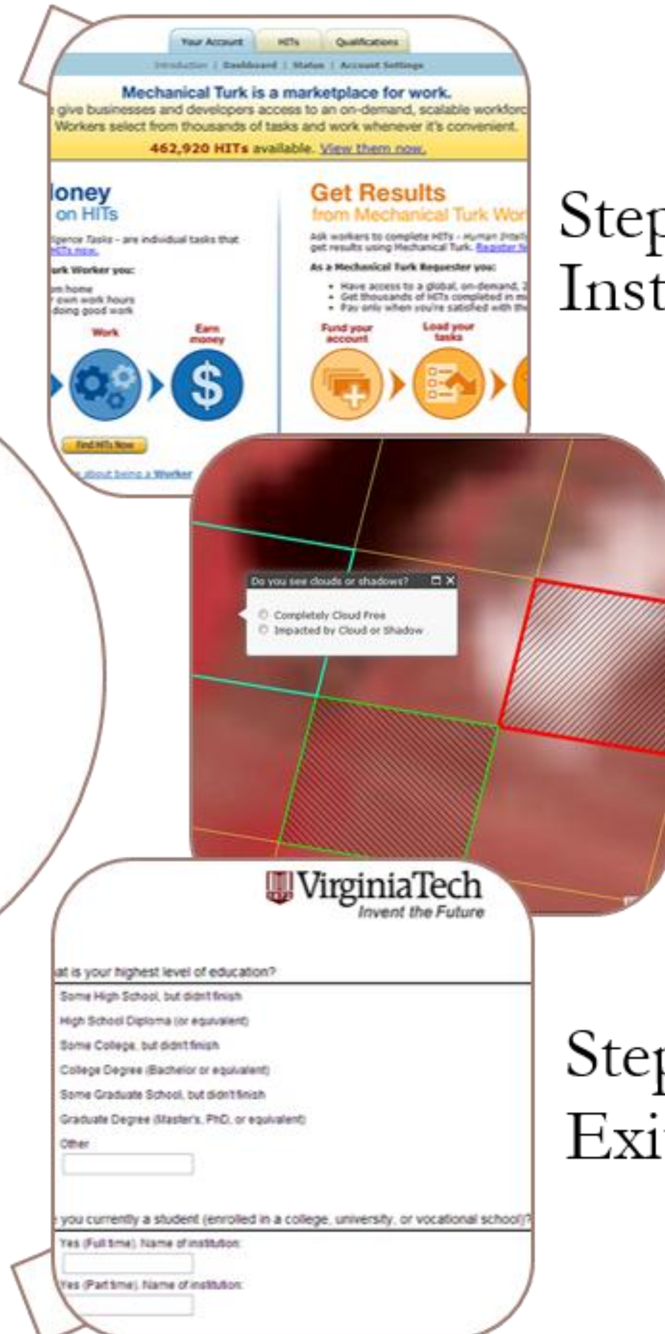
# Crowds, Clouds, and High Performance Computing of Subtle Forest Change Using Landsat

Randolph H. Wynne and Valerie A. Thomas

C.E. Blinn, E.B. Brooks, J.O. Coulston,
T.R. Fox, S. Ghannam, M.N. House,
K. Moeltner, A.J. Oliphant, S.S. Peery,
R. Saxena, L.T. Watson, L. Yu

