingleby et al., 2000	MLR
Organic matter	> 0.72			Lab (1603 - 2598 nm)	Hummel et al., 2001	SMLR
Total C	0.87*	X	X	Lab (1300 - 2500 nm)	Chang et al., 2001	PCR
Total N	0.85*					
Mehlich-3 P	0.40*					
Total C	0.93*	X	X	Lab (MIR: 4000 - 400 cm^{-1})	Reeves et al., 2001	PLS
Total N	0.95*					
Organic C	0.96		X	Lab (MIR: 4000 - 400 cm^{-1})	McCarty et al., 2002	PLS
	0.90			Lab (1100 - 2498 nm)		
Total C	0.92*	X	X	Lab (400 - 2500 nm)	Ludwig et al., 2002	PLS
Total N	0.86*					
Olsen P	0.59*					
Organic C	0.60		X	Lab (400 - 2500 nm)	Udelhoven et al., 2003	PLS
Total N	0.62					
Available P	0.54					
Total Mg	0.69			Field (400 - 2500 nm)		
Total P	0.92*	X	X	Lab Spectra (UV-VIS-NIR)	Bogrekci et al., 2005a	PLS
Water-soluble P	0.78*					
Mehlich-1 P	0.91*					
Organic C	0.88*	X	X	Lab (1100 - 2500 nm)	Chang et al., 2005	PLS
	0.91*		X			
Total N	0.91*	X	X			
Organic C (coarse-sand)	0.80	X	X	Lab spectra (400 - 2500 nm)	Cozzolino et al., 2006	PLS
Organic C (find-sand)	0.85					
Organic C (clay + silt)	0.90					
Organic matter	0.92*	X	X	Lab spectra (350 - 2500 nm)	He et al., 2007	PLS
Total N	0.88*					
Available P	0.54*					
Bray P	0.58	X	X	Lab Spectra (VIS-NIR)	Moron et al., 2007	PLS
Resins P	0.61					
Organic C	0.93			Aerial photograph (red, green & blue)	Chen et al., 2000	Linear equation
Organic matter	> 0.70			Aerial imagery	Fox et al., 2002	SLED
Organic matter	0.54			IKONOS	Ray et al., 2004	Stepwise regression
Available N	0.28					
Organic C	0.81*			ATLAS	Chen et al., 2006	Stepwise regression
Organic matter	0.4*			Lansat-7 ETM+	Dematte et al., 2007	MLR

Organic C	> 0.77			ATLAS	Chen et al., 2008	Stepwise regression
Organic matter	--			DAIS-7915 (400-2500 nm)	Ben-Dor et al., 2002	MLR
Organic matter	0.54*			RDACS/H3 (471 - 828 nm)	Bajwa et al., 2005	PLS
Organic matter	0.56			CASI (408 - 947 nm)	Uno et al., 2005	SMLR
	0.35					ANN
Organic C	0.85*			CASI-2 (405 - 950 nm)	Stevens et al., 2006	PLS
Organic C	0.89*			HyMap (450 - 2500 nm)	Selige, 2006	MLR
Total N	0.91*					
Salinity	> 0.61*			HyMap	Farifteh et al., 2007	PLS
	> 0.60*					ANN

MR: moisture removal; PSR: particle size removal; * R^2 for validation, otherwise, R^2 for calibration; -- : data unavailable; MLR: multiple linear regression; SMLR: stepwise multiple linear regression; PCR: principle component regression; PLS: partial least squares; SLED: Soil line Euclidean distance; ANN: artificial neural network.

1.4.4 Multispectral remote sensing of soil properties

Soil properties have been mapped with remotely sensed multispectral images (Chen et al., 2000; Fox et al., 2002; Ray et al., 2004; Dematte et al., 2007). Chen et al. (2000) mapped soil organic C distribution using an aerial color photograph ($R^2 = 0.926$). Fox et al. (2002) presented the soil line Euclidean distance (SLED) technique to estimate SOM with aerial color images. They built a relationship between SOM and a pixel's Euclidean distance along the soil line determined from the scatter plots of image red and NIR bands, and reported coefficients of determination 0.70 and 0.78 between observed and predicted SOM for two study sites. Ray et al. (2004) estimated soil organic C and N from an IKONOS multispectral image, but the resultant correlation between measured and predicted values was very low, 0.54 for organic C and 0.28 for N. Dematte et al. (2007) used Lansat-7 ETM+ images for mapping several soil properties across a large geographical area (43,000 hectares) and resulted in a low validation R^2 of 0.4 for SOM

via MLR. Although both Chen et al. (2000) and Fox et al. (2002) achieved a relatively high correlation between measured and predicted SOM, their prediction equations only worked for a single farm field. Chen et al. (2006, 2008) proposed a method to map OC of multiple fields. They proposed to use Euclidean distance for statistical clustering and the Ward neural network system to select fields with similar image properties to ensure the success of mapping OC for a group of fields. Therefore, mapping soil properties with multispectral imagery is problematic, either prediction accuracy is low or the prediction equation is not transferable across a large geographical area.

2 MATERIALS AND METHODS

2.1 Study Site and Soil Sampling

The study area covers Eagle Creek, Cicero Creek and Fall Creek Watersheds, Central Indiana (Fig. 1). Eagle Creek Watershed is 420 km^2 in area including Marion, Hendricks, Boone, and Hamilton Counties. Cicero Creek Watershed is 588 km^2 in area encompassing Hamilton, Boone, Clinton and Tipton Counties, while Fall Creek Watershed has an area of 824 km^2 including Hancock, Henry, Hamilton, Madison and Delaware Counties. These areas are relatively flat with row-crop agriculture being dominant type of land use. In 2000, agricultural land use was about 55% for Eagle Creek Watershed and Fall Creek Watershed and 77% for Cicero Creek Watershed (Tedesco et al., 2005).

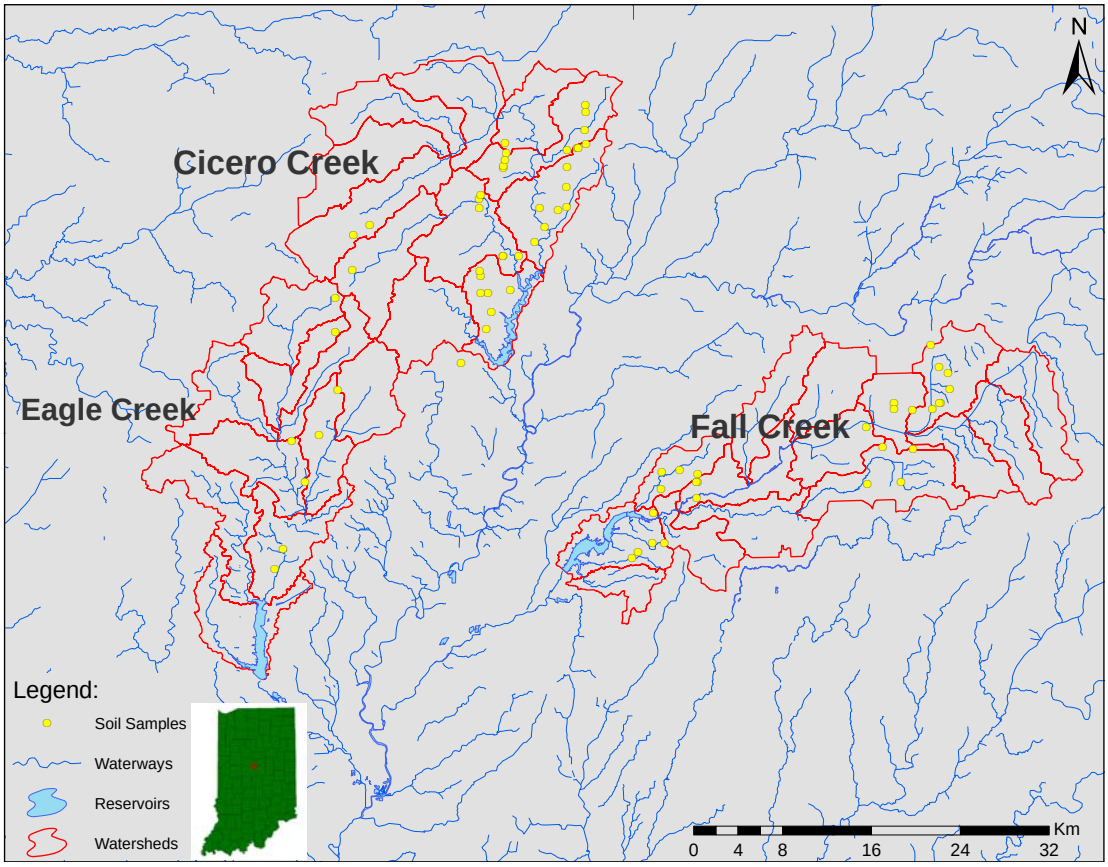


Figure 1. Location of study areas and sampling sites

Soil orders in the study area are Alfisols, Mollisols and Inceptisols (NRCS, 2008b). According to State Soil Geographic (STATSGO) data base, Crosby-Treaty-Miami, Miami-Crosby-Treaty and Sawmill-Lawson-Genesee are the main soil associations in Eagle Creek Watershed, Crosby-Treaty-Miami, Miami-Crosby-Treaty, Patton-Del Rey-Crosby and Drummer-Toronto-Wingate in Cicero Creek Watershed and Fox-Ockley-Westland, Miami-Crosby-Treaty, and Crosby-Treaty-Miami in Fall Creek Watershed (Fig. 2). The Crosby-Treaty-Miami association consists of a deep, poorly drained, nearly level to gently sloping soil. The Miami-Crosby-Treaty association consists of deep well drained to somewhat poorly drained, moderately sloping soils. Fox-Ockley-Westland is well-drained, moderately deep soils. Patton-Del Rey-Crosby is nearly level, poorly drained and somewhat poorly drained soils.

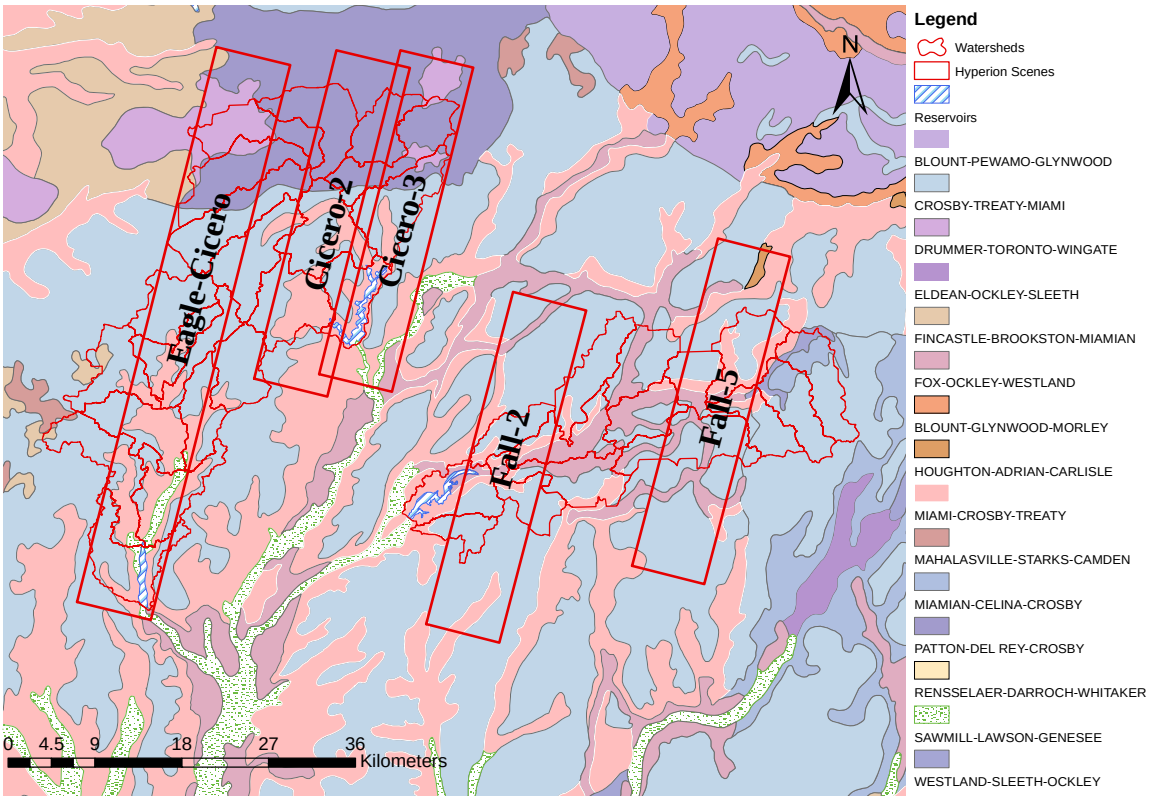


Figure 2. Soil associations and flight lines of Hyperion images

Soil samples were collected simultaneously with the acquisition of satellite hyperspectral imagery with the exception of Cicero-2 (Table 2). Soil sampling for the Cicero-2 area was conducted four days after the Cicero-2 image acquisition. A total of 70 surface (0 - 2 cm) agricultural soils were sampled and each soil sample was collected from about a 20 × 20 cm area. The sampling sites were randomly selected and they were sparsely distributed. By the time of sampling at Eagle-Cicero, the crops grew to 20 to 40 cm heights, affecting the spectral signal of soil. In order to yield representative soil samples and to obtain a broad range of SOM, N and P concentrations, about five soil samples were collected for each soil association. Sampled soil series are mainly Crosby, Brookston and Patton. Other soil series including Miami, Williamstown, Shoals, Del Rey, Eel, Ockley, Camden, Fox, Kokomo, Celina and Westland (NRCS, 2007). Farm fields

with high residue cover were avoided in sampling. *In situ* soil reflectance was measured at each sampling site using a field portable ASD spectrometer (Analytical Spectral Devices, Boulder, Colorado). The geographic coordinates of each sampling site were recorded with one meter accuracy using a global positioning system (GPS) instrument. The GPS coordinates were used to locate each site on the Hyperion images and the corresponding image spectrum can be extracted for creating spectral-compositional models.

Soil samples were kept fresh in Zip-loc bags (17 cm × 20 cm), transported over ice in a cooler to the laboratory and stored in refrigerators (4°C) before soil property analyses. Table 2 summarizes the time for the imagery acquisition, soil sampling and the total number of samples for each scene.

Table 2. Data acquisition information

Watershed	Scene name	Scene No.	Imagery acquisition time	Date of soil sampling	No. of soil sample
Cicero Creek	Cicero-2	1	04/24/2007	04/29/2007	15
	Cicero-3	2	05/07/2007	05/07/2007	18
Fall Creek	Fall-2	3	05/12/2007	05/12/2007	11
	Fall-5	4	05/22/2007	05/22/2007	15
Eagle Creek & Cicero Creek	Eagle-Cicero	5	06/09/2007	06/09/2007	11

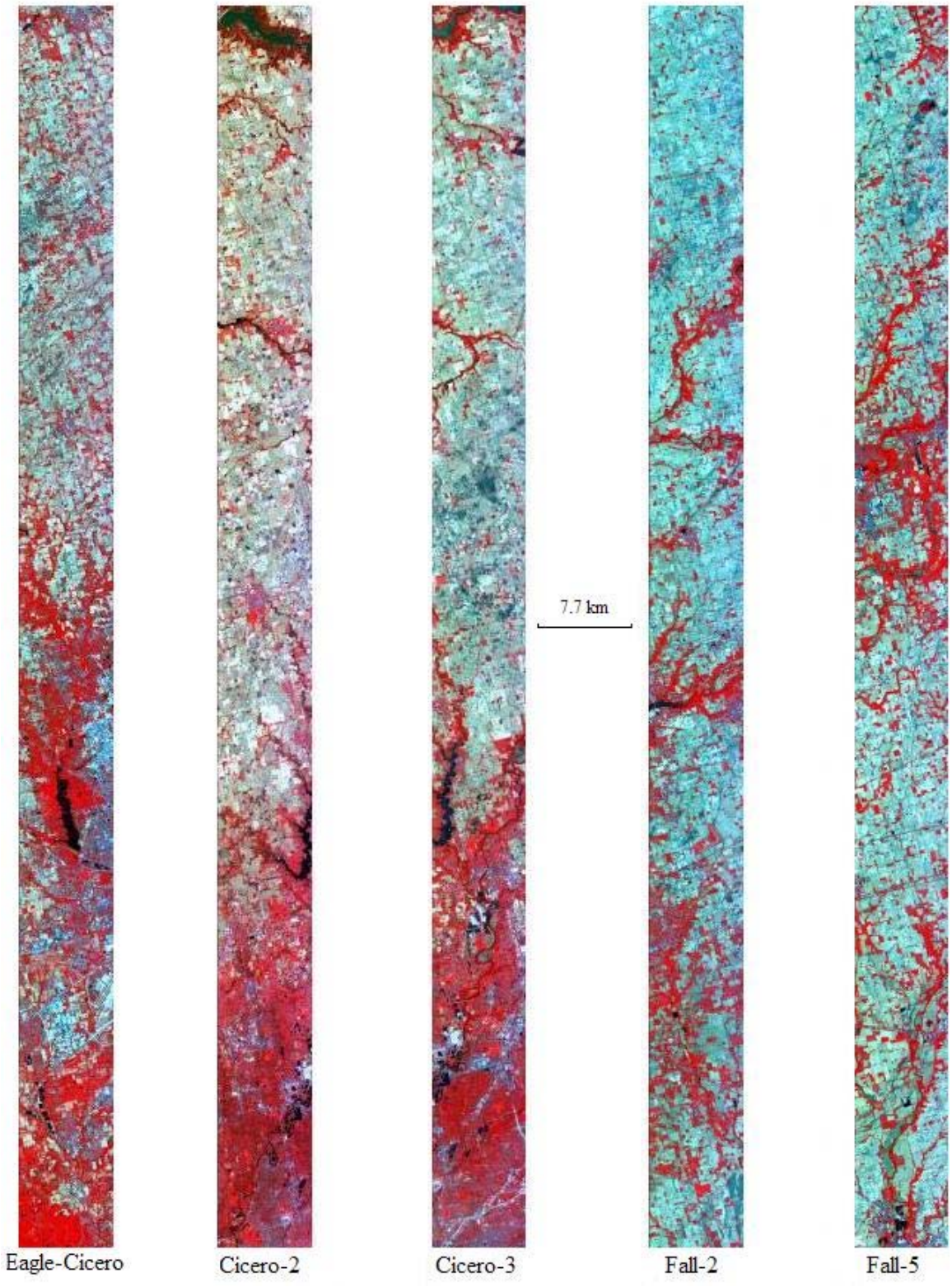


Figure 3. Hyperion false color images

2.2 Lab Analysis for Soil Properties

The soil lab analysis methods are summarized in Table 3 and the procedure is shown in Figure 4. Soil organic matter was determined by the Loss on Ignition (LOI) method, total N by dry combustion and total P using strong acid digestion.

Each soil sample was split into two portions: one for soil property analysis and laboratory spectral measurement, and the other for backup. Soil property analyses were conducted after completing spectral measurements twice: one with field-moist samples and the second with oven-dried samples. After the spectral measurement of oven-dried soils, soils were ground to pass through a < 2 mm sieve.

Organic matter content (OMC) was measured by calculating the weight loss after 550°C combustion of moisture-free soils. Since the oven-dried soil was exposed to air during grinding and sieving, 1.05 g of the soils (< 2 mm) was weighed and put into a 105°C oven for at least three hours to further remove soil moisture. Later, the samples were moved into desiccators for cooling down to room temperature and weighed to record initial weight (W_1). The oven-dried samples were then put into the 550°C furnace for 2.5 hours, and left in the furnace for another 30 minutes after turning the furnace off. When cooling down to room temperature in desiccators, the final weight of the samples was recorded as W_2. OM was calculated as $(W_1-W_2)/ W_1$.

The first step for measuring total P was to convert organic phosphorus into inorganic by 550°C combustion. Later, 2M hydrochloric acid was added into the combusted soils and shaken overnight so that all phosphate ions were released into the solution. Color-Developing Reagent (CDR) was added into the solution. The more phosphate contains in the solution, the more molybdate-phosphate complexes are formed

and the darker blue the color is. Phosphate concentrations were measured with a spectrophotometer which can detect the complex absorption at 880 nm. Absorbance readings were converted into concentrations using a standard curve.

The measurement of total nitrogen was conducted in the department of Geological Sciences, Indiana University-Bloomington. Total N was measured using Costech elemental analyzer. Soil samples were required to be ground overnight using a rolling grinder and sieved into 250 μm before weighing 10 mg of the soil for the analysis. The elemental analyzer generates high temperature (1020°C) to break sample down into elemental components, CO_2, N_2, H_2O and SO_2. Sequentially, the gases are separated and detected by thermal conductivity detector (TCD).

Table 3. Soil properties analytical methods

Soil property	Method
Organic matter content	Loss on Ignition
Total N	Dry combustion (Costech elemental analyzer)
Total P	Strong acid digestion (Spectrophotometer)

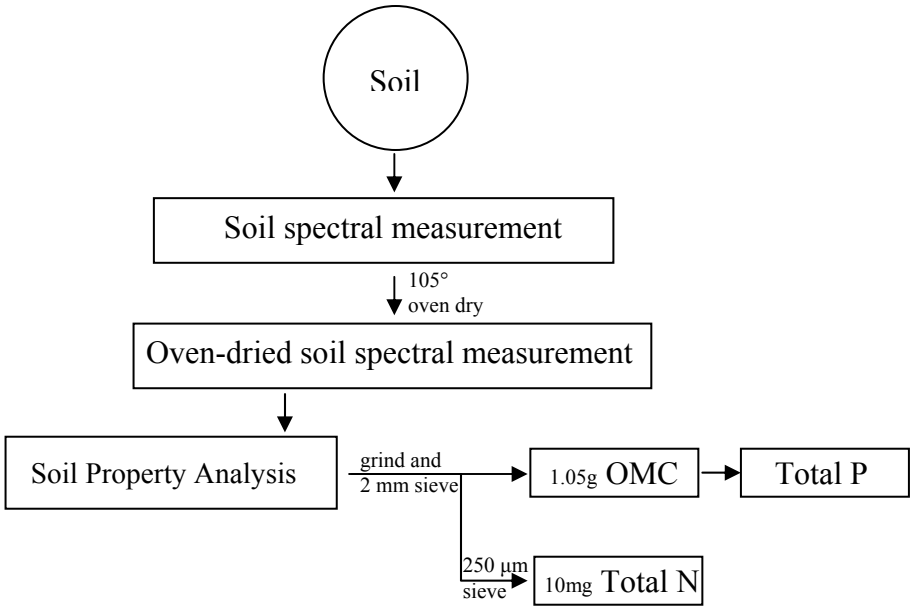


Figure 4. A flowchart for laboratory soil analysis

Each soil constituent shows a wide range of concentration: 2.5 to 13% for SOM, 0.04 to 0.52% for total N and 220 to 1504 mg kg^{-1} soil for total P. Figures 5 and 6 show the correlation between SOM and N, SOM and P respectively. Total N has high correlation with SOM ($R^2 = 0.92$), while total P has relatively low correlation ($R^2 = 0.63$).

