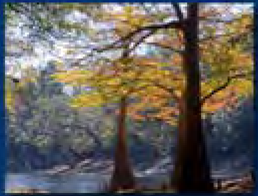
Florida Wetland Integrity Dataset (FWID)

Wetland Mapping Consortium
July 15, 2015

John Humphreys
AmirSassan Mahjoor




Florida Wetland Integrity Dataset (FWID)

Presentation Outline

1. Project Scope and Goals
2. Predicting Wetland Presence
   - A. Bayesian Vs. Frequentist Methods
   - B. Identifying Explanatory Variables
     - i. Interrogating (bulk) Soils Data
     - ii. Estimating Available Water Capacity (AWC)
     - iii. Topographic Indices
     - iv. Spatial Correlation – The Neglected Variable
   - C. Model Construction
     - i. Model Selection
     - ii. The "Final" Model
     - iii. Results
3. Summary and Closing
4. Q & A

Project Scope and Goals

1. Approximate the locations and extents of wetlands and other surface waters throughout Florida.

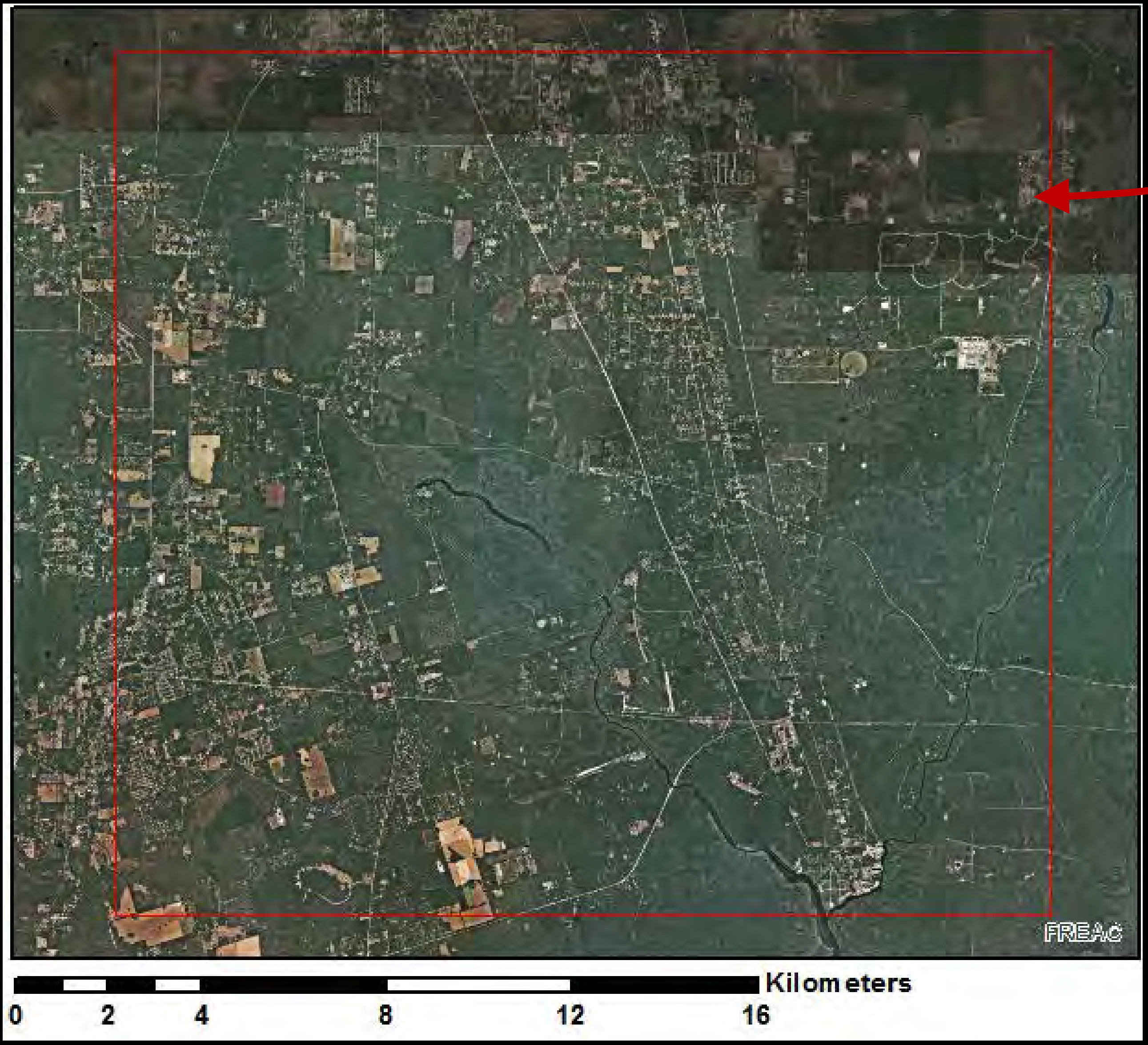
2. Approximate the *integrity* and condition of natural communities throughout Florida.

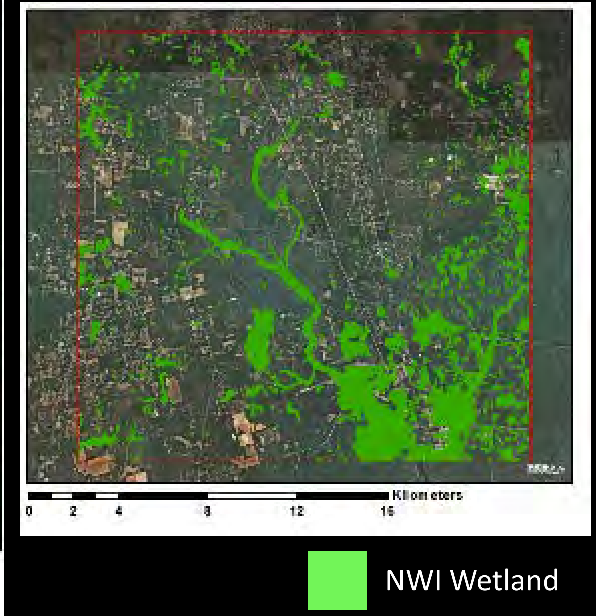
3. Develop products that will remain valuable into the future.

4. Apply scientifically rigorous methods.

Predicting Wetland Presence

Example Study Area for Today's Talk





Bayesian Vs Frequentist

Statistical Inference

Frequentist

- Probability of the data given the hypothesis $P(Y|H_0)$
- Use of a P-value
- Standard "significance" cut-off of P-value is the Neyman–Pearson acceptable probability of committing a Type-I statistical error ($\alpha = 0.05$)
- If P-value is "small," reject H_0 - a "pass or fail" significance test
- Probability is a frequency dependent concept in which the "true value" is realized only with the "true" population (∞)
- Includes a "confidence interval" that is also frequency dependent: ratio of events of interest to total events observed

Bayesian

- Probability of the hypothesis given the available data $P(H_0|Y)$
- No P-value
- Results are reported as a continuous probability ("posterior distribution") rather than a pass/fail test
- Setting the α at 0.05 or any other value is arbitrary
- Probability is *subjective* in that it quantifies a *degree of belief* based on prior knowledge ("prior distribution") and likelihood using the data at hand rather than an assumption of an infinite population
- Includes a "credible interval" that is not frequency dependent but rather reflects the *belief* that the "true value" falls within a particular interval

Interrogating Soils Data

Steps

1. Differentiate those soil mapping units (SMU) associated with wetlands from those associated with uplands.
2. Decompose SMU attributes from categorical data to a unique and continuous numeric index scaled by the amount of variation that each explains.
3. Evaluate the decomposed values ability to predict the wetland or upland association identified in Step #1 and discard those that do not significantly predict their association. Retain those that do predict their association for possible model inclusion.

Objective 1: Differentiate by Wetland/Upland Association

Natural Wetland Land Cover

Natural Upland Land Cover

Soil Map Unit (NRCS SSURGO)

1 = Soil Associated with Wetland

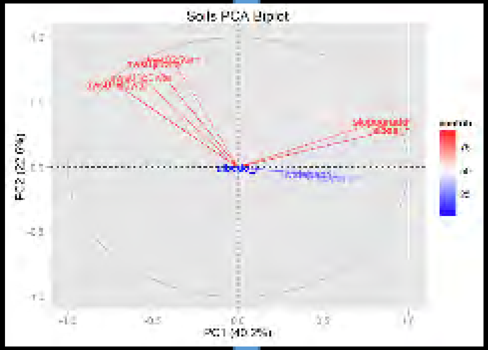
0 = Soil Associated with Upland

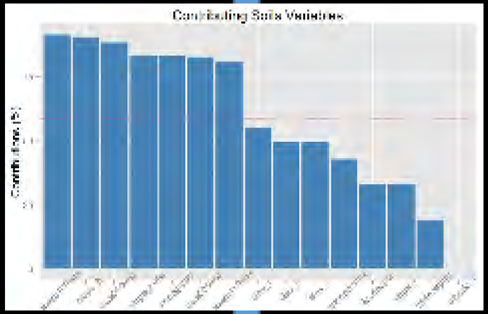
Designation	MUSYM	MUNAME	FARMLNDCL	WTDEPANNMI	PONDFREQPR	HYDCLPRS	COMPPCT_R	MAJCOMPFL	SLOPE_L	SLOPE_H	TFACT	WEG
1	3	Alpin sand, 0 to 5 percent slo	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
1	44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
0	35	Scranton sand	Not prime farmland	8	0-14%	Partially hydric	65	Yes	1	2	5	1
0	4	Alpin sand, 0 to 5 percent slo	Not prime farmland	0	0-14%	Not hydric	88	Yes	3	5	5	1
0	26	Alpin fine sand, 0 to 5 percen	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
0	44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
1	35	Scranton sand	Not prime farmland	8	0-14%	Partially hydric	65	Yes	1	2	5	1
1	44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
1	44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
0	18	Kershaw sand, 0 to 5 percent	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
1	8	Chipley fine sand, 0 to 2 perc	Not prime farmland	76	0-14%	Partially hydric	80	Yes	1	2	5	1
1	8	Chipley fine sand, 0 to 2 perc	Not prime farmland	76	0-14%	Partially hydric	80	Yes	1	2	5	1
0	19	Kershaw sand, 5 to 8 percent	Not prime farmland	0	0-14%	Not hydric	80	Yes	7	8	5	1
0	3	Alpin sand, 0 to 5 percent slo	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1

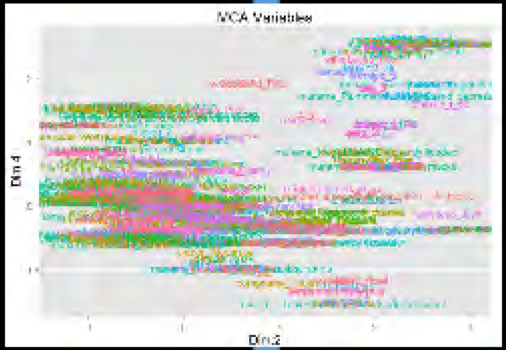
Objective 2: Explain Variation in Soil Data

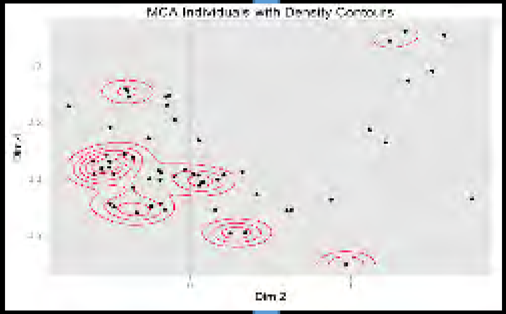
Soil Dataset

Continuous Numeric Attributes	Categorical Numeric & Nominal Attributes
Principle Components Analysis (PCA)	Multiple Correspondence Analysis (MCA)









Singular Value Decomposition (SVD)

$$\underset{m \times n}{\overset{\hat{X}}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} \approx \underset{m \times r}{\overset{U}{\begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \\ u_{m1} & & u_{mr} \end{pmatrix}}} \underset{r \times r}{\overset{S}{\begin{pmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{pmatrix}}} \underset{r \times n}{\overset{V^{\mathsf{T}}}{\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{r1} & & v_{rn} \end{pmatrix}}}$$

A technique from linear algebra that breaks down a rectangular matrix into the product of three matrices: an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V.

- Orders attributes by the amount of variation they explain
- Relationships between attributes is preserved
- Ensures that attribute indices are not correlated

Objective 2: Convert to Continuous Index Scaled by Variation (continued)

MUSYM	MUNAME	FARMLNDCL	WTDEPANNMI	PONDFREQPR	HYDCLPRS	COMPPCT_R	MAJCOMPFL	SLOPE_R	SLOPE_H	TFACT	WEG
3	Alpin sand, 0 to 5 percent slop	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
38	Scranton sand	Not prime farmland	5	0-14%	Partially hydric	65	Yes	1	2	5	1
4	Alpin sand, 0 to 5 percent slop	Not prime farmland	0	0-14%	Not hydric	88	Yes	3	5	5	1
28	Alpin fine sand, 0 to 5 percent	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
38	Scranton sand	Not prime farmland	5	0-14%	Partially hydric	85	Yes	1	2	5	1
44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
44	Pickney soils, occasionally flo	Not prime farmland	0	75-100%	Partially hydric	85	Yes	1	2	5	2
18	Kershaw sand, 0 to 5 percent :	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1
8	Chipley fine sand, 0 to 2 perce	Not prime farmland	75	0-14%	Partially hydric	80	Yes	1	2	5	1
8	Chipley fine sand, 0 to 2 perce	Not prime farmland	75	0-14%	Partially hydric	80	Yes	1	2	5	1
19	Kershaw sand, 5 to 8 percent :	Not prime farmland	0	0-14%	Not hydric	80	Yes	7	8	5	1
3	Alpin sand, 0 to 5 percent slop	Not prime farmland	0	0-14%	Not hydric	85	Yes	3	5	5	1

Soil Attribute Table
(a rectangular matrix)

$$\underbrace{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}_{m \times n}^{\hat{X}} \approx \underbrace{\begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \\ u_{m1} & & u_{mr} \end{pmatrix}}_{m \times r}^{U} \underbrace{\begin{pmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{pmatrix}}_{r \times r}^{S} \underbrace{\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{r1} & & v_{rn} \end{pmatrix}}_{r \times n}^{V^{\top}}$$

Conduct SVD

	V1	V2	V3	V4	V5	V6
1	-0.1856695	0.1140680	-0.42420481	-0.08014229	-0.28120475	2.0482152
2	-0.1698416	-0.9475450	0.09416006	0.18175889	-0.04574227	-0.8454406
3	-0.1655010	-0.6869339	0.19898838	1.58933469	-0.04306314	-0.3930183
4	-0.1810848	0.1327860	-0.43814525	-0.06470997	-0.23329958	2.0505967
5	-0.1655010	-0.6869339	0.19898838	1.58933469	-0.04306314	-0.3930183
6	-0.1592390	-0.8623658	0.16102280	0.30185739	-0.07451060	-0.5019071

