Remote-sensing Supported Digital Soil Mapping in South Florida

Jongsung Kim, Sabine Grunwald (P.I.), Todd Z. Osborne and Matt Norton
Soil and Water Science Department
University of Florida

Rick Robins, Howard Yamataki, Martin Figueroa, Greg Brannon and Tom Weber
USDA-Natural Resources Conservation Service, Florida

Rationale and Significance

- Soil Survey – Greater Everglades: University of Florida Agricultural Experiment Station (H. Mowry) and USDA-SCS (H.H. Bennett) "Soils, Geology and Water Control in the Everglades Region" Bulletin 442, March 1948
- SSURGO / Soil Data Mart incomplete

- Improve our understanding of the heterogeneity in soils in an unique freshwater marsh ecosystem

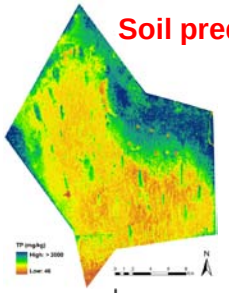
Introduction

Conceptual Soil-landscape Modeling

- Jenny (1941) – CLORPT model
- McBratney et al. (2003) – SCORPAN model

$$S_c \text{ or } S_a\,[x,y,\sim t] = f\,(S[x,y,\sim t],\ C[x,y,\sim t],\ O[x,y,\sim t],\ R[x,y,\sim t],\ P[x,y,\sim t],\ A[x,y],\ N)$$

S_c = soil class; S_a = soil attributes; S = soil, other properties of the soil at an observation point; C = climate
O = organisms including humans; R = relief; P = parent material; A = age (time factor); N = space
x, y = x and y coordinates; t=*time*



Soil predictions

Soil data

Climate data (temperature, precipitation etc.)

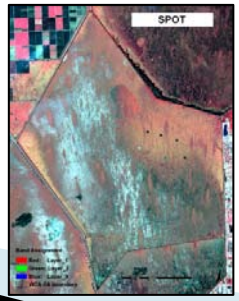
Relief (Topographic information, DEM, etc.)

$$S_c \text{ or } S_a\,[x,y,\sim t] = f\,(S[x,y,\sim t],\ C[x,y,\sim t],\ O[x,y,\sim t],\ R[x,y,\sim t],\ P[x,y,\sim t],\ A[x,y],\ N)$$

Organisms, land use, vegetation etc.

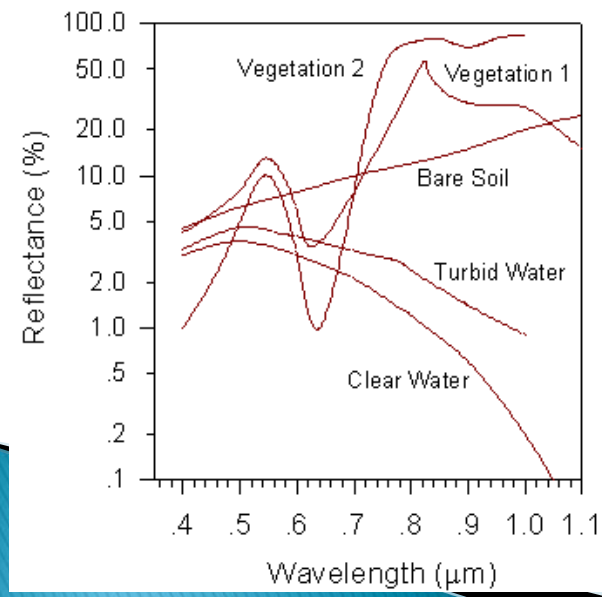
Parent material

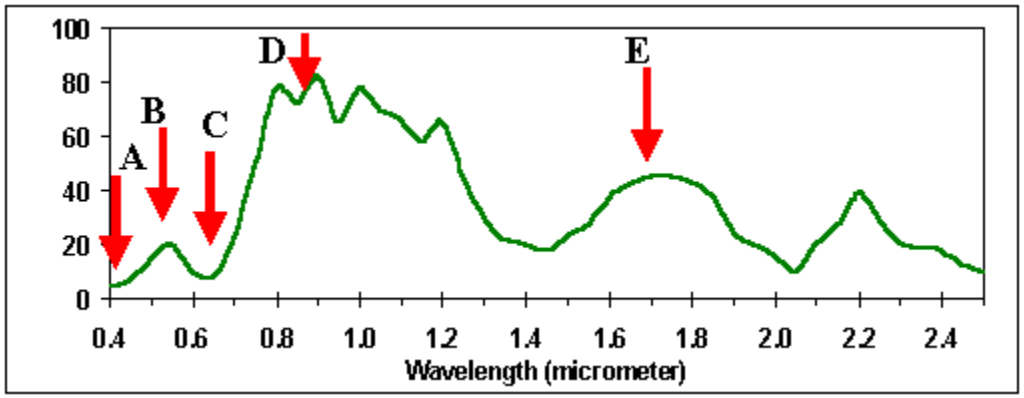
Age and geographic landscape position



Digital Soil Mapping (DSM)
Geostatistics and Remote-sensing

- Geostatistics
  - Quantify spatial autocorrelation and predict soil properties
- Remote sensing
  - Measurement of the electromagnetic energy
  - Different materials have different reflectance characteristics.
  - Can play a role in the identification, inventory, and mapping of soils that are on the surface of the Earth.
  - Inferential statistical modeling of soil properties/classes





Reflectance spectrum of five types of land-cover (left) and Typical reflectance spectrum of vegetation (right)
Figures from Center for remote imaging, sensing & processing
http://www.crisp.nus.edu.sg/~research/tutorial/optical.htm