Mean and standard deviation of repeated measurements of the same sample for SOM and total P are listed in Table 4. The high correlation between SOM and total carbon ($R^2 = 0.98$) suggests that SOM dataset is reliable (Appendix I). The total N measurement accuracy and precision were calculated using the multiple measurements of the standard. The actual value of the standard is 10.36% total N and the mean of the measured standard is 10.46%. The measurement accuracy is 0.91% and precision is 0.32 for total N.

Mineral N (nitrate: NO_3-N; ammonium: NH_4-N) and inorganic P represent the nutrient pools immediately available for plant uptake. In the laboratory, available P is usually analyzed by Bray 1, Mehlich 3 or Olsen extraction, while nitrate-nitrogen is generally measured by KCl extraction. For detailed procedures for these methods, please refer to Soil Survey Laboratory Methods Manual (2004). The agronomic and environmental significance of these mineral N and P forms is recognized. However, as a first step toward development of the PLS model, the present study focussed on organic N and P.

Table 4. Mean and standard deviation of selected soil samples

Soil property	Sample	F501	F212	C210	C214	C301	C307	C318	EC01	EC07
OMC %	Mean	3.51	5.19	3.55	3.64	3.22	3.67	4.99	3.67	4.12
	Standard deviation	0.14	0.02	0.08	0.07	0.13	0.14	0.01	0.02	0.08
Total P (mg kg^{-1} soil)	Mean	454	623	591	467	350	638	578	307	281
	Standard deviation	11	3.6	16	9.6	12	7.6	36	13	7.1

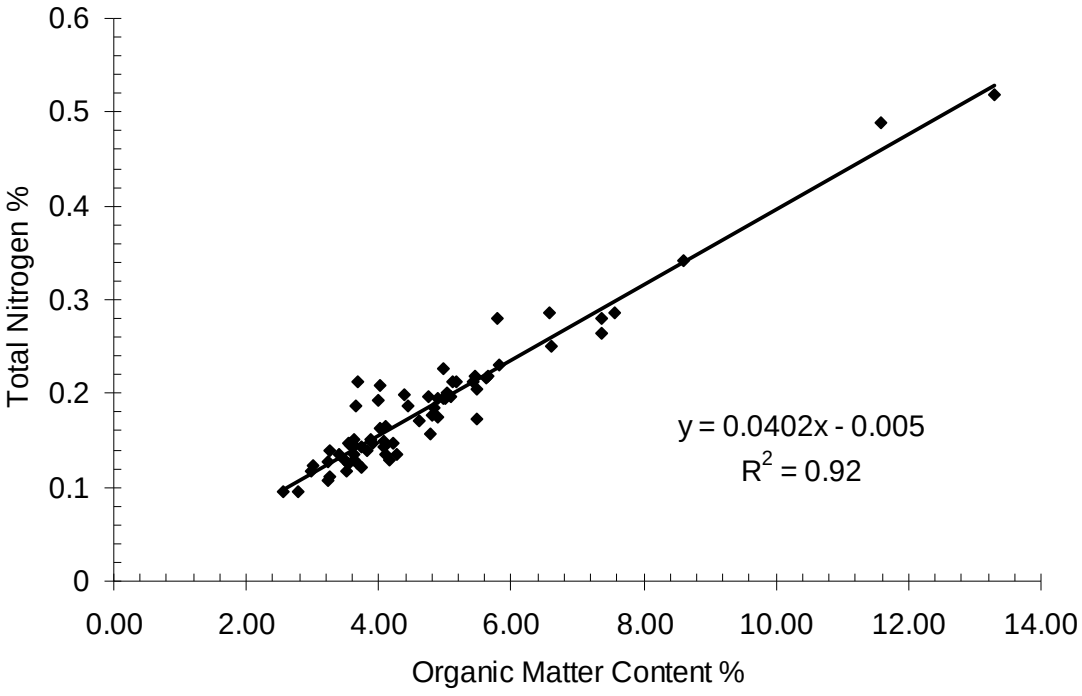


Figure 5. The correlation between soil organic matter (SOM) and total N

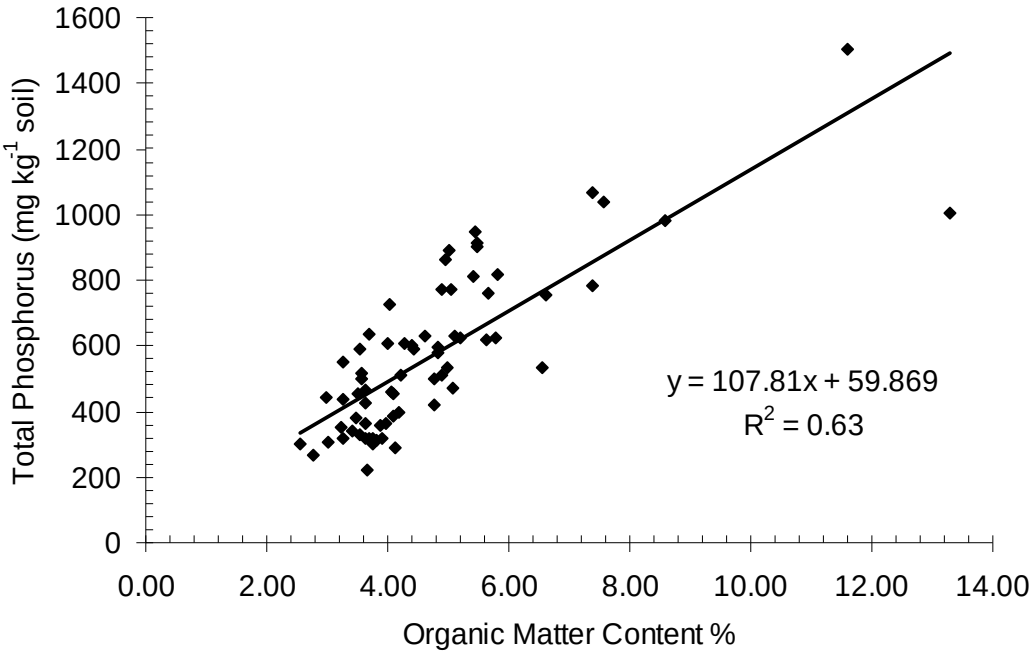


Figure 6. The correlation between soil organic matter (SOM) and total P

2.3 Soil Spectra Measurement

2.3.1 Field spectra

Field reflectance of soils was measured at each sampling site using an ASD spectrometer. The fiber-optical head pointed at a nadir viewing angle at approximately a height of 1 m above the soil surface. The ASD spectrometer has a spectral range from 350 to 2500 nm with 1 nm spectral intervals. Each reflectance measurement is the average of 20 spectral scans. A white reference panel was used as the reference in order to minimize the effect of changing atmosphere on soil spectra.

Figure 7 shows the field spectral curves of some soil samples with the removal of some noisy bands centered at around 1.4 and 1.9 μm, generating 1920 bands totally. Some patches of noise in the regions of 1.4 and 1.9 μm are due to the atmospheric water

absorption in these wavelength regions. The spectrum of sample EC05 is affected by vegetation as shown by chlorophyll absorption at 0.68 μm.

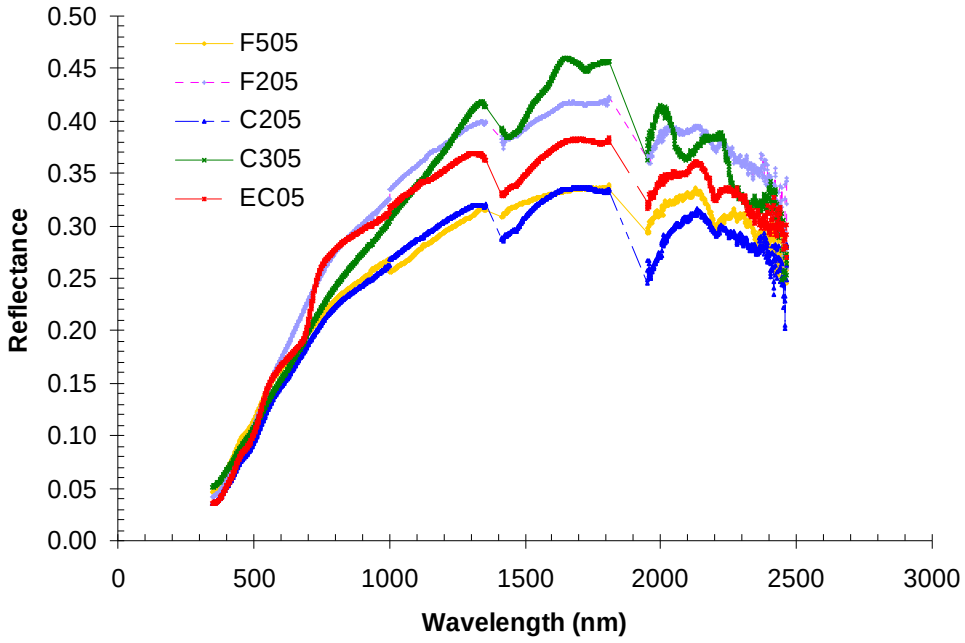


Figure 7. Field spectra of soils

2.3.2 Laboratory spectra

The reflectance spectra of field-moist and oven-dried soils were obtained in the laboratory by ASD spectrometer. Pro lamp was used as artificial illumination with approximate 30° angle from the nadir. The fiber-optical head pointed at the nadir viewing angle at approximately a height of 30 cm above the soil surface. Each soil sample was measured three times with about 120 degree rotation for both field-moist and oven-dried samples. To reduce the effect of soil texture on spectra, the average of three measurements of each sample was used in spectral-compositional modeling.

Figure 8 shows laboratory-measured spectral reflectance of three selected samples with a total of 2150 bands. SOM results in the decrease of reflectance almost uniformly

across the visible and near-infrared spectra (Fig. 8A). Figure 8B shows how moisture affects soil spectral reflectance when soils contain similar amounts of OM content. Soil moisture can mask the contribution of other soil parameters to soil reflectance, and impedes soil property estimation from remote sensing data. Therefore, in most circumstances, soil properties can be estimated from spectra of moisture free soil more accurately than those of moist soil.

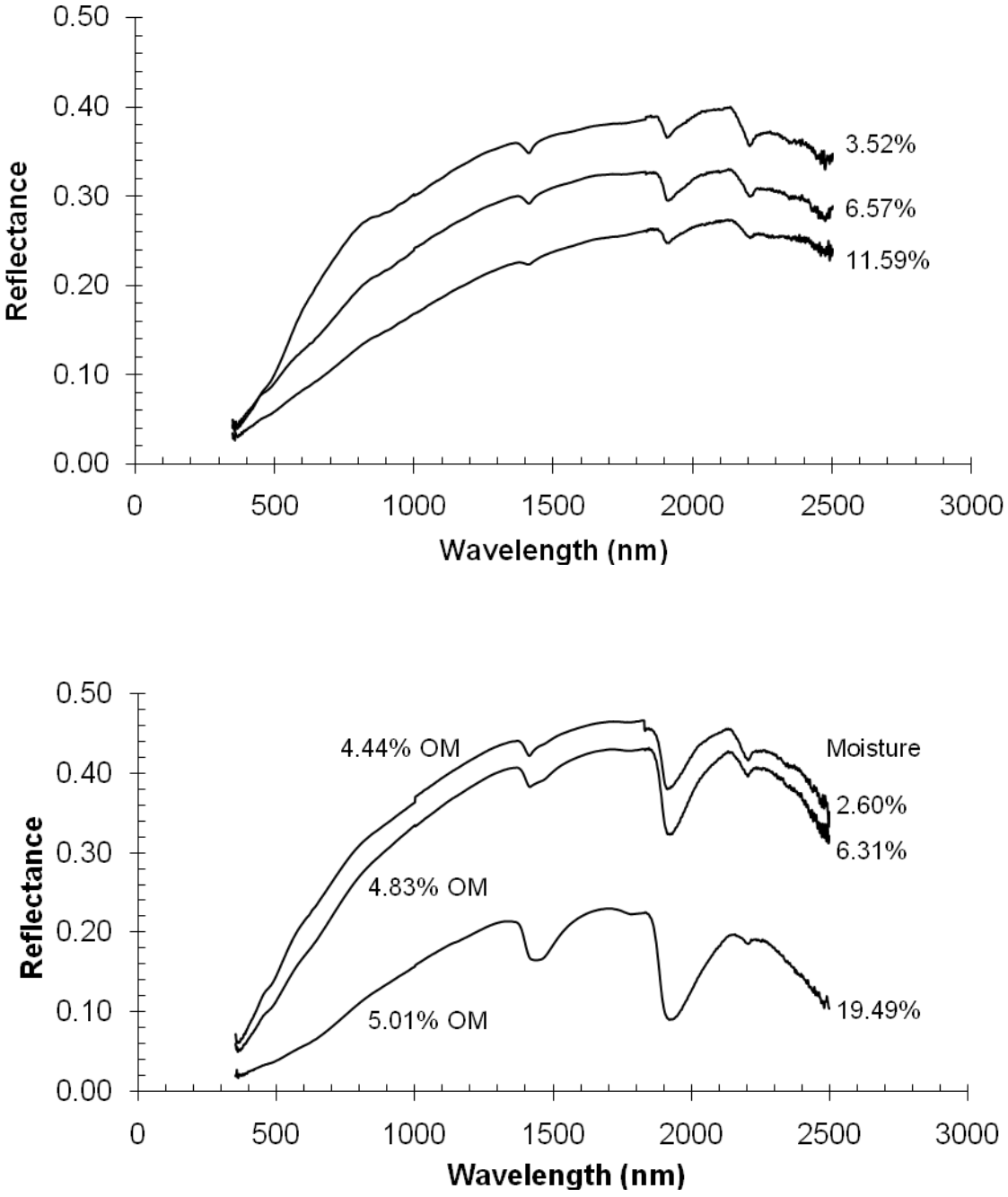


Figure 8. Spectra of oven-dried soil with different content of organic matter (A); Moisture effect on reflectance spectra of soils with similar amounts of organic matter and different moisture content (B)

2.3.3 Hyperion satellite imagery

The launch of EO-1 satellite on November 21, 2000 makes the dream of spaceborne hyperspectral remote sensing become true. A Hyperion scene has 7.7 km cross-track width with 42 km or 185 km along-track length depending on the data acquisition request. The Hyperion data with a 10 nm sampling interval over the spectral region of 356 - 2577 nm are acquired by two separate pushbroom imaging spectrometers: one for the VIS-NIR region and the other for the SWIR. The Hyperion data are characterized by a total of 242 channels, 30 meters spatial resolution, and approximate 50:1 signal-to-noise ratio (SNR).

To cover the three watersheds, at least 13 scenes of Hyperion imagery are needed with the 16% overlap between adjacent scenes. EO-1 satellite does not acquire data continuously and its sensors are only activated to collect specific scenes upon request. Data Acquisition Request (DAR) was submitted in January, 2007 and the time window for the data acquisition ranged from March 15th to June 9th, 2007. In this time window, the soil in agricultural fields was mostly bare. Each flight took place at around 10 am local time (16:00 GMT). Figure 2 shows the acquisitions of Hyperion images that have been fulfilled. Due to the unfavorable weather condition, only five scenes of Hyperion images with cloud cover less than 5% were acquired. The Cicero-2 and Cicero-3 images covered the eastern Cicero Creek Watershed area, the Fall-2 and Fall-5 flight lines covered the Fall Creek Watershed area, and the Eagle-Cicero flight line passed over both the Eagle Creek and western Cicero Creek Watersheds area. The acquisition time for the corresponding image and scene number are summarized in Table 2. False color Hyperion images are shown in Figure 3.

Figure 9 shows the soil spectral curves extracted from the Hyperion images of the watersheds. A total of 165 bands were used for modeling after removing noisy bands centered at 1.4 and 1.9 μm. Note that the Hyperion image spectra are noisier than field spectra.

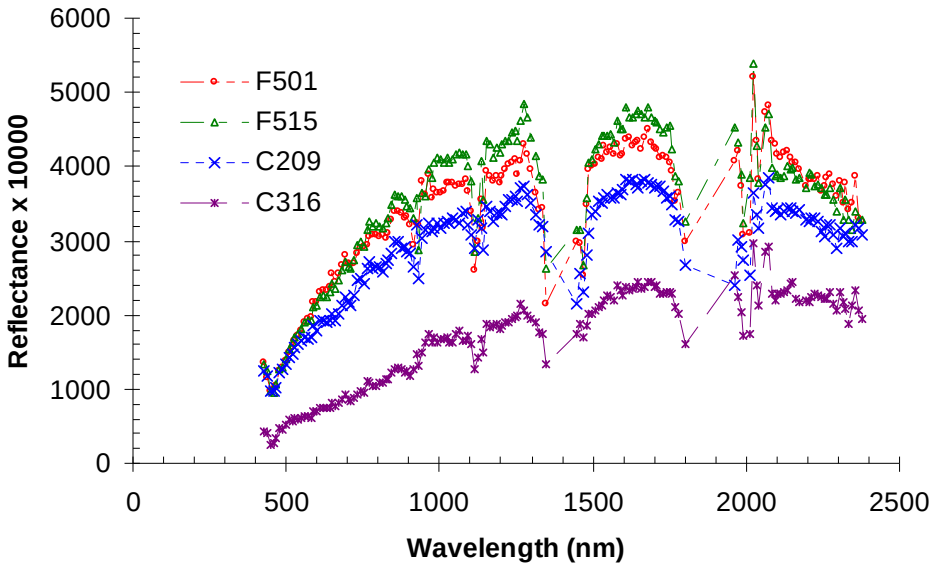


Figure 9. Examples of soil reflectance spectra extracted from Hyperion imagery

2.4 Image Processing

2.4.1 Atmospheric correction

The Hyperion data purchased from USGS are in radiance. Radiance is a physical quantity that describes the amount of light that is emitted from an object and then falls within a given solid angle in a specified direction. It also can be viewed as the amount of light that gets into the remote sensing instruments. However, the sensed radiance by the instruments contains radiation from both the object being observed and the atmosphere. Thus, atmospheric correction is required to convert radiance into reflectance.

Reflectance is the ratio of the amount of light leaving an object to the amount of light incident on the object.

The Hyperion images were atmospherically corrected using ENVI FLAASH module, a calibration program used to convert radiance into reflectance. The atmospheric correction was conducted by following the instructions from EO-1 User Guide (Beck, 2003). Out of the 242 collected channels, there were a total of 196 unique spectral bands provided by the level 1R data which is radiometrically corrected without geometric correction. After atmospheric correction, there are 178 bands for Fall-5 and Cicero-3, 172 bands for Fall-2, 176 bands for Cicero-2. Since an output scale factor of 10,000 was applied, the digital number (DN) of calibrated images ranges from 0 to 10,000 instead of from 0 to 1. To obtain consistent spectral data across different watersheds, Hyperion spectra were all subset into 172 bands.

2.4.2 Georeferencing

Each scene of Hyperion images was georeferenced using 2006 aerial photographs of the Hamilton and Madison counties in Indiana. The projection is Universal Transverse Mercator (UTM) Zone 16, NAD-1983 Datum. High spatial resolution (2 m) aerial photographs were degraded to match the 30 meter spatial resolution of Hyperion images. The georeferencing used pairs of ground control point (GCPs) manually selected from both images. A linear warping method and bilinear resampling method were chosen to warp Hyperion images. Registration accuracy was assessed by linking the warped to the base images and using dynamic overlays. Sampling sites were projected onto the

atmospherically corrected and georeferenced images and their corresponding spectra were extracted out for building spectral-chemical compositional models.

2.4.3 Image masking

Non-soil areas in the images such as vegetated and water bodies, were masked out. A binary mask image was first created using a band threshold operation that assigns 1 for users' area of interest and 0 for other areas. The target image for the band threshold operation was soil fraction images generated by linear spectral unmixing of Hyperion images with three endmembers (vegetation, soil and water). Assuming the reflectance of each pixel of an image is a linear combination of the reflectance of each material (or endmember), linear spectral unmixing can solve for the areal abundance of each endmember for every pixel based on the spectral characteristics of these materials. In the soil fraction image, a soil area has higher value than the other areas; hence the minimum of soil fraction can be determined. This fractional minimum was used as a threshold; a mask was then created where soil regions have the DN of 1 and non-soil areas are represented by the DN of 0. After the mask was multiplied with Hyperion images during the masking process, only the soil areas were kept in the Hyperion images. Figure 10 shows masked Hyperion images covering a portion of Cicero Creek Watershed and Fall Creek Watershed.

Figure 10. Hyperion images after masking and spatial subset
(upper-left: Cicero-2; upper-right: Cicero 3; lower-left: Fall-2; lower-right: Fall-5)

2.5 Partial Least Squares (PLS) Modeling

The PLS method was chosen to establish the correlation between soil spectral and chemical concentration data. PLS modeling is a multivariate statistical tool, combining some of the advantage of both classical least squares (CLS) and inverse least squares (ILS) methods (Haaland and Thomas, 1988b).

PLS consists of eigendecompositions of spectral matrix (X) and chemical concentration matrix (Y) and links score matrices T and U together through a regression:

$$X = TP^T + E \qquad (1)$$

$$Y = UQ^T + F \qquad (2)$$

$$U = BT \qquad (3)$$

Where E and F are residual matrices, P and Q are the loading matrices and B is the regression coefficient. Y can then be estimated from B:

$$Y = TBQ^T + F \qquad (4)$$

Selecting the optimal number of latent variables is essential to build a robust PLS model. Choosing more latent variables always leads to an improved performance of a calibration model, but the model might overfit the concentration data. The leave-one-out cross-validation method was used to select the optimal number of latent variables. Given a set of m samples, $m - 1$ samples are used to develop a calibration model and the concentration of the left out sample is predicted using the resulting calibration model. This process is repeated m times until each sample has been left out once. Then the predicted error sum of squares (PRESS) can be obtained:

$$\text{PRESS} = \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \qquad (5)$$

Where $\hat{y}_i$ and y_i are the estimated and actual concentration respectively. Root mean square error of cross-validation (RMSECV) for each PLS model with a given number of latent variables is calculated by

$$\text{RMSECV} = \sqrt{PRESS_j / m} \qquad (6)$$

where j is the number of latent variables. In general, the number of latent variables is considered optimal when it yields the minimum RMSECV.

In this study, PLS was first applied to laboratory and field spectral data to test its performance, and then to the Hyperion data. Samples were split into two subsets: one for calibration and the other for validation. Both spectral and chemical data were preprocessed using mean centering which often reduces the complexity of the model. In mean centering operation, the average calibration spectrum was subtracted from each spectrum and the average calibration concentration was subtracted from each concentration. PLS modeling with laboratory, field, and Hyperion spectra were compared to assess the effects of interfering factors that might reduce the robustness of the models and provide possible strategies for model improvement in the future. Finally, the model calibrated for Hyperion image spectra was applied to the calibrated and processed Hyperion imagery to generate the spatial distribution maps of SOM, N and P.