Decomposed Values

Objective 3: Evaluate the ability of SVDs to Predict Wetland/Upland Association

Bayesian Independent Random Noise Model

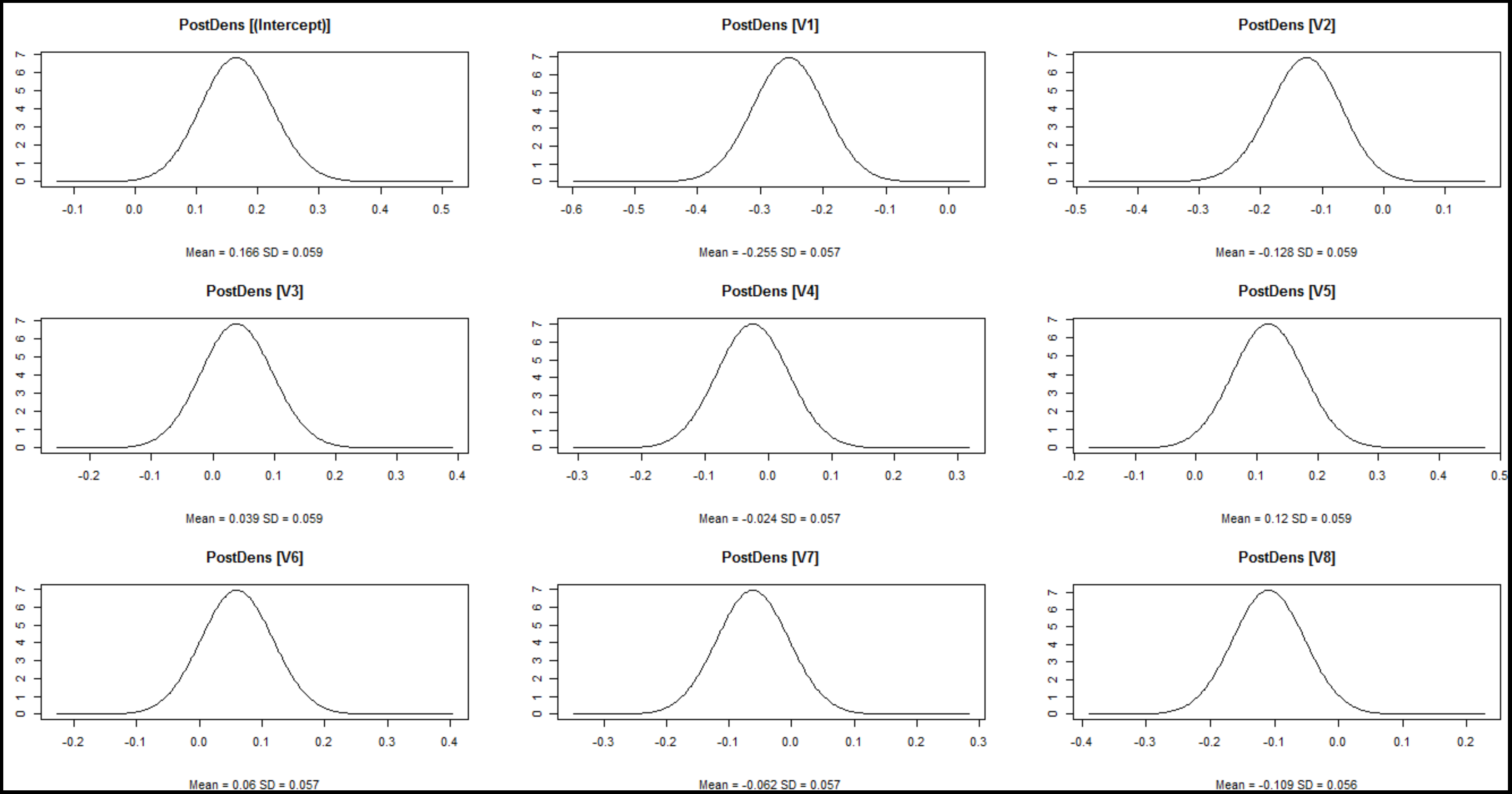
$$\text{Designation} \sim SVD_{MCA_{V1}}, SVD_{MCA_{V2}}, SVD_{MCA_{V3}}, \ldots, SVD_{MCA_{V50}}$$

$$\text{Designation} \sim SVD_{V1}, SVD_{V2}, SVD_{V3}, \ldots, SVD_{V25}$$

Decomposed Soil Attributes

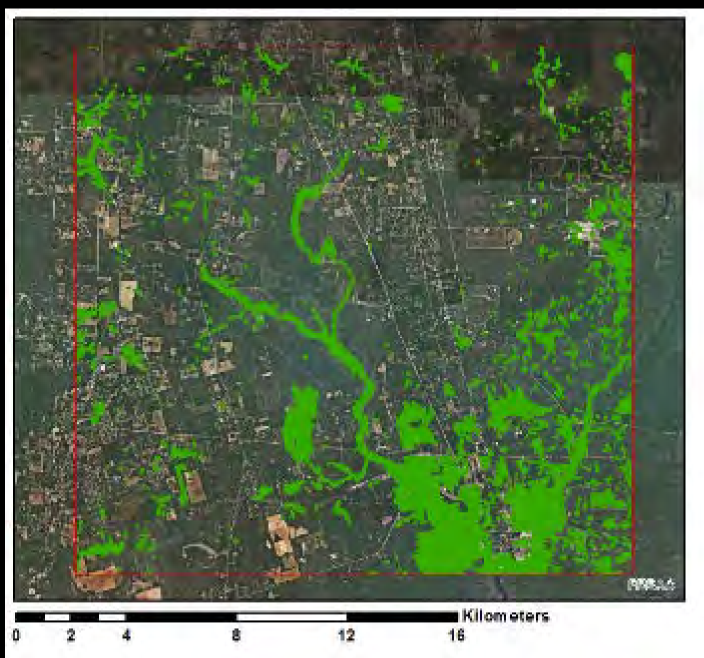
Binomial indicating Wetland or Upland Association (“1” or “0”)

Objective 3: Evaluate the ability of SVDs to Predict Wetland/Upland Association (cont'd)

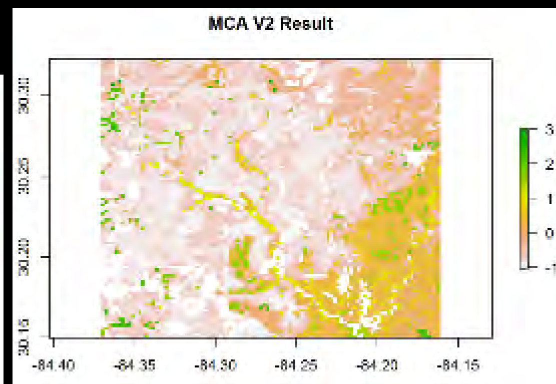


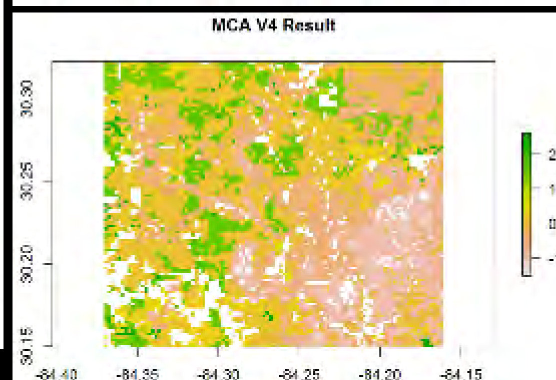
Since results were derived from georeferenced vector data, they can be plotted; here, as rasters with 250x250m cells.

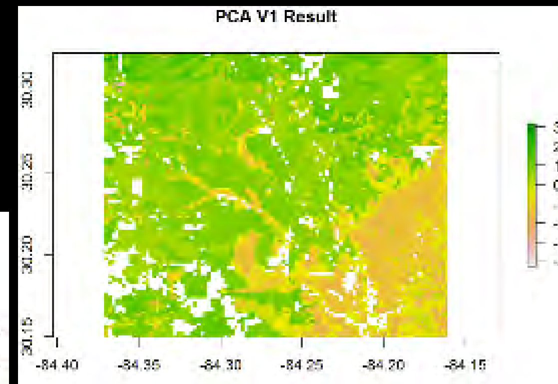
(94 SSURGO attributes have been converted to 5 variables.)

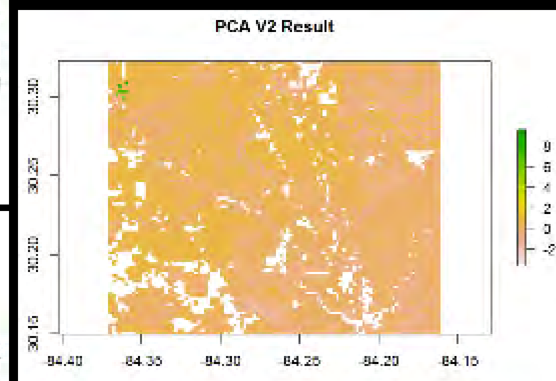


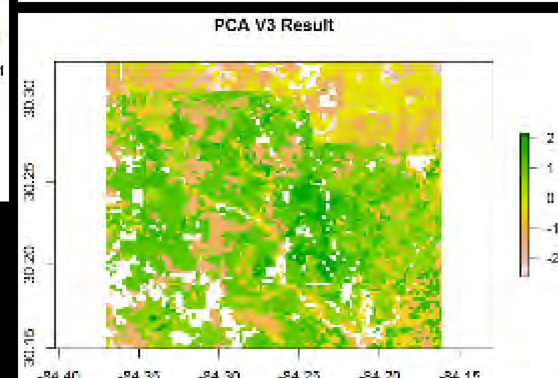
NWI Wetland







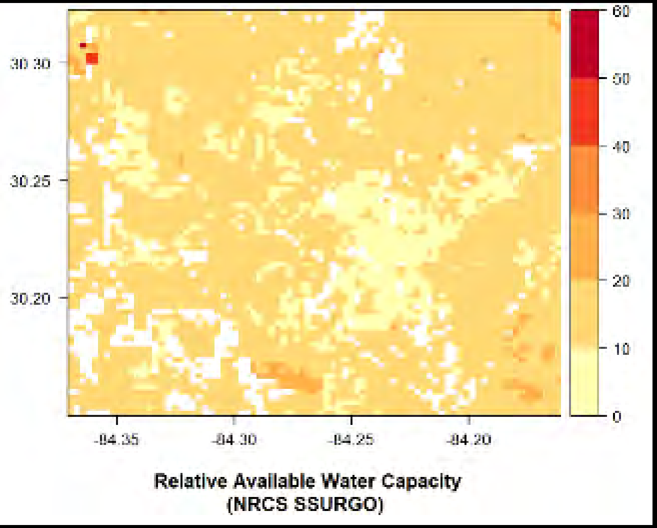




Available Water Capacity (AWC)

Steps

1. Rasterize SSURGO MUKEY for the area of interest (AOI).
2. Submit a data SQL query to the NRCS Soil Data Access server (SDA); requesting horizon-level available water capacity data for the AOI.
3. Aggregate data by profiling total water storage by soil horizon.
4. Calculate the average total water storage within each SMU (weighted by component percentage).
5. Join results to the MUKEY raster.
6. Verify response with a BIRN Model



Topographic Indices (TPI & CTI)

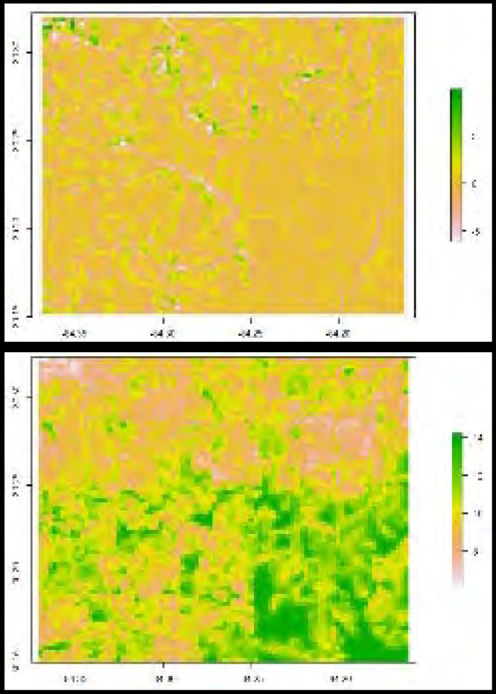
Topographic Position Index (TPI)

1. Often used to classify landscape morphologies (Mountains Vs. canyons Vs. plains, etc…)
2. Type of “ruggedness” or “roughness” index
3. Difference between target cell and mean of its eight neighbors

Compound Topographic Index (CTI)

1. “Wetness Index”
2. Higher values represent “wetter” areas
3. ln(a/tan B), a = Contributing area; B = Slope

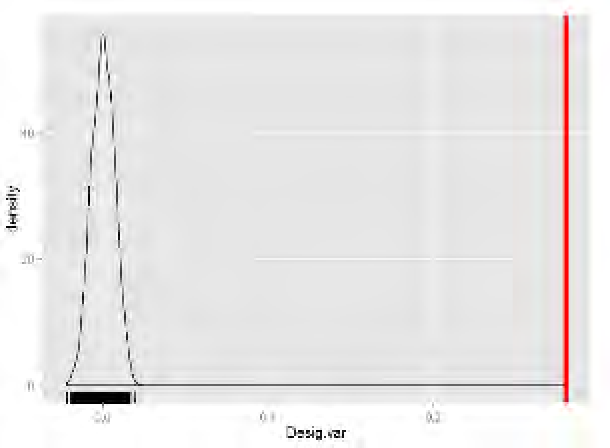
Both evaluated for co-linearity and applied to a BIRN Model prior to being considered for model inclusion.



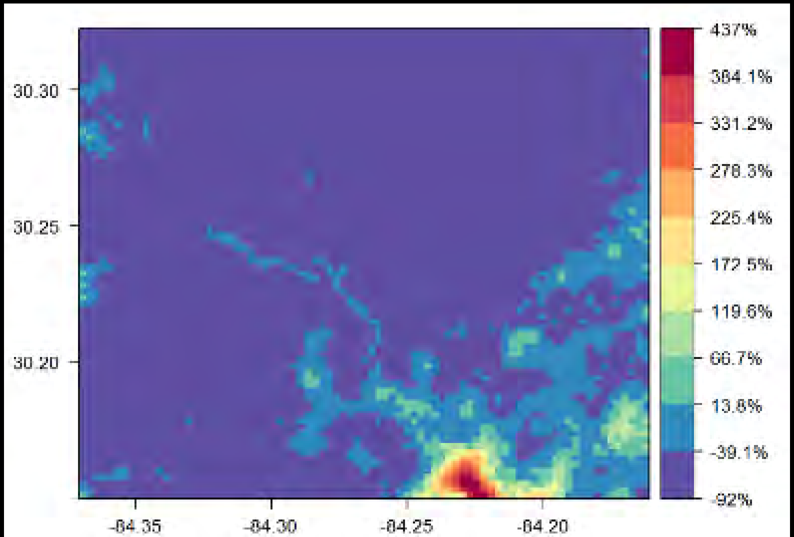
Top: TPI Raster
Bottom: CTI Raster
(250m resolution)

Spatial Correlation

The Neglected Variable



Non-parametric inference on Moran's I
(Monte-Carlo Simulation)



Model "Prediction" Based on Spatial Structure Alone
Legend Values are Relative to Mean Density of Wetlands
(250m resolution)

Model Construction

Model Construction

Designation ~ {

- Spatial Correlation
- MCA_{V2}
- MCA_{V4}
- PCA_{V1}
- PCA_{V2}
- PCA_{V3}
- AWC
- TPI
- CTI

The integrals cannot be solved analytically, so the integrated nested Laplace approximation (INLA) method is used. INLA provides a fast alternative to Markov Chain Monte Carlo simulation for models that have a latent Gaussian structure [Rue et al., 2009].

Rue, H., S. Martino, and N. Chopin (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 319-392.

Model Selection

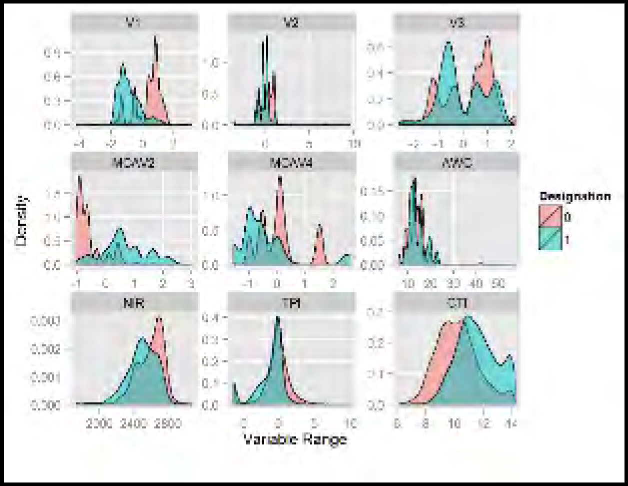
More than a dozen models were fitted:

1. Deviance Information Criterion (DIC)
    - Similar to the Akaike Information Criterion (AIC), but adapted to Bayesian hierarchical models.