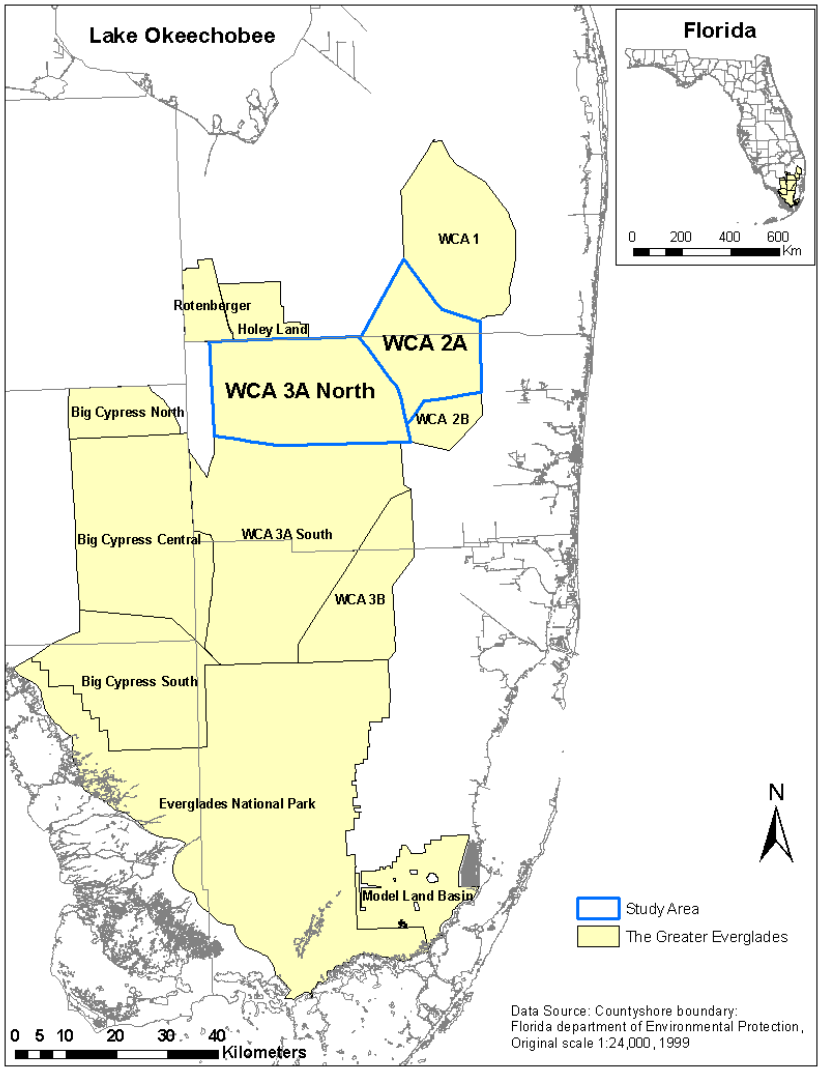
Objectives

- To develop procedures for remote sensing supported DSM of various soil properties and classes.
  - Objective 1
    - Employ univariate and multivariate statistical and geostatistical methods to predict soil properties (**TP – total phosphorus, TN – total nitrogen**, TC, Ca, and Mg) and cross-validate prediction models
  - Objective 2
    - Develop prediction models for **soil series** and cross-validate prediction models

Study Area

- WCA-2A (41, 773 ha) and WCA-3AN (72,160 ha)

  Northern Everglades
- Dominant vegetation
  - Sawgrass (*Cladium Jamaicense*)
  - Cattail (*Typha Domingensis*)
- Majority of the soils
  - Histosols
- Field sampling
  - 108 sites from WCA-2A
  - 74 sites from WCA-3AN
  - 0 – 10, 10 –20 cm soil depth
  - Pedon descriptions

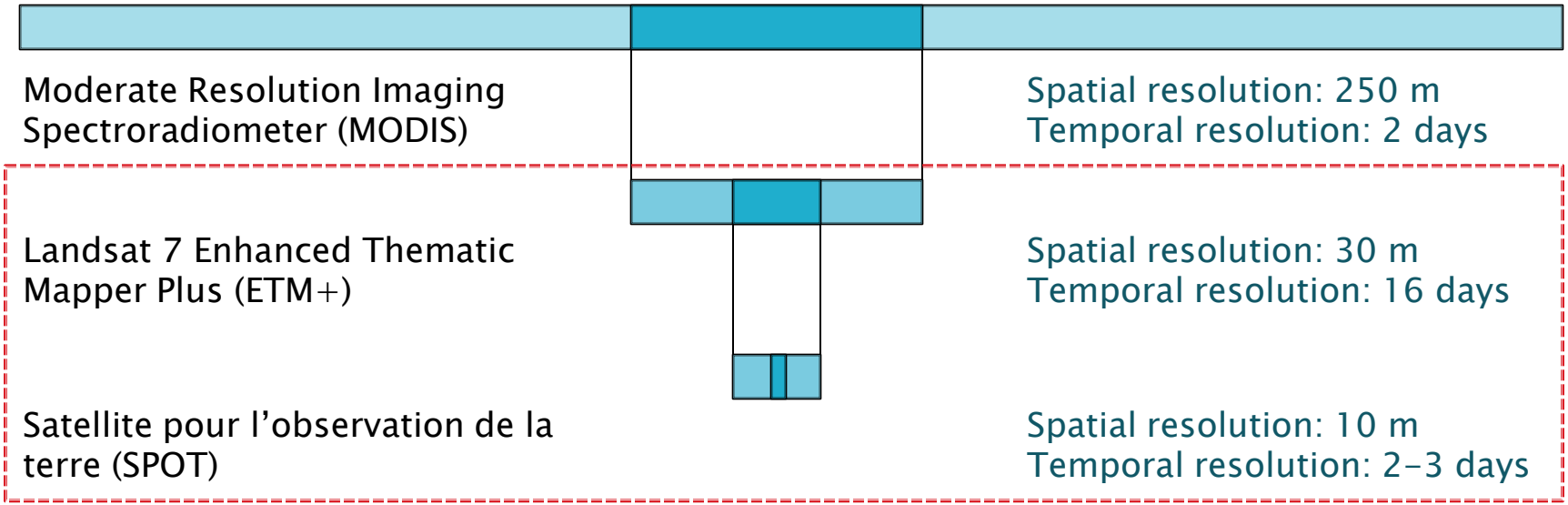


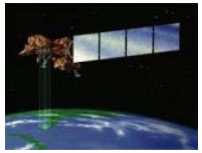
Remote Sensing and GIS Data

- Imagery: MODIS, Landsat ETM+, and SPOT.
- Indices: Normalized Difference Vegetation Index (NDVI), NDVIgreen, Normalized Difference Water Index (NDWI), Enhanced Vegetation Index (EVI), Transformed Vegetation Index (TVI), Mid-infrared index (MidIR), Reduced simple ratio (RSR).

e.g. Normalized Difference Vegetation Index (NDVI)= $\frac{(\text{Near Infrared} - \text{Red})}{(\text{Near Infrared} + \text{Red})}$

- GIS data: Distance to water control structure, elevation, and geophysical data.

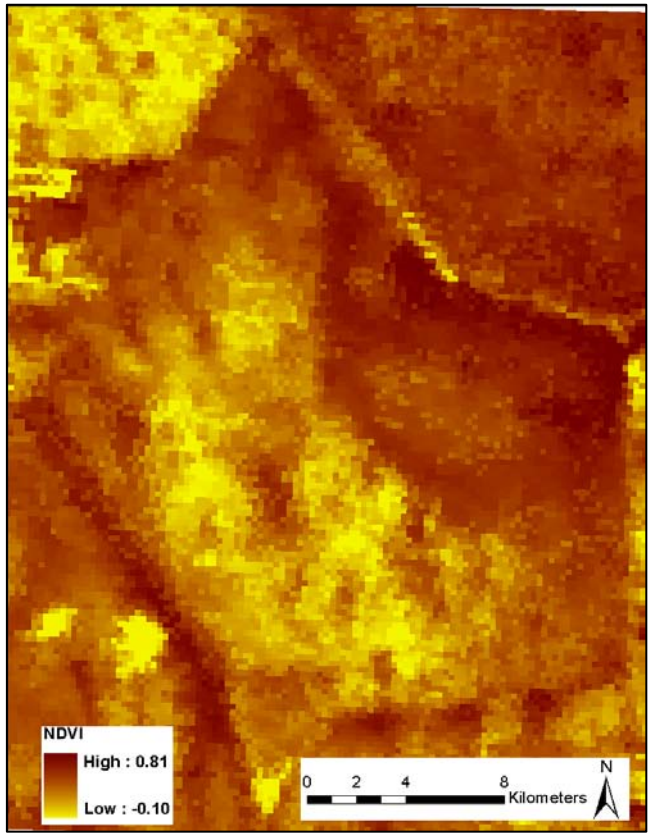




Landsat 7 ETM+ satellite platform (left) and SPOT 5 satellite platform (right)
Figures from http://www.satimagingcorp.com/satellite-sensors/landsat.htm (left), www.satimagingcorp.com/satellite-sensors/SPOT-5.htm (right).