3 RESULTS AND DISCUSSION

3.1 PLS Modeling with Laboratory Spectra

A total of 70 soil samples were divided into a calibration data set (n = 50) and a validation data set (n = 20). Laboratory spectra range from 350 to 2500 nm with 1 nm increment, generating a total of 2150 bands.

3.1.1 SOM

The result from PLS modeling of the calibration data demonstrates its capability of estimating SOM from laboratory spectral reflectance. PLS yielded a coefficient of determination, R^2, of 0.92 for field-moist soils and 0.89 for oven-dried soils, implying that the explained variance is 92.95% for field-moist samples, and 89.20% for oven-dried. PLS modeling with validation data set resulted in an R^2 of 0.76, a root mean square error of prediction (RMSEP) 1.536 for field-moist samples and R^2 of 0.81, a RMSEP 1.0922 for oven-dried samples (Fig. 11). The PLS model for oven-dried soils has the low prediction error and high R^2 value as compared to those for field-moist soils, suggesting soil moisture has an adverse influence on the prediction accuracy of the PLS model.

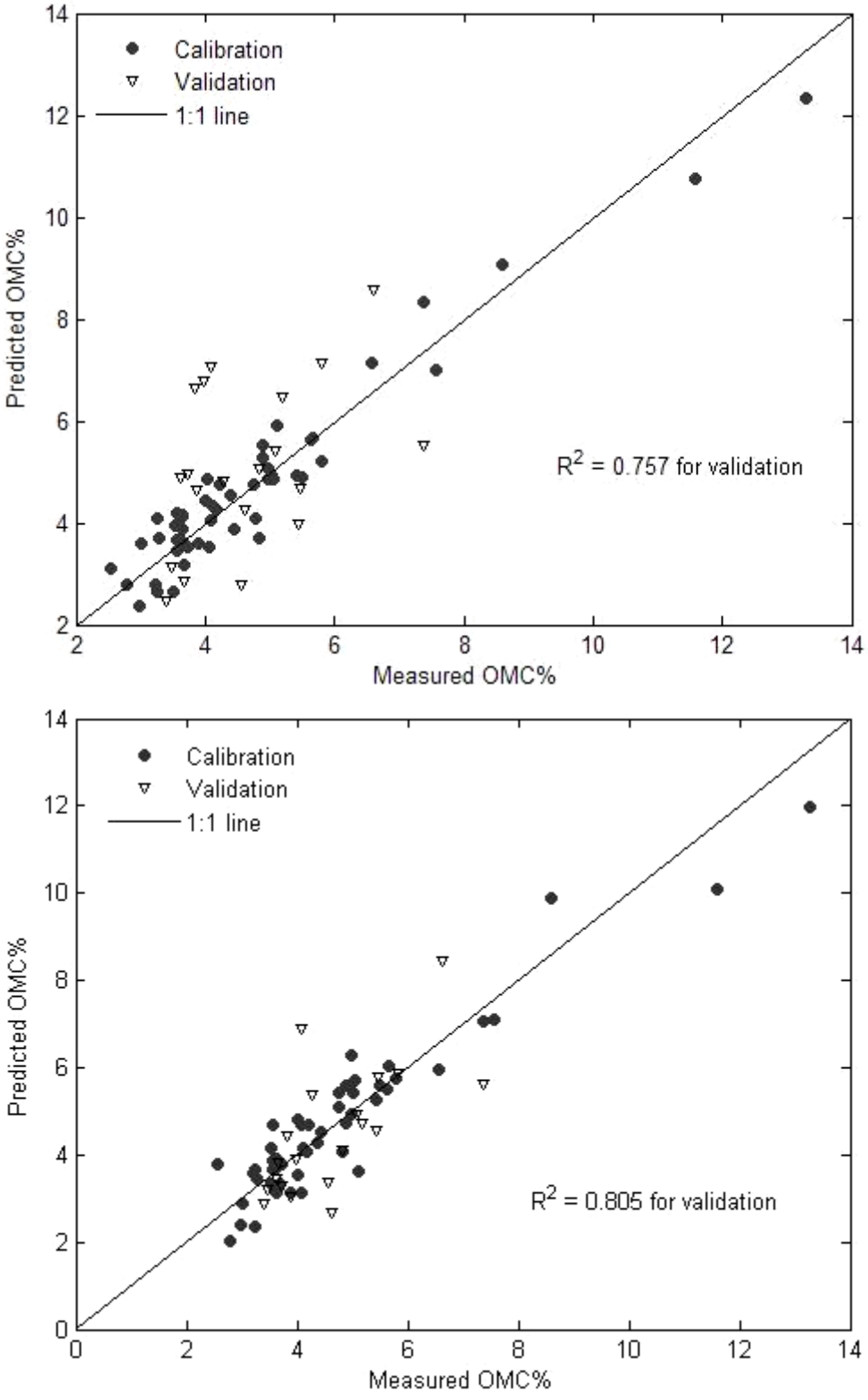


Figure 11. Comparison of measured OMC with OMC predicted by PLS modeling of field-moist soils (up) and oven-dried soils (bottom)

Table 5. Results from PLS modeling for oven-dried and field-moist samples

Model	Soil properties	No. of LVs	Validation		Calibration	
			R^2	RMSEP	R^2	RMSEC
Oven-dried	OMC %	8	0.81	1.09	0.89	0.66
Field-moist		9	0.76	1.54	0.93	0.53
Oven-dried	Total N %	9	0.82	0.045	0.91	0.025
Field-moist		10	0.75	0.073	0.96	0.016
Oven-dried	Total P	6	0.73	148.12	0.79	95.98
Field-moist	(mg kg^{-1} soil)	5	0.67	137.59	0.68	118.87

Examination of PLS loading vectors indicates that the optimal PLS model for field-moist soils needs nine latent variables (LV), one more than the model does for the oven-dried soils (Table 5). This is attributed to the effect of soil moisture. Figure 12 shows loading vectors of the first three latent variables derived by PLS modeling of soil spectral reflectance before and after oven drying. Spectral and SOM variations explained by the first three latent variables are listed in Table 6 for oven dried soils and Table 7 for field-moist soils. In Figure 12, the first LV reflects albedo variations, and shows similar amounts of spectral variance in the models for both types of samples (98.50% for field-moist model and 98.45% for oven-dried model) but different amounts of SOM variance. The first LV of the model for oven dried soils explains 55.27% SOM variance (Table 6), while the first LV of the model for field-moist soils only explains 21.42% SOM variance due to soil moisture effects (Table 7). In other words, soil moisture masks out approximate by 33% of SOM variance. The scatter plots between the first LV scores and moisture content (Fig. 13) show that the first LV of the PLS model for field-moist soils describes 64% moisture variance, while the one for oven-dried soil only explains 20% moisture variance. The second LV of the PLS model for oven-dried soils shows the diagnostic feature in the region 0.5 to 1.2 μm indicative of organic matter (Ben-Dor,

1997). While the narrow absorption in the region 0.6 to 0.8 μm due to organic matter degradation is very strong in the 2nd LV of the PLS model for oven-dried soils, it reflects only 8.33% SOM variance. The 2nd LV of the model for field-moist soils captures a flat and broad spectral feature from 0.7 to 1.3 μm and water absorptions at 1.4 and 1.9 μm. Because a flat curve suggests a capture of overall shape and a large variance, the flat absorption region (0.7 to 1.3 μm) indicates that the 2nd LV of the model for field-moist soils captures large variance of organic matter (46.46%). The third LV in the models for both types of soil samples show water absorption features at 1.4 μm and 1.9 μm and an unknown absorption at 0.6 nm. The second and the third LVs for field-moist samples exhibit stronger water absorption features at 1.4 μm and 1.9 μm than those for oven-dried samples do. These strong water absorption bands at around 1.4 μm and 1.9 μm appearing in the second and third LVs of the model for field-moist soils are primarily due to hygroscopic water, while the relatively weak water absorption features in the second and third LVs of oven-dried model are mainly due to hydration water (lattice OH).

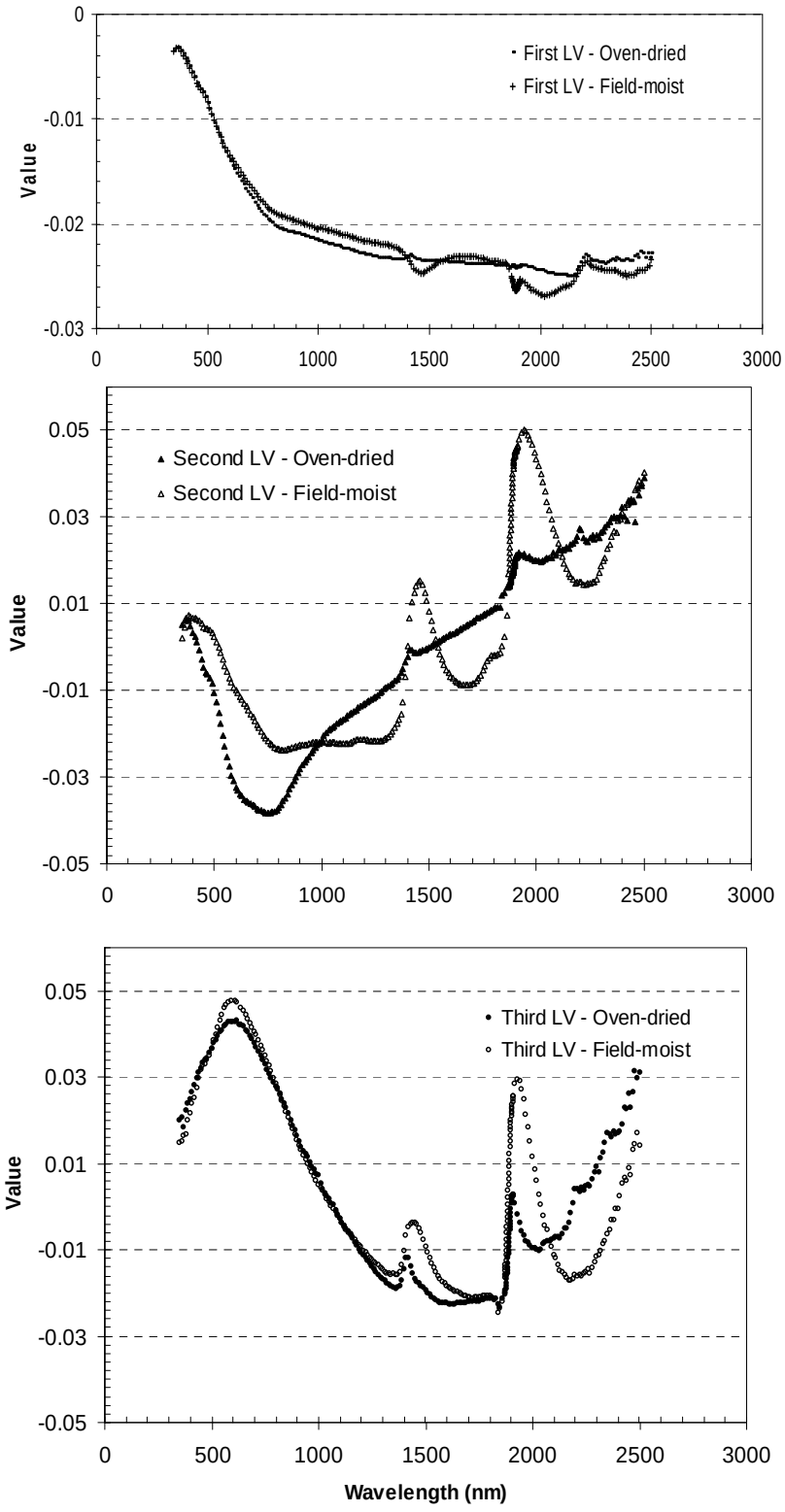


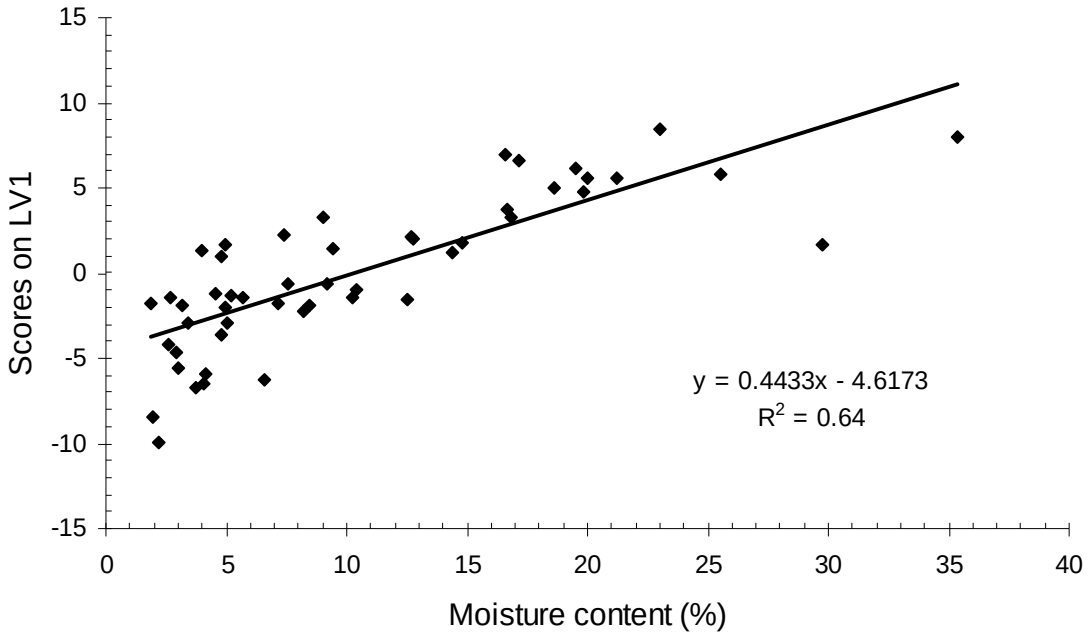
Figure 12. Comparison of the first three latent variables (LVs) for PLS models of field-moist and oven-dried soils

Table 6. Percent variance captured by regression model for oven-dried soils

Latent variable	Spectra		SOM	
	Percent variance	Cumulative	Percent variance	Cumulative
1	98.45	98.45	55.27	55.27
2	0.89	99.33	8.33	63.60
3	0.42	99.75	3.52	67.12
4	0.06	99.81	6.30	73.42
5	0.09	99.90	3.20	76.63
6	0.02	99.93	6.24	82.87
7	0.03	99.96	2.36	85.22
8	0.01	99.97	3.97	89.20

Table 7. Percent variance captured by regression model for field-moist soils

Latent variable	Spectra		SOM	
	Percent variance	Cumulative	Percent variance	Cumulative
1	98.50	98.50	21.42	21.42
2	0.93	99.42	46.46	67.88
3	0.36	99.79	0.59	68.47
4	0.09	99.88	1.63	70.10
5	0.03	99.90	5.41	75.51
6	0.04	99.95	3.30	78.82
7	0.03	99.98	2.84	81.66
8	0.01	99.98	5.14	86.80
9	0.00	99.98	6.15	92.95



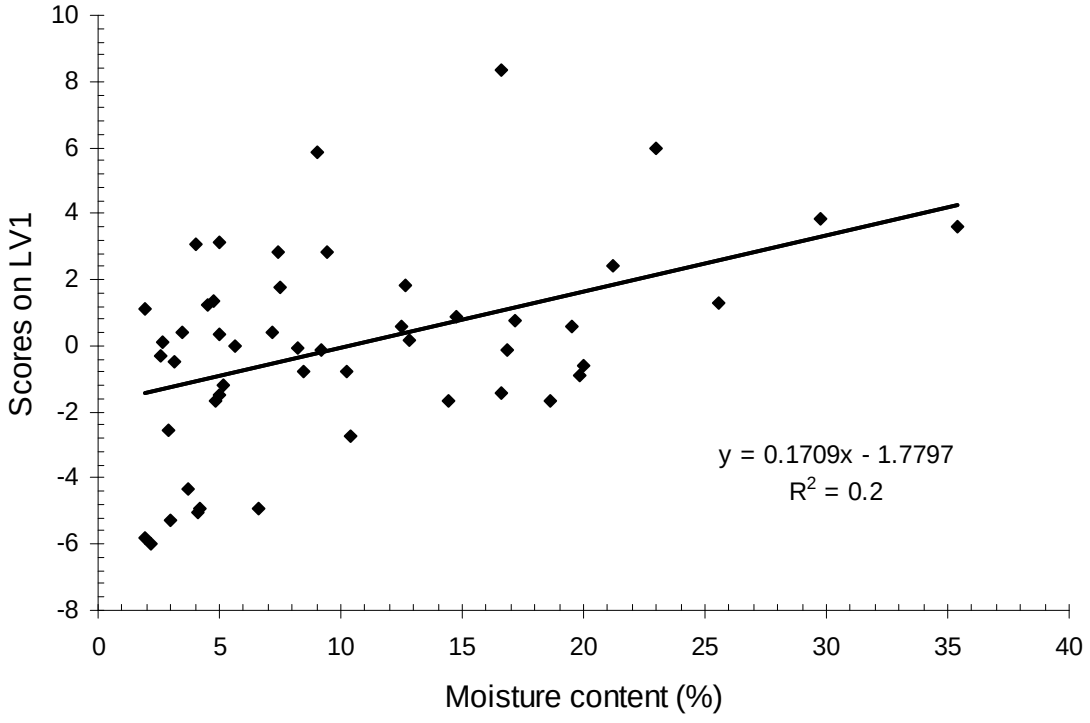
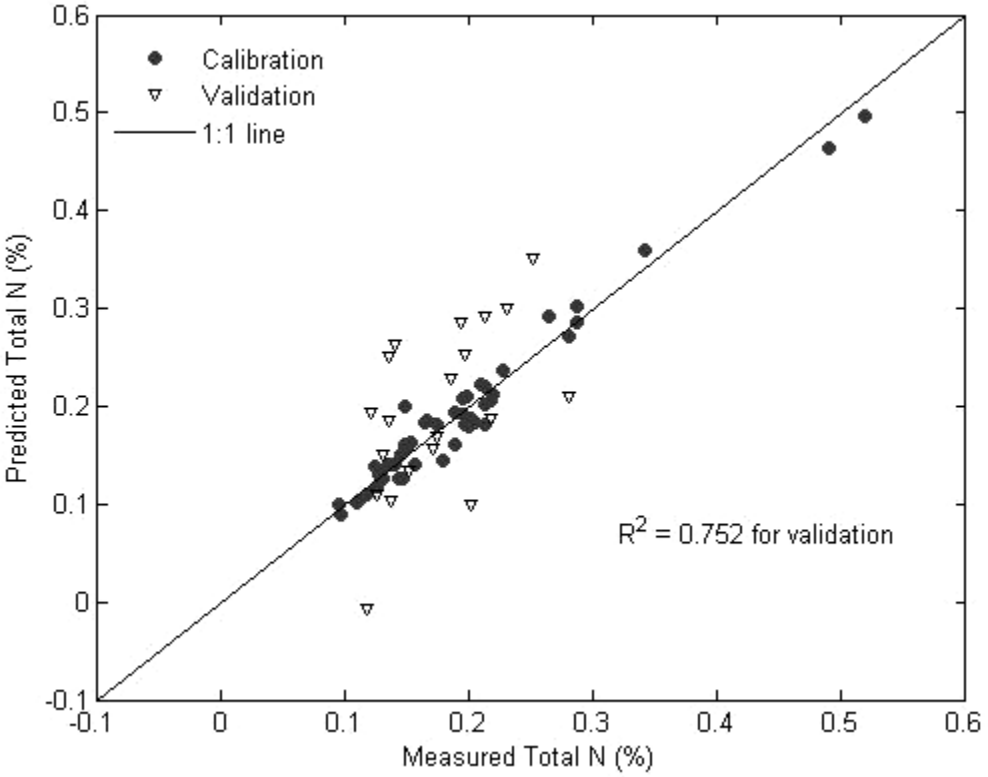


Figure 13. Moisture content (%) vs. scores on LV1 for the model of field-moist soils (up) and the model of oven-dried soils (bottom)

3.1.2 Total nitrogen

The PLS model with the calibration data resulted in an R^2 of 0.96 for field-moist soils and 0.91 for oven-dried soils, implying 96.11% of total N variance for field-moist samples, and 90.67% for the oven-dried is explained. The PLS model with the validation data set resulted in a coefficient of determination, R^2, of 0.75 for field-moist samples and an R^2 of 0.82 for oven-dried samples (Fig. 14). Like the PLS models for estimating SOM, the optimal PLS model for estimating N of field-moist soil samples used one more latent variable than that for oven-dried soils did. The oven-dried soil model resulted in a lower prediction error (RMSEP = 0.045) than the field-moist model did (RMSEP = 0.073). These results again support the conclusion that the performance of the PLS model is affected by soil moisture.

Figure 15 shows the first three loading vectors of the PLS model for oven-dried soils, which totally capture more than 60% of the total N variance. The first LV reflects albedo variations, the second LV shows the diagnostic feature due to organic matter (0.5 ～1.0 μm) and the water absorption features at 1.4 and 1.9 μm, and the third LV shows water absorption features at 1.4 and 1.9 μm with an unknown diagnostic area at 0.6 μm. It is surprising that these three loading vectors are the exactly same as those for SOM. This is attributed to the high correlation between total N and SOM (correlation coefficient, R = 0.96). This also explains why total N can be estimated from soil spectra though total N has no diagnostic spectral characteristics.