2. Watanabe-Akaike Information Criteria (WAIC)
    - A more contemporary version of the DIC.

3. Log Conditional Predictive Ordinances (LCPO)
    - “Leave one out” Cross-validation Method.

4. Brier Score (BS)
    - Similar to a Root Mean Squared Error (RMSE) comparing results to original LC/LU Wetlands.



The "Final" Model

$$W_s \mid \mu_s, n \sim \text{NegBin}(\mu_s, n)$$

$$\mu_s = \exp(\nu_s \cdot \text{area}_s)$$

$$\nu_s = \beta_0 + \beta_{PCA_{V2}} \cdot PCA_{V2} + \beta_{PCA_{V3}} \cdot PCA_{V3} + \beta_{MCA_{V2}} \cdot MCA_{V2} + \beta_{AWC} \cdot AWC + \beta_{TPI} \cdot TPI + \beta_{CTI} \cdot CTI + u_s$$

The random effect (***u_s***) follows a Besag formulation [Besag, 1975]:

$$u_i \setminus u_j \,,\, i \neq j \,,\, \tau \sim N\left(\frac{1}{m_i} \sum_{i \sim j} u_j \,, \frac{1}{m_i} \tau\right)$$

Where ***N*** is the normal distribution with mean $\frac{1}{m_i} * \sum_{i \sim j} u_j$ and variance $\frac{1}{m_i} * \tau$ where m_i is the number of neighboring cells of cell ***í*** and τ is the precision; ***í~j*** indicates cells ***í*** and ***j*** are neighbors.

Besag, J. (1975), Statistical analysis of non-lattice data, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24, 179-195.

Results

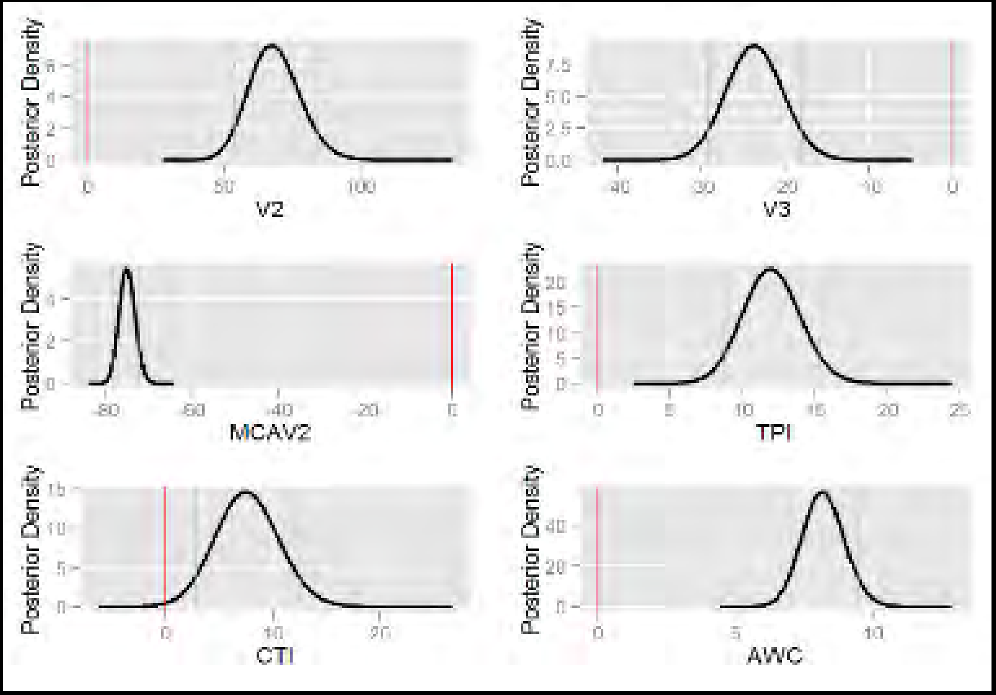
Model Summary

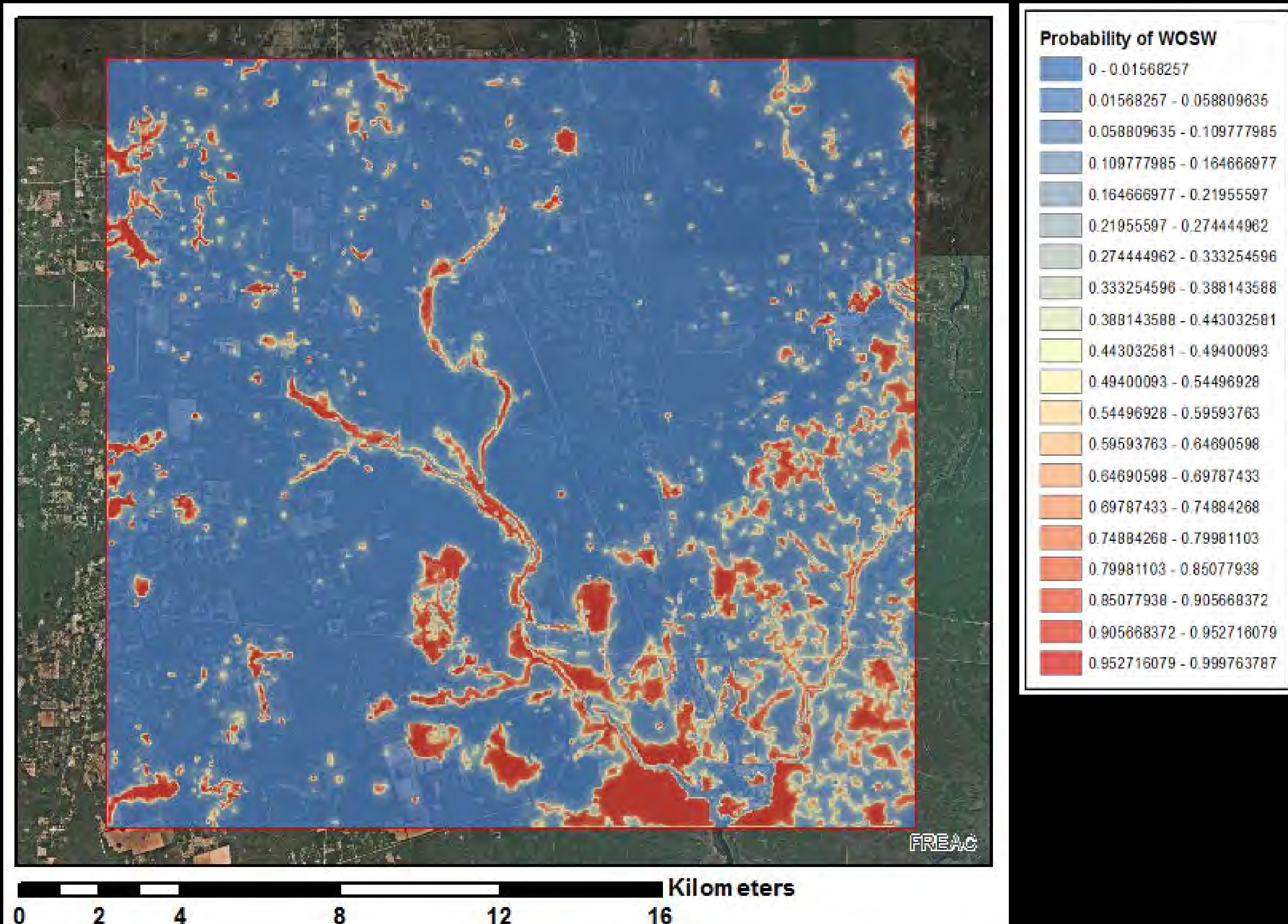
Results Summary

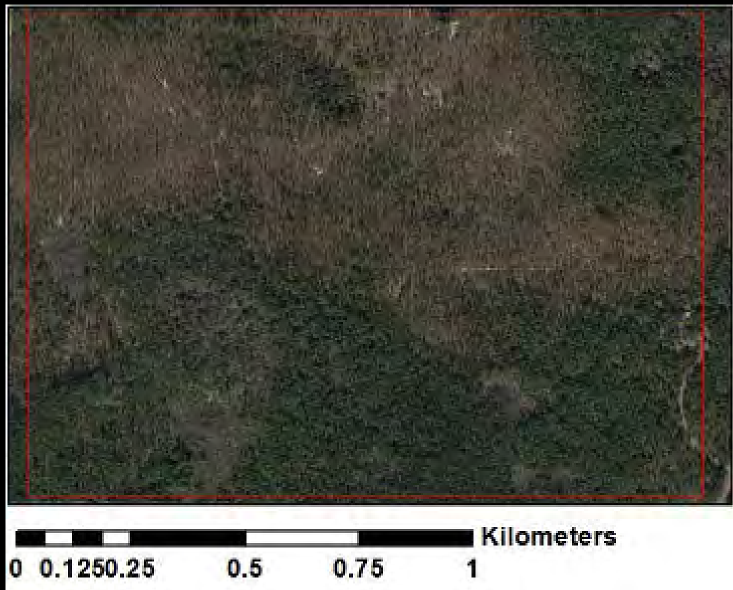
Effect	Posterior Mean	Credible Interval	
		0.025 Quant	0.975 Quant
PCAV2	-0.5171	-0.6278	-0.5164
PCAV3	0.2722	0.1863	0.2716
MCAV2	1.3972	1.2579	1.396
AWC	0.0787	0.0653	0.0786
TPI	-0.1128	-0.1479	-0.1127
CTI	0.0722	0.0186	0.0721

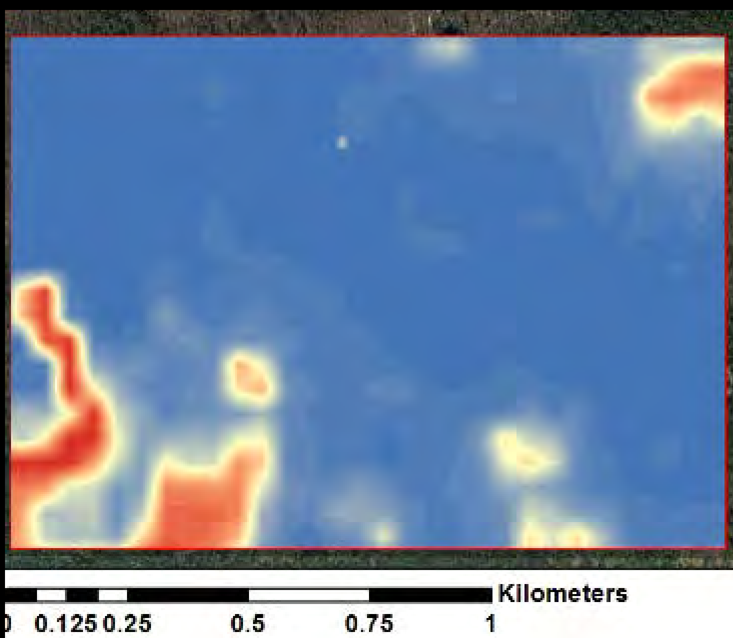
Brier Score (BS)
0.0723

Posterior Distributions



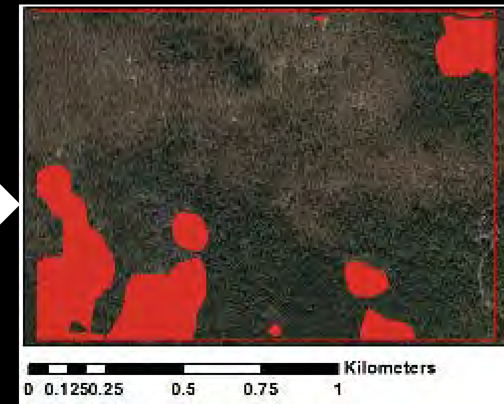

Probability of WOSW
0 - 0.01568257
0.01568257 - 0.058809635
0.058809635 - 0.109777985
0.109777985 - 0.164666977
0.164666977 - 0.21955597
0.21955597 - 0.274444962
0.274444962 - 0.333254596
0.333254596 - 0.388143588
0.388143588 - 0.443032581
0.443032581 - 0.49400093
0.49400093 - 0.54496928
0.54496928 - 0.59593763
0.59593763 - 0.64690598
0.64690598 - 0.69787433
0.69787433 - 0.74884268
0.74884268 - 0.79981103
0.79981103 - 0.85077938
0.85077938 - 0.905668372
0.905668372 - 0.952716079
0.952716079 - 0.999763787
FREAC
Kilometers
0 2 4 8 12 16





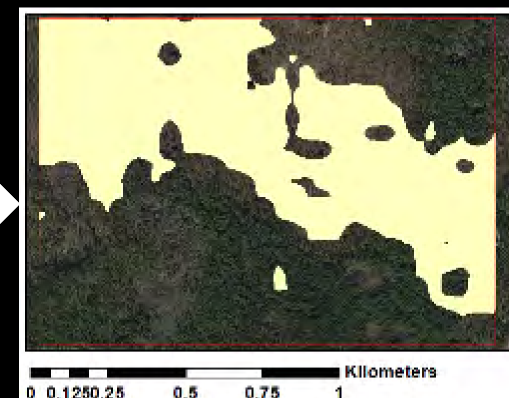
Wetland Probability =

0.50 or Greater



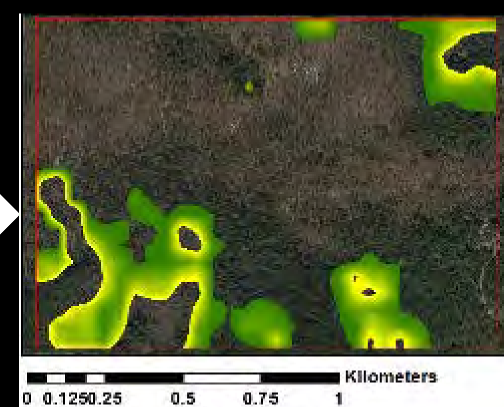
Wetland Probability =

0.02 or Less

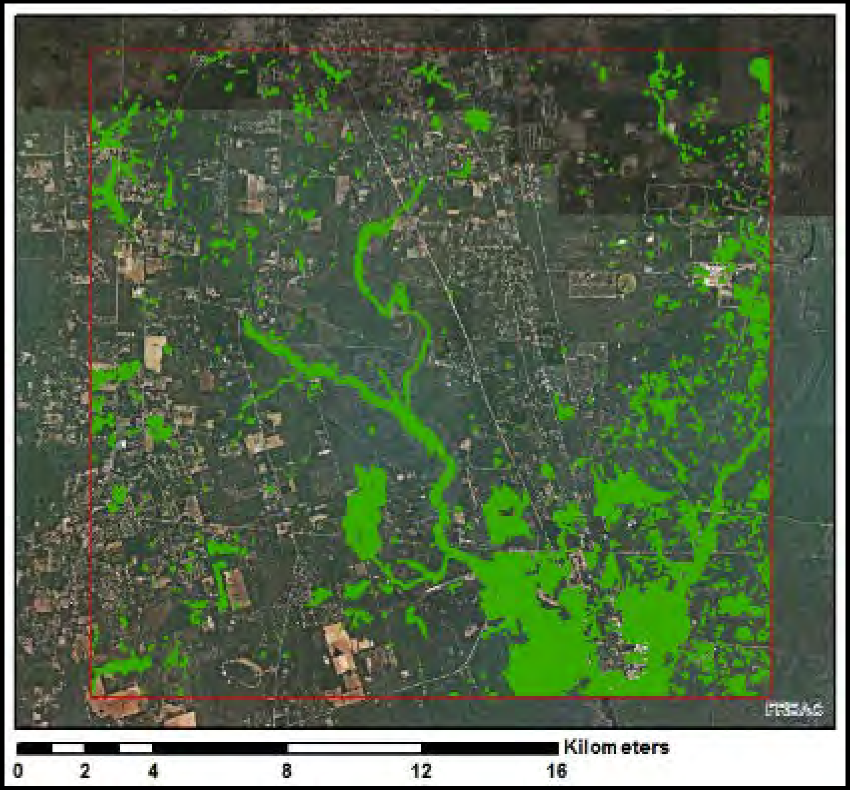


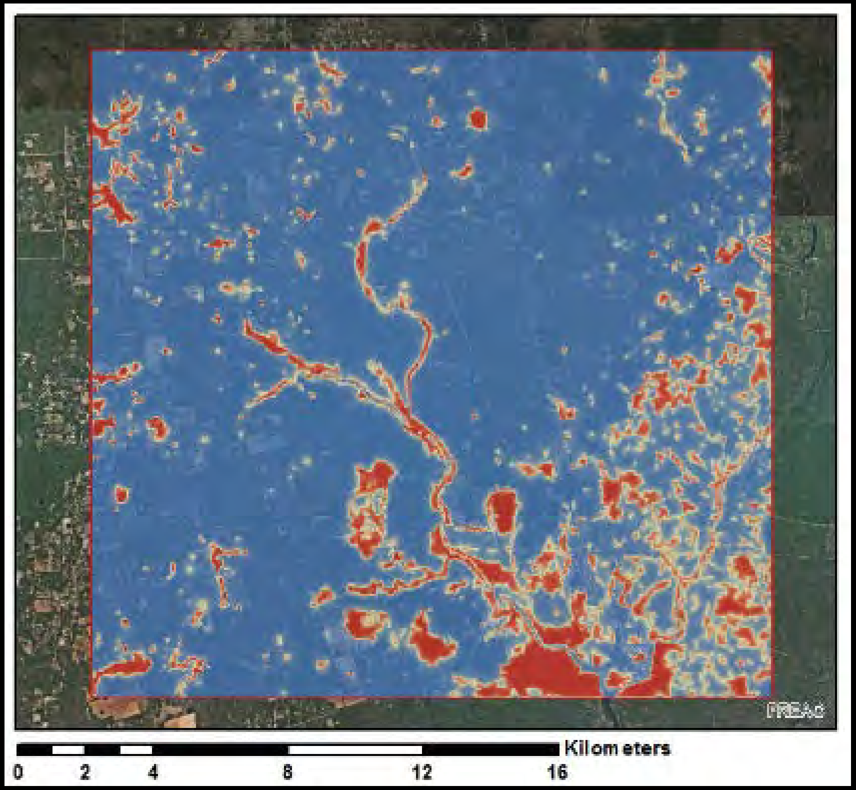
Wetland Probability =

0.10 – 0.49



Comparison to the National Wetlands Inventory (NWI)