NOISE TO SIGNAL
Rob Cottingham

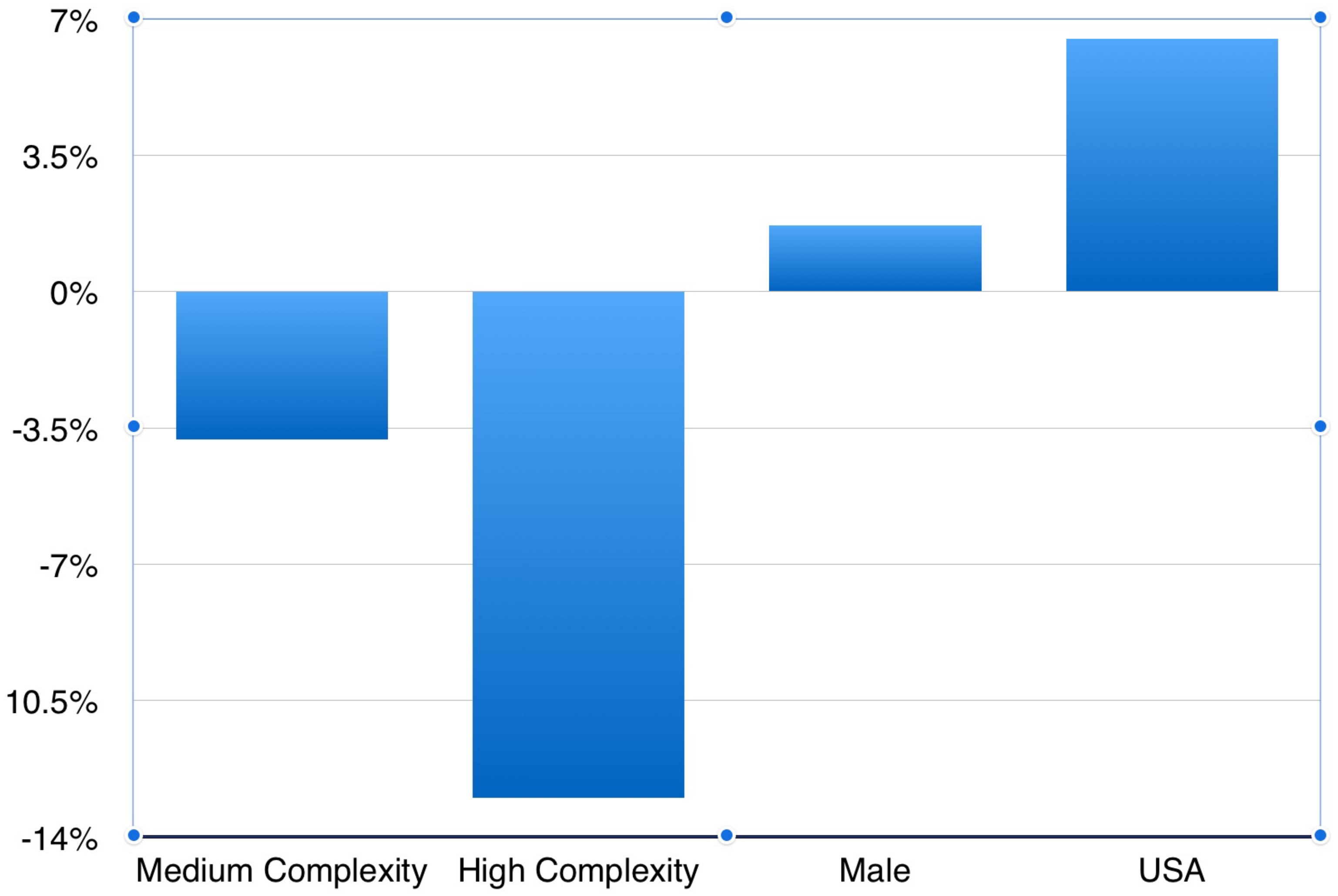
HIT



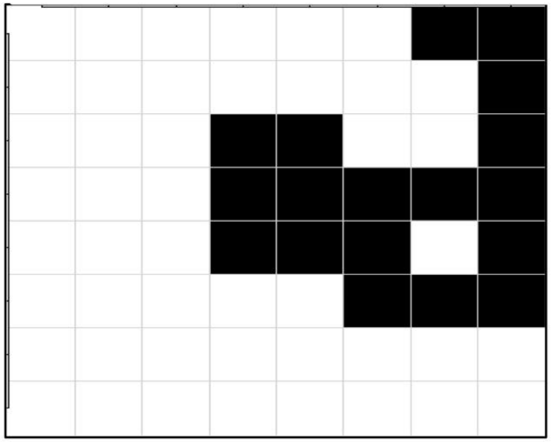

Step 1:
Instruction

Step 2:
Cloud interpretation
on 3 satellite images

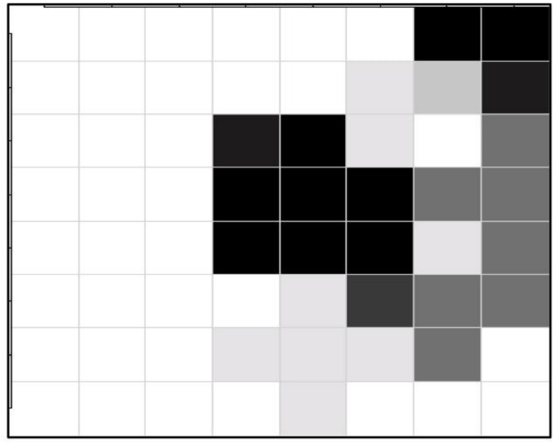
Step 3:
Exit Survey


7%
3.5%
0%
-3.5%
-7%
10.5%
-14%
Medium Complexity
High Complexity
Male
USA

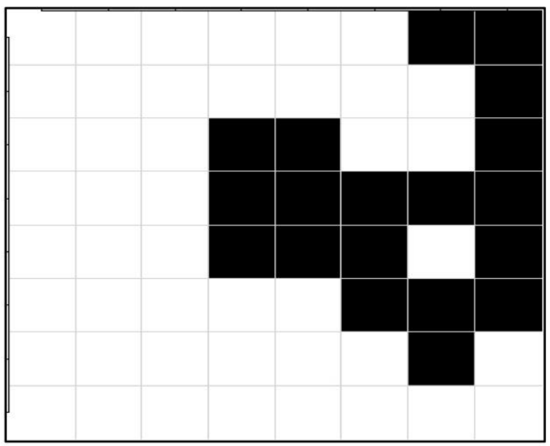
# Easy



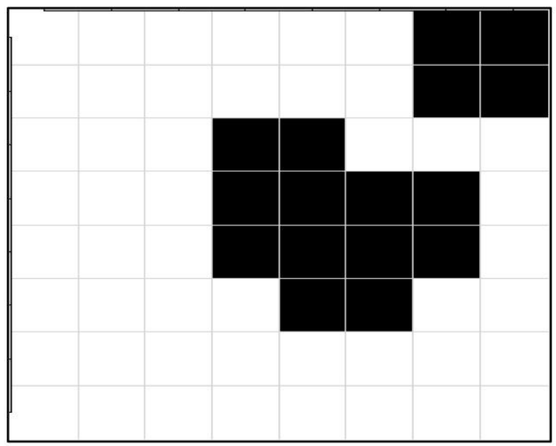
Expert benchmark
(black = CCS impacted)