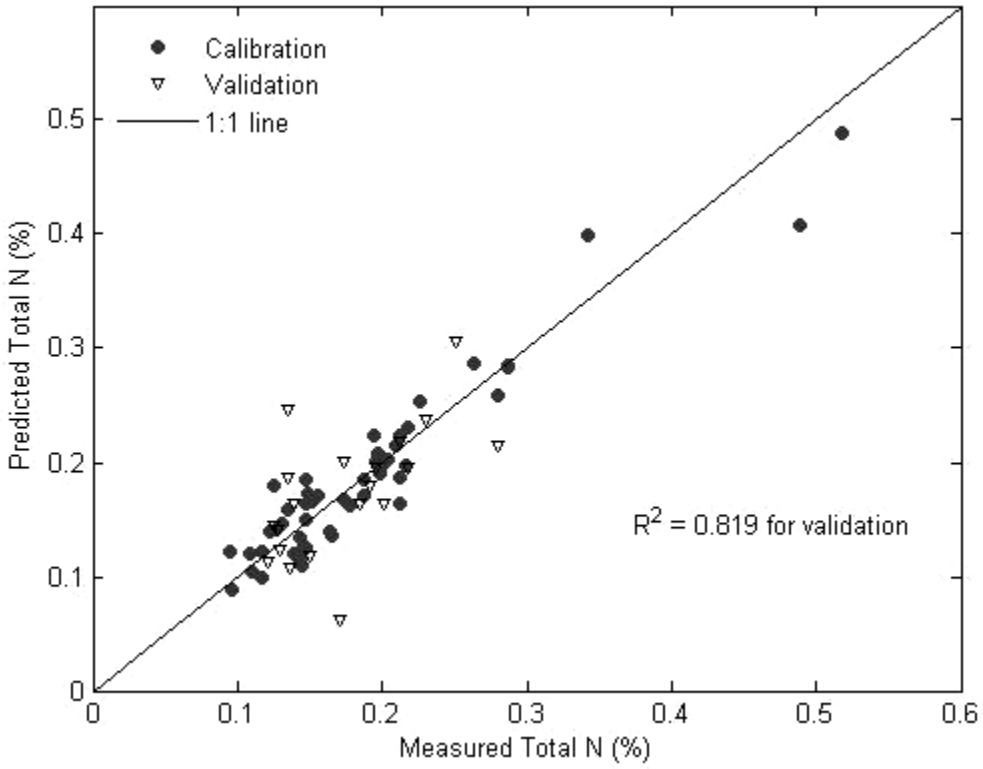


Figure 14. Comparison of measured total N with predicted total N by PLS modeling of field-moist soils (up) and oven-dried soils (bottom)

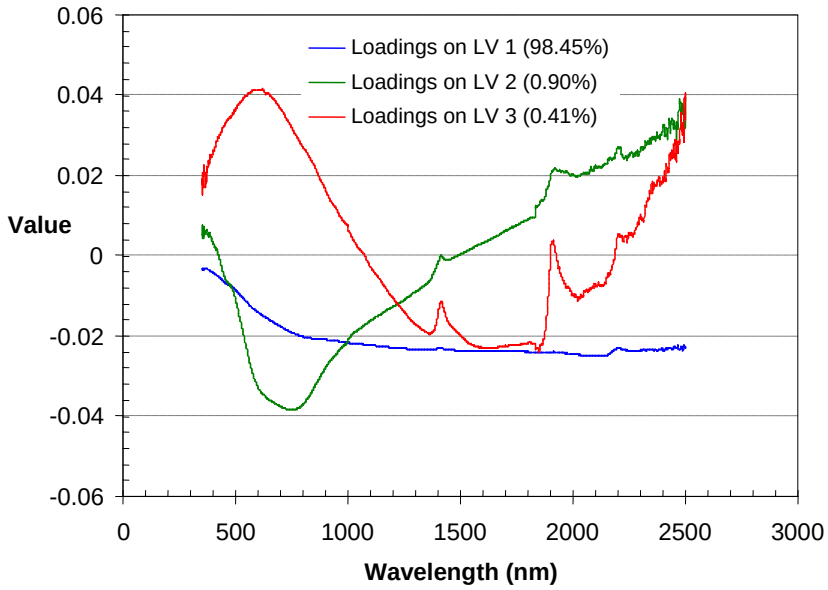


Figure 15. Loadings of the first three latent variables (LVs) for PLS modeling of total N with oven-dried soil spectra

3.1.3 Total phosphorous

The PLS model with the calibration data set resulted in a coefficient of determination, R^2, of 0.68 for field-moist soils and an R^2 of 0.79 for oven-dried soils, indicating that 68.29% of total P variance for field-moist samples, and 79.13% of total P variance for oven-dried samples is modeled. The PLS model with the validation data set resulted in an R^2 of 0.67 for field-moist soil samples and R^2 of 0.73 for oven-dried soil samples (Fig. 16). Differing from the PLS models for estimating organic matter and total N, the optimal oven-dried model requires one more latent variable than that for the field-moist soils does, possibly because more spectral information can be captured after removing the masking effect of soil moisture. The field-moist model resulted in a lower prediction error (RMSEP = 137.59 mg kg^{-1} soil) than did the oven-dried model (RMSEP

= 148.12 mg kg^{-1} soil). However, a strong relationship between oven-dried soil spectra and total P for the validation data set was obtained.

The first three LVs reflecting more than 60% of total P variance are shown in Figure 17. The first two LVs in the PLS models for both field-moist and oven-dried soils show the similar spectral characteristics to those for SOM models, but the third LVs are different from those for the SOM PLS models. The third LVs of the PLS model for estimating field-moist soil P show water absorption features at 1.4 and 1.9 μm, but in an opposite direction compared to the third LV of the SOM PLS model. The 3rd LV of SOM model compensated for moisture effect, whereas the 3rd LV of P model actually captures the water absorbance feature. The third LV in oven-dried model shows a distinct spectral signature at 0.5 μm, which is coincident with the absorbance spectra of inorganic phosphate. This spectral signature is present in the absorbance spectrum of iron (III) phosphate dehydrate shown in Figure 18 along with the absorbance spectra of three other inorganic P compounds (Bogrekci et al., 2005b). The third LV shares several common spectral signatures with the absorbance spectra of those inorganic phosphates: the peaks at 1.4, 1.9 and 2.2 μm and the absorptions at 2.1 and 2.3 μm. The difference in spectral features reflected by the third LVs between field-moist model and oven-dried model indicates that soil moisture impedes PLS to capture spectral response of P. The higher coefficient of determination for oven-dried soil suggests that the prediction of P with PLS can be improved by removing moisture effect.

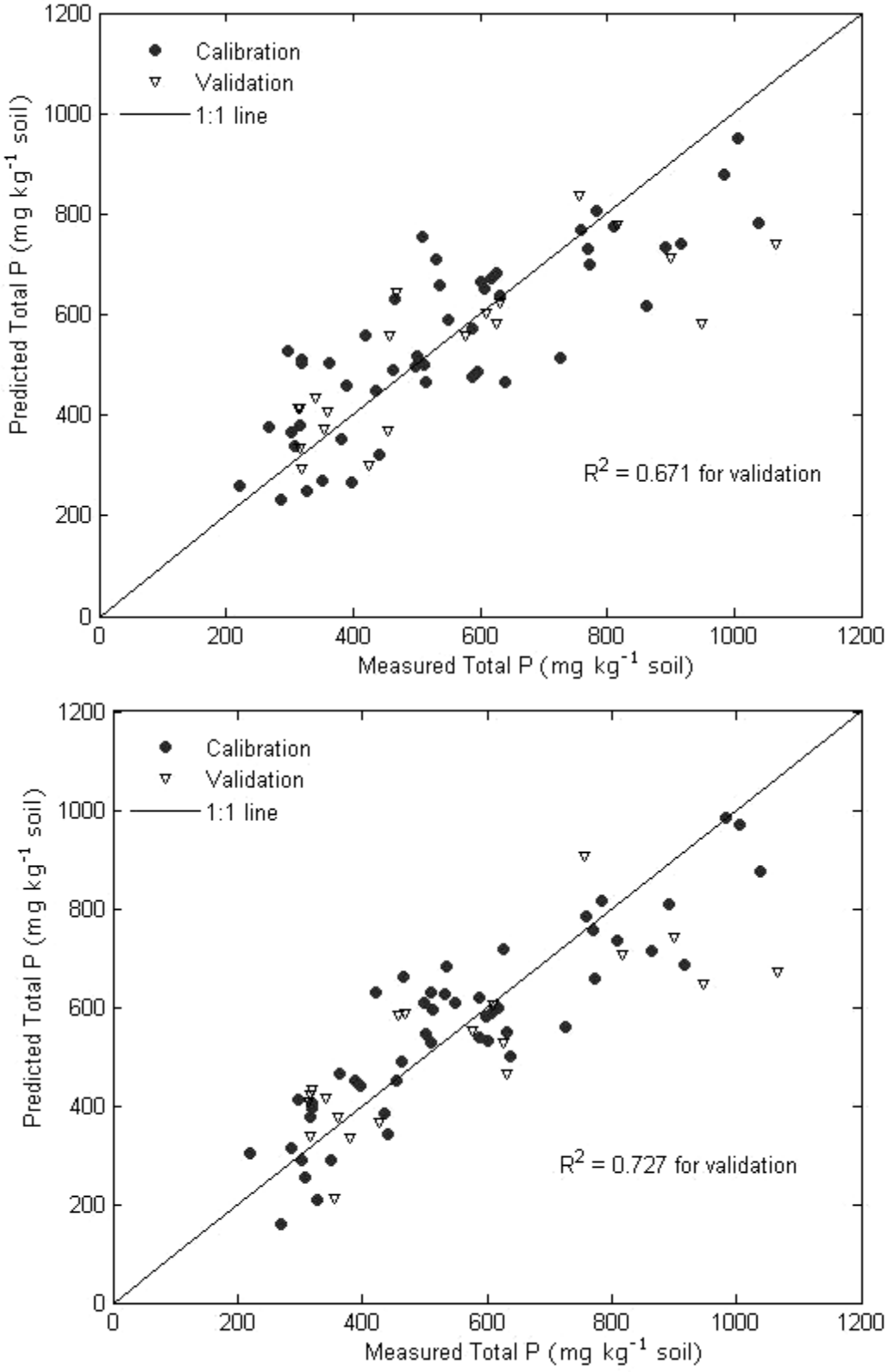


Figure 16. Comparison of measured total P with predicted total P by PLS modeling of field-moist soils (up) and oven-dried soils (bottom)

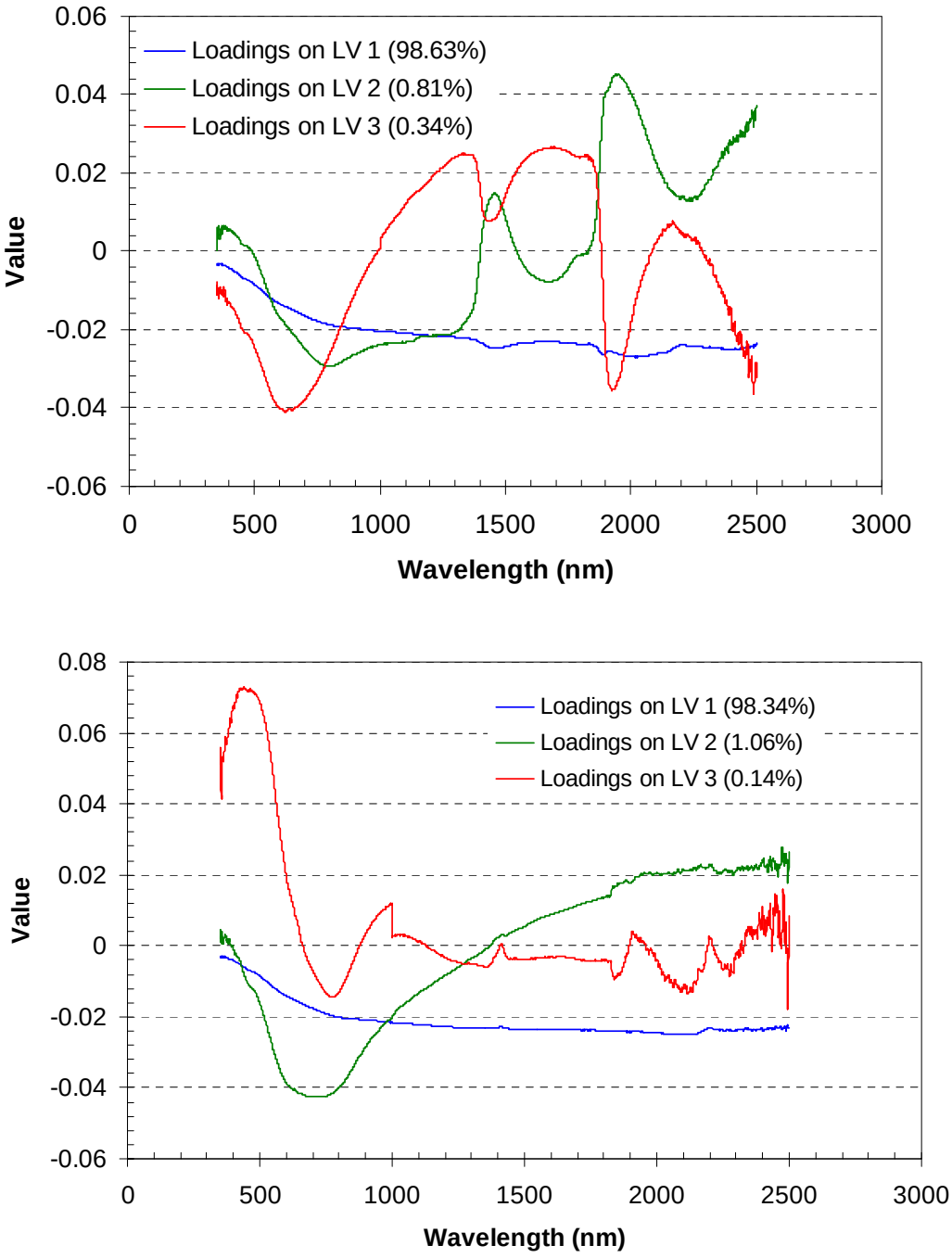


Figure 17. Loadings of the first three latent variables (LVs) for PLS modeling of total P with field-moist soil spectra (up) and oven-dried spectra (bottom)

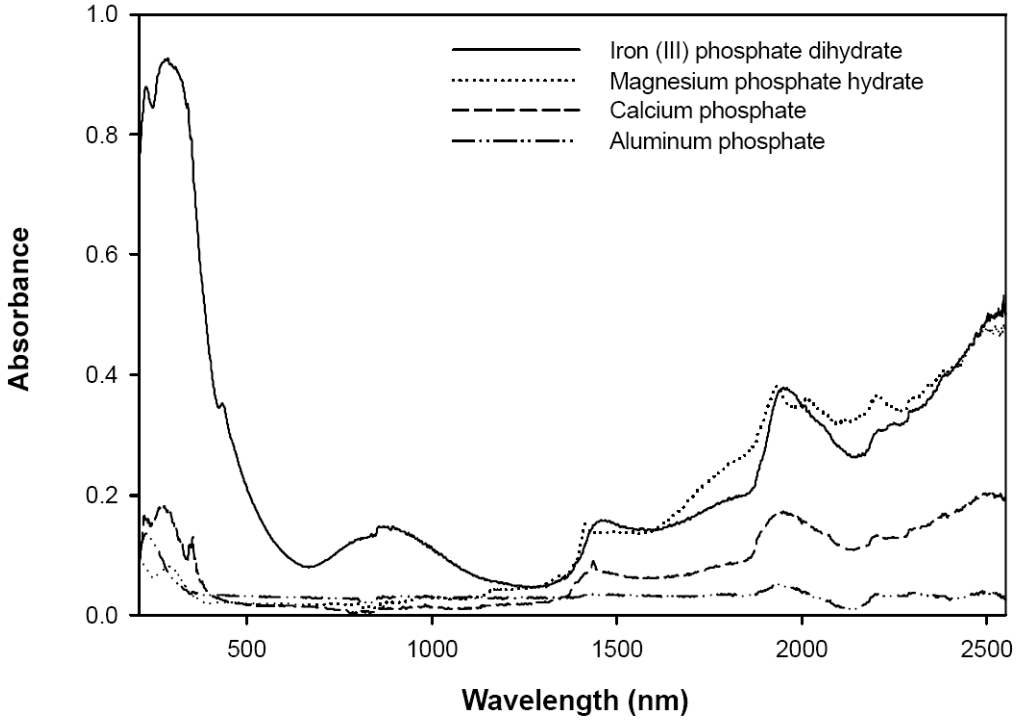


Figure 18. Absorbance of four P compounds in 208 - 2550 nm (from Brogrekci et al., 2005b)

3.2 PLS Modeling with Field Spectra

Unlike those measured in well controlled laboratory conditions, reflectance spectra measured in the field are affected by illumination changes, soil roughness and so forth, adding difficulties in mapping soil properties with the field spectra. The PLS capability of estimating soil properties with the field spectra is tested so that the potential of imagery for mapping soil properties can be known. PLS modeling of field spectra was based on 58 out of 70 samples because the 11 soil samples collected in Eagle-Cicero were excluded due to the effect of vegetation on their spectra, and one sample (C204) was eliminated as a spectral outlier. The field spectra have a total of 1920 bands after the

removal of some noisy bands centered at 1.4 and 1.9 μm. Fifty-eight samples were split into a calibration data set (44 samples) and a validation data set (14 samples).

Results from PLS modeling for SOM, total N and total P are summarized in Table 8. These results demonstrate the potential of PLS method for remotely estimating SOM, total N and P of agricultural soils from Hyperion images.

3.2.1 SOM

PLS modeling of field spectra resulted in an R^2 of 0.77 for calibration, an r^2 of 0.74 for validation (Fig. 19) and a RMSEP of 0.84. The R^2 (0.74) resulting from the field spectra model is slightly lower than that from the lab spectra model of field-moist soils ($R^2 = 0.76$).

Table 8. Results from PLS modeling with field spectra

Soil properties	No. of LVs	Validation		Calibration	
		R^2	RMSEP	R^2	RMSEC
OMC %	6	0.74	0.839	0.77	0.978
Total N %	7	0.79	0.039	0.83	0.038
Total P (mg kg^{-1} soil)	6	0.60	188.12	0.68	133.31

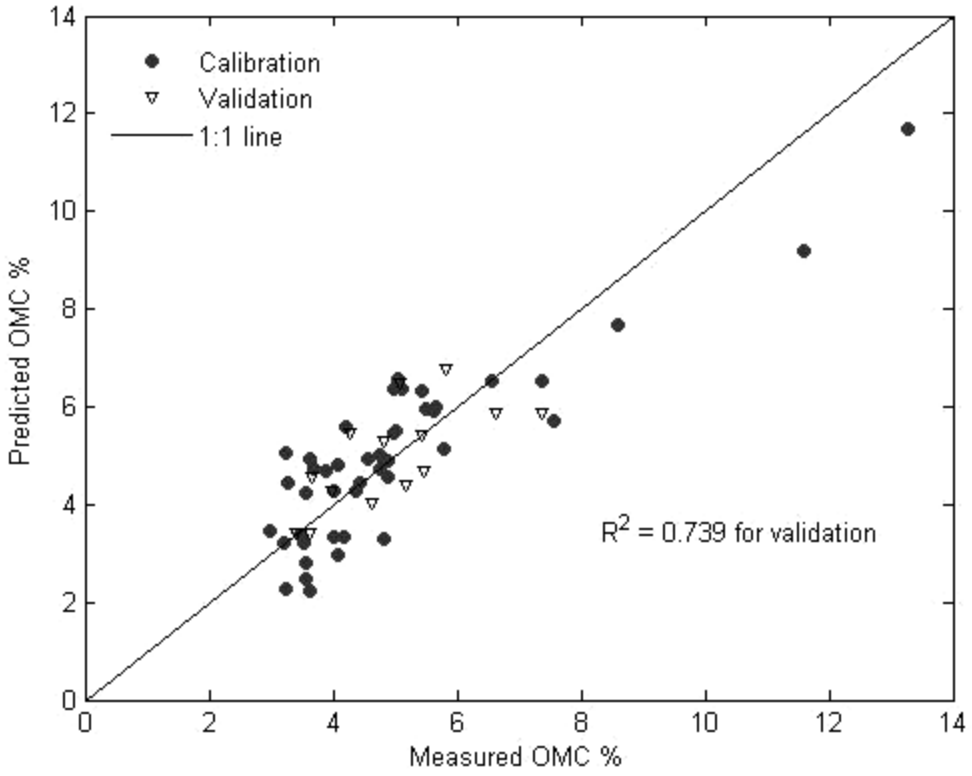


Figure 19. Measured OMC vs. OMC predicted by PLS modeling with field spectra

3.2.2 Total nitrogen

PLS modeling of field spectra resulted in an R^2 of 0.83 for calibration, an R^2 of 0.79 for validation (Fig. 20) and a RMSEP of 0.039. For validation, the R^2 (0.79) resulting from PLS modeling of field spectra is slightly higher than that from PLS modeling of lab spectra of field-moist soils ($R^2 = 0.75$).

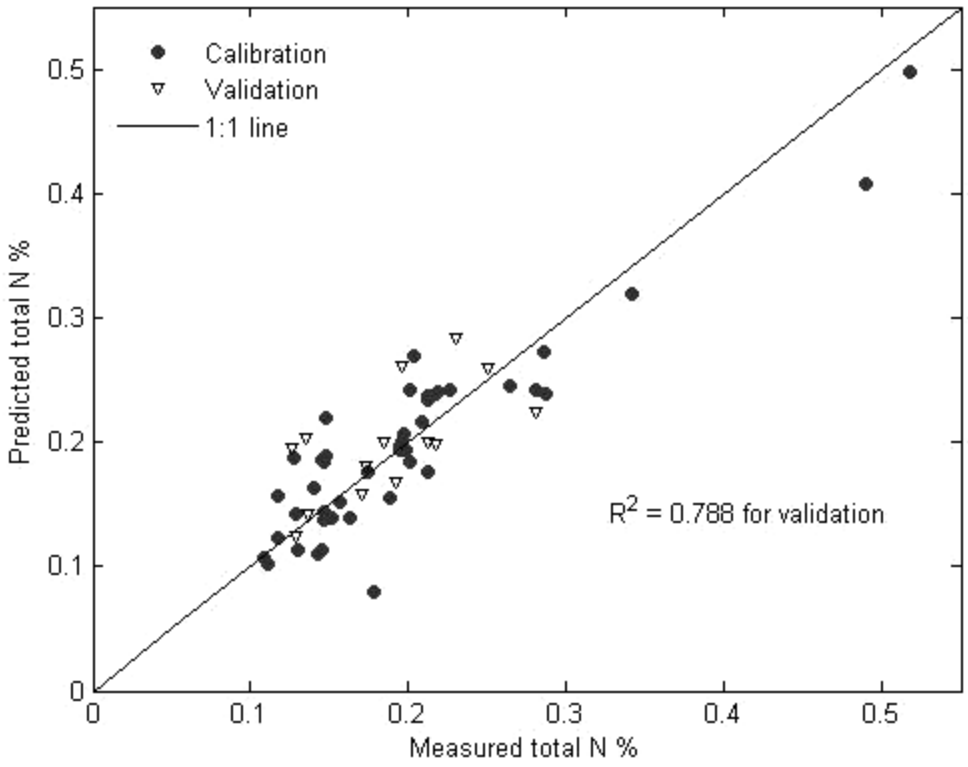


Figure 20. Measured total N vs. total N predicted by PLS modeling with field spectra

3.2.3 Total phosphorous

PLS modeling of field spectra resulted in an R^2 of 0.68 for calibration, an R^2 of 0.60 for validation (Fig. 21) and a RMSEP of 188.12 mg kg^{-1} soil. The R^2 (0.60) resulting from PLS modeling of field spectra is slightly lower than that from PLS modeling of lab spectra of field moist soils (R^2 = 0.67).