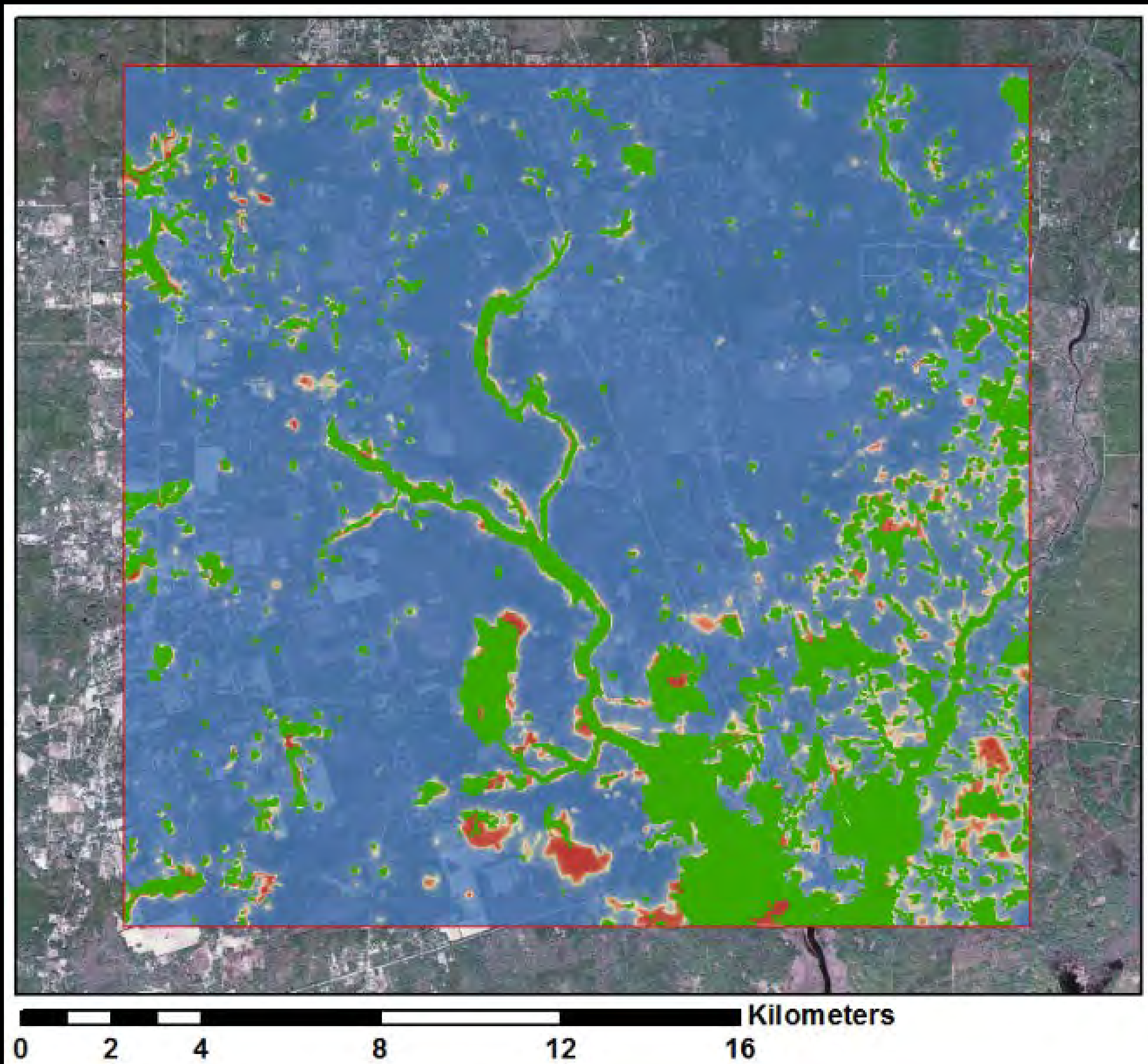


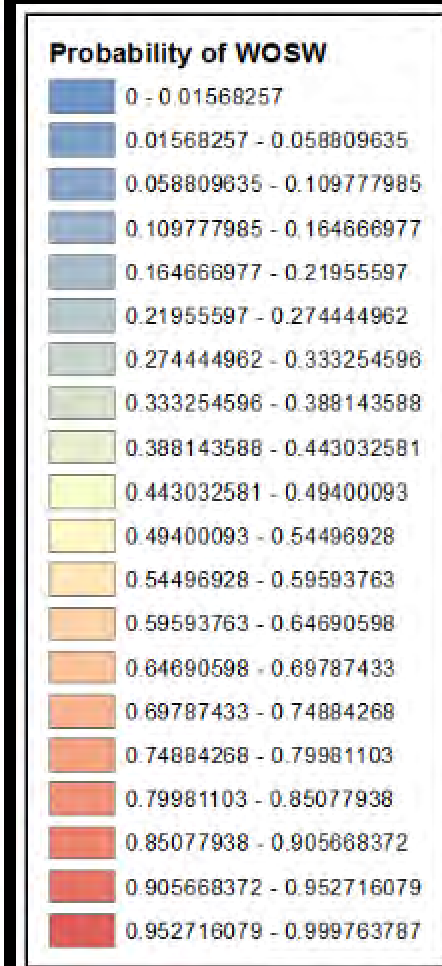


NWI Wetland

~40mn Run Time

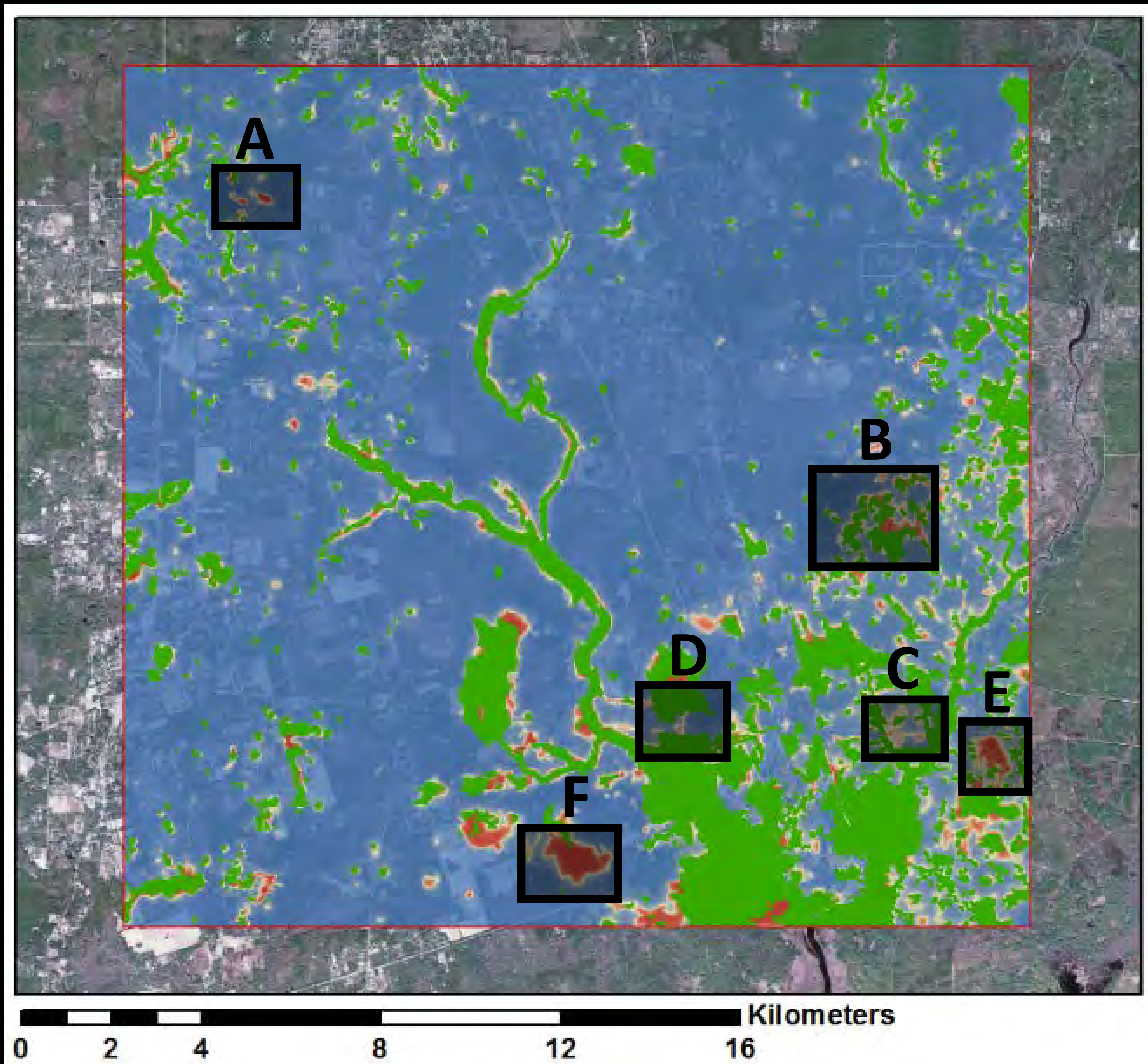
~30% More Wetland Area (@0.50 Probability)





NWI Wetland

~30% More Wetland Area (@0.50 Probability)

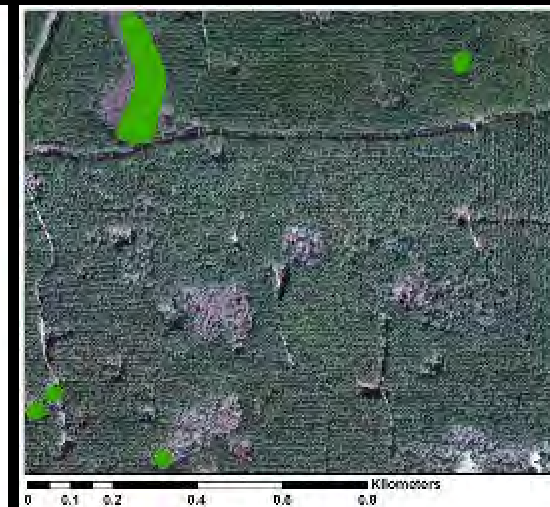


Probability of WOSW

- 0 - 0.01568257
- 0.01568257 - 0.058809635
- 0.058809635 - 0.109777985
- 0.109777985 - 0.164666977
- 0.164666977 - 0.21955597
- 0.21955597 - 0.274444962
- 0.274444962 - 0.333254596
- 0.333254596 - 0.388143588
- 0.388143588 - 0.443032581
- 0.443032581 - 0.49400093
- 0.49400093 - 0.54496928
- 0.54496928 - 0.59593763
- 0.59593763 - 0.64690598
- 0.64690598 - 0.69787433
- 0.69787433 - 0.74884268
- 0.74884268 - 0.79981103
- 0.79981103 - 0.85077938
- 0.85077938 - 0.905668372
- 0.905668372 - 0.952716079
- 0.952716079 - 0.999763787