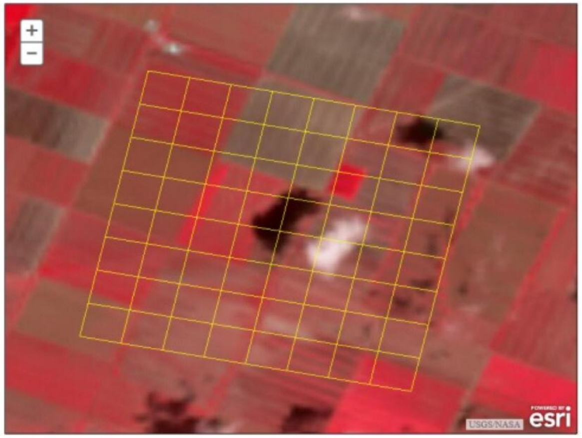
MT results
(darker = higher % of CCS impact choices)

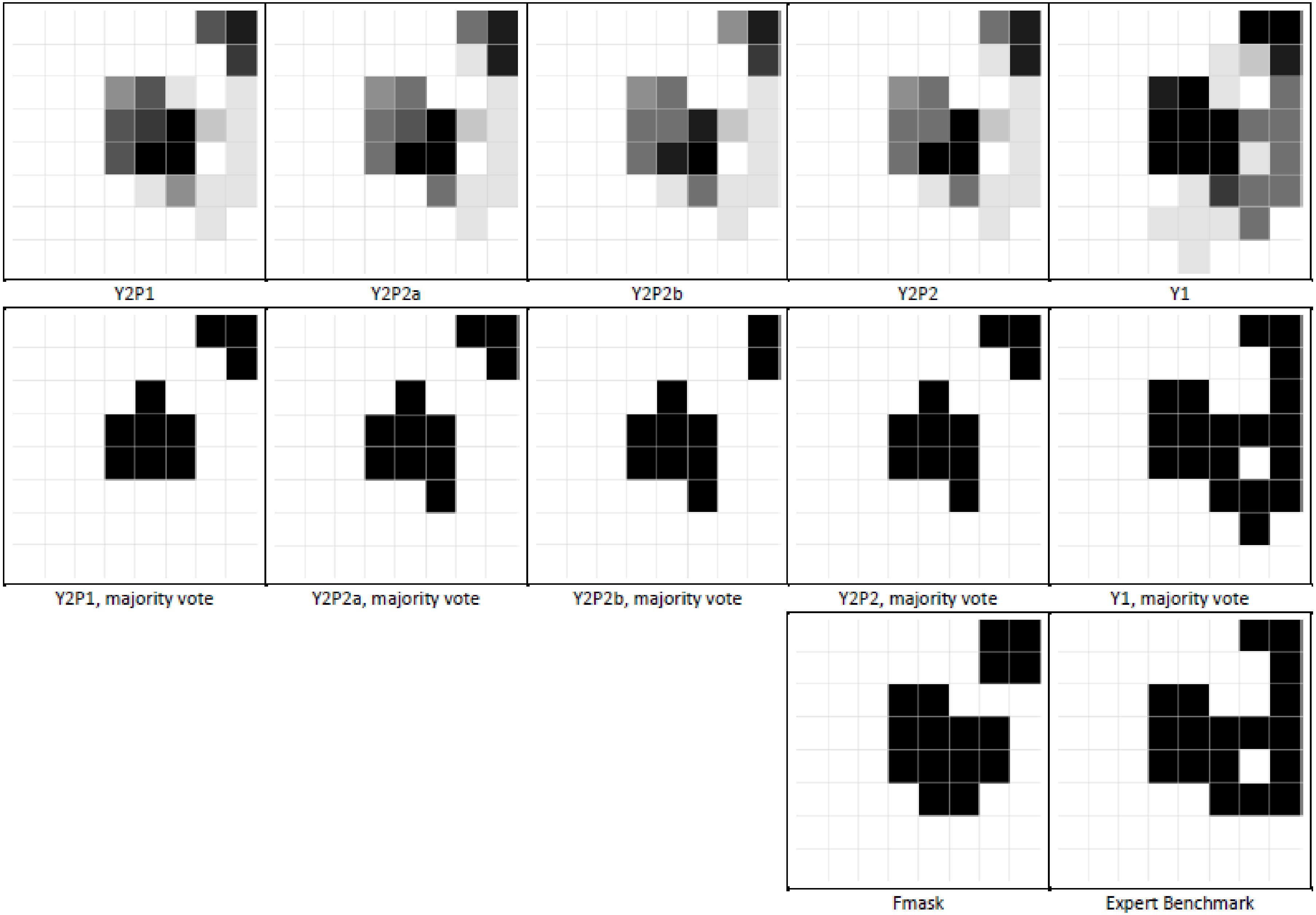


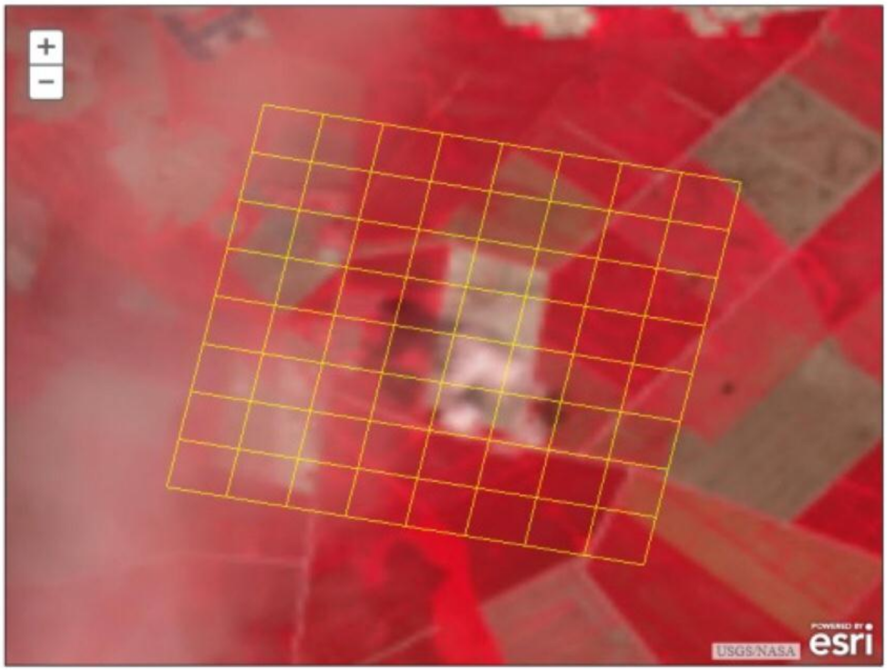