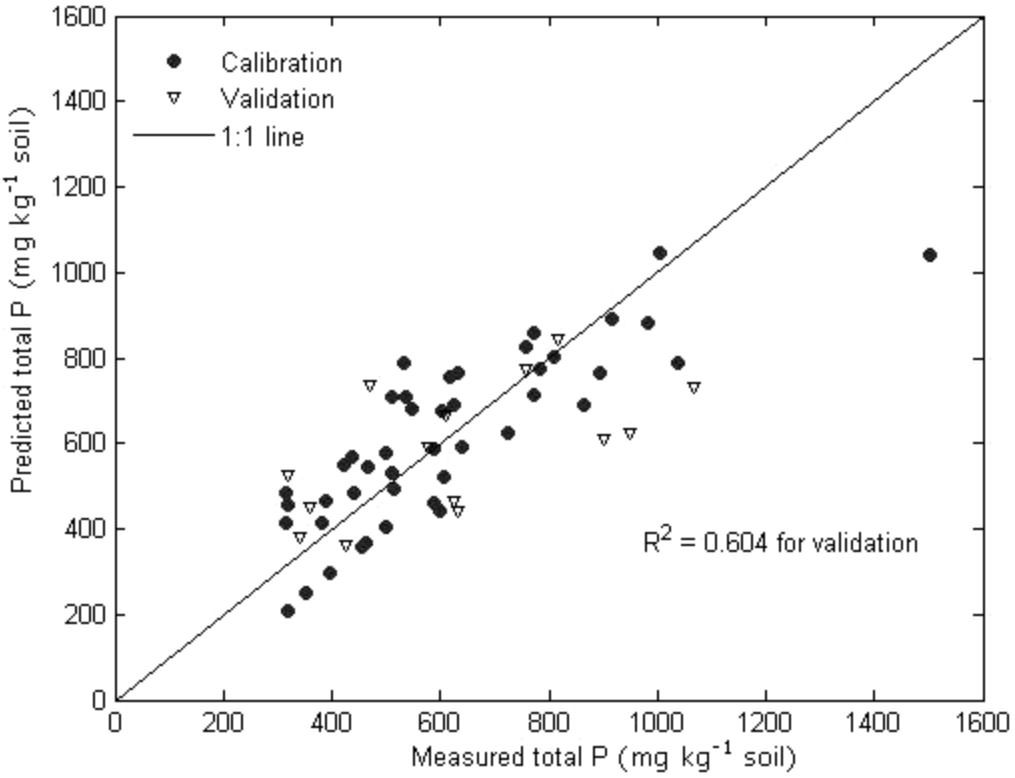


Figure 21. Measured total P vs. total P predicted by PLS modeling with field spectra

3.3 PLS Modeling with Hyperion Spectra

Based on the PLS performance with laboratory and field spectral data, the next step was taken to test the PLS feasibility with Hyperion data. PLS modeling of Hyperion image spectra was based on 58 out of 70 samples because the 11 soil samples collected in Eagle-Cicero were excluded, and the spectrum for the sample F504 was unavailable because its sampling location was outside of the image. The resultant PLS models were applied to the Hypersion images to generate spatial distribution maps of soil OM, total N and total P. The soil properties distribution maps of Cicero-2 and Cicero-3 were mosaiced together after they had been generated. This was carried out using ENVI mosaicing tool. Cicero-3 is the bottom image, while Cicero-2 is the top image.

3.3.1 SOM

The PLS model with Hyperion spectra resulted in an R^2 of 0.84 for calibration, R^2 of 0.74 for validation (Fig. 22) and a RMSEP of 1.42. PLS performed equivalently well on the Hyperion spectra ($R^2 = 0.74$) and field spectra ($R^2 = 0.74$).

In order to test the performance and the transferability of the PLS model, the data collected for three out of four scenes were combined to form a calibration data set and the remaining data were used for validation. The PLS modeling results for each validation test are summarized in Table 9 and shown in Figure 23. The highest R^2 is 0.87 for calibration and 0.74 for validation. The relatively low R^2 values for validation ($R^2 = 0.65$, 0.66) suggest that the corresponding calibration data set is not representative enough for Fall Creek soil samples, but the performance of the model could be improved if more soil samples collected from Fall Creek Watershed were included in the calibration model. In addition, the model could be more robust if the model was built separately for each watershed soils since the soil series in the two watersheds (Cicero Creek and Fall Creek) are not similar to each other. Fall Creek Watershed doesn't have Patton which is one of the dominant soil series in Cicero Creek Watershed, while the soil association map shows that Fox-Ockley-Westland is present in Fall Creek Watershed, not in Cicero Creek Watershed. Patton and Westland soils are poorly drained, while Fox and Ockley soils are well drained (NRCS, 2008b).

Figure 23 shows that the PLS models for scene 1 and scene 4 under predicted SOM (A and D), while the models for scene 2 and scene 3 over estimated SOM (B and C). One of possible reasons for this over/under estimation is relevant to changes in image acquisition conditions among Hyperion scenes such as illumination and instrumental

stability. The effect of these changes in imaging conditions on the estimation of SOM can be eliminated by normalization. Normalization can force spectral data of one image (slave image) to match those of master image. Normalization between adjacent images can potentially improve the model's performance. However, normalization requires that adjacent images are overlapping. In this study, only two images are adjacent and overlapped, but other scenes are not. Therefore normalization was not implemented in image preprocessing.

The PLS calibration model with 58 soil samples was applied to the Hyperion images from which SOM spatial distribution maps were generated. The SOM distribution maps are dominated by orange color which indicates 2 to 4% of SOM (Fig. 24 and 25). In Fall Creek Watershed, its upper portion, in general, has a higher amount of SOM than the lower portion.

Table 9. PLS results for estimating SOM with different calibration datasets

Calibration			Validation		
Data	R^2	RMSEC	Data	R^2	RMSEP
44 (Random)	0.84	0.83	14	0.74	1.42
Scene 2, 3, 4	0.78	0.93	Scene 1	0.72	1.21
Scene 1, 3, 4	0.87	0.61	Scene 2	0.71	1.65
Scene 1, 2, 4	0.77	0.95	Scene 3	0.65	1.97
Scene 1, 2, 3	0.83	0.74	Scene 4	0.66	1.87

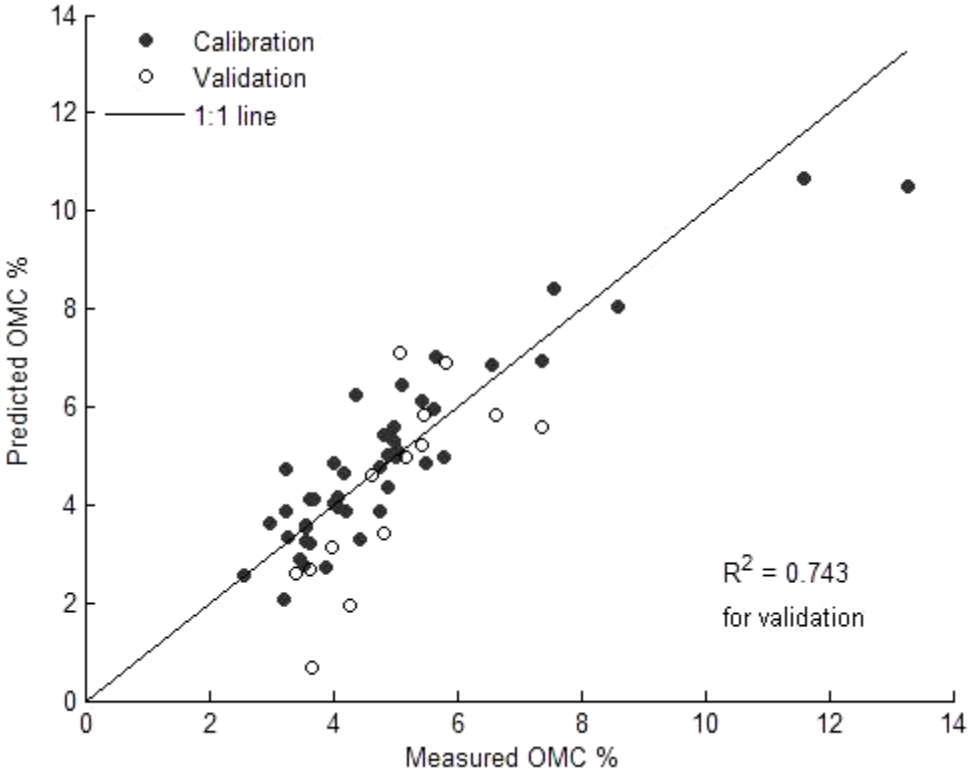


Figure 22. Measured OMC vs. OMC predicted by PLS modeling with Hyperion spectra (randomly split)

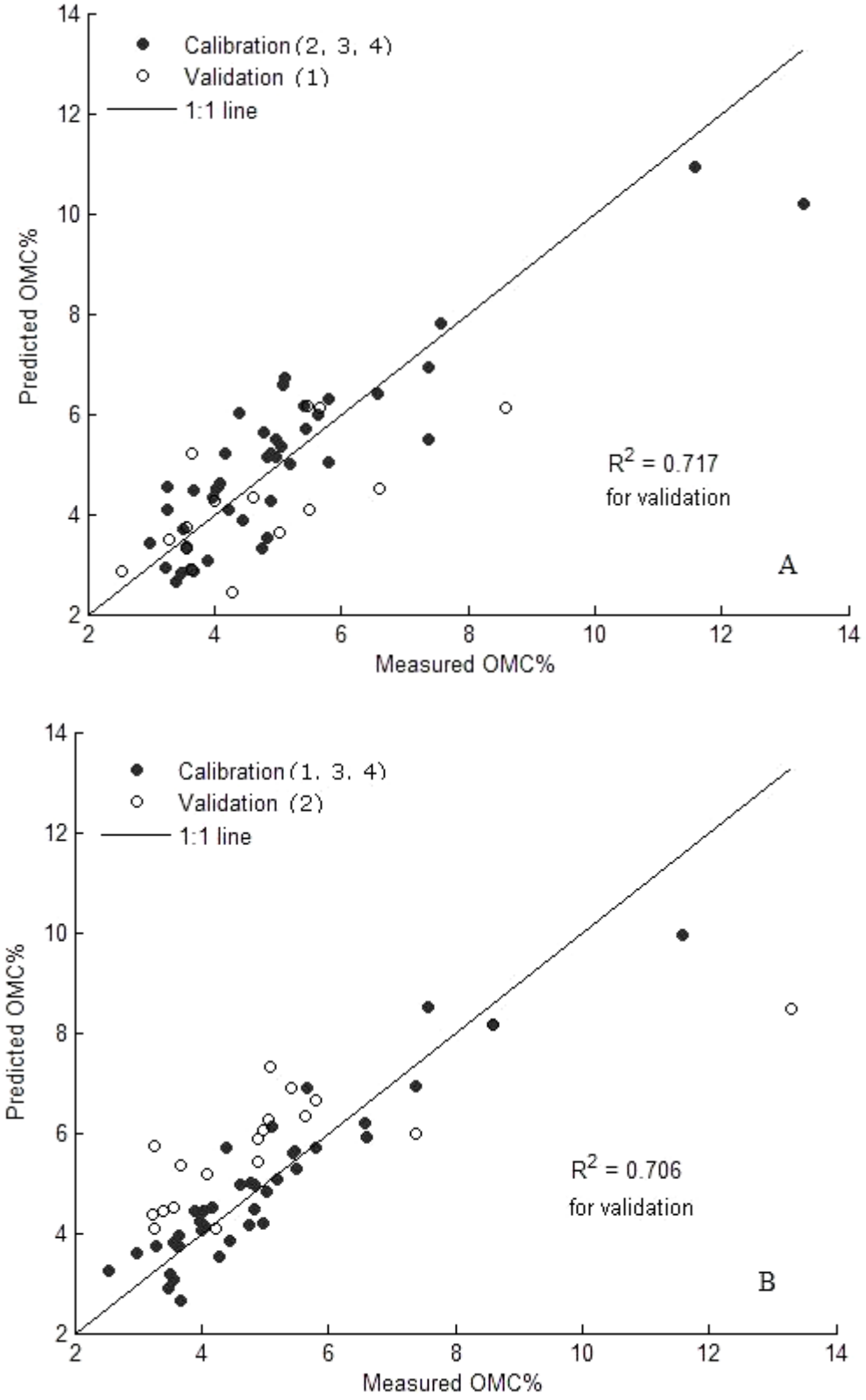
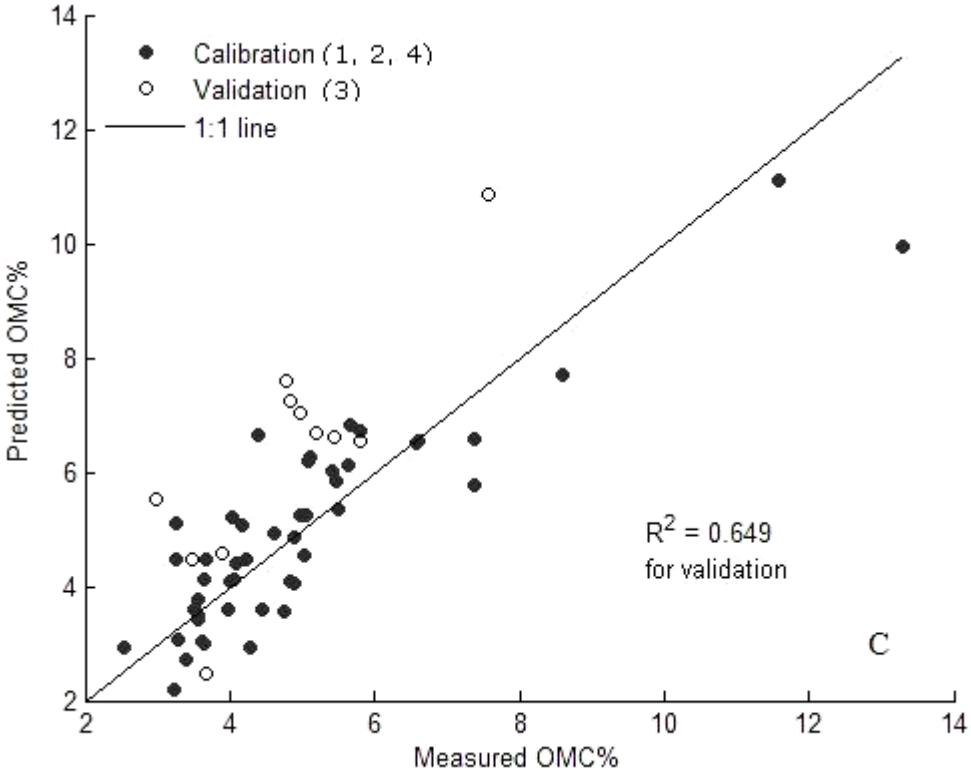


Figure 23. Measured OMC vs. OMC predicted by PLS modeling with Hyperion spectra: A) scene 1 for validation, B) scene 2 for validation



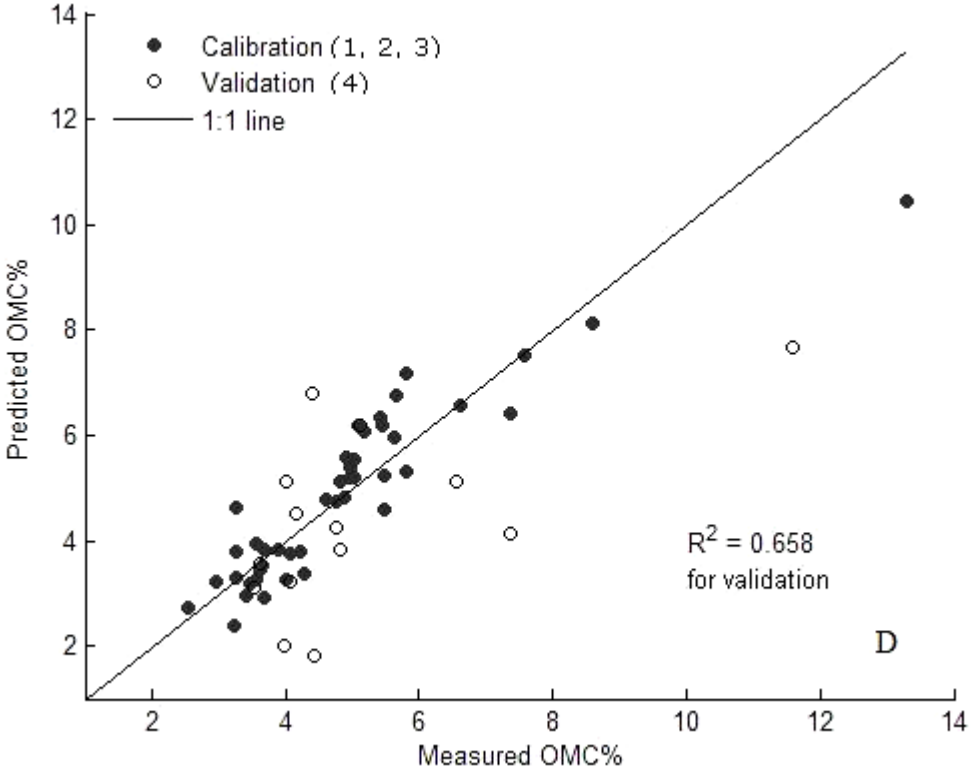


Figure 23 (cont.). Measured OMC vs. OMC predicted by PLS modeling with Hyperion spectra: C) scene 3 for validation, and D) scene 4 for validation

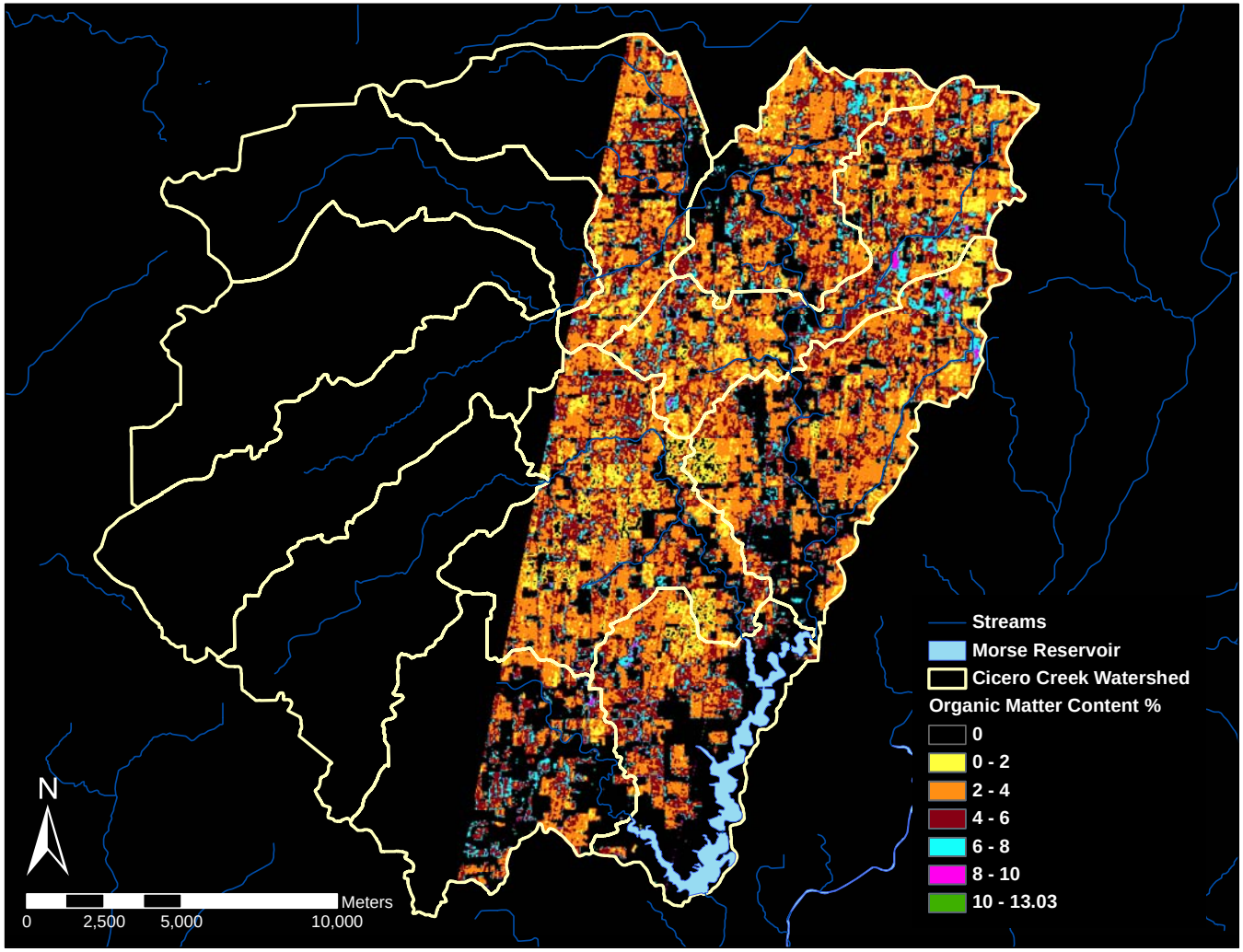


Figure 24. Predicted soil OM distribution map in Cicero Creek Watershed

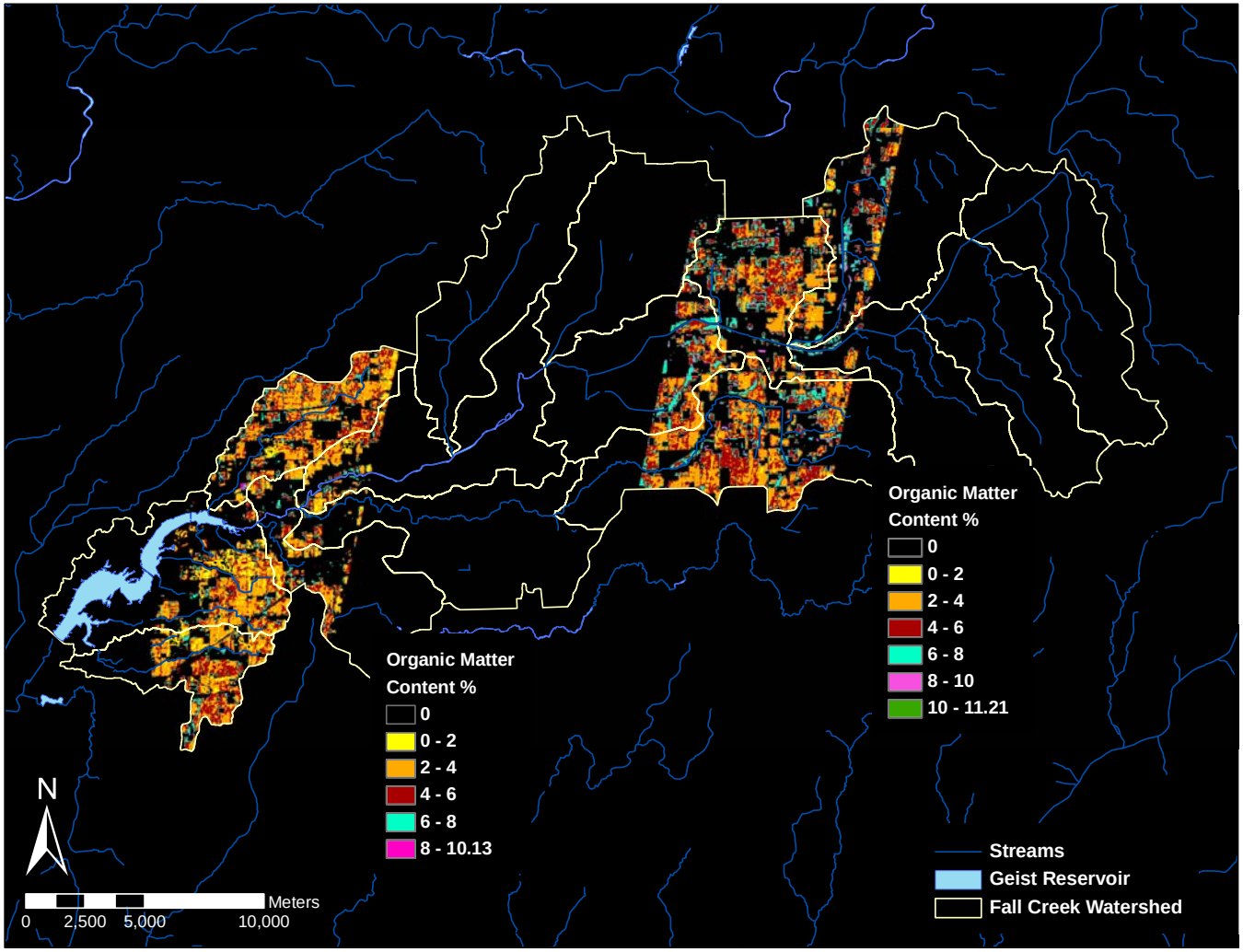


Figure 25. Predicted soil OM distribution map in Fall Creek Watershed

3.3.2 Total nitrogen

PLS modeling of Hyperion spectra resulted in an R^2 of 0.81 for calibration, R^2 of 0.72 for validation (Fig. 26) and a RMSEP of 0.058. For validation, the PLS model with Hyperion data (R^2 = 0.72) performed marginally poor compared with the PLS model using field spectra (R^2 = 0.79).