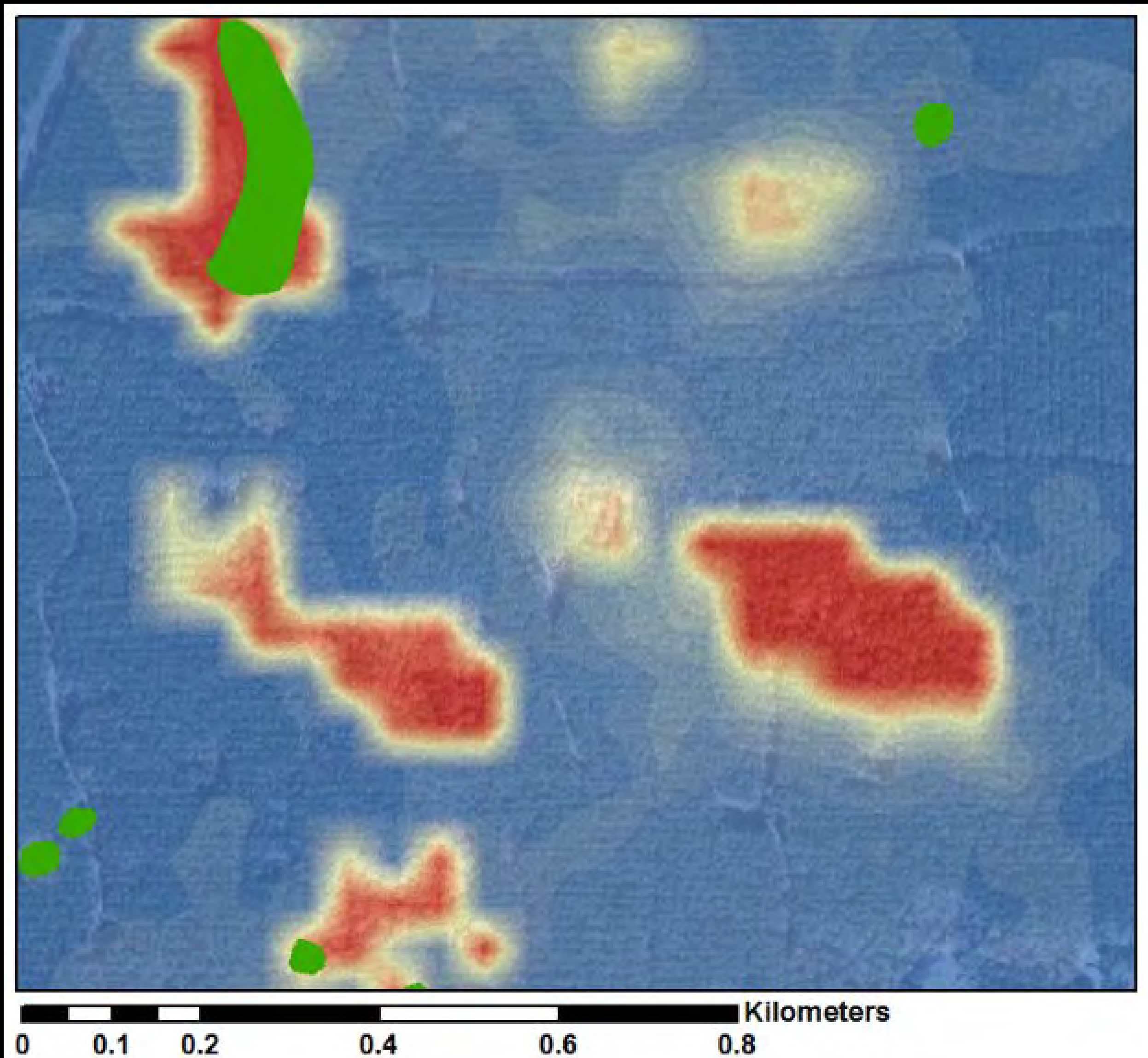
NWI Wetland

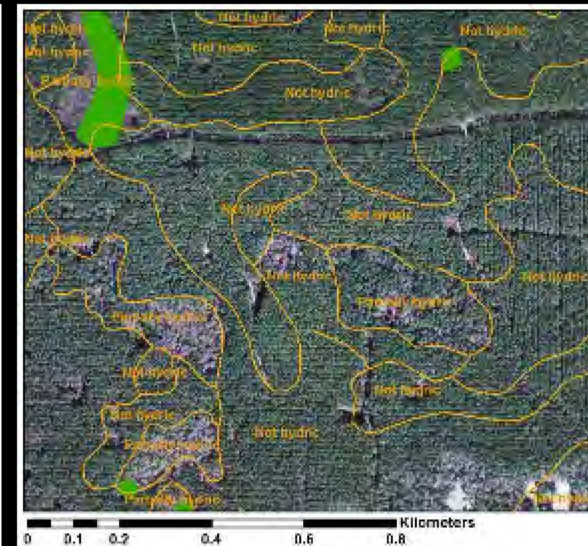
A



NWI Wetland

A

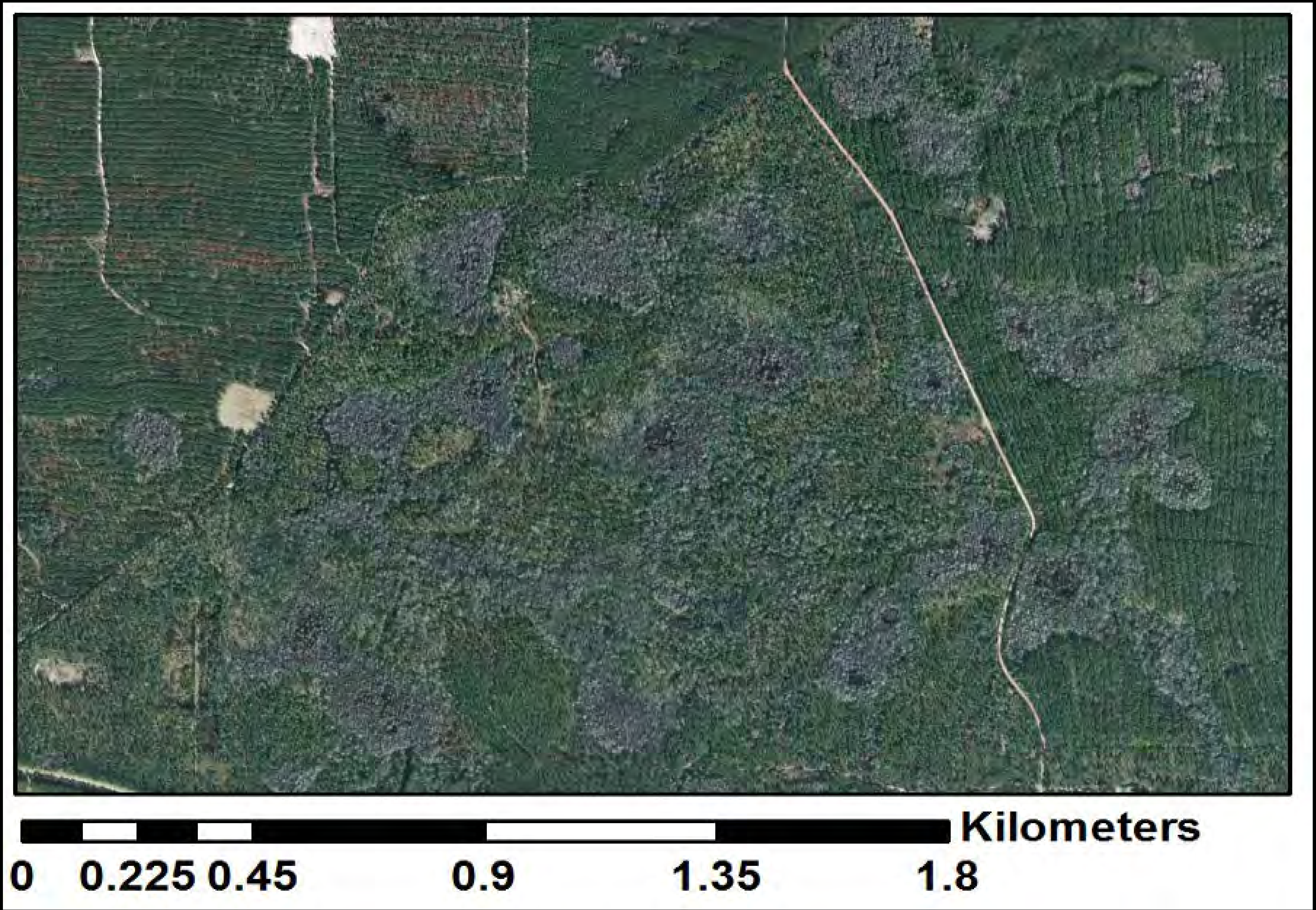




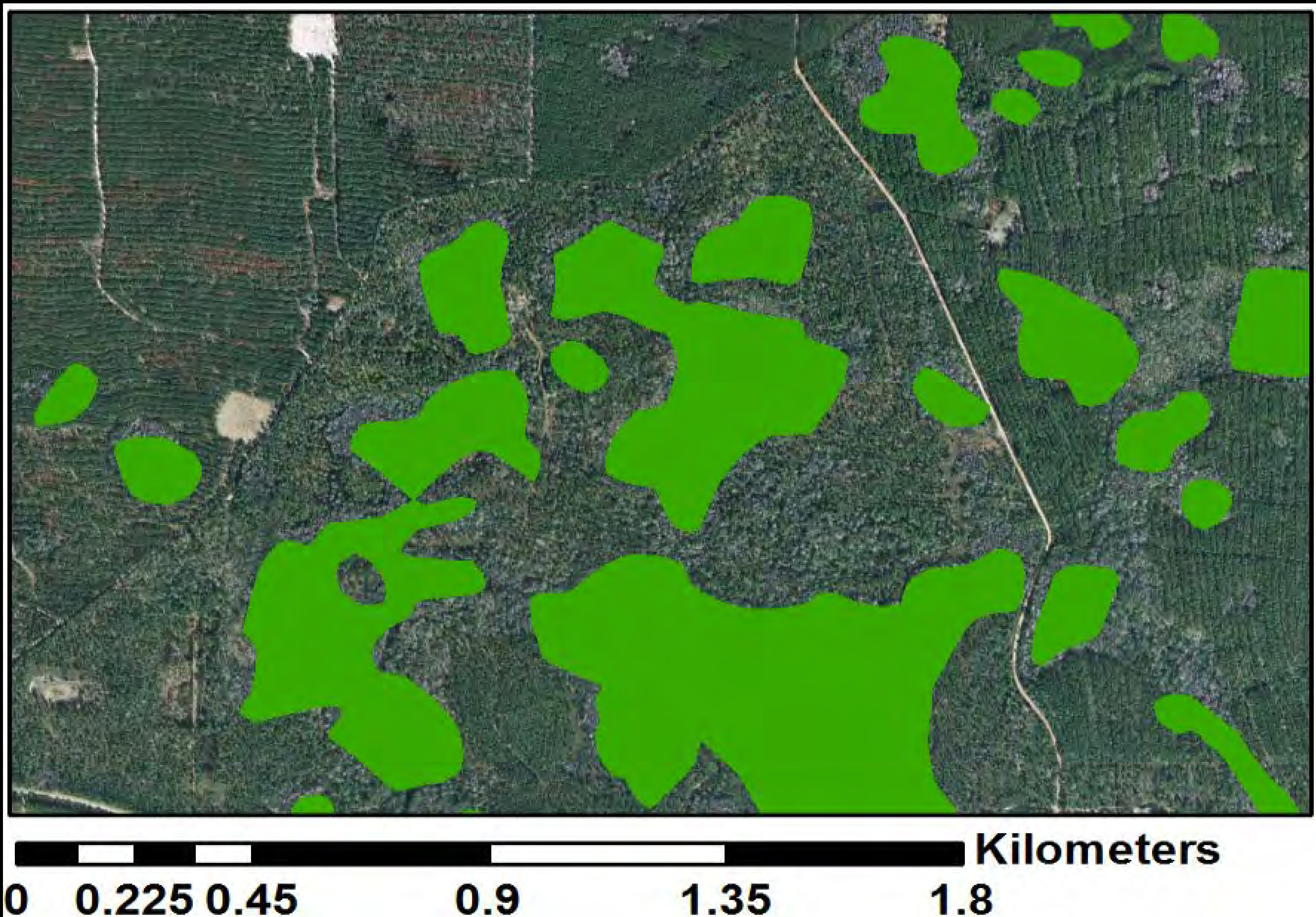
NWI Wetland

SSURGO Soils

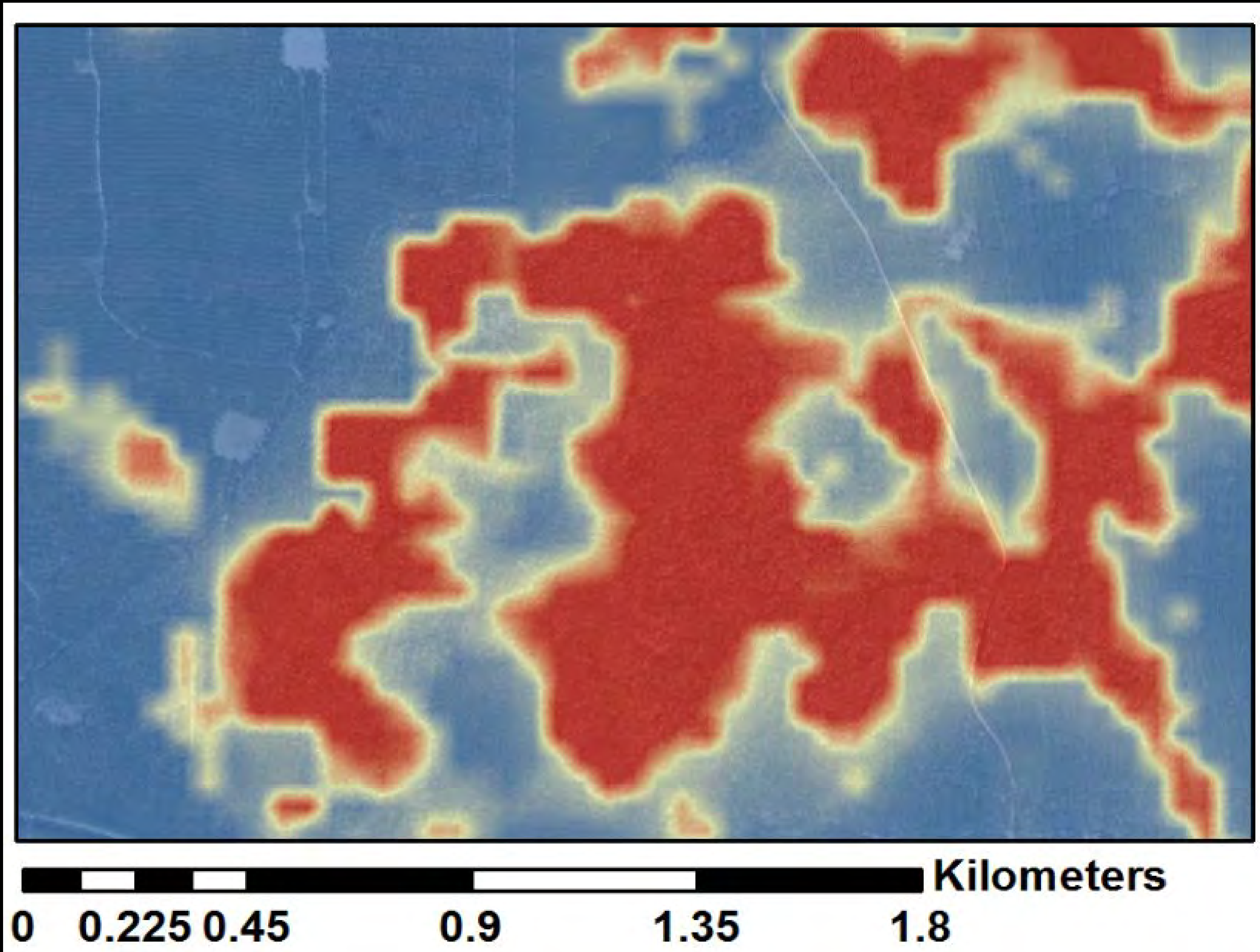
B



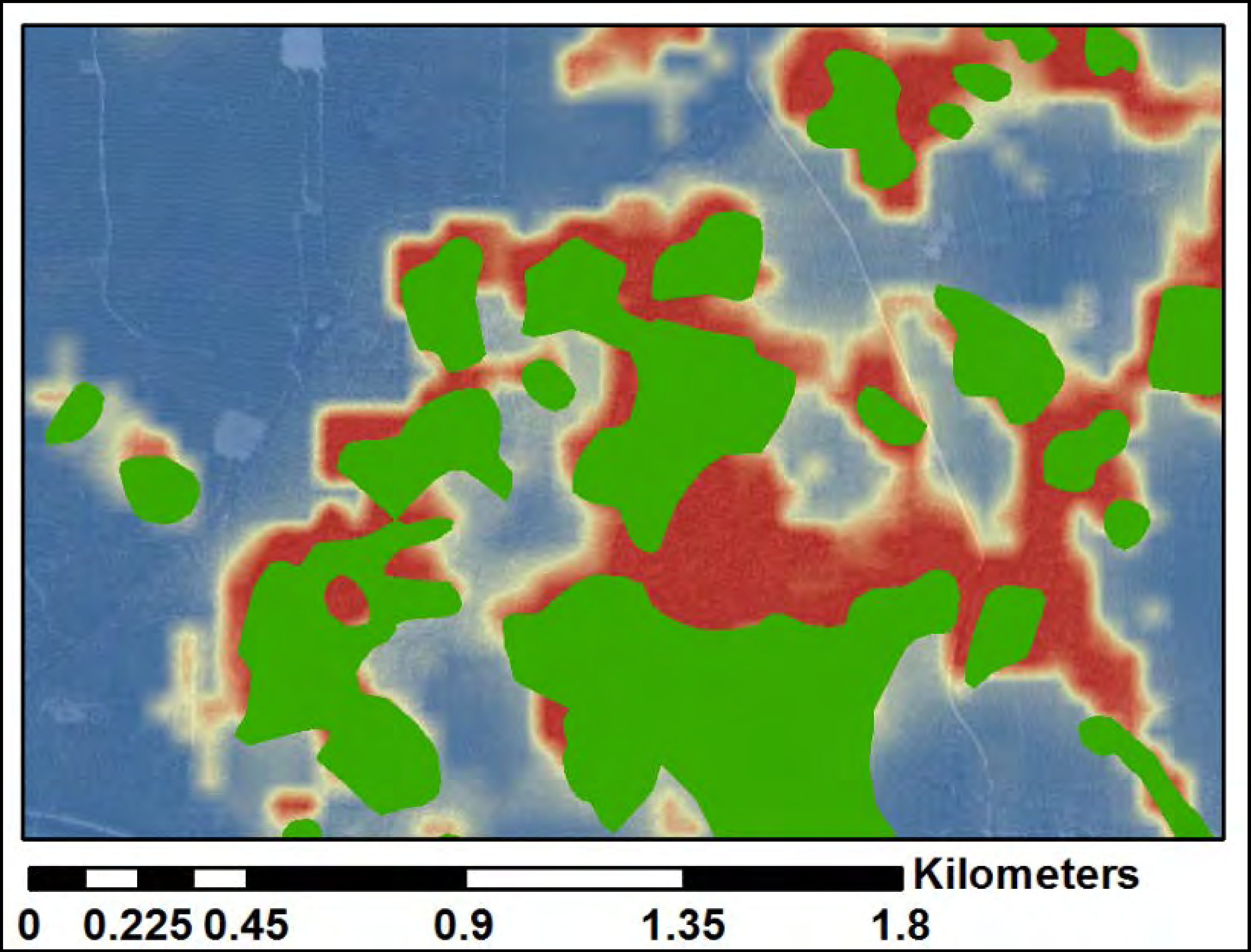
B



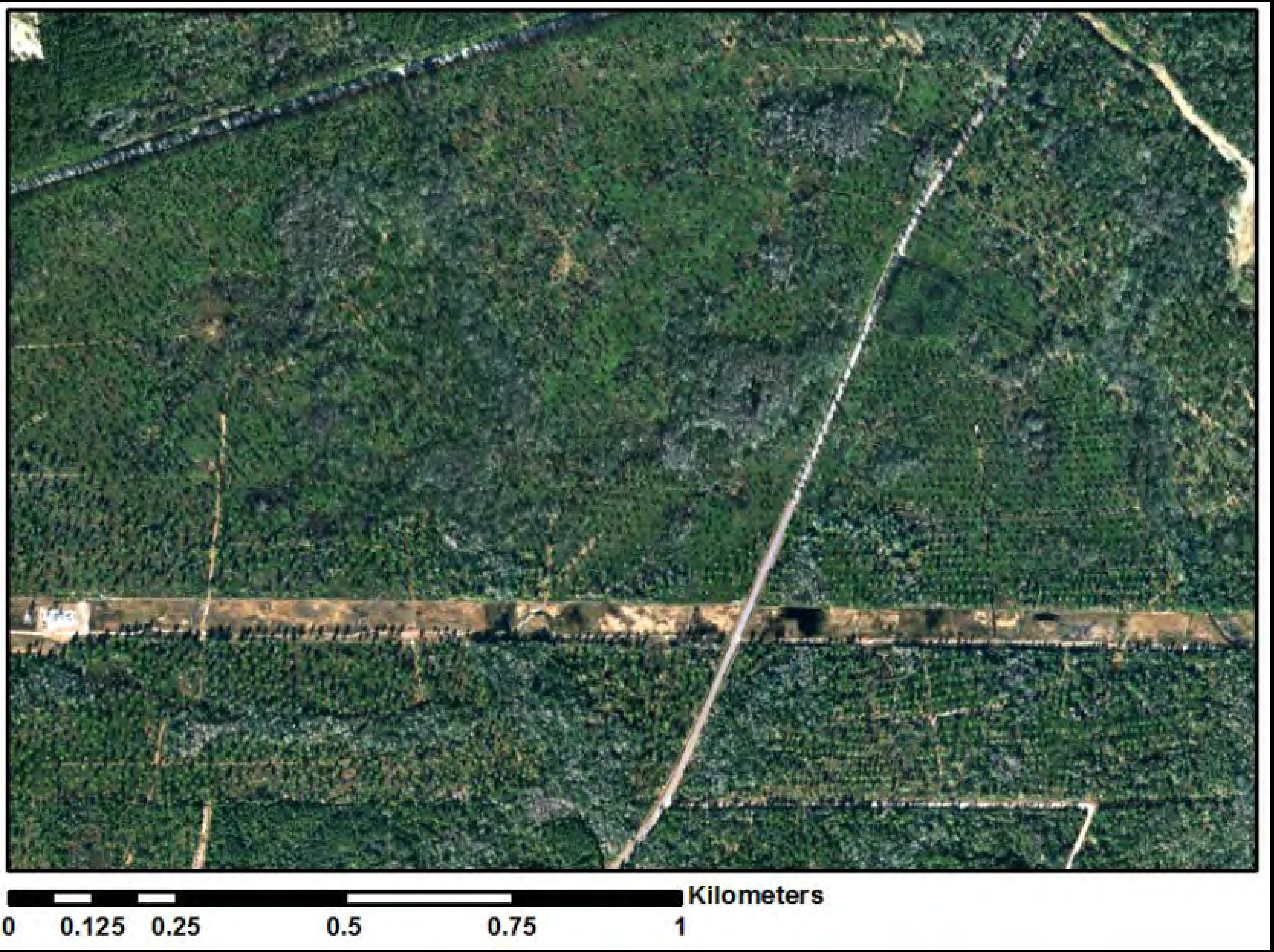
B



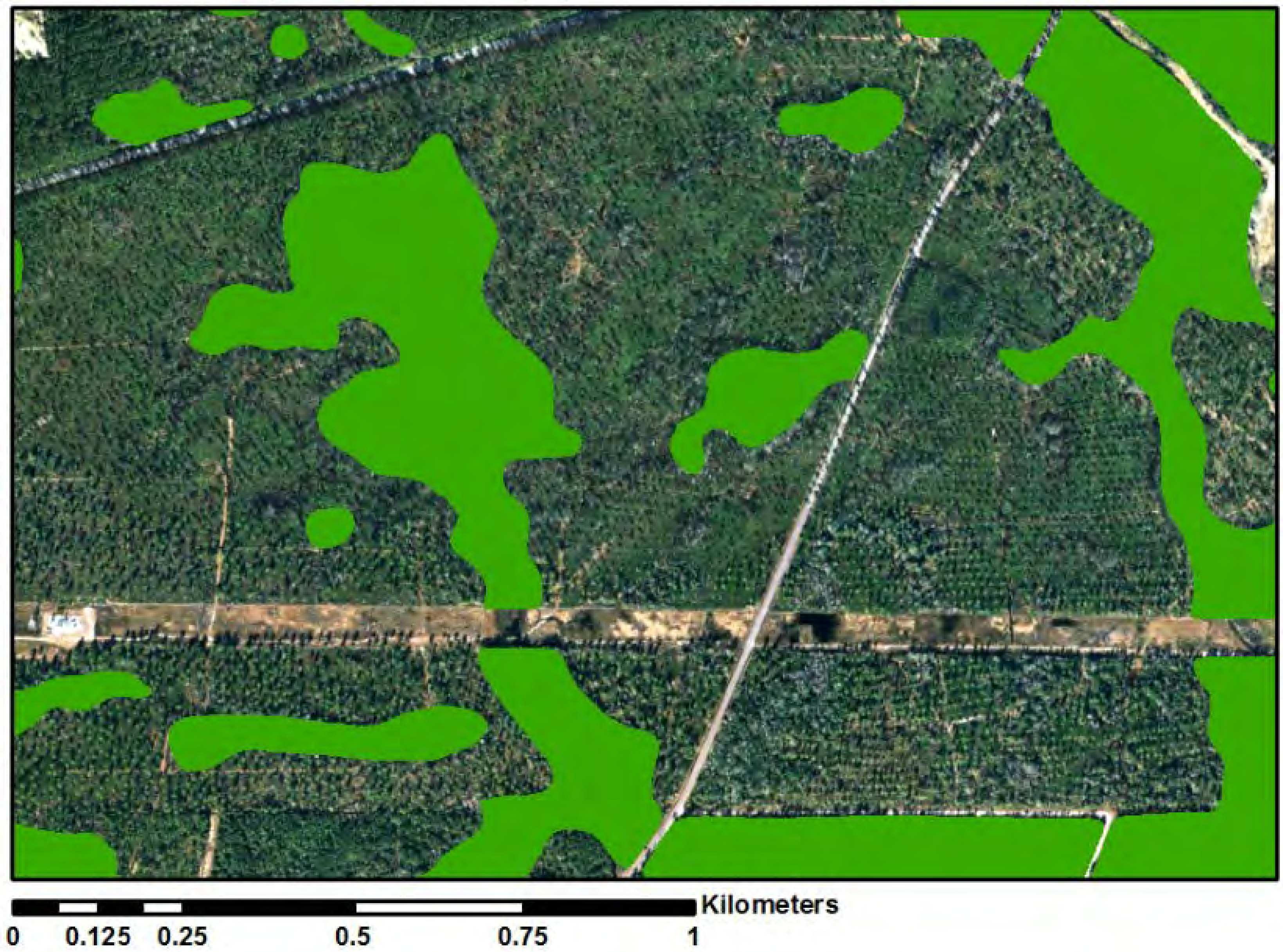
B



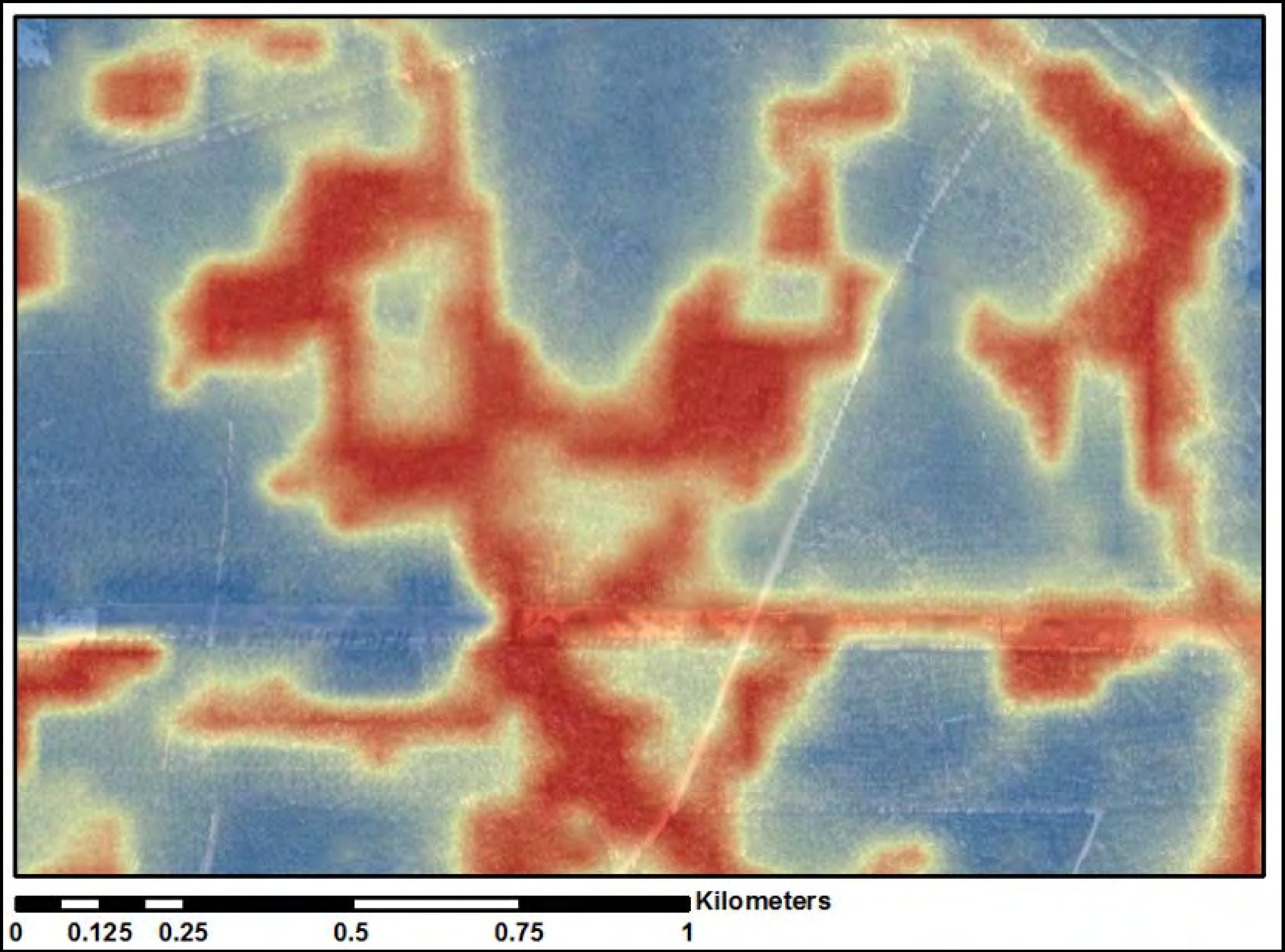
C



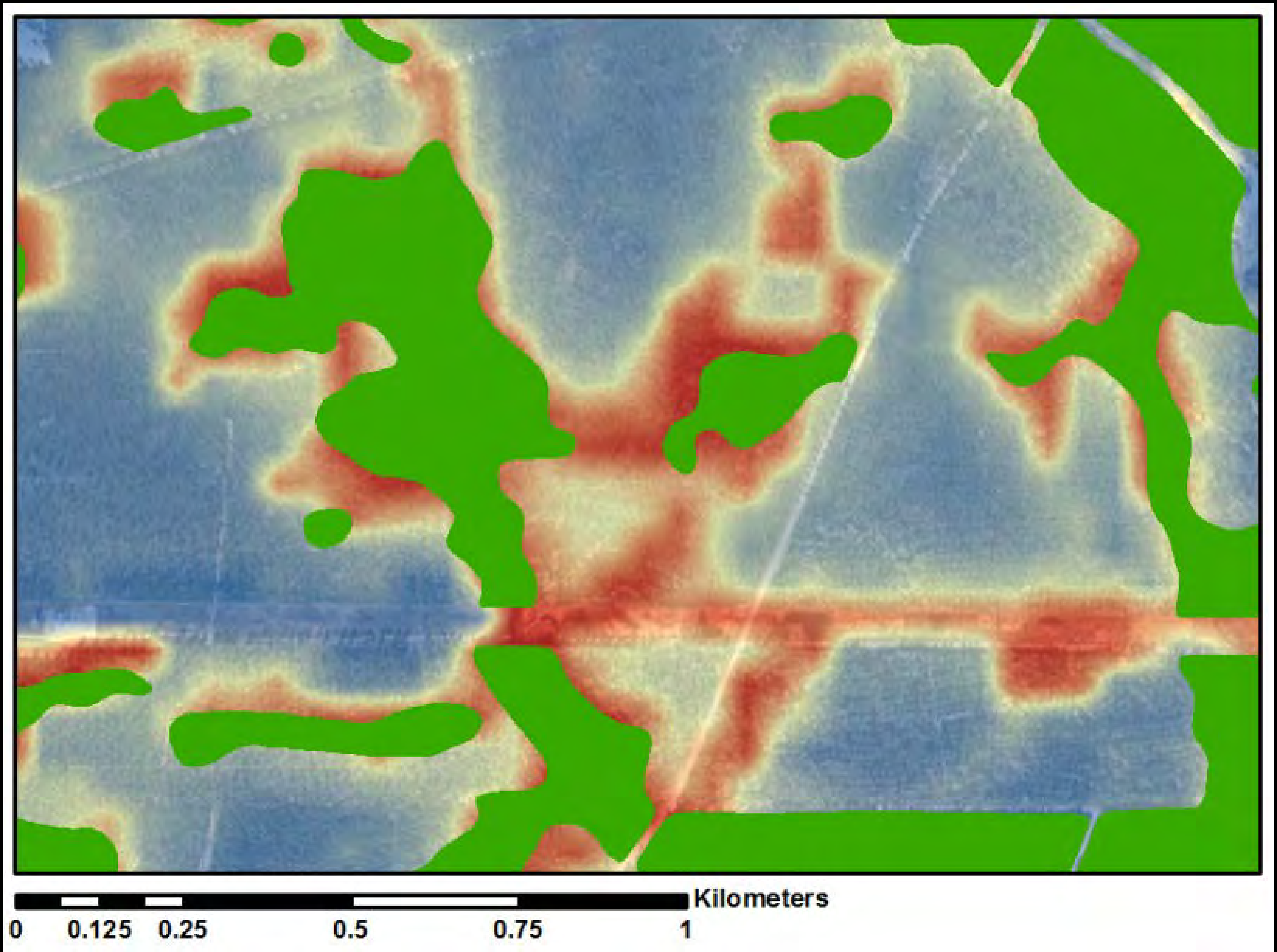
C



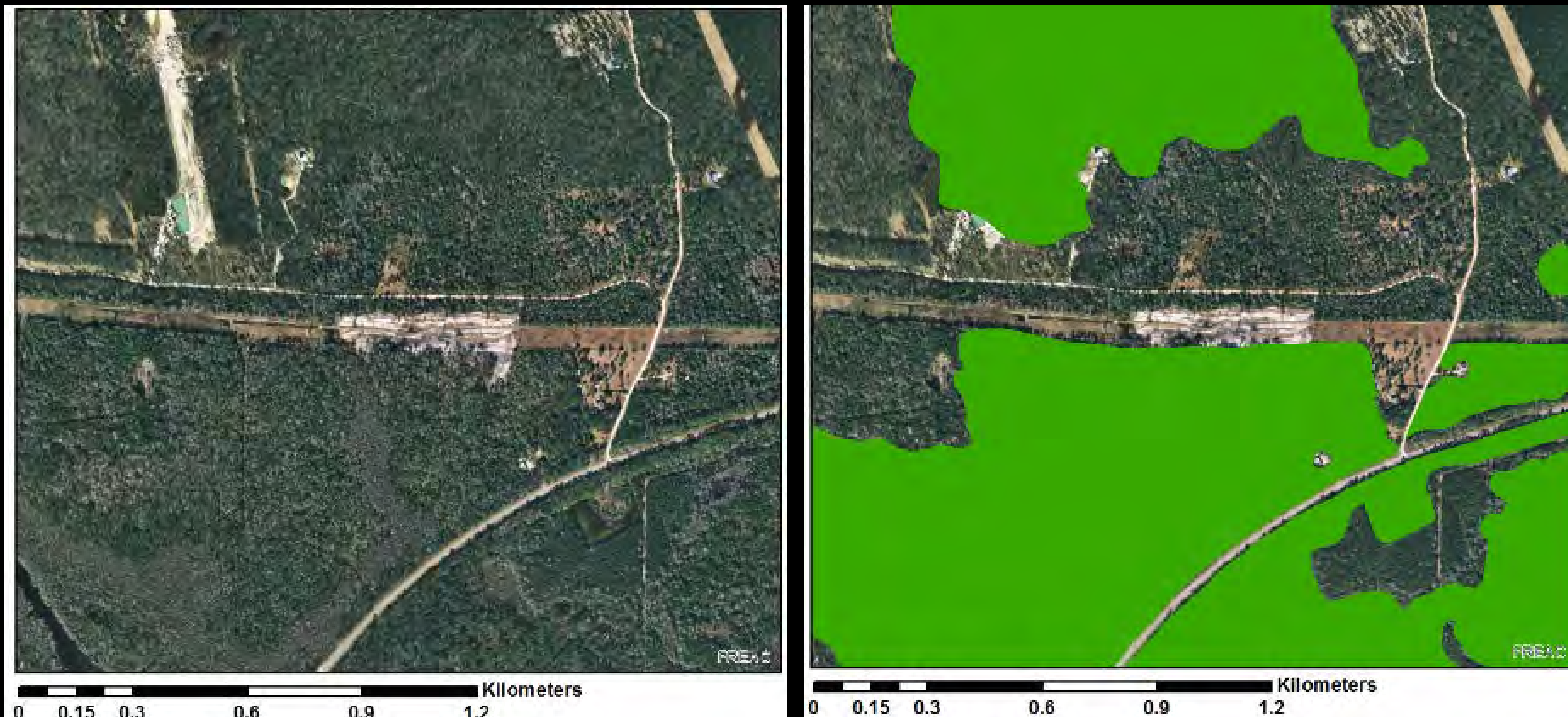
C



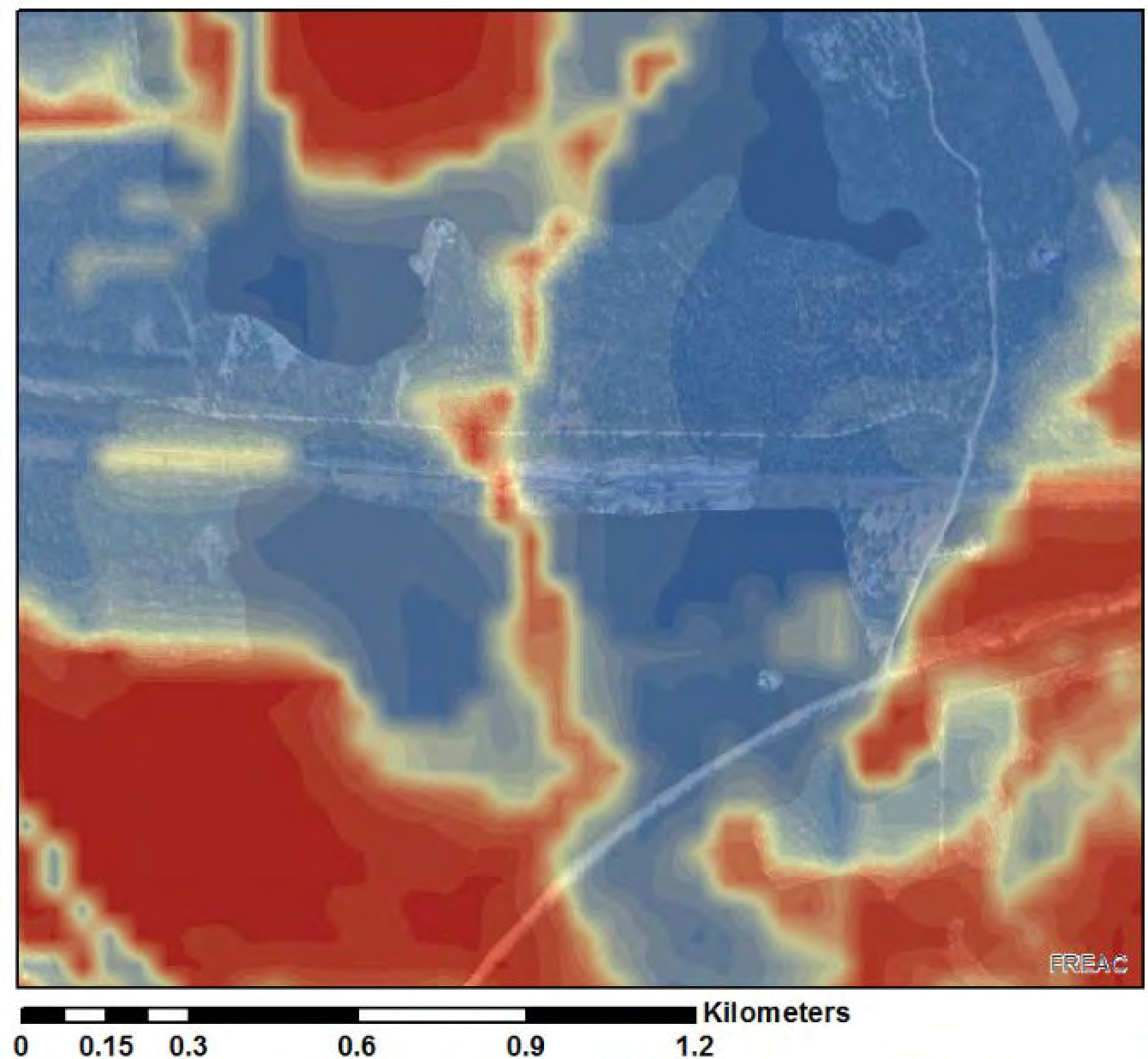
C



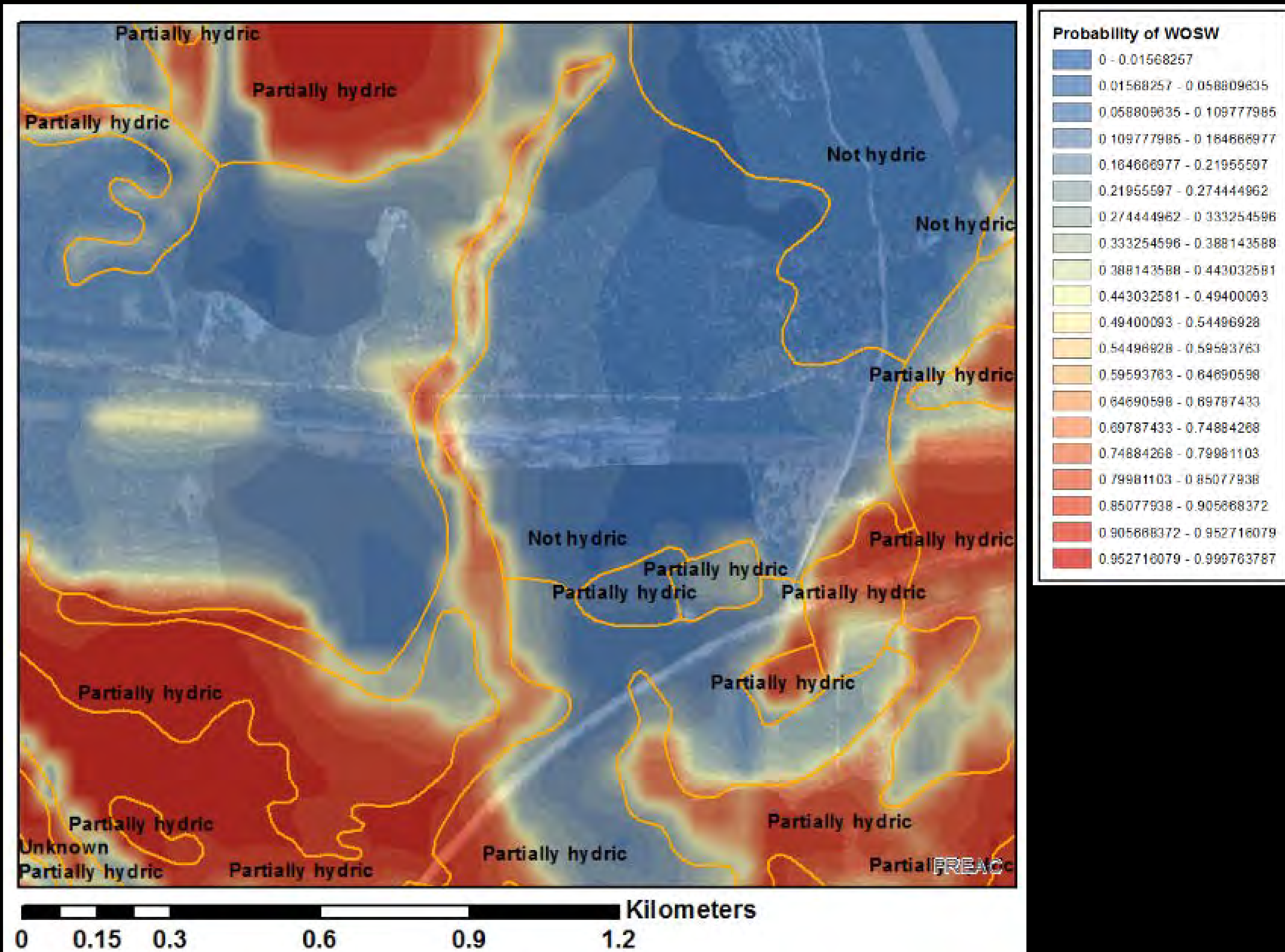
D



D



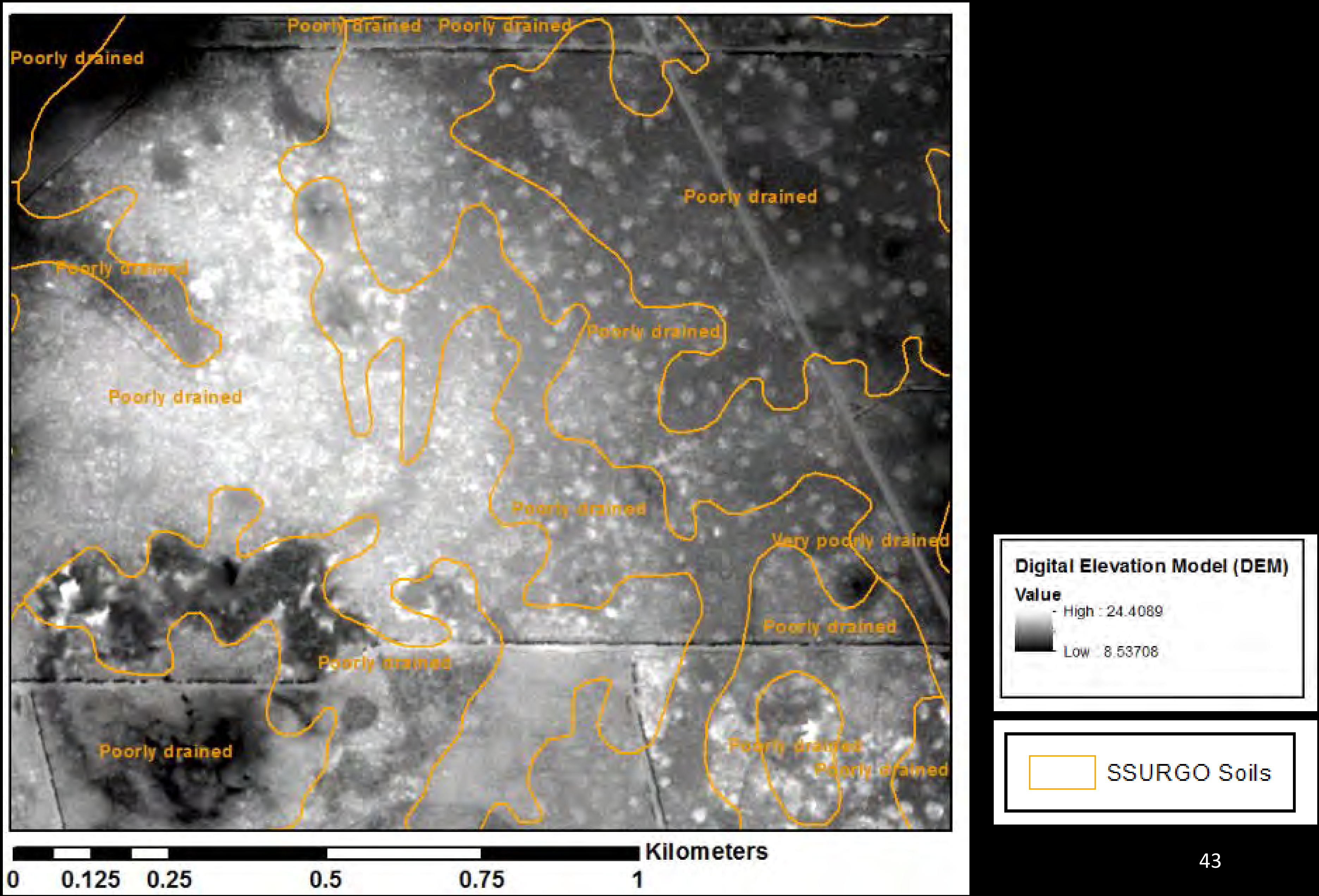
D



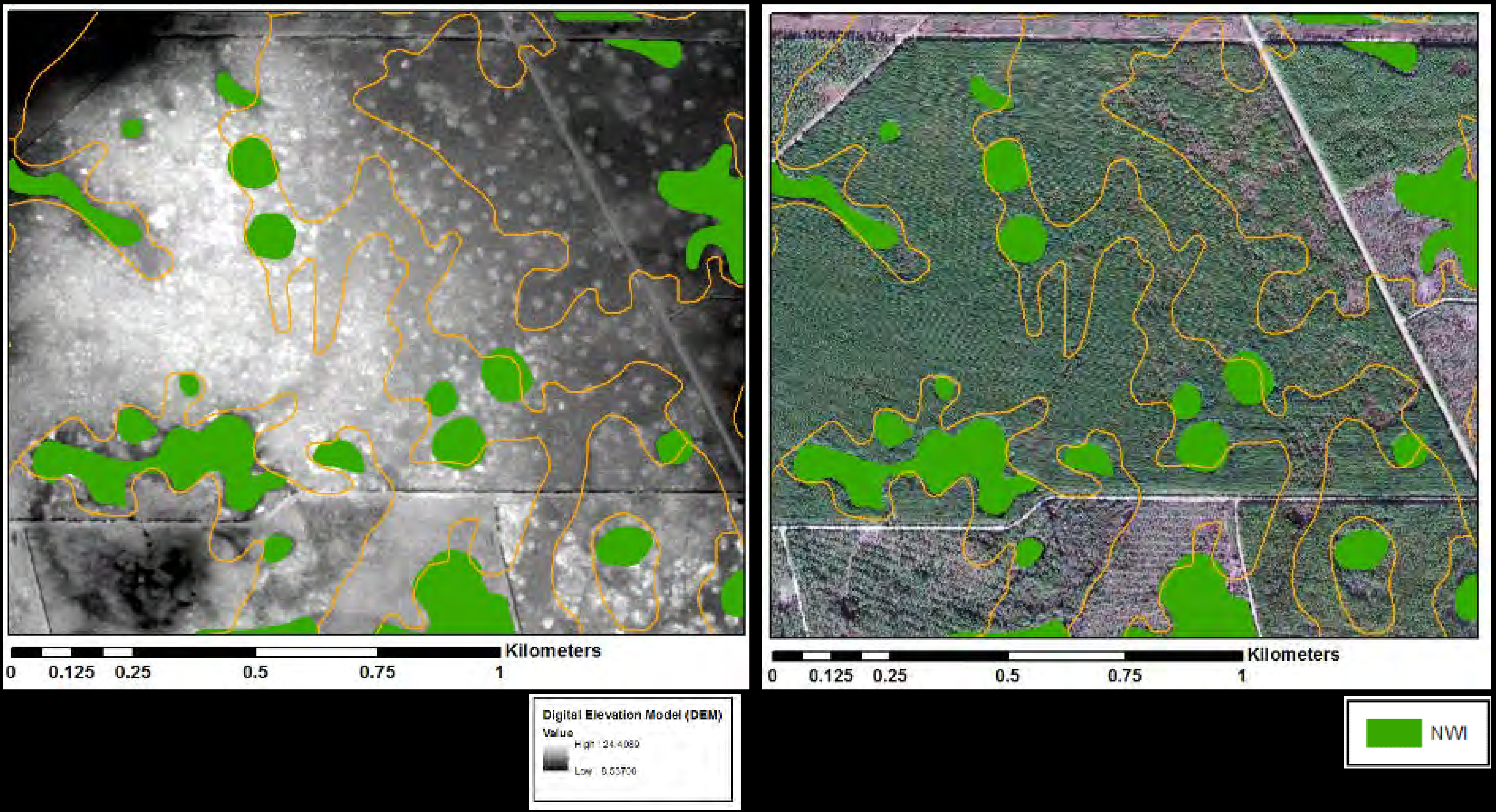
Partially hydric
Partially hydric
Partially hydric
Not hydric
Not hydric