MT results – majority vote
(black = >50 % of CCS impact choices)

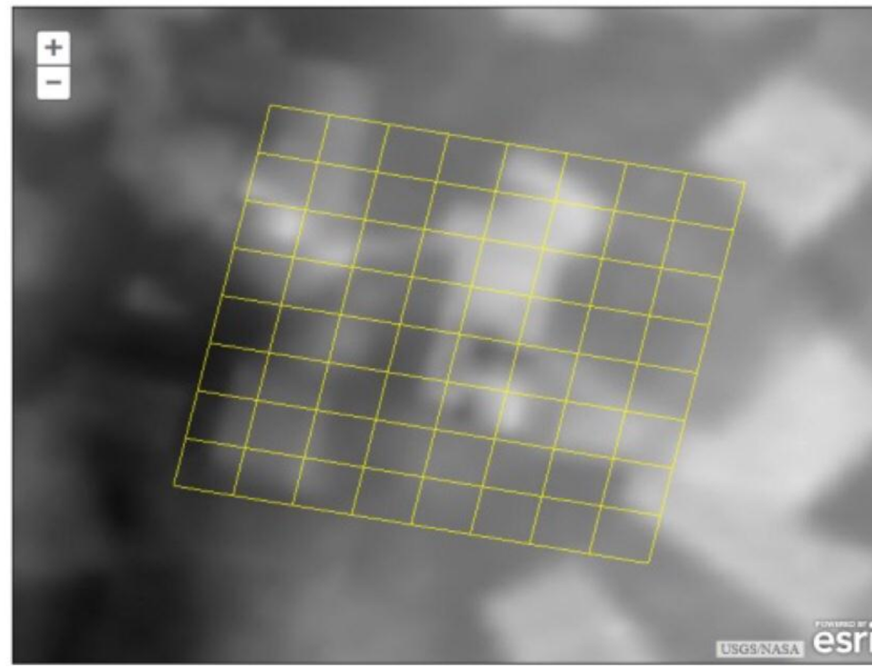


Fmask results
(black = identified as CCS impacted)