The PLS modeling results for each validation test are summarized in Table 10. The highest R^2 is 0.86 for calibration and 0.76 for validation. Similar to the case for estimating SOM, the PLS model for the estimation of total N also resulted in a low R^2 value (R^2 = 0.65) when the model was validated onto the datasets of the scenes 3 and 4. The samples from scene 2 and scene 3 were mostly over predicted (Fig. 27B and C) for total N, while the soil samples from scene 4 were mostly under predicted (Fig. 27D).

The PLS calibration model with 58 soil samples was applied to the Hyperion images and total N spatial distribution maps were generated. Total N distribution maps show that soils mostly have 0.1 to 0.2% of total N (Fig. 28 and 29). Like SOM distribution maps, total N spatial distribution maps show that upper portion of Fall Creek Watershed has a higher amount of total N than the lower portion.

Table 10. PLS results for estimating total N with different calibration datasets

Calibration			Validation		
Data	R^2	RMSEC	Data	R^2	RMSEP
44 (Random)	0.81	0.037	14	0.72	0.058
Scene 2, 3, 4	0.77	0.040	Scene 1	0.73	0.041
Scene 1, 3, 4	0.86	0.024	Scene 2	0.76	0.060
Scene 1, 2, 4	0.79	0.038	Scene 3	0.65	0.082
Scene 1, 2, 3	0.82	0.031	Scene 4	0.65	0.078

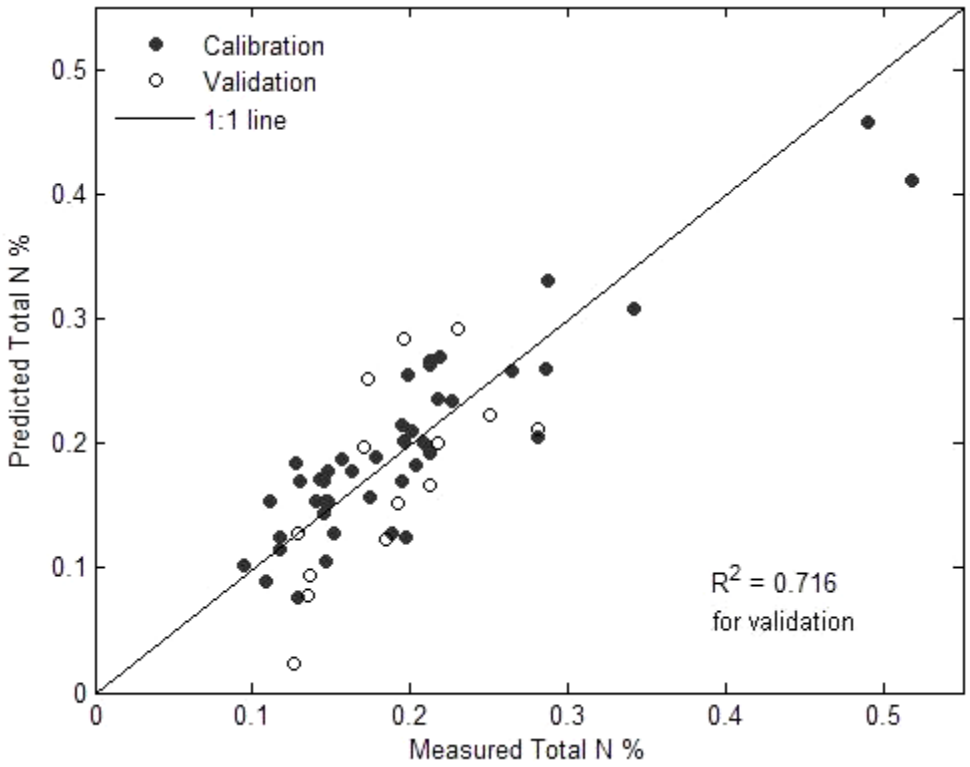


Figure 26. Measured total N vs. total N predicted by PLS modeling with Hyperion spectra (randomly split)

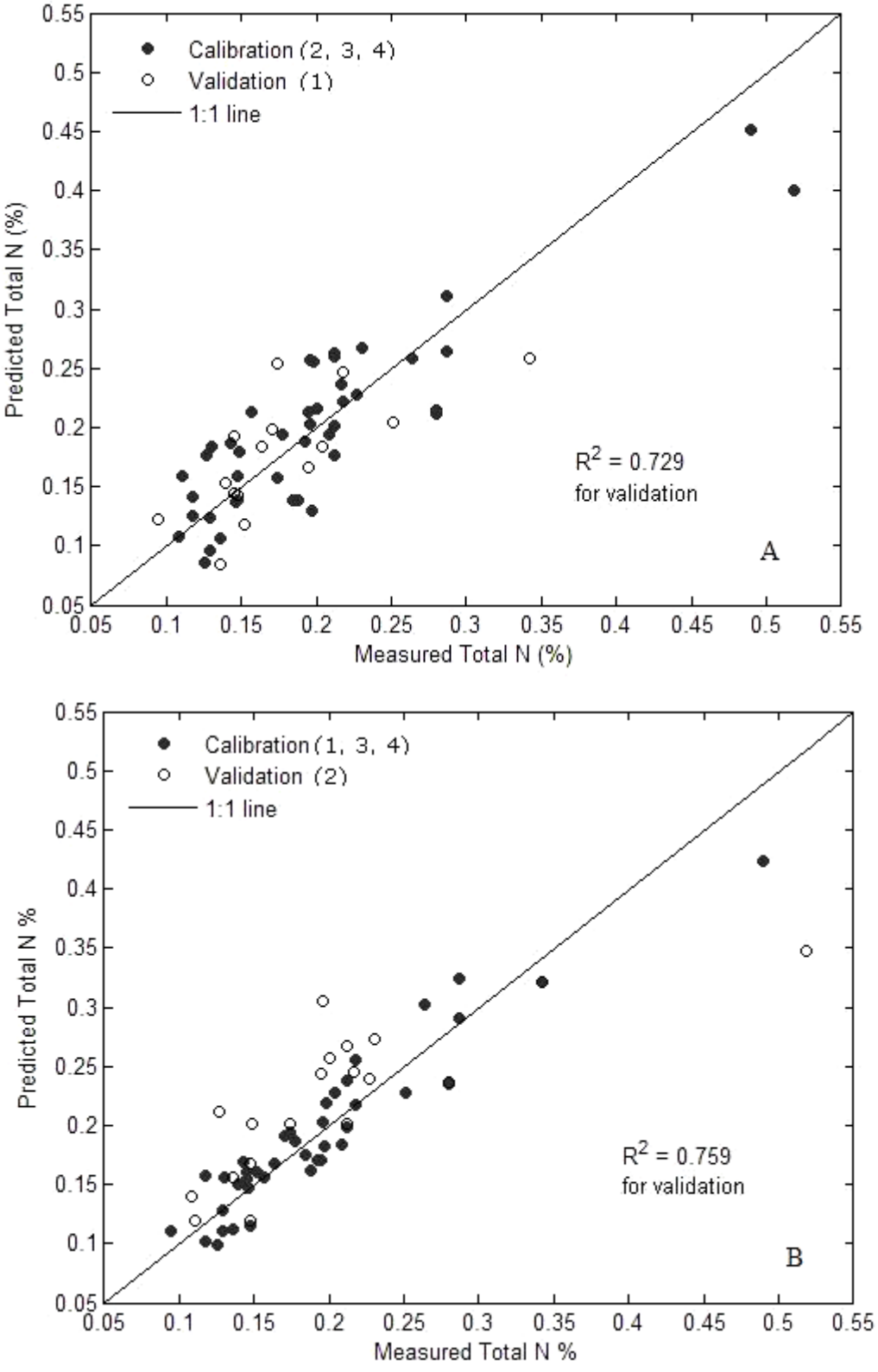
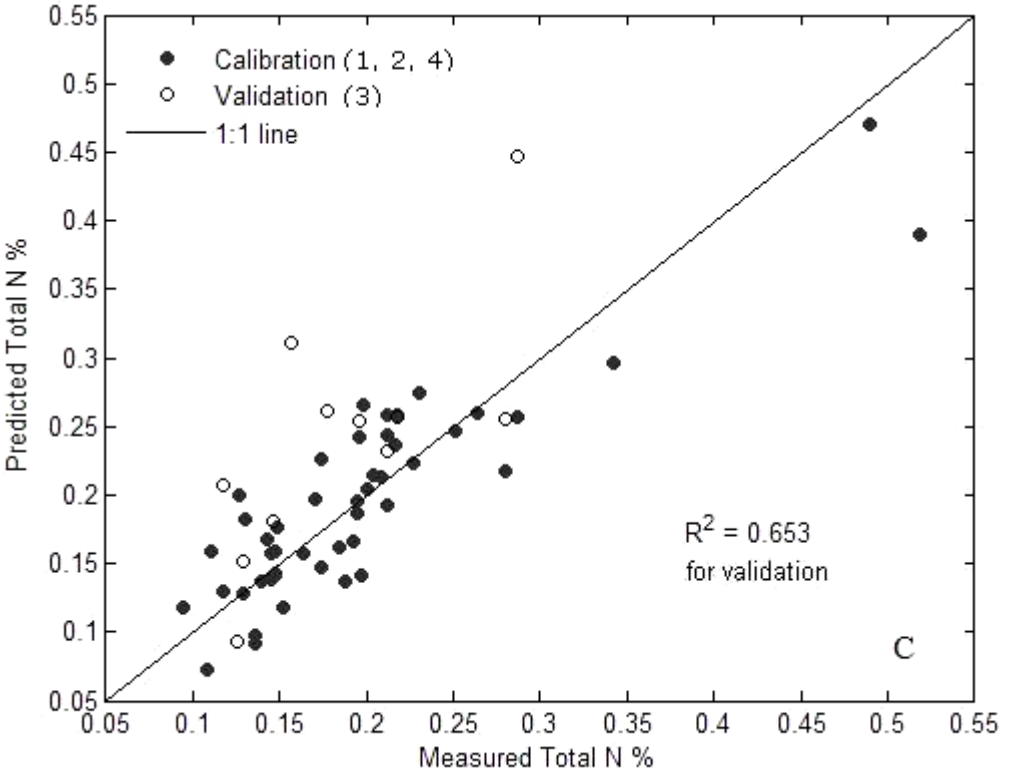


Figure 27. Measured total N vs. total N predicted by PLS modeling with Hyperion spectra: A) scene 1 for validation, B) scene 2 for validation



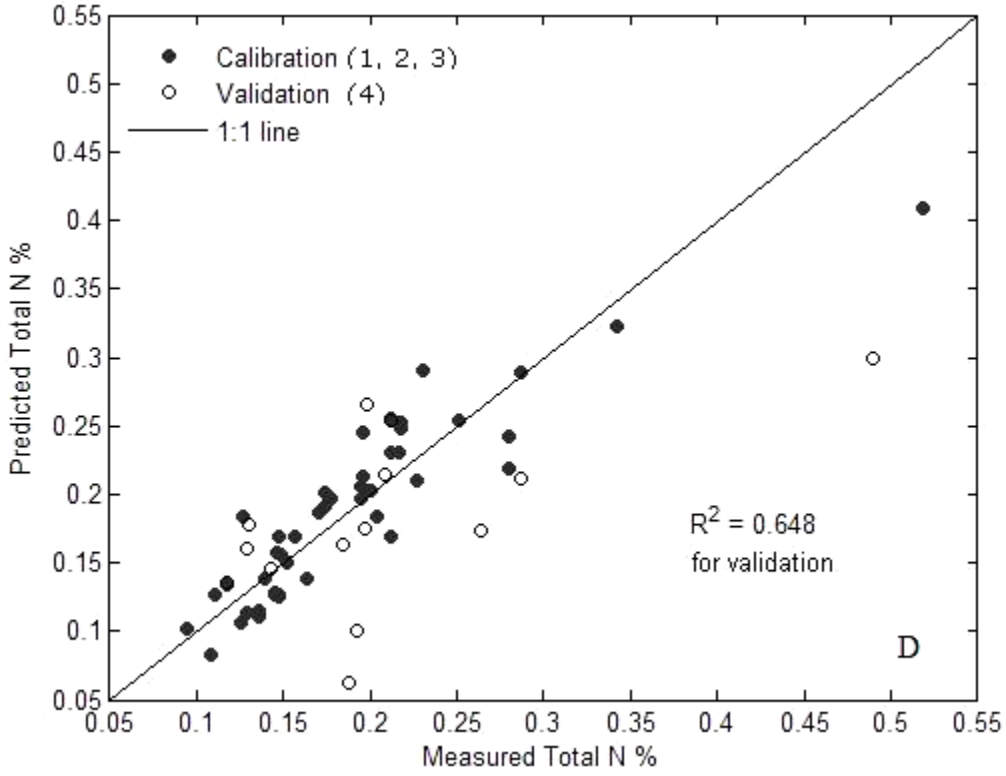


Figure 27 (cont.). Measured total N vs. total N predicted by PLS modeling with Hyperion spectra: C) scene 3 for validation, and D) scene 4 for validation

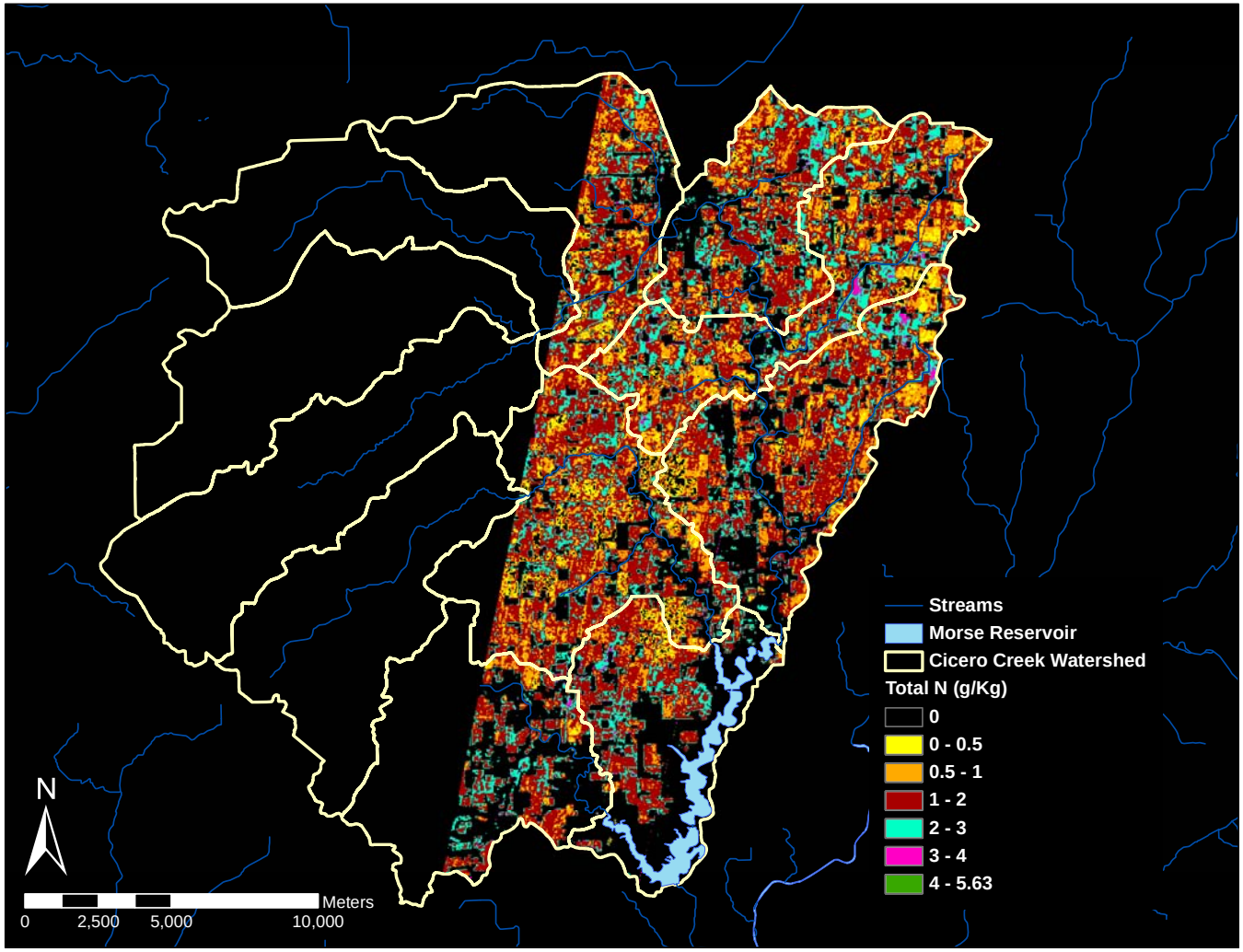


Figure 28. Predicted soil total N distribution map in Cicero Creek Watershed

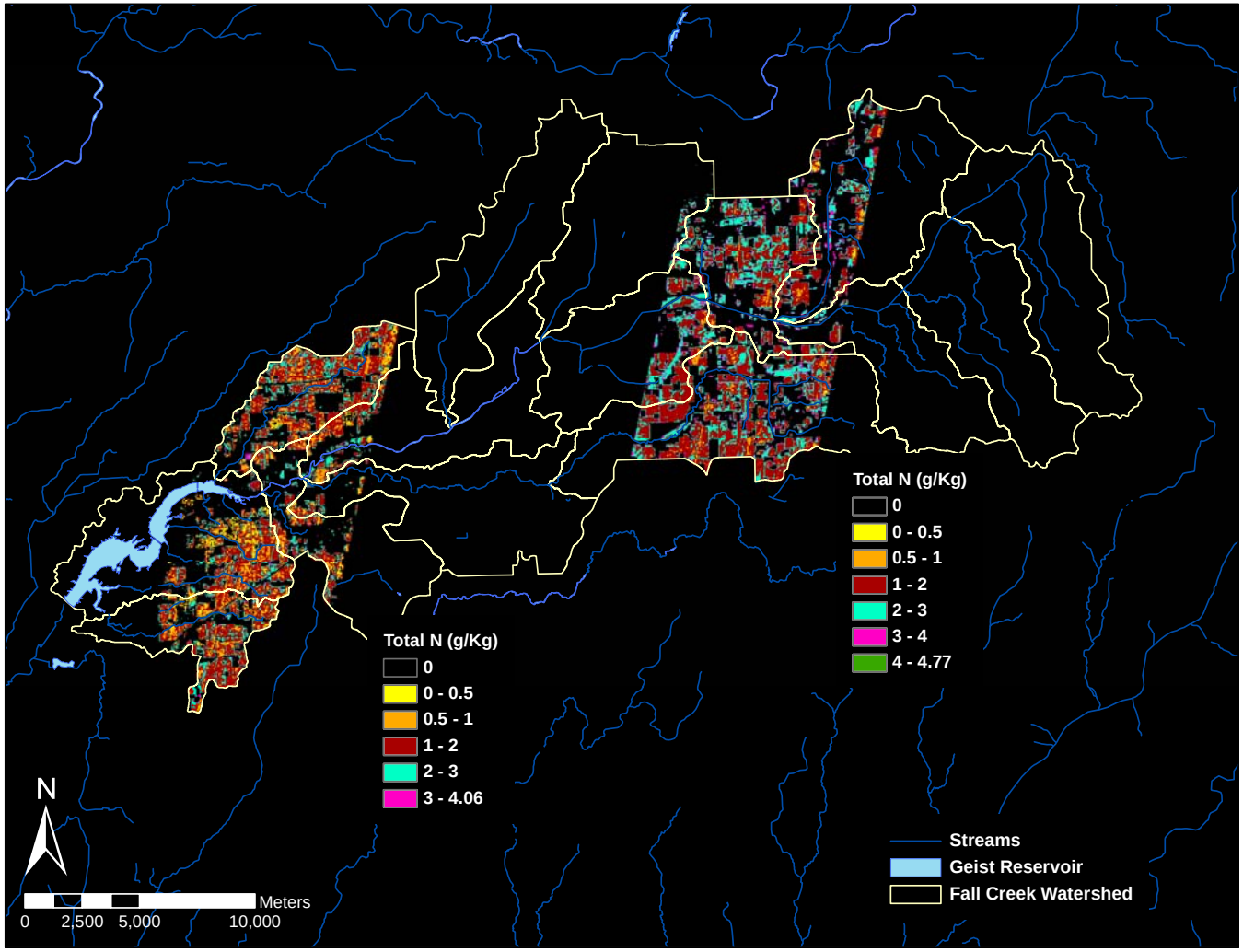


Figure 29. Predicted soil total N distribution map in Fall Creek Watershed

3.3.3 Total phosphorous

PLS modeling of Hyperion spectra resulted in an R^2 of 0.82 for calibration, R^2 of 0.67 for validation (Fig. 30) and a RMSEP of 228.17. The R^2 (0.67) resulting from PLS modeling of Hyperion spectra is slightly higher than that of field spectra model (R^2 = 0.60).

The PLS modeling results for each validation test are summarized in Table 11. The highest R^2 is 0.82 for calibration and 0.67 for validation. PLS modeling for estimating total P yielded a relatively low R^2 value (0.52) when PLS was validated onto the datasets of scene 3. The samples from scene 2 and scene 3 are mostly over predicted (Fig. 31B and C) for total P, while the samples from scene 1 are mostly under predicted (Fig. 31A). The relatively low R^2 value (0.52) for the validation on the dataset of scene 3 implies the lack of representative of calibration data set.