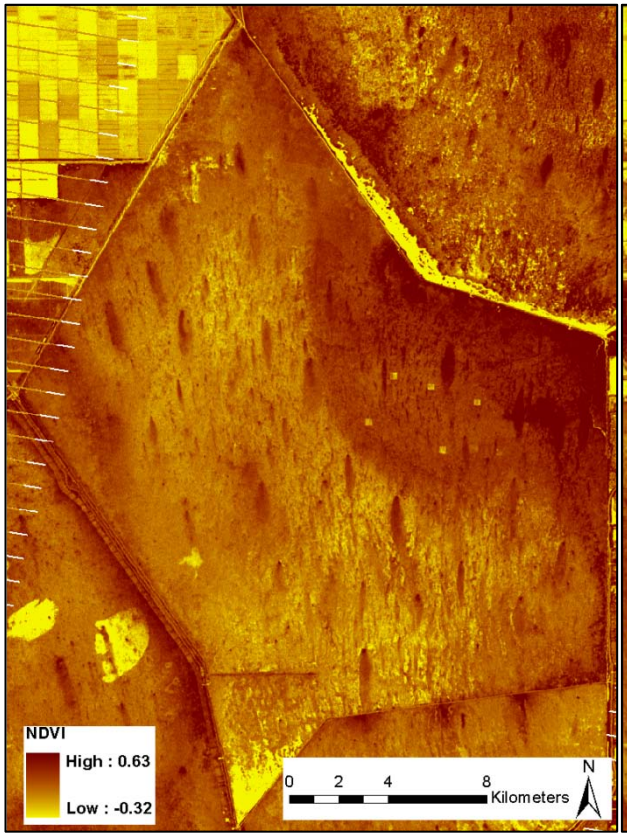
Remote Sensing Imagery Comparison: NDVI

MODIS (250 m)



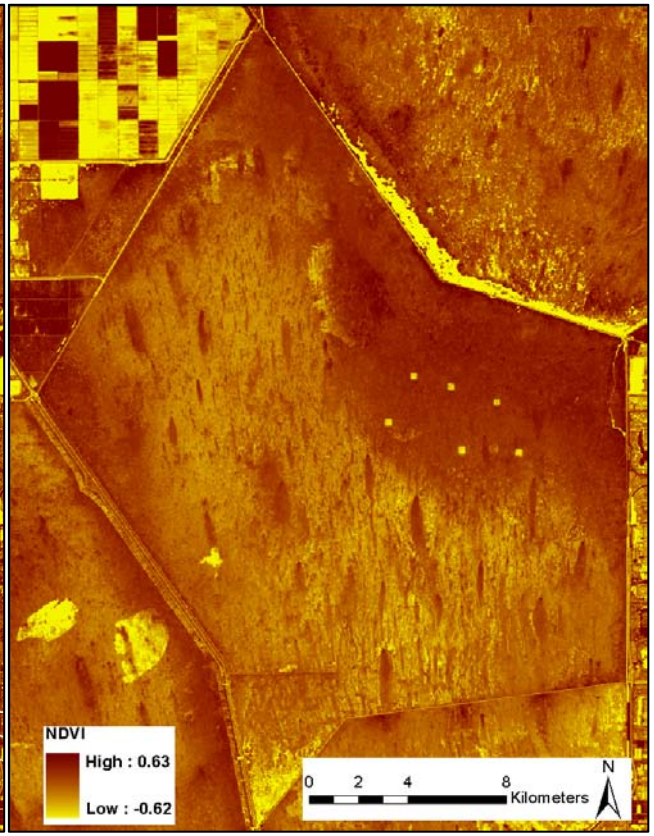
Feb. 18, 2010

Landsat ETM+ (30 m)



Feb. 16, 2010

SPOT (10 m)

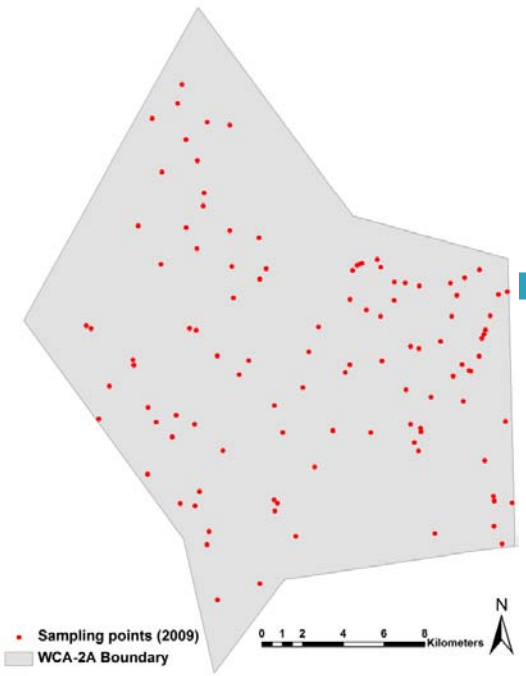


Jan. 09, 2009

$$\text{Normalized Difference Vegetation Index (NDVI)} = \frac{(\text{Near Infrared} - \text{Red})}{(\text{Near Infrared} + \text{Red})}$$

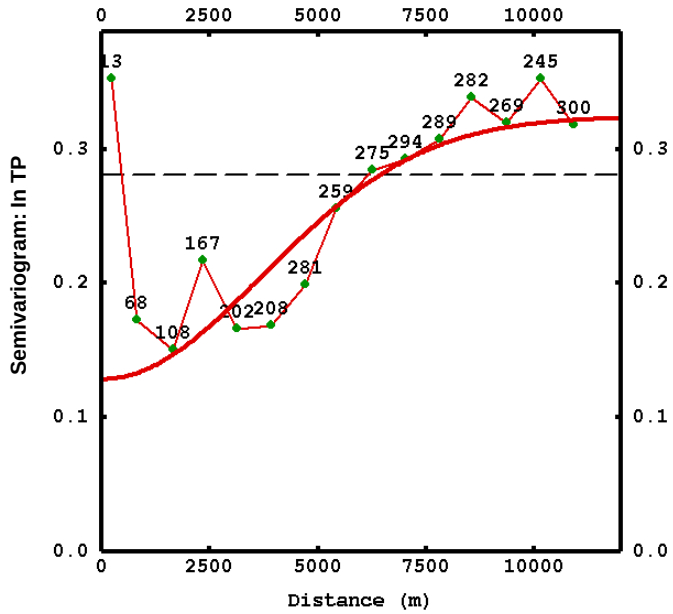
Model development: TP (and TN)

Field Observations



Point observations
(TP, TC, TN, Ca, Mg)

Semivariograms (TP)

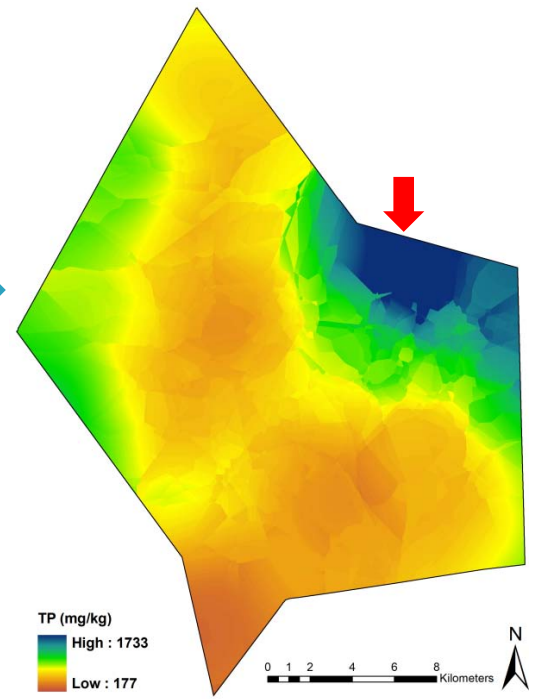


Property	Model	Nugget*	Sill*	Range (m)
ln TP (mg·kg^{-1})	Gaussian	0.129	0.325	9064.68