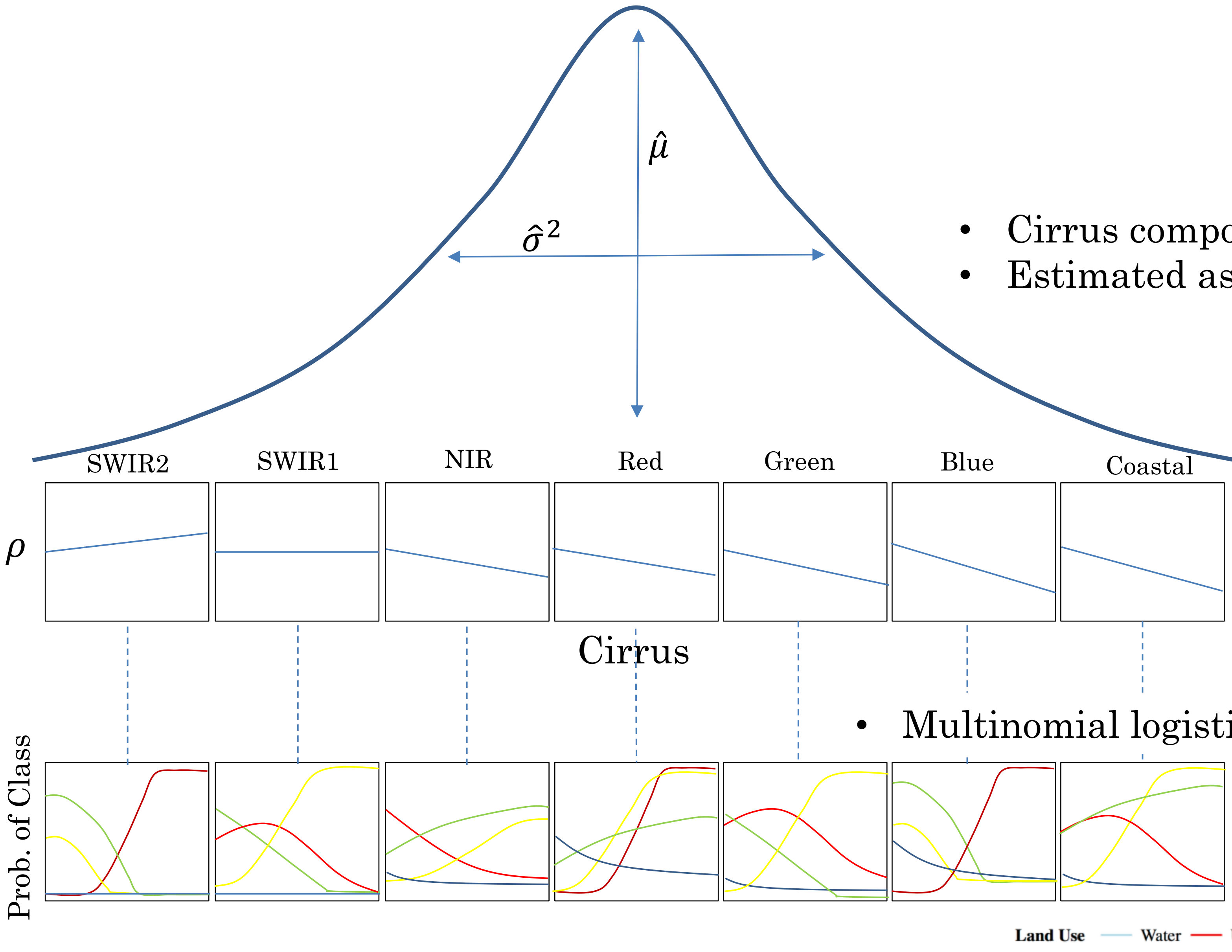

Y2P1
Y2P2a
Y2P2b
Y2P2
Y1
Y2P1, majority vote
Y2P2a, majority vote
Y2P2b, majority vote
Y2P2, majority vote
Y1, majority vote
Fmask
Expert Benchmark