PLS calibration model with 58 soil samples was applied to the all the Hyperion images from which total P spatial distribution maps were generated. The P distribution maps are mostly shown in orange indicating a range of 400 and 600 mg kg^{-1} soil of total P (Fig. 32 and 33).

Table 11. PLS results for estimating total P with different calibration datasets

Calibration			Validation		
Data	R^2	RMSEC	Data	R^2	RMSEP
44 (Random)	0.82	100.88	14	0.67	228.17
Scene 2, 3, 4	0.71	130.78	Scene 1	0.66	159.02
Scene 1, 3, 4	0.74	123.73	Scene 2	0.62	204.89
Scene 1, 2, 4	0.75	117.43	Scene 3	0.52	334.41
Scene 1, 2, 3	0.75	110.56	Scene 4	0.64	209.34

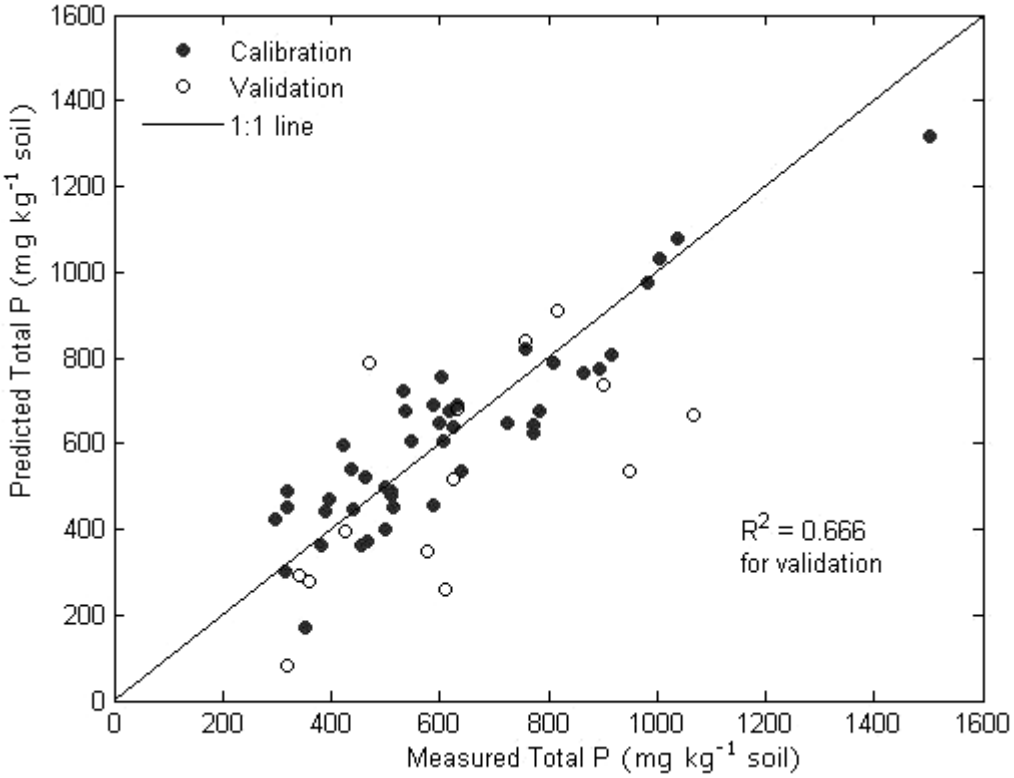


Figure 30. Measured total P vs. total P predicted by PLS modeling with Hyperion spectra (randomly split)

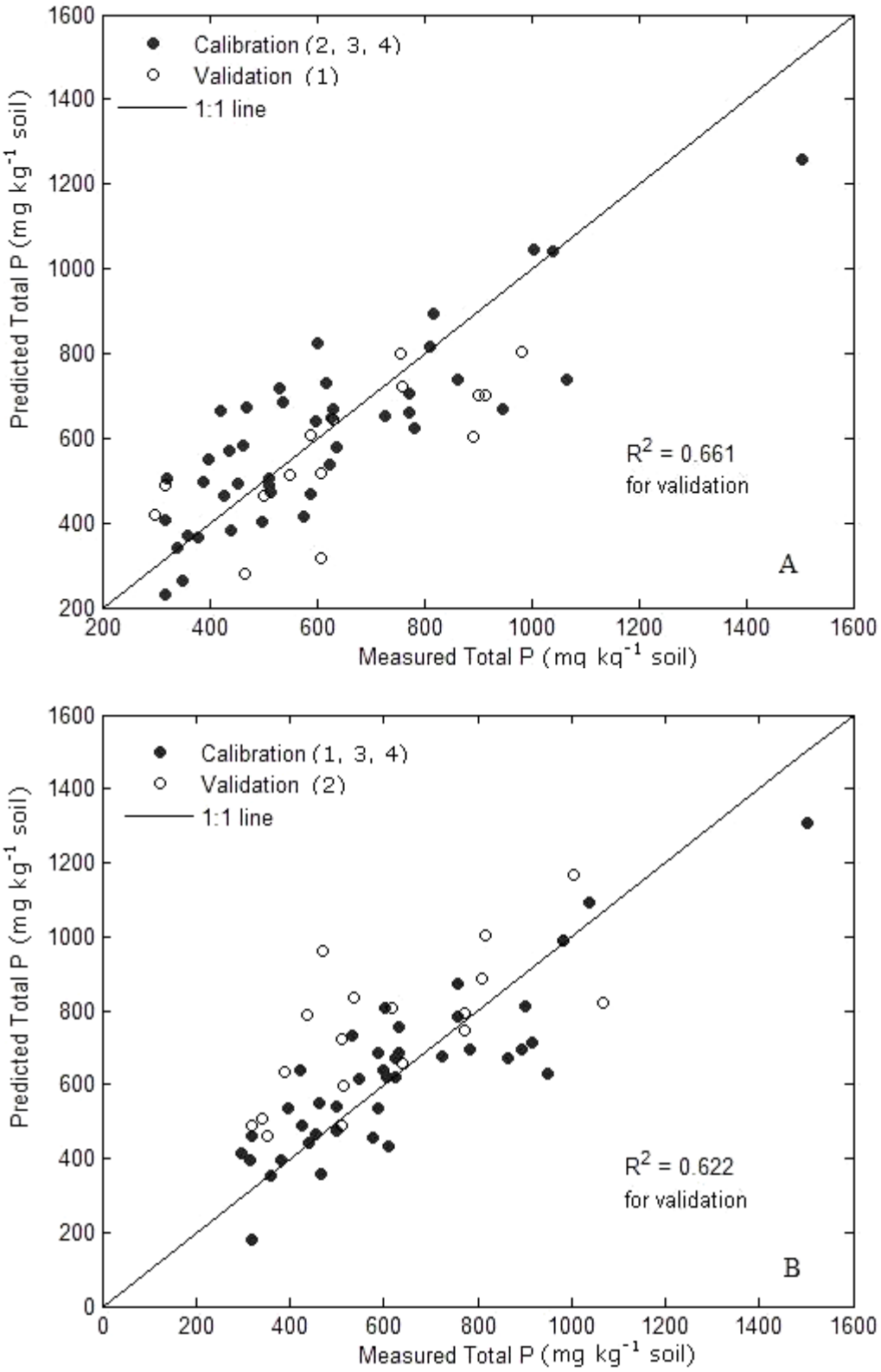
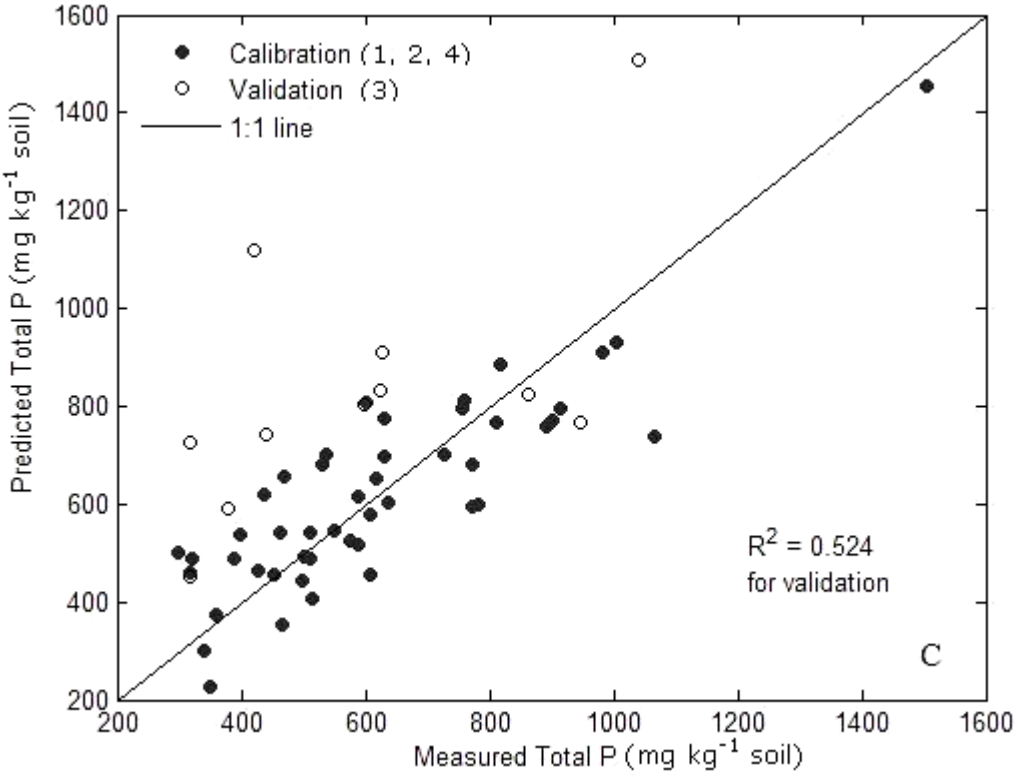


Figure 31. Measured total P vs. total P predicted by PLS modeling with Hyperion spectra: A) scene 1 for validation, B) scene 2 for validation



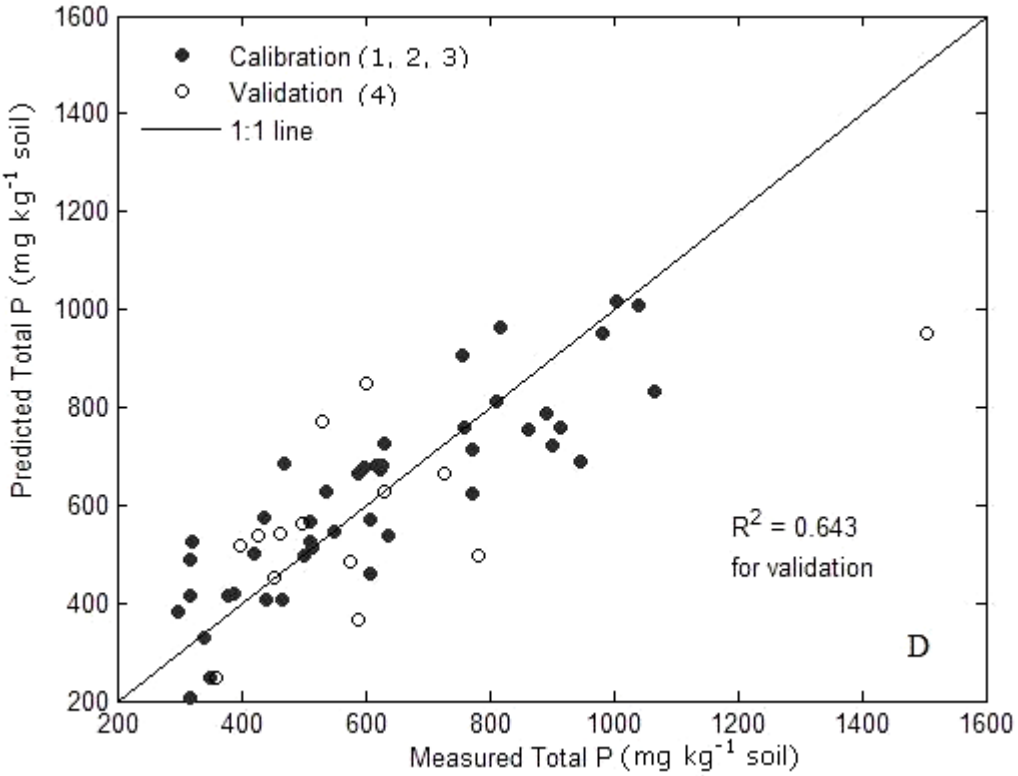


Figure 31 (cont.). Measured total P vs. total P predicted by PLS modeling with Hyperion spectra: C) scene 3 for validation, and D) scene 4 for validation

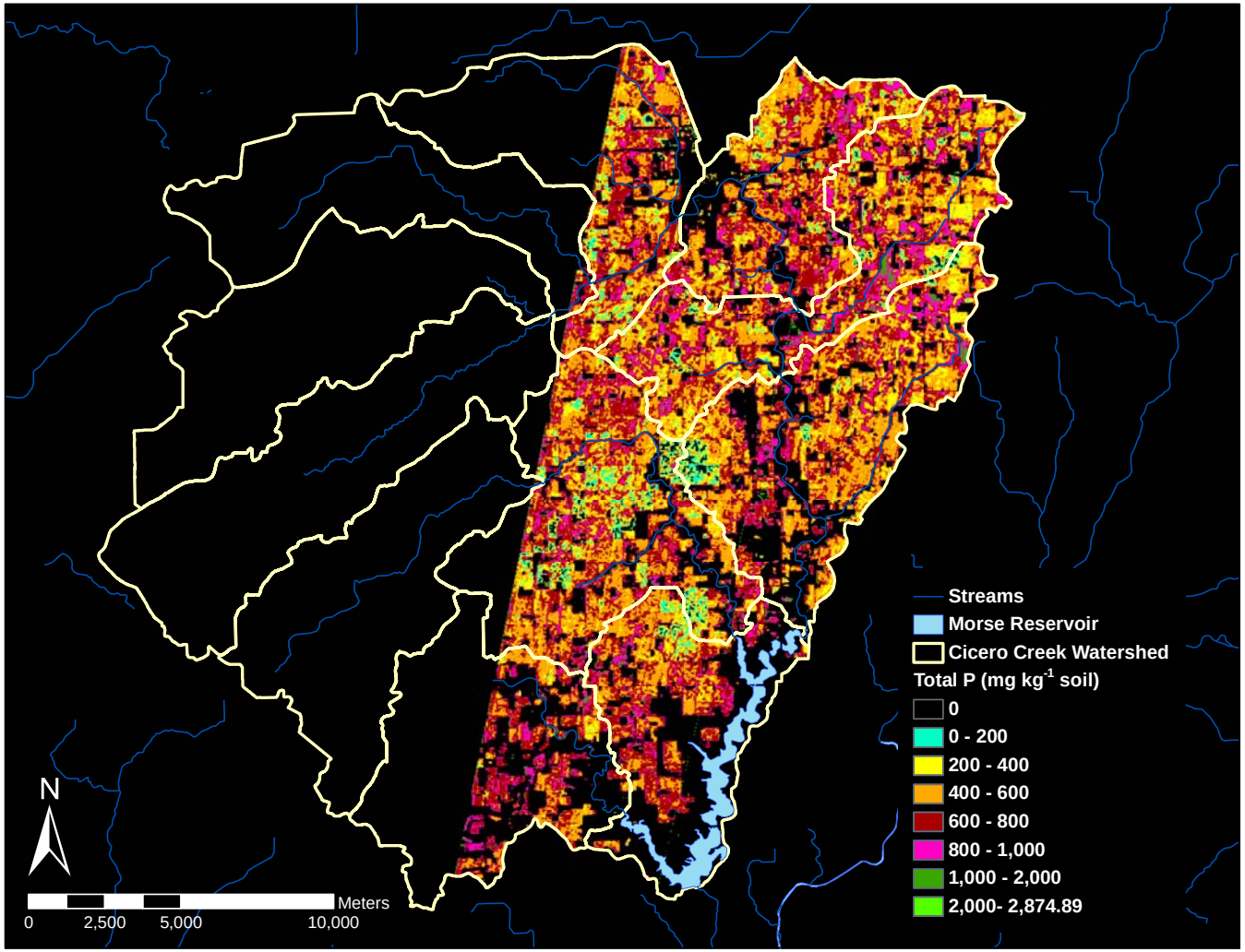


Figure 32. Predicted soil total P distribution map in Cicero Creek Watershed

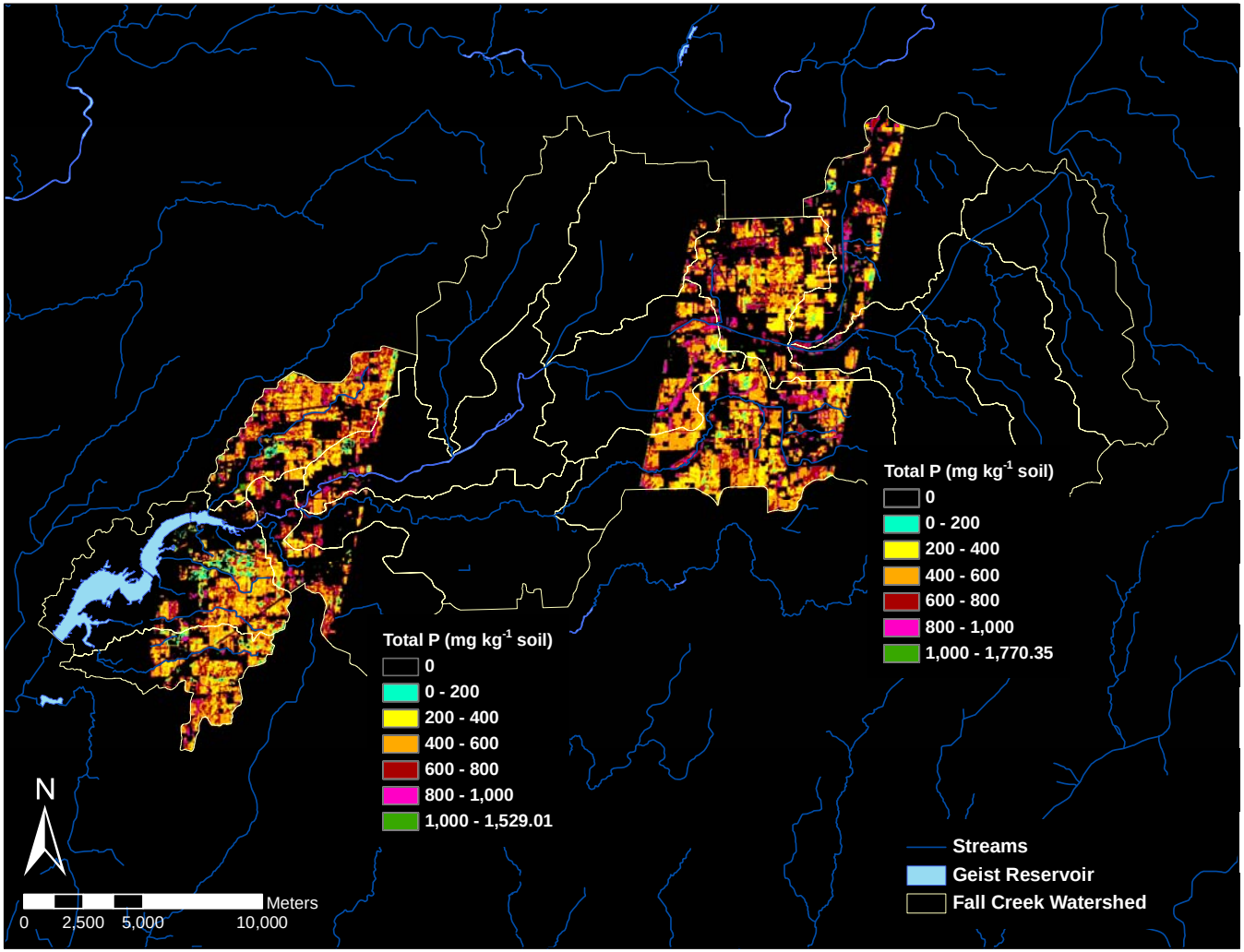


Figure 33. Predicted soil total P distribution map in Fall Creek Watershed

3.4 Comparison between PLS Modeling with Laboratory, Field and Hyperion Spectra

Table 12 summarizes the PLS modeling results with the laboratory, field and Hyperion spectra. Overall, the laboratory spectra are shown to have stronger relationships with SOM, total N and total P than the field spectra due to well-controlled measurement conditions in the laboratory. A larger number of latent variables were used in PLS modeling of the laboratory spectra for estimating SOM and total N because, compared to the field measured spectra, the laboratory measured spectra have less noise and more spectral signals related to soil properties. For instance, the PLS model for estimating SOM of field-moist soil samples used three more LVs than both field spectral and Hyperion spectral models did. The larger amount of prediction errors resulting from PLS modeling of the laboratory spectra for estimating SOM and total N than those from the models with the field spectra is likely due to six more soil samples in the validation data set. The PLS models for estimating total P with three types of spectral datasets resulted in similar number of latent variables and similar R^2 values. However, the PLS model with laboratory spectra for estimating total P does have the lowest prediction error (RMSEP = 137.59 mg kg^{-1} soil). PLS modeling of Hyperion image spectra has similar performance to that of field spectra, indicating the spectral resolution doesn't exert significant effects on the model performance. Nevertheless, a higher RMSEP resulted from PLS modeling of the imagery spectra is possibly due to the low signal to noise ratio (SNR) of Hyperion spectra.

Table 12. Results from PLS modeling

with laboratory (field-moist soils), field and Hyperion spectra

Model	Soil properties	No. of LVs	Validation		Calibration	
			R^2	RMSEP	R^2	RMSEC
Laboratory	OMC %	9	0.76	1.54	0.93	0.53
Field		6	0.74	0.84	0.77	0.98
Hyperion		6	0.74	1.42	0.84	0.83
Laboratory	Total N %	10	0.75	0.073	0.96	0.016
Field		7	0.79	0.039	0.83	0.038
Hyperion		5	0.72	0.058	0.81	0.037
Laboratory	Total P (mg kg^{-1} soil)	5	0.67	137.59	0.68	118.87
Field		6	0.60	188.12	0.68	133.31
Hyperion		5	0.67	228.17	0.82	100.88

In summary, organic matter, total N and total P can be mapped with Hyperion imagery with a relatively high R^2 for prediction models (R^2 = 0.74 for SOM; R^2 = 0.72 for total N; R^2 = 0.67 for total P). Based on these spatial distribution maps of the soil constituents, areas with low or relatively high soil chemical concentration can be located (Fig. 24, 25, 28, 29, 32 and 33). These maps can bring economic and environmental benefits. In agricultural applications, farmers can apply adequate amounts of crop-production input to areas with low amount of SOM to increase crop productivity, such as the yellow area with 0 to 2% of organic matter on the map (Fig. 24). In environmental applications, the risk of environmental pollution can be reduced by applying proper amounts of fertilizers. For instance, high application rates should be avoided in the areas with high SOM, N and P concentration. Although a relatively lower correlation and higher prediction errors (RMSEP = 228.17 mg kg^{-1} soil) were obtained for total P, the spatial distribution map of total P can be served as a guide, identifying where soils with a high or low amount of total P are present.