*Unit for Nugget and Sill: (mg·kg^{-1})^2 for TP

Ordinary Kriging
Block Kriging
Regression Kriging

Soil Property Prediction

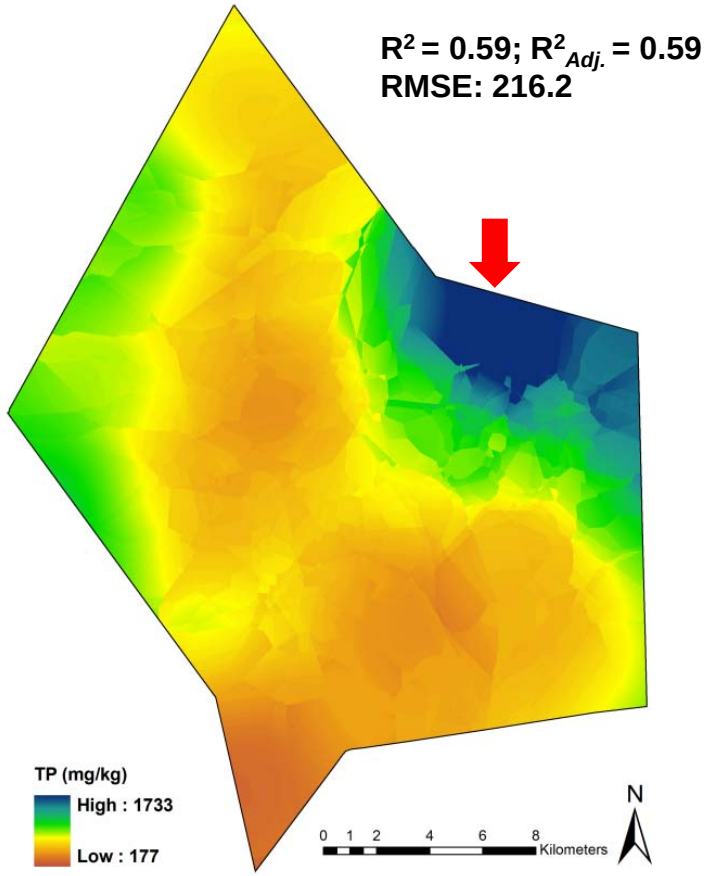


Mean Error
Root Mean Squared Error
Coefficient of determination

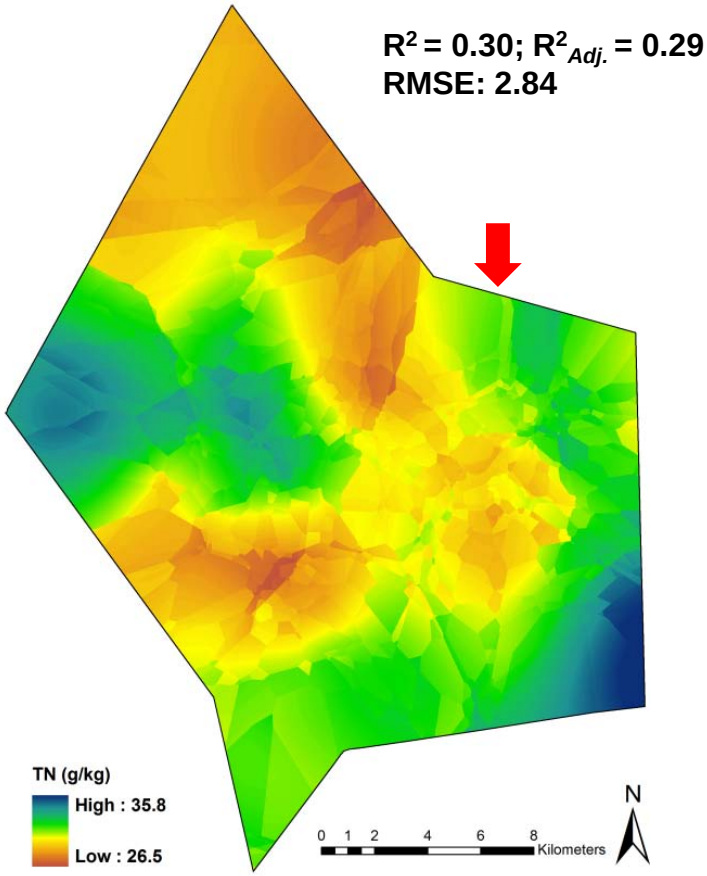
Model Results: Soil TP and TN (0–10 cm)

- Statistical and geostatistical approach: Block Kriging

TP: Block (10X10 m)



TN: Block (10X10 m)

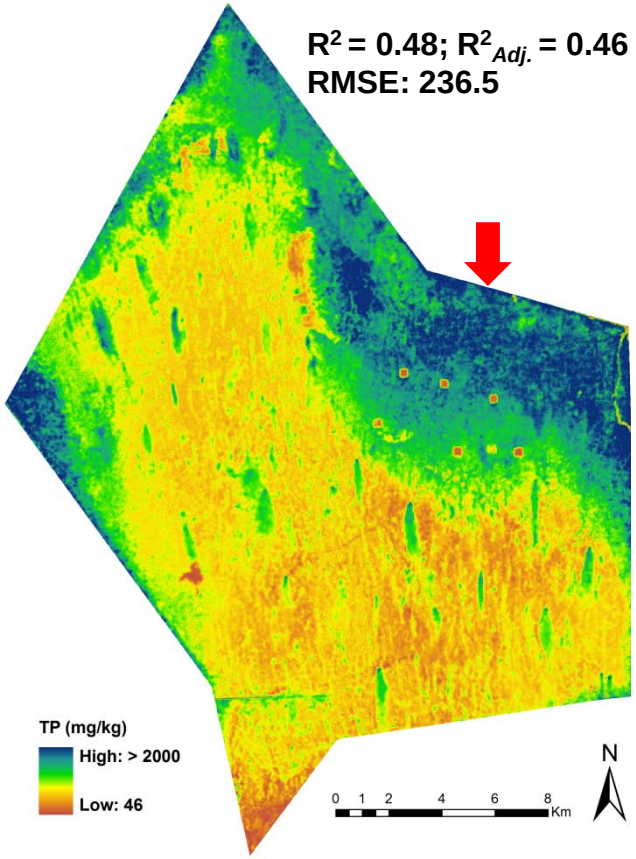


There were no major differences among the block-size of 10 X 10, 30 X 30, 60 X 60, 120 X 120, and 250 X 250 m since the search neighborhoods which control the kriging estimations were same for all different blocks.