USGS/NASA
POWERED BY
esri


USGS/NASA
POWERED BY
esri

# An Hierarchical Approach to Dealing with Cirrus Clouds

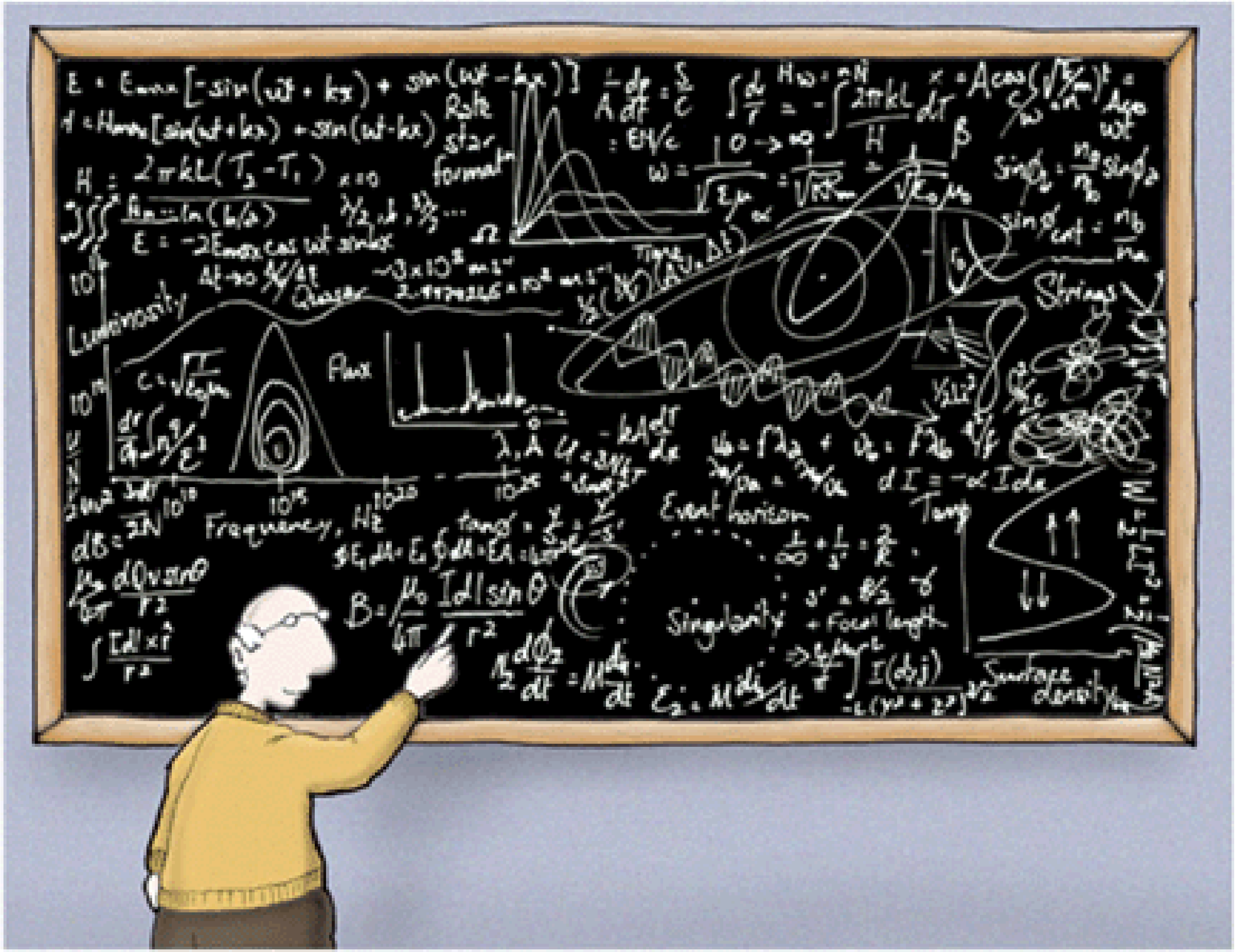
G.B. Anderson



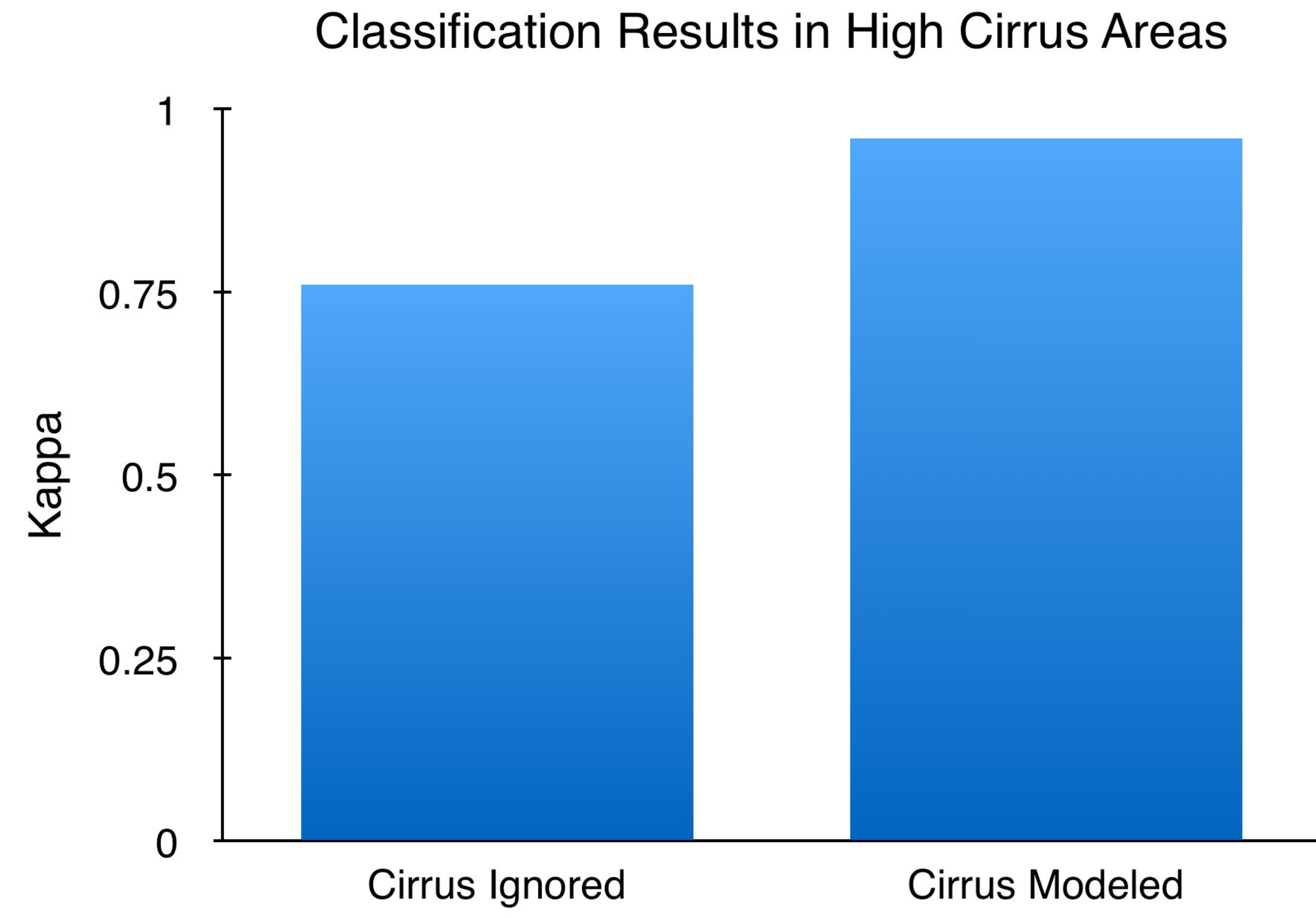

- Cirrus component of model
- Estimated as random effects

- Multinomial logistic regression

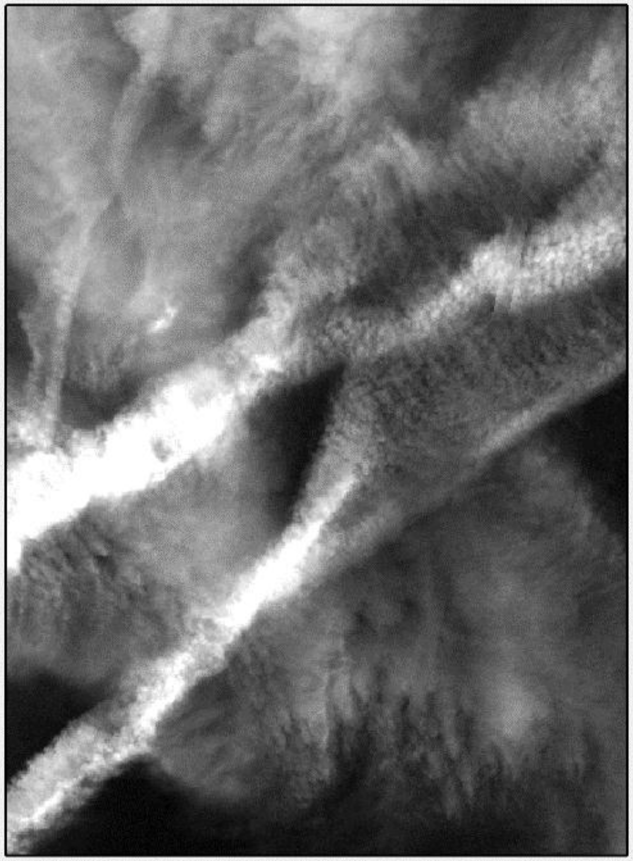