4 CONCLUSIONS

This study applied PLS to reflectance spectra of soils collected from Eagle Creek watershed, Cicero Creek Watershed, and Fall Creek Watershed in Central Indiana in order to examine the performance of PLS in estimating SOM, total N and total P from laboratory and field measured reflectance spectra. Built upon the results for laboratory and field measured spectra, PLS was further assessed with Hyperion image spectral to demonstrate its capability to map spatial distribution of the soil constituents with Hyperion images. The results from PLS modeling indicate that PLS is an effective tool for linking soil properties with laboratory, field and Hyperion image spectra to quantify soil nutrients.

With laboratory measured reflectance spectra, PLS resulted in the estimates of SOM with an R^2 of 0.76 and RMSEP 1.54, N with an R^2 of 0.75 and RMSEP 0.073, and P with an R^2 of 0.67 and RMSEP 137.59 mg kg^{-1} soil for field-moist soils; the estimates of SOM with an R^2 of 0.81 and RMSEP 1.09, N with an R^2 of 0.82 and RMSEP 0.045, and P with an R^2 of 0.73 and RMSEP 148.12 for oven-dried soils. Higher R^2 and lower RMSEP for oven-dried soils than those for field-moist soils demonstrate that soil moisture degrades the performance of PLS in estimating soil constituents with spectral reflectance. It is observed that the number of LVs used by the PLS models for estimating SOM and total N is one more for field-moist soils than the number of LVs for oven dried soils, this is also attributed to the effect of soil moisture. However, the opposite case was found for the estimation soil P. Although N has no diagnostic spectral characteristics in the VIS-NIR-SWIR region, total N can be quantified with PLS modeling of laboratory measured spectra and this is attributed to high correlation between SOM and total N. A

degraded model performance happened for estimating total P of field-moist and oven dried soils, which is due to low correlation between SOM and total P. For both types of soils, the third LV of the PLS models for estimating total P exhibits different spectral features from the third LV for estimating SOM. For field-moist soils, the third LV for estimating P is opposite to the direction of that for estimating SOM, but for oven dried soil the third LV for estimating P shows diagnostic spectral features consistent with that of ferric phosphate dehydrate shown by Bogrekci et al. (2005b). The implication of these third LVs for estimating P cannot be fully understand without additional soil analysis data. However, additional spectral bands from the ultraviolet spectral region should be helpful to improve PLS prediction of total P. Building PLS models for a specific phase of P such as organic P and inorganic P is worth testing.

With field measured reflectance spectra, PLS estimated SOM with an R^2 of 0.74 and RMSEP 0.839, N with an R^2 of 0.79 and RMSEP 0.039, and P with an R^2 of 0.60 and RMSEP 188.12 mg kg^{-1} soil. With Hyperion image spectra, PLS estimated SOM with an R^2 of 0.74 and RMSEP 1.42, N with a R^2 of 0.72 and RMSEP 0.058, P with a R^2 of 0.67 and RMSEP 228.17 mg kg^{-1} soil. The PLS results for the field measured and Hyperion image spectra show a slight decrease of model performance from laboratory, field to imagery data. One must be aware of the differences in the number of soil samples and spectral bands used in modeling of laboratory, field and image spectra. In practical application, it is expected that PLS modeling of image spectral data should result in higher prediction error and lower R^2 values.

5 LIMITATIONS AND FUTURE WORK

One of the limitations to this study is the relatively low representativity of soil samples due to the 30 m spatial resolution of Hyperion imagery. The assumption that soils in surrounding areas of each sampling point are homogeneous is not necessarily valid. Soil samples at each sampling site may not represent the bulk soil corresponding an image pixel because the image pixel contains spectral information of an area of 30 × 30 m instead of 20 × 20 cm. Therefore, soils should have been taken from several points within the area (900 m^2) covered by one pixel, and mixed together to yield one composite soil sample.

There are several ways to improve PLS models. One is to expand spectral datasets by including soils with high amounts of SOM (6% to 13%). Because the soil samples for this study mainly fall in the range of 2% to 6% in SOM, high prediction errors may happen to the soils with higher amounts of SOM. Thus, using more representative soil samples will improve the performance of PLS models. For example additional soil samples from Fall Creek Watershed are required to improve our models. The prediction accuracy can also be increased if PLS is applied to individual watersheds, and normalization between adjacent images can potentially increase the robustness of the models. Genetic algorithms which are used to select relevant spectral variables could be incorporated into PLS to improve predictive ability. Several studies have shown that the robustness of PLS models is enhanced with removal of irrelevant spectral bands (Leardi et al., 1998; Swierenga, et al., 1998). Other statistical methods are also worth testing, such as artificial neural network (ANN) and boosted regression trees (BRT), because studies showed that the ANN modeling slightly outperforms the PLS modeling for

image-scale data sets (Farifteh et al., 2007) and BRT outperformed PLS when the spectra of soil samples were collected from all around the world (Brown et al., 2006).

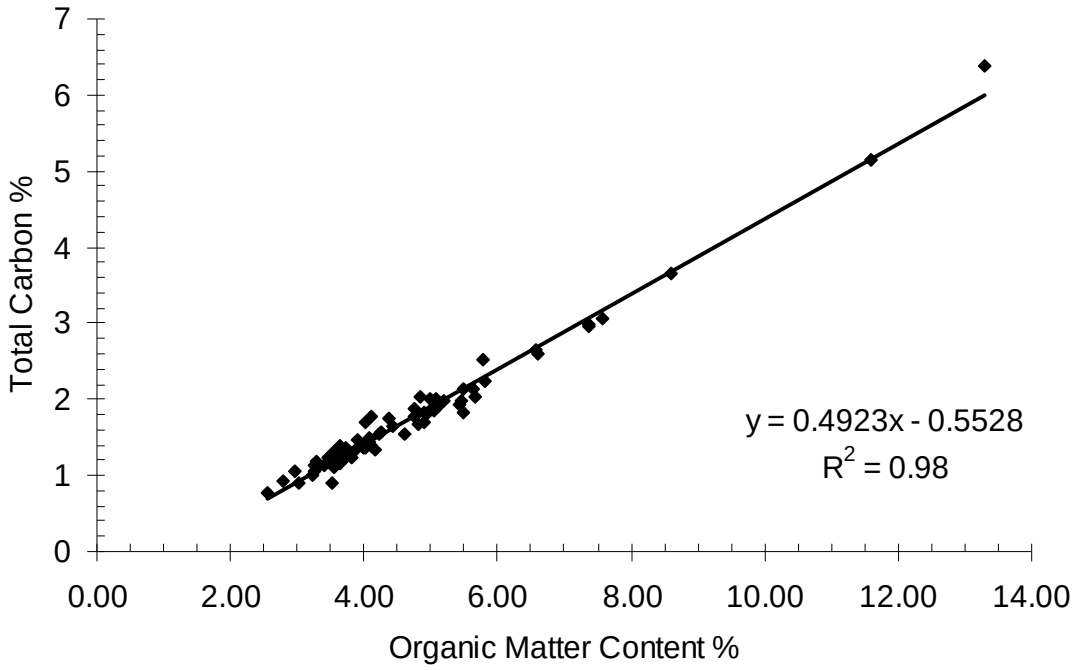
In the future, the PLS method could be extended to airborne hyperspectral data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) because the data have a higher SNR and spatial resolution (Table 13). The high SNR and spatial information of AVIRIS data are expected to help improve prediction accuracy.

Table 13. AVIRIS and Hyperion sensor characteristics

	AVIRIS	Hyperion
Spectral range	0.4 - 2.5 μm	0.4 - 2.5 μm
No. of spectral bands	224	220
Spectral resolution	10 nm	10 nm
Spatial resolution	4 - 20 m	30 m
Signal to noise ratio (SNR)	～500:1	～50:1

6 APPENDIX I

The Correlation between Organic Matter Content Determined by the Loss on Ignition (LOI) Procedure and Total Carbon Obtained by Dry Combustion (Elemental Analyzer)



7 REFERENCES

Bajwa, S. G. and L. F. Tian (2005). Soil fertility characterization in agricultural fields using hyperspectral remote sensing. *Trans. ASAE, 48*(6): 2399-2406.

Beck, R. (2003). EO-1 User Guide USGS. Retrieved December 2007 from http://eo1.usgs.gov/documents.php

Beck, R. H., B. F. Robinson, W. W. McFee and J. B. Peterson (1976). Spectral characteristics of soils related to the interaction of soil moisture, organic carbon and clay content. LARS Information Note 081176. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN.

Ben-Dor, E., Y. Inbar, and Y. Chen (1997). The Reflectance Spectra of Organic Matter in the Visible Near-Infrared and Short Wave Infrared Region (400 – 500 nm) during a Controlled Decomposition Process. *Remote Sens. Environ., 61*: 1-15.

Ben-Dor, E., J. R. Irons and G. F. Epema (1999). Soil reflectance. *Remote Sensing for the Earth Sciences: Manual of remote sensing* (pp. 111-188). New York: Wiley.

Ben-Dor, E. (2002). Quantitative remote sensing of soil properties. *Adv. Agron.*, *75*: 173-243.

Ben-Dor, E., K. Patkin, A. Banin, and A. Karnieli (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data - a case study over clayey soils in Israel. *Int. J. Remote Sens., 23*(6): 1043-1062.

Bogrekci, I. and W. S. Lee (2005a). Spectral Phosphorus Mapping using Diffuse Reflectance of soils and Grass. *Biosyst. Eng.*, *91*(3): 305-312.

Bogrekci, I. and W. S. Lee (2005b). Spectral measurement of common soil phosphates. *Trans. ASAE, 48*(6): 2371-2378.

Bogrekci, I. and W. S. Lee (2006). Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorus concentrations using UV-VIS-NIR spectroscopy. *Trans. ASABE, 49*(4): 1175-1180.

Brown, D. J., K. D. Shepherd, M. G. Walsh, M. D. Mays, and T. G. Reinsch (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, *132*(3-4): 273-290.

Chang, C. W., D. A. Laird, M. J. Mausbach, and C. R. Hurburgh (2001). Near-Infrared Reflectance Spectroscopy—Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J., 6*: 480-490.

Chang, C. W., D. A. Laird, and C. R. Hurburgh (2005). Influence of Soil Moisture on Near-Infrared Reflectance Spectroscopic Measurement of Soil Properties. *Soil Sci., 170*(4): 244-255.

Chen, F., D. E Kissel, L. T. West, and W. Adkins (2000). Field-Scale Mapping of Surface Soil Organic Carbon Using Remotely Sensed Imagery. *Soil Sci. Soc. Am. J., 64*: 746-753.

Chen, F., D. E. Kissel, L. T. West, W. Adkins, D. Rickman, and J. C. Luvall (2006). Feature selection and similarity analysis of crop fields for mapping organic carbon concentration in soil. *Comput. Electron. Agr., 54*(1): 8-21.

Chen, F., D. E. Kissel, L. T. West, W. Adkins, D. Rickman, and J. C. Luvall (2008). Mapping soil organic carbon concentration for multiple fields with image similarity analysis. *Soil Sci. Soc. Am. J., 72*(1): 186-193.

Coleman, T. L., P. A. Agbu, O. L. Montomery, T. Gao, and S. Prasad (1991). Spectral Band Selection for Quantifying Selected Properties in Highly Weathered Soils. *Soil Sci., 151*(5): 355-361.

Cozzolino, D. and A. Moron (2006). Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil Till. Res.*, *85*(1-2): 78-85.

Dalal, R. C. and R. J. Henry (1986). Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Sci. Soc. Am. J., 50*: 120-123.

Dematte, J. A. M., M. V. Galdos, R. V. Guimaraes, A. M. Genu, M. R. Nanni, and J. Zullo (2007). Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. *Int. J. Remote Sens., 28*(17): 3813-3829.

Eagle Creek Watershed Plan (2005). Retrieved May 2007, from http://www.cees.iupui.edu/Research/Water_Resources/CIWRP/Publications/Reports/ECW-Management_Plan-2005/

Ehsani, M. R., S. K. Upadhyaya, W. R. Fawcett, L. V. Protsailo, and D. Slaughter (2001). Feasibility of detecting soil nitrate content using a mid-infrared technique. *Trans. ASAE, 44*(6): 1931-1940.

Farifteh, J., F. D. Van der Meer, C. Atzberger, and J. M. Carranza (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ., 110*(1): 59-78.

Fox, G. A. and G. J. Sabbagh (2002). Estimation of soil organic matter from red and near-infrared remotely sensed data using a soil line Euclidean distance technique. *Soil Sci. Soc. Am. J., 66*(6): 1922-1929.

Ge, Y., J. A. Thomasson, and R. Sui (2006). Remote Sensing of Soil Properties in Precision Agriculture: A Review. 2006 *ASABE Annual International Meeting No. 061176.* Portland, Oregon.

Ge, Y., C. L. S. Morgan, J. A. Thomasson, and T. Waiser (2007). A new perspective to near-infrared reflectance spectroscopy: A wavelet approach. *Trans. ASABE, 50*(1): 303-311.

Grisso, R., M. Alley, P. McClellan, D. Brann and S. Donohue (2002). Precision Farimg: A Comprehensive Approach. Retrieved March 29, 2008, from http://www.ext.vt.edu/pubs/bse/442-500/442-500.html

Haaland, D. M. and E. V. Thomas (1988a). Partial Least-Squares Methods for spectral Analyses. 2. Application to Simulated and Glass Spectral Data. *Anal. Chem.*, *60*: 1202-1208.

Haaland, D. M. and E. V. Thomas (1988b). Partial Least-Squares Methods for spectral Analyses. 1. Relation to other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem., 60*: 1193-1202.

Horsley, S. W. (1995). Precision Farming - Farmers Using Satellites, Computers, and Soils Tests to Protect Ground-Water. *Ground Water Monit. R., 15*(4): 66-66.

Hummel, J. W., K. A. Sudduth, and S. E. Hollinger (2001). Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Comput. Electron. Agr., 32*(2): 149-165.

He, Y., M. Huang, A. Garcia, A. Hernandez, and H. Song (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. *Comput. Electron. Agr., 58*(2): 144-153.

Indiana Department of Environmental Managment. (2002). *Indiana Integrated Water Quality Monitoring and Assessment Report IDEM/34/02/004/2002*. Retrieved from http://www.in.gov/idem/programs/water/303d/index.html

Indiana Department of Environmental Managment. (2004). *Indiana Integrated Water Quality Monitoring and Assessment Report*. Retrieved March 6, 2008 from http://www.in.gov/idem/programs/water/303d/index.html

Indiana Department of Environmental Management. (2006). *Indiana Integrated Water Quality Monitoring and Assessment Report*. Retrieved March 6, 2008, from http://www.in.gov/idem/programs/water/303d/index.html

Ingleby, H. R. and T. G. Crowe (2000). Reflectance models for predicting organic carbon in Saskatchewan soils. *Can. Agr. Eng., 42*(2): 57-63.

Leardi, R. and A. L. Gonzalez (1998). Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometr. Intell. Lab., 41*(2): 195-207.

Li, L. (2006), Partial least squares modeling to quantify lunar soil composition with hyperspectral reflectance measurements, *J. Geophys. Res., 111,* E04002, doi: 10.1029/2005JE002598.

Ludwig, B., P. K. Khanna, J. Bauhus, and P. Hopmans (2002). Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. *Forest Ecol. Manag., 171*(1-2): 121-132.

Mandal, D. and S. K. Ghosh (2000). Precision farming - The emerging concept of agriculture for today and tomorrow. *Curr. Sci. India, 79*(12): 1644-1647.

Masserschmidt, I., C. J. Cuelbas, R. J.Poppi, J. C. De Andrade, C. A. De Abreu, and C. U. Davanzo (1999). Determination of Organic Matter in Soils by FTIR/Diffuse Reflectance and Multivariate Calibration. *J. Chemomet., 13*(3-4): 265-273.

Mathews, H. L., R. L. Cunningham, and G. W. Peterson (1973). Spectral reflectance of selected Pennsylvania soils. *Proc. Soil Sci. Soc. Am. J., 37*: 421-424.

McCarty, G. W., J. B. Reeves, V. B. Reeves, R. F. Follett, and J. M. Kimble (2002). Mid-infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. *Soil. Sci. Soc. Am. J., 66*(2): 640-646.

Moron, A. and D. Cozzolino (2007). Measurement of phosphorus in soils by near infrared reflectance spectroscopy: Effect of reference method on calibration. *Commun. Soil Sci. Plan., 38*(15-16): 1965-1974.

Morra, M. J., M. H. Hall and L. L. Freeborn (1991). Carbon and Nitrogen Analysis of Soil Fractions Using Near-Infrared Reflectance Spectroscopy. *Soil Sci. Soc. Am. J., 55*: 288-291.

National Assessment Database (2004). *Environmental Protection Agency.* Retrieved March 8, 2008, from http://www.epa.gov/waters/305b/index_2004.html

[NRCS] Natural Resources Conservation Service (2008a). Organic Matter in Soil. Retrieved May 7, 2008, from http://www.nrcs.usda.gov/Feature/backyard/orgmtrsl.html

[NRCS] Natural Resources Conservation Service (2008b). Soil Series Classification Database [Online WWW]. Retrieved March 8, 2008, USDA-NRCS, Lincoln, NE, from http://soils.usda.gov/soils/technical/classification/scfile/index.html

[NRCS] Natural Resources Conservation Service (2007). Web Soil Survey. Retrieved March 8, 2008, from http://websoilsurvey.nrcs.usda.gov/

Ray S.S., J. P. Singh, G. Das, S. Panigrahy (2004). Use of High Resolution Remote Sensing Data for Generating Site-specific Soil Mangement Plan. *XX ISPRS Congress, Commission 7*. Istanbul, Turkey The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: 127-131.

Reeves, J.B., McCarty, G.W., and Reeves, V.B. (2001). Mid-infrared Diffuse Reflectance Spectroscopy for the Quantitative Analysis of Agricultural Soils. *J. Agr. Food Chem., 49*(2): 766-772.

Selige, T., J. Bohner, and U. Schmidhalter (2006). High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma, 136*(1-2): 235-244.

Soil Survey Laboratory Methods Manual (2004). *Soil Survey Investigations Report No. 42*. Retrieved June 28, 2008, from ftp://ftp-fc.sc.egov.usda.gov/NSSC/Lab_Methods_Manual/SSIR42_2004_view.pdf

Stevens, A., B. Van Wesemael, G. Vandenschrick, S. Toure, and B. Tychon (2006). Detection of carbon stock change in agricultural soils using spectroscopic techniques. *Soil Sci. Soc. Am. J., 70*(3): 844-850.

Sudduth, K. A. and J. W. Hummel (1996). Geographic operating range evaluation of a NIR soil sensor. *Trans. ASAE 39*(5): 1599-1604.

Swierenga, H., P. J. de Groot, A. P. de Weijer, M. W. J. Derksen, and L. M. C. Buydens (1998). Improvement of PLS model transferability by robust wavelength selection. *Chemometr. Intell. Lab., 41*(2): 237-248.

Tedesco, L.P., D.L. Pascual, L.K. Shrake, L.R. Casey, P.G.F. Vidon, F.V. Hernly, K.A. Salazar, R.C. Barr, J. Ulmer, and D. Pershing (2005). Eagle Creek Watershed Management Plan: An Integrated Approach to Improved Water Quality. CEES Publication 2005-07, IUPUI, Indianapolis, 182. Retrieved March, 2008 from http://cees.iupui.edu/Research/Water_Resources/CIWRP/Publications/Reports/ECW-Management_Plan-2005/

Terhoeven-Urselmans, T., K. Michel, M. Helfrich, H. Flessa, and B. Ludwig (2006). Near-infrared spectroscopy can predict the composition of organic matter in soil and litter. *J. Plant Nutr. Soil Sc.-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde, 169*(2): 168-174.

Udelhoven, T., C. Emmerling, and T. Jarmer (2003). Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant Soil, 251*(2): 319-329.

Uno, Y., S. O. Prasher, R. M. Patel, I. B. Strachan, E. Pattey, and Y. Karimi. (2005). Development of field-scale soil organic matter content estimation models in Eastern Canada using airborne hyperspectral imagery. *Can. Biosyst. Eng., 45*: 1.9-1.14.

Van Waes, C., I. Mestdagh, P. Lootens, and L. Carlier (2005). Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. *J. Agr. Sci., 143*: 487-492.

CURRICULUM VITAE

Baojuan Zheng

EDUCATION

- 2008 M.S. in Geology, Indiana University, IUPUI. Department of Earth Sciences. GPA: 4.0
- 2006 B.E. in Geology, Jilin University, China. College of Earth Sciences.

FELLOWSHIPS AND AWARDS

2008 University Travel Fellowship Fund, IUPUI
2008 School of Science Travel Award, IUPUI
2006 - 2007 Research Investment Fund Fellowship, IUPUI
2007 Geological Society of America Travel Fund
2004 Third-best Scholarship & Excellent English Award, Jilin University

RESEARCH EXPERIENCE AND SKILLS

- Proficiency with Image Analysis Software (ENVI, ERDAS IMAGINE)
- Experience with ArcMap Software
- Skill in MATLAB PLS Toolbox
- Designed and Implemented a Research Project
  - Designed sampling site locations using ArcMap
  - Collected laboratory and field spectra using an ASD field spectrometer
  - Acquired and processed Hyperion imagery
  - Analyzed soil chemical properties: pH, moisture content, organic matter content, total carbon, nitrogen and phosphorus, particle size, carbonate content and clay minerals (X-ray)

PROFESSIONAL EXPERIENCE

2008 Research Assistant, Department of Earth Sciences, IUPUI
2007 Completed Chemometrics Courses, Eigenvector University.

PUBLICATIONS

- Zheng, B., and L. Li. In progress. Estimating Organic Matter Content of Agricultural Soils in Central Indiana Using Visible/Near-infrared Spectroscopy.
- Zheng, B., L. Li., G.M. Filippelli, and P.A. Jacinthe. In progress. Mapping Soil Organic Matter, Total Nitrogen and Total Phosphorus Using Satellite Hyperspectral Imagery.

CONFERENCE PRESENTATIONS

- Zheng, B., L. Li, and G. M. Filippelli, Using Hyperspectral Remote Sensing to Estimate Soil Organic Matter in Agricultural Fields, Geological Society of America 2007 Annual Meeting, Oct. 28 - 31, 2007, Denver, CO.
- Zheng, B., L. Li, G.M. Filippelli, and P.A. Jacinthe, Estimating Organic Matter, Total Nitrogen and Total Phosphorus of Agricultural Soils Using Hyperspectral Reflectance, Association of American Geographers 2008 Annual Meeting, Apr. 15 - 19, 2008, Boston, MA.

SOCIETY AFFILIATIONS

Association of American Geographers (Remote Sensing Specialty Group)
American Society for Photogrammetry and Remote Sensing