Model Results: Soil TP (0–10 cm)

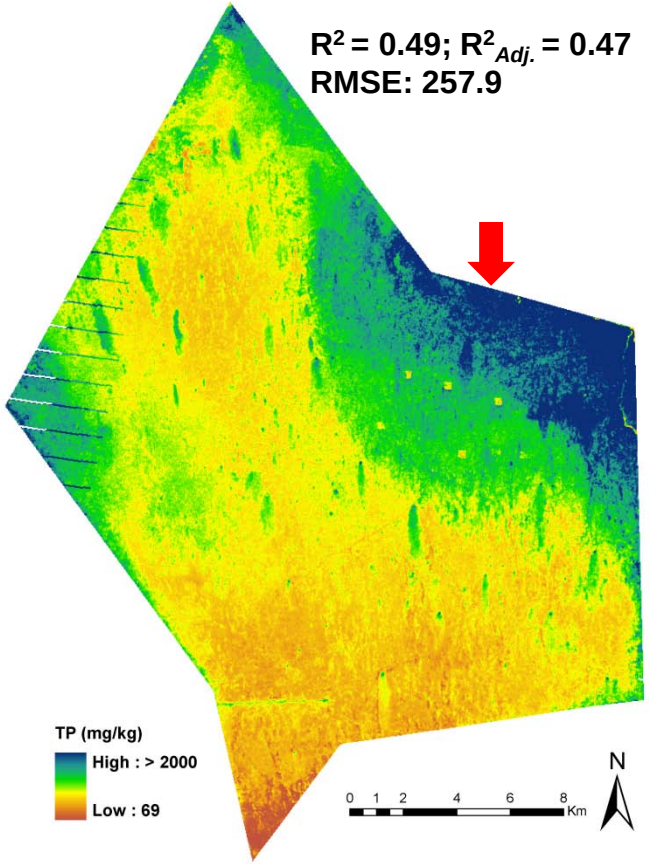
- Mixed statistical/geostatistical modeling approach: Regression Kriging

GIS + SPOT→ TP



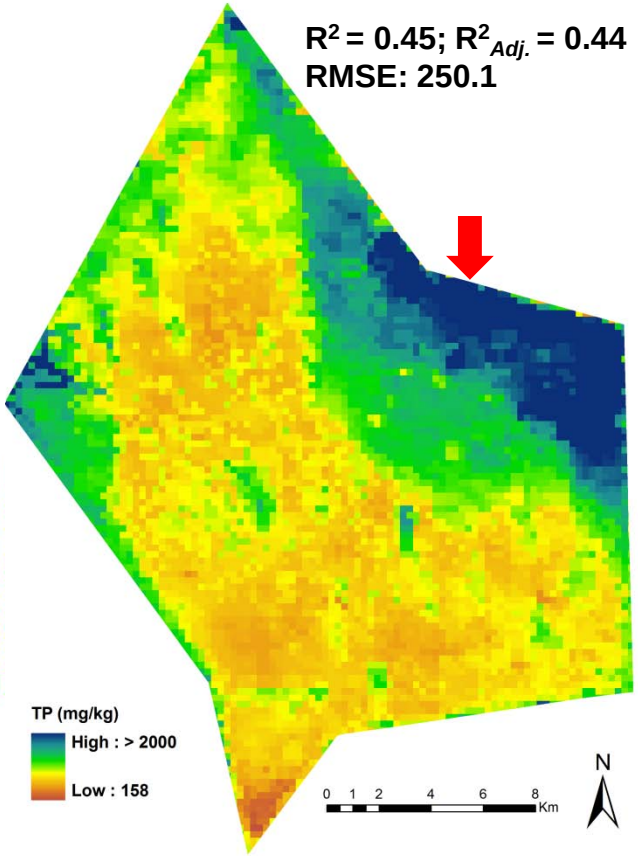
ln TP = 5.28+0.0073*NIR+8.34*NDVI-12.69 *NDVI3X3+8.03*NDVI9X9
(*p*-value < 0.05)

GIS + Landsat ETM+ → TP



ln TP = 6.19-0.00003*Dis.WCS+0.012*NIR +3.49*NDVIg-0.007*PCA1 (*p*-value < 0.05)

GIS + MODIS → TP



ln TP = 5.119+5.221*EVI (*p*-value < 0.05)

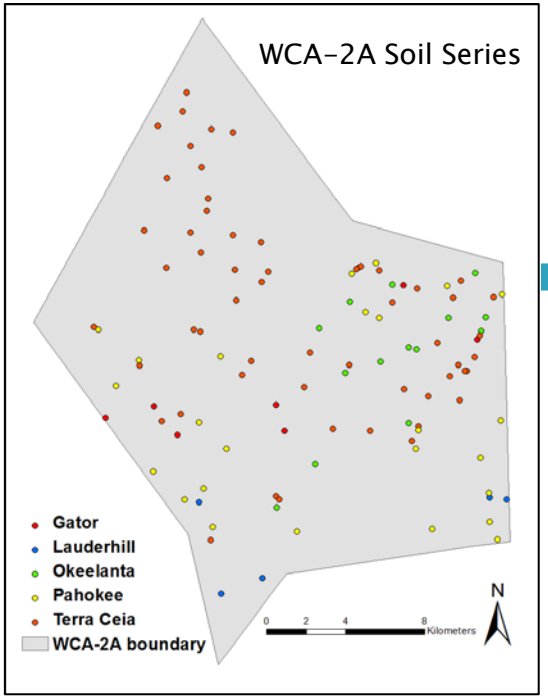
RMSE: Root-mean squared error

Results: TP and TN

- TP and TN predictions
  - Soil TP showed strong spatial autocorrelation, whereas TN had a less pronounced spatial structure as indicated by semivariograms.
  - Remote sensing derived properties explained about 50% of the variability of soil TP. Spectral indices infer on the chlorophyll content, stress, and composition of vegetation, which are correlated with soil phosphorus content.
  - Although the spatial resolution of remote sensing images differed (10, 30, and 250 m) the R^2 to estimate TP was similar. However, there were profound differences in the roughness and distribution of spatial TP patterns visible in TP grids. Landscape features such as tree islands and hammocks were smoothed out by the TP model relying on coarse-grained MODIS spectral data as predictor variable.

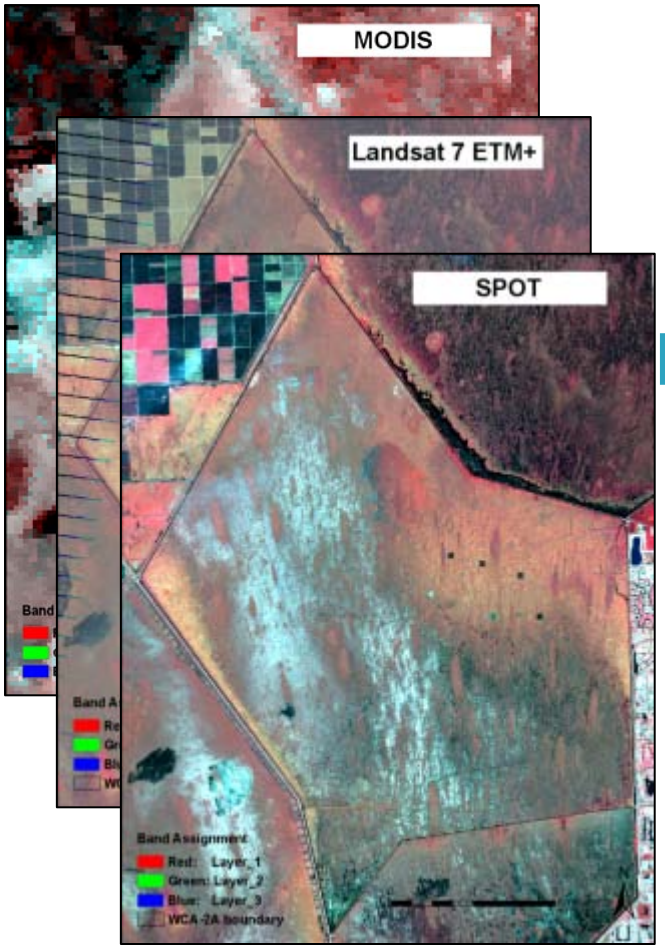
Model development: Soil Series

Field Observations



Soil Series

Remote Sensing Imagery



Classification and Regression Tree

Model Development

Soil class $= f$(RS Imagery, GIS Layers)

Confusion Matrix

Soil Series – WCA-2A, Everglades

Lauderhill
(Euic, hyperthermic **Lithic** Haplosaprists)
-thickness of OM* ranges **16 to 36 inches**.
-**have limestone** within the control section.
-Oa1/Oa2/Oa3 (~ 31 inches)/2R $.