Partially hydric
Not hydric
Partially hydric
Partially hydric
Partially hydric
Partially hydric
Partially hydric
Partially hydric
Partially hydric
Unknown
Partially hydric
Partially hydric
Partially hydric
Partially hydric
FREAC
Kilometers
0
0.15
0.3
0.6
0.9
1.2
Probability of WOSW
0 - 0.01568257
0.01568257 - 0.058809635
0.058809635 - 0.109777985
0.109777985 - 0.164666977
0.164666977 - 0.21955597
0.21955597 - 0.274444962
0.274444962 - 0.333254596
0.333254596 - 0.388143588
0.388143588 - 0.443032581
0.443032581 - 0.49400093
0.49400093 - 0.54496928
0.54496928 - 0.59593763
0.59593763 - 0.64690598
0.64690598 - 0.69787433
0.69787433 - 0.74884268
0.74884268 - 0.79981103
0.79981103 - 0.85077938
0.85077938 - 0.905668372
0.905668372 - 0.952716079
0.952716079 - 0.999763787

E

E

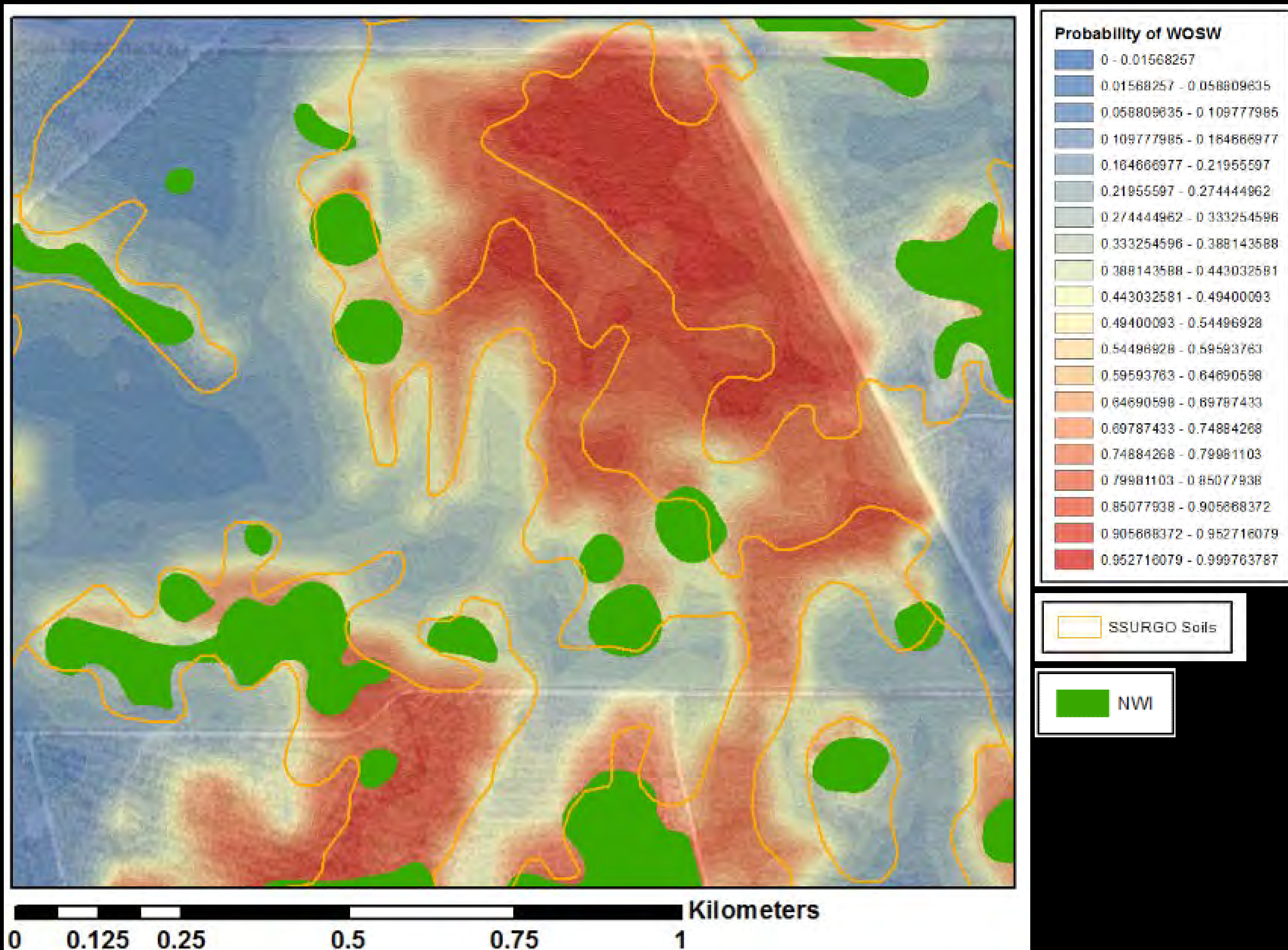

Poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Very poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Poorly drained
Kilometers
0
0.125
0.25
0.5
0.75
1
Digital Elevation Model (DEM)
Value
High : 24.4089
Low : 8.53708
SSURGO Soils

E



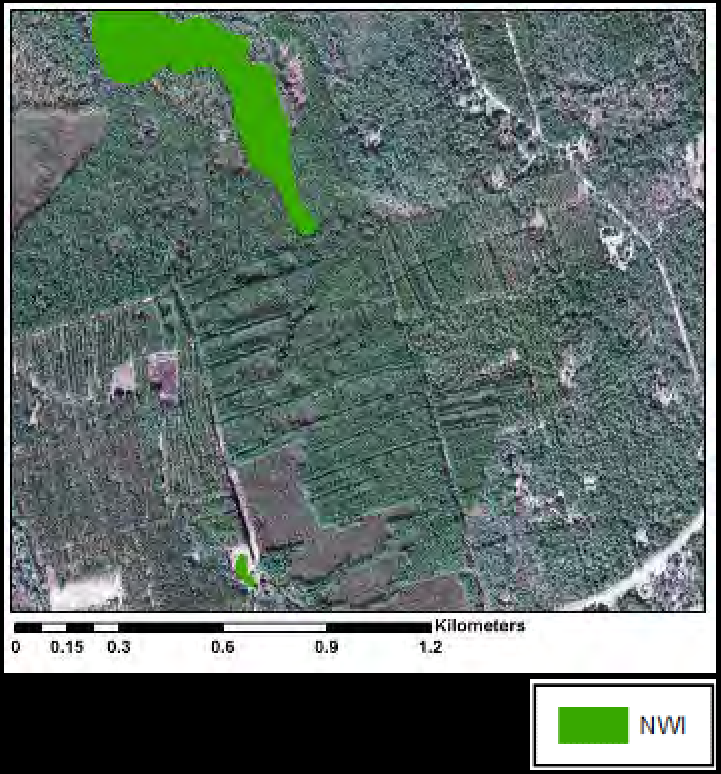
Kilometers
0 0.125 0.25 0.5 0.75 1
Digital Elevation Model (DEM)
Value
High : 24.4089
Low
Kilometers
0 0.125 0.25 0.5 0.75 1
NWI

E

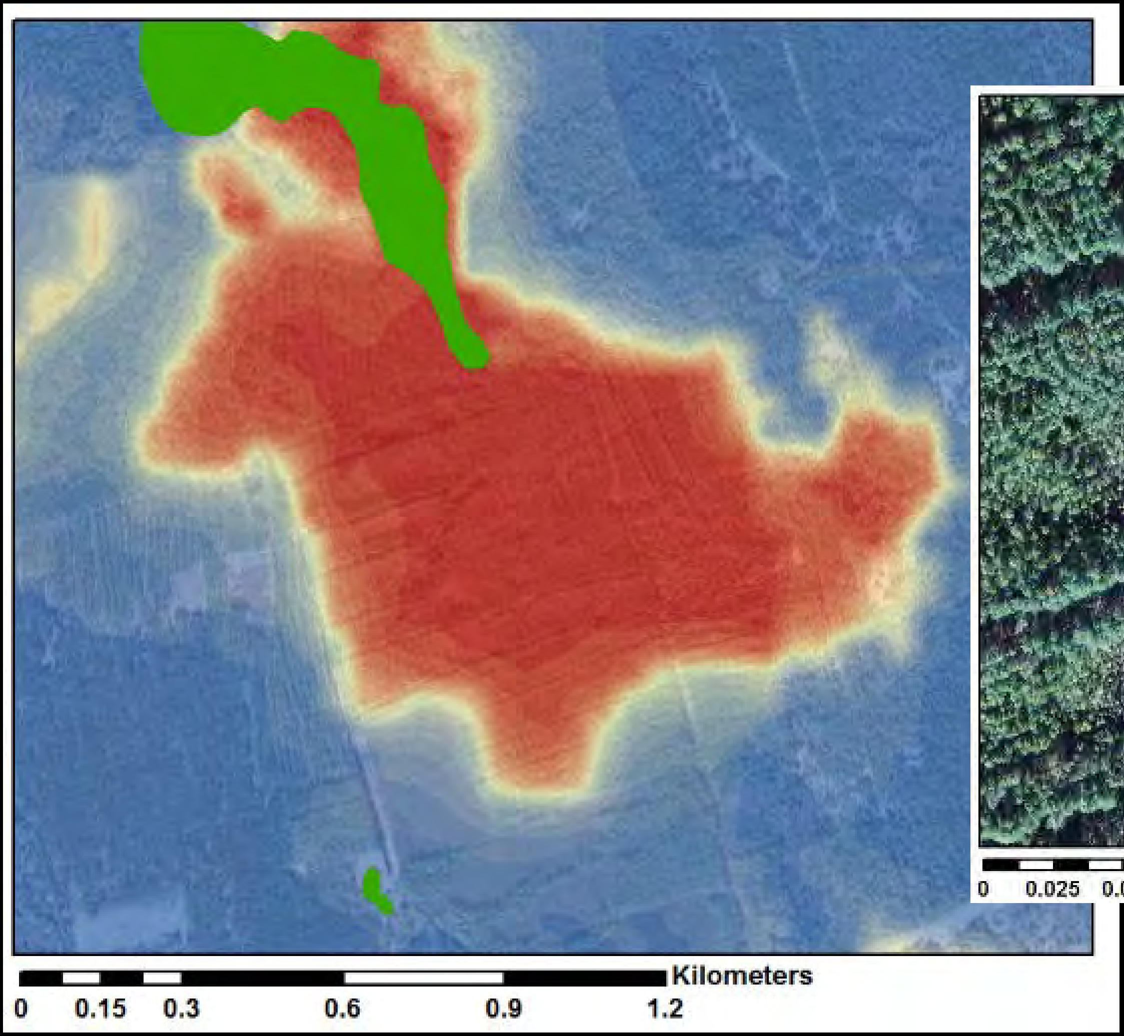


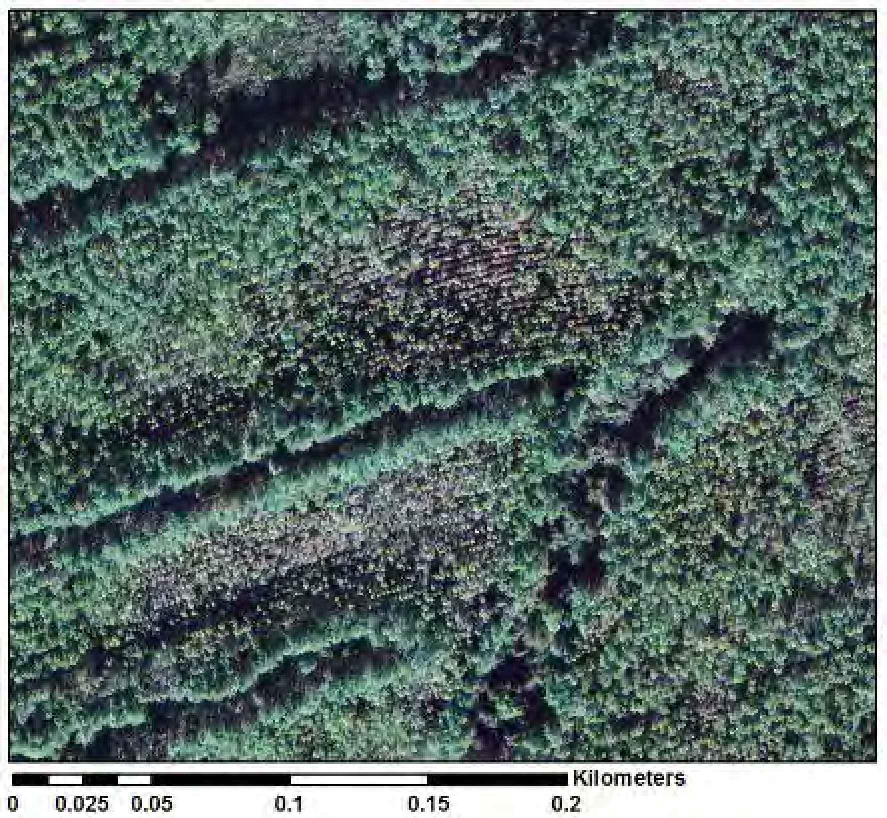
Probability of WOSW
0 - 0.01568257
0.01568257 - 0.058809635
0.058809635 - 0.109777985
0.109777985 - 0.164666977
0.164666977 - 0.21955597
0.21955597 - 0.274444962
0.274444962 - 0.333254596
0.333254596 - 0.388143588
0.388143588 - 0.443032581
0.443032581 - 0.49400093
0.49400093 - 0.54496928
0.54496928 - 0.59593763
0.59593763 - 0.64690598
0.64690598 - 0.69787433
0.69787433 - 0.74884268
0.74884268 - 0.79981103
0.79981103 - 0.85077938
0.85077938 - 0.905668372
0.905668372 - 0.952716079
0.952716079 - 0.999763787
SSURGO Soils
NWI
Kilometers
0
0.125
0.25
0.5
0.75
1

F



F





Summary and Closing

Funded by EPA Cost-Share Grant
Wetland Program Development Grant 00D14313

Peer Advisory Group Members

Tim Rach, FDEP
Andy May, FDEP
Dr. Daniel Irick, FDEP
Dr. Tingting Zhao, FSU
Dr. James Elsner, FSU
Dr. Daniel McLaughlin, UF
Dr. Thomas Hoctor, UF
Jonathan Oetting, FNAI
Amy Knight, FNAI
Kathleen Okeife, FWC
Mark Barrett, FWC

Contact Information

Questions?

John Humphreys

Phone: (850) 245-8487

Email: John.humphreys@dep.state.fl.us

AmirSasan Mahjoor

Phone: (850) 245-8817

Email: AmirSassan.Mahjoor@dep.state.fl.us