$$\left(\frac{\rho - \bar{\rho}}{\sigma_\rho}\right) - \left(\gamma \times \frac{\rho_{Ci} - \overline{\rho_{Ci}}}{\sigma_{\rho_{Ci}}}\right)$$




STATISTICS MADE EASY

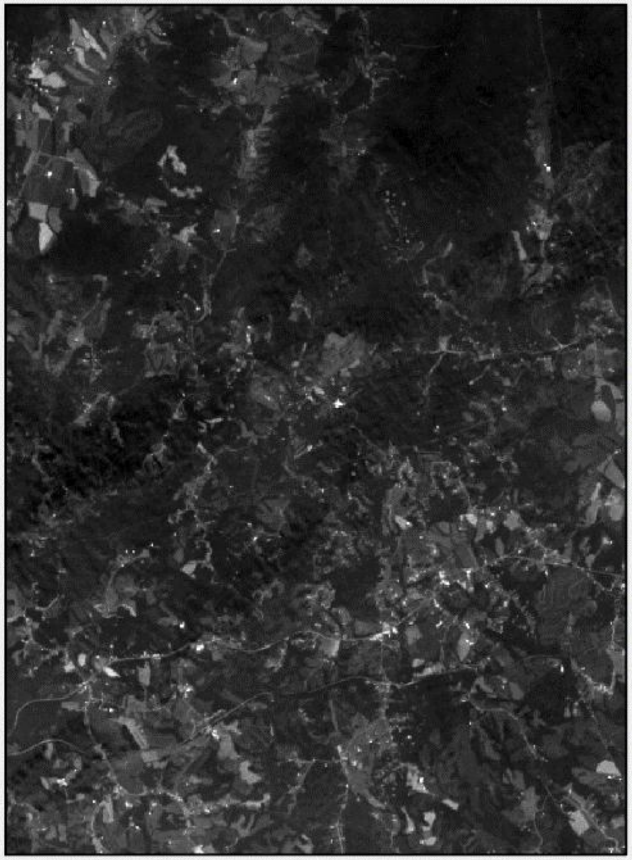

Classification Results in High Cirrus Areas
1
0.75
0.5
0.25
0
Kappa
Cirrus Ignored
Cirrus Modeled