Pahokee
(Euic, hyperthermic **Lithic** Haplosaprists)
-thickness of OM ranges **36 to 51 inches**.
-**have limestone** within the control section.
-Oap/Oa1/Oa2 (~ 42 inches)/2R.

Terra Ceia
(Euic, hyperthermic **Typic** Haplosaprists)
-thickness of OM **more than 51 inches**.
-**do not have mineral horizons** within the control section.
-Oap/Oa (~ 65 inches).

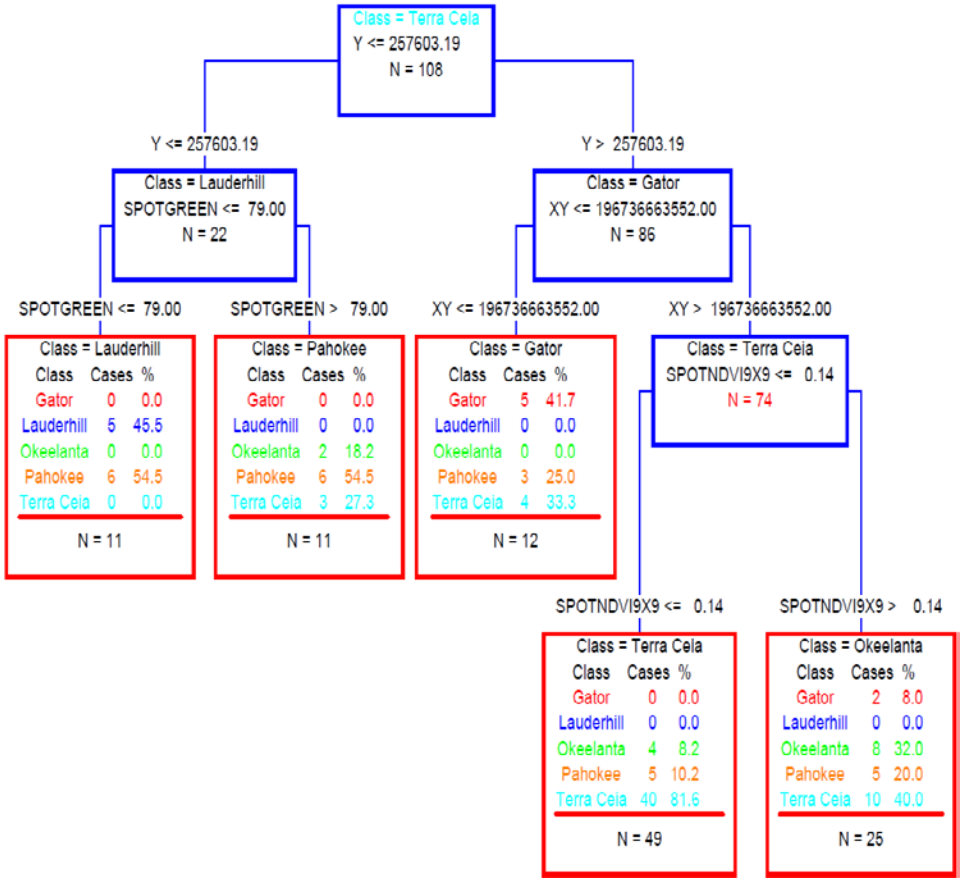
Gator
(Loamy, siliceous, euic, hyperthermic **Terric** Haplosaprists)
-thickness of OM ranges **16 to 50 inches**.
-**loamy material** within the control section.
-Oa (~ 34 inches)/Cg1/Cg2/Cg3.

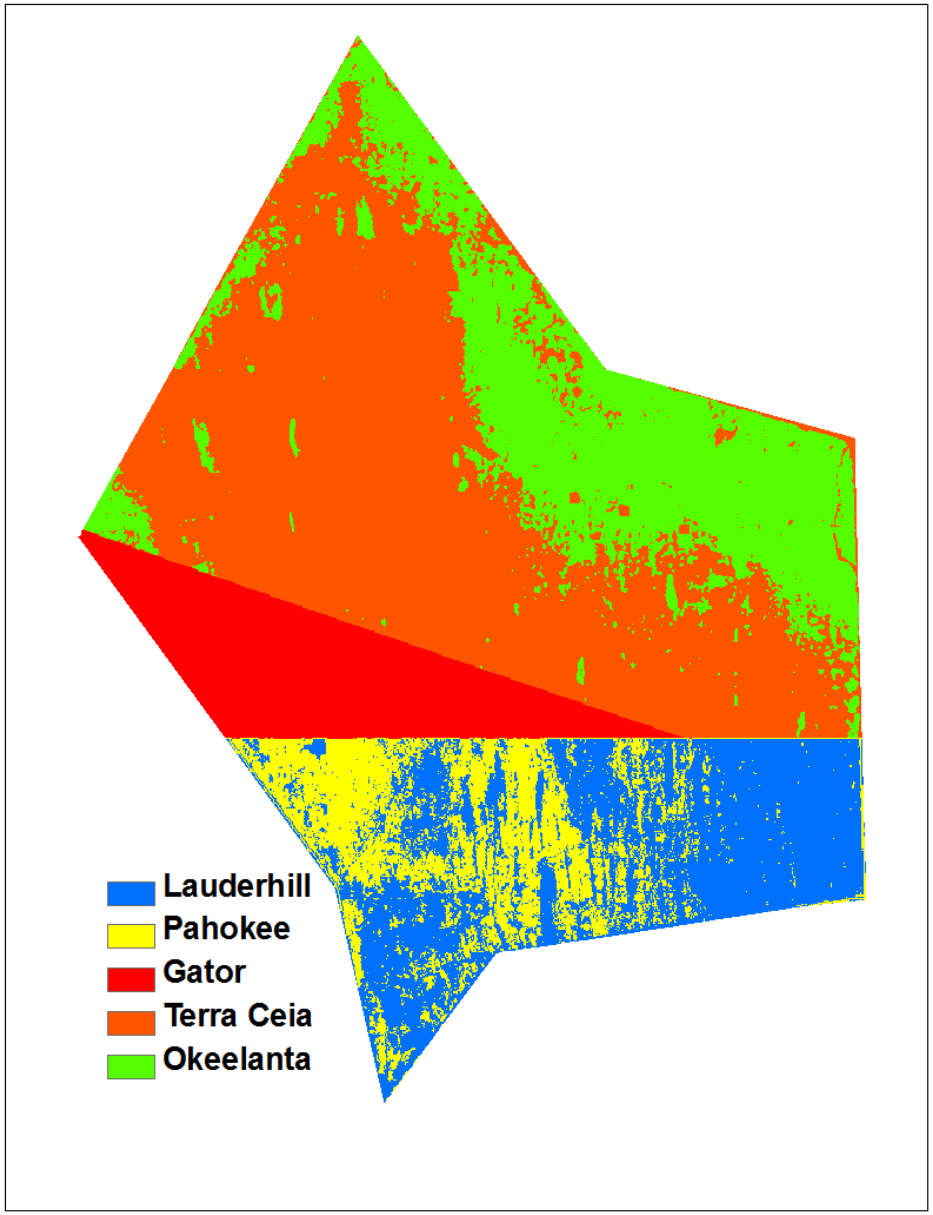
Okeelanta
(Sandy or sandy-skeletal, siliceous, euic, hyperthermic **Terric** Haplosaprists)
-thickness of OM ranges **16 to 50 inches**.
-**sandy material** within the control section.
-Oap/Oa (~ 31 inches)/C1/C2

*OM: Organic material
$ limestone bedrock

Model Results: Soil Series

Clasification Tree Model using SPOT Image





Model Results: Soil Series

- CART: Single classification tree model

Tree models	Methods	Minimal Cost (w/bedrock depth)	Average accuracy (w/bedrock depth)	Overall accuracy (w/bedrock depth)	$\hat{K}$ coefficient of agreement (w/bedrock depth)	Variable importance (only the important ones are listed)	Variable importance w/bedrock depth (only the important ones are listed)
SPOT image	Entropy	0.56 (0.39)	64.6 % (80.6 %)	**59.3 %** (68.5 %)	**41.4 %** (58.6 %)	y coordinate (100) Elevation (95.0) Distance to WCS (88.0) xy coordinate (45.2) SPOT PCA1 (45.1) SPOT SWIR (32.0)	Bedrock depth 2009 (100) y coordinate (54.0) Distance to WCS (50.5) Elevation (50.4) SPOT PCA1 (25.8) SPOT NIR (24.0)
Landsat ETM+ image	Twoing	0.61 (0.43)	59.0 % (93.0 %)	57.4 % **(84.3 %)**	36.4 % **(77.8 %)**	y coordinate (100) Elevation (77.4) xy coordinate (65.2) Distance to WCS (62.6) ETM NDVIg (46.4) ETM PCA 1 (37.3)	Bedrock depth 2009 (100) y coordinate (66.6) Elevation (59.5) Distance to WCS (53.8) xy coordinate (39.0) ETM PCA 1 (25.4)
MODIS image	Twoing	0.61 (0.37)	65.8 % (80.9 %)	50.0 % (68.5 %)	36.4 % (58.6 %)	y coordinate (100) Elevation (83.1) Distance to WCS (60.6) xy coordinate (59.3) x coordinate (44.0) MODIS EVI (23.8)	Bedrock depth 2009 (100) y coordinate (64.1) Elevation (55.3) Distance to WCS (46.4) Bedrock depth 1948 (33.5) xy coordinate (30.6)
All images	Gini	0.65 (0.41)	65.7 % (80.9 %)	50.9 % (68.5 %)	36.4 % (58.6 %)	y coordinate (100) Elevation (79.7) Distance to WCS (74.1) xy coordinate (54.2) ETM TC1 (53.0) ETM NIR (30.5)	Bedrock depth 2009 (100) Elevation (77.1) y coordinate (74.6) Distance to WCS (63.9) Bedrock depth 1948 (24.5) SPOT PCA1 (20.4)

Model Results: Soil Series

- CART: Single classification tree model

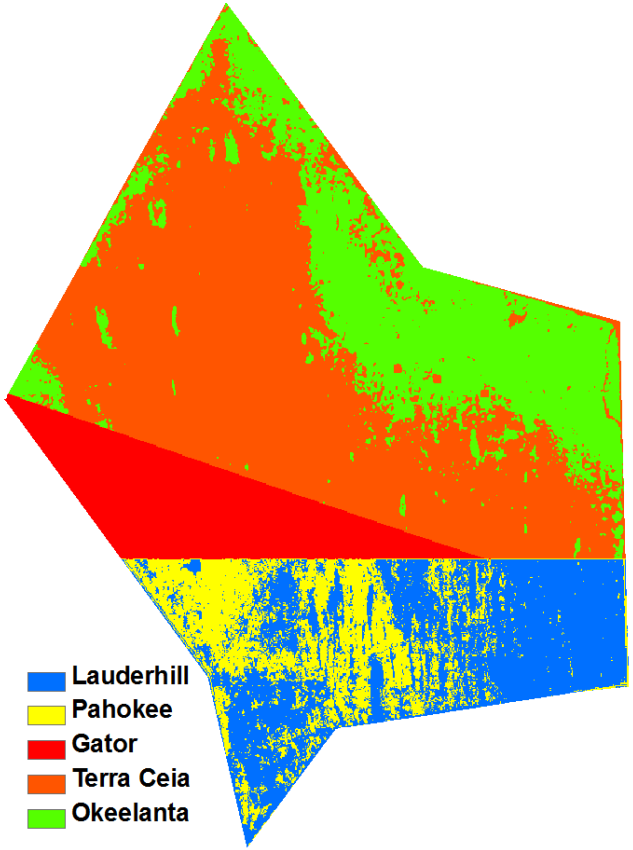
Tree models	Methods	Minimal Cost (w/bedrock depth)	Average accuracy (w/bedrock depth)	Overall accuracy (w/bedrock depth)	$\hat{K}$ coefficient of agreement (w/bedrock depth)	Variable importance (only the important ones are listed)	Variable importance w/bedrock depth (only the important ones are listed)
SPOT							(100)).5)
						SPOT SWIR (32.0)	SPOT NIR (24.0)
Landsat ETM+ image	Twoing	0.61 (0.43)	59.0 % (93.0 %)	57.4 % **(84.3 %)**	36.4 % **(77.8 %)**	y coordinate (100) Elevation (77.4) xy coordinate (65.2) Distance to WCS (62.6) ETM NDVIg (46.4) ETM PCA 1 (37.3)	Bedrock depth 2009 (100) y coordinate (66.6) Elevation (59.5) Distance to WCS (53.8) xy coordinate (39.0) ETM PCA 1 (25.4)
MODIS image	Twoing	0.61 (0.37)	65.8 % (80.9 %)	50.0 % (68.5 %)	36.4 % (58.6 %)	y coordinate (100) Elevation (83.1) Distance to WCS (60.6) xy coordinate (59.3) x coordinate (44.0) MODIS EVI (23.8)	Bedrock depth 2009 (100) y coordinate (64.1) Elevation (55.3) Distance to WCS (46.4) Bedrock depth 1948 (33.5) xy coordinate (30.6)
All images	Gini	0.65 (0.41)	65.7 % (80.9 %)	50.9 % (68.5 %)	36.4 % (58.6 %)	y coordinate (100) Elevation (79.7) Distance to WCS (74.1) xy coordinate (54.2) ETM TC1 (53.0) ETM NIR (30.5)	Bedrock depth 2009 (100) Elevation (77.1) y coordinate (74.6) Distance to WCS (63.9) Bedrock depth 1948 (24.5) SPOT PCA1 (20.4)

Predictor variables:

- Bedrock depth and geographic/topographic properties ranked high, whereas remote sensing derived properties which capture biophysical and phenological properties of vegetation ranked lower