# An Hierarchical Approach to Dealing with Cirrus Clouds



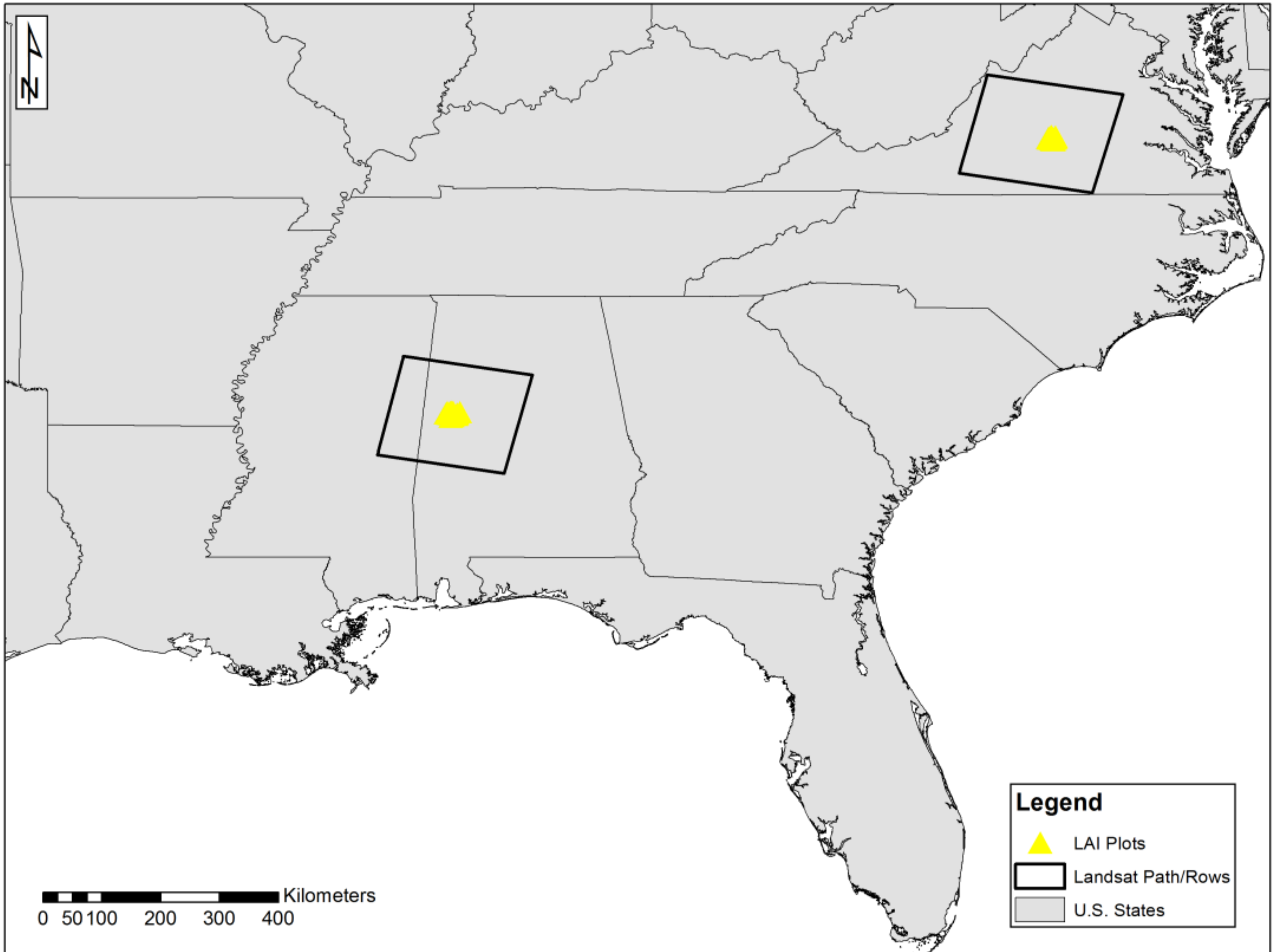
Cirrus
in May

Costal Aerosol
in May

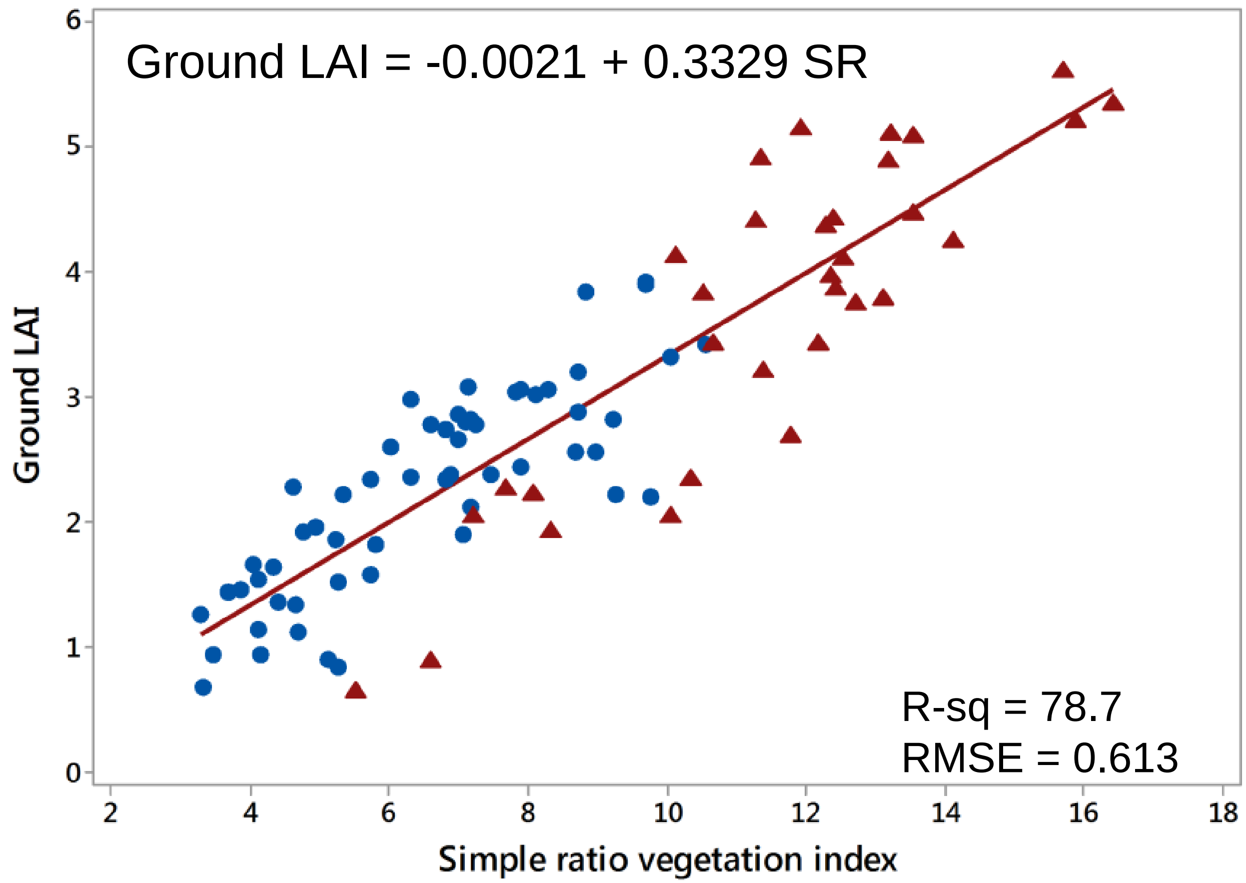


Predicted Costal Aerosol
in May

# LAI Study Site Locations