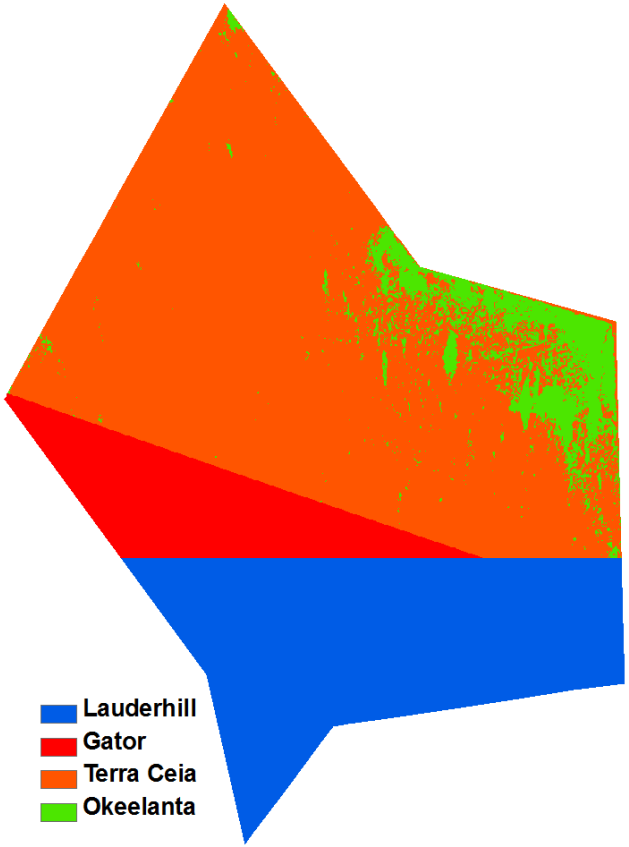
Model Results: Soil Series

GIS + SPOT→ Series



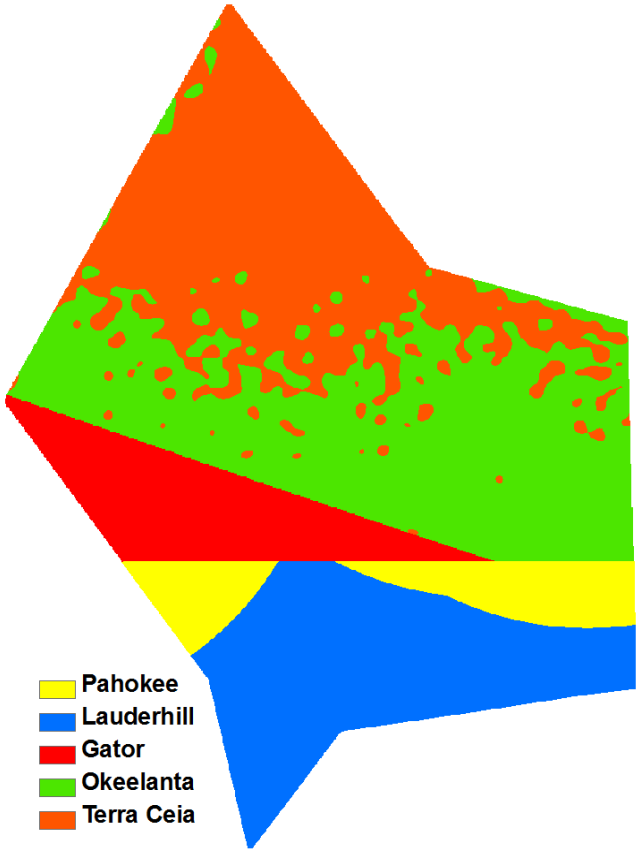
- Accuracy: Average 64.55 %, Overall 59.26 % (map shown)
- **Accuracy w/ Bedrock depth data: Average 80.55 %, Overall 68.52 %**

GIS + Landsat ETM+ → Series



- Accuracy: Average 59.00 %, Overall 57.41 % (map shown)
- **Accuracy w/ Bedrock depth data: Average 92.96 %, Overall 84.26 %**

GIS + MODIS → Series



- Accuracy: Average 65.80 %, Overall 50.00 % (map shown)
- **Accuracy w/ Bedrock depth data: Average 80.90 %, Overall 68.52 %**

Results: Soil Series

Confusion error matrix (derived from SPOT images)

Class	Observed Case	Predicted (Classified) Soil Series				
		Gator N=12	Lauderhill N=11	Okeelanta N=25	Pahokee N=11	Terra Ceia N=49
Gator	7	**5**	0	2	0	0
Lauderhill	5	0	**5**	0	0	0
Okeelanta	14	0	0	**8**	2	4
Pahokee	25	3	6	5	**6**	5
Terra Ceia	57	4	0	10	3	**40**

Confusion error matrix (derived from MODIS images)

Class	Observed Case	Predicted (Classified) Soil Series				
		Gator N=12	Lauderhill N=15	Okeelanta N=48	Pahokee N=7	Terra Ceia N=26
Gator	7	**5**	0	2	0	0
Lauderhill	5	0	**5**	0	0	0
Okeelanta	14	0	2	**12**	0	0
Pahokee	25	3	5	9	**7**	1
Terra Ceia	57	4	3	25	0	**25**

Tutorial for remote-sensing supported DSM

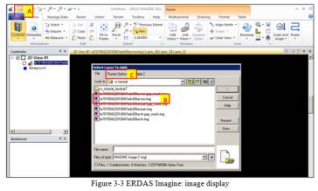


Figure 3-3 ERDAS Imagine: image display

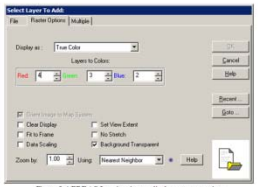


Figure 3-4 ERDAS Imagine: image display – raster option

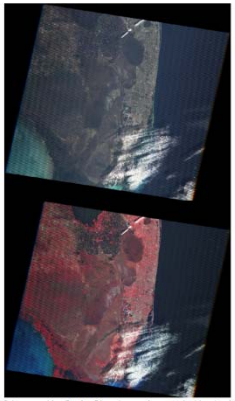


Figure 3-5 Color composition (Landsat 7 image): true-color composite (above) and false-color infrared (below) image

g. For the experimental semivariogram set, go to 'Application' in the variograms window (D in Figure 10-8) and then the 'Variogram calculation parameters' window should appear (Left in Figure 10-9). Set the variogram parameters (A in Figure 10-9) and click 'OK'. Go back to variogram window and click 'Application' again. Select 'Save in parameter file' under application and name the experimental variogram (e.g. variogram_TP_tutorial). Click 'Add' and 'OK'.

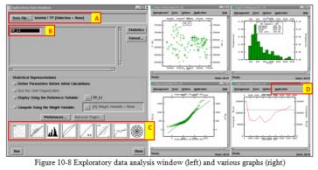


Figure 10-8 Exploratory data analysis window (left) and various graphs (right)

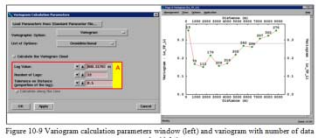


Figure 10-9 Variogram calculation parameters window (left) and variogram with number of data pairs (right)

h. Go to 'Statistics' and click 'Variogram fitting'. Once the variogram fitting window appears, bring in the experimental variogram (A in Figure 10-10) which was saved in step (g). Click 'Variogram model' and name the output file (fitted variogram model; B in Figure 10-10). To see the fitted variogram, check 'Fitting window' in Graphic windows (D in Figure 10-10).

i. Click 'Edit' button (C in Figure 10-10) and then the edit window should appear. In 'Edit' window, develop variogram models. Play with several parameters. Try to find the "best" variogram model for the variable. You may want to add a 'Nugget effect' into your variogram. Select 'Nugget effect' and click 'Add'. Set 'Sill'.

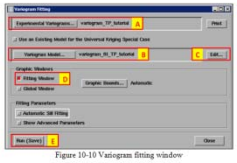


Figure 10-10 Variogram fitting window

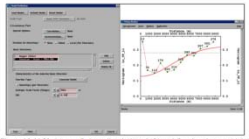


Figure 10-11 Variogram fitting edit window (left) and fitted variogram (right)

<u>Acknowledgement & funding sources</u>:

Cooperative Ecosystem Studies Unit (CESU) – Natural Resources Conservation Service (NRCS)– NRCS, Gainesville, FL (phase 2); National Geospatial Development Center, Morgantown, WV (phase 1)

and

Planet Action (provided SPOT satellite images);

National Aeronautics and Space Administration (NASA) (provided Landsat ETM+ and MODIS satellite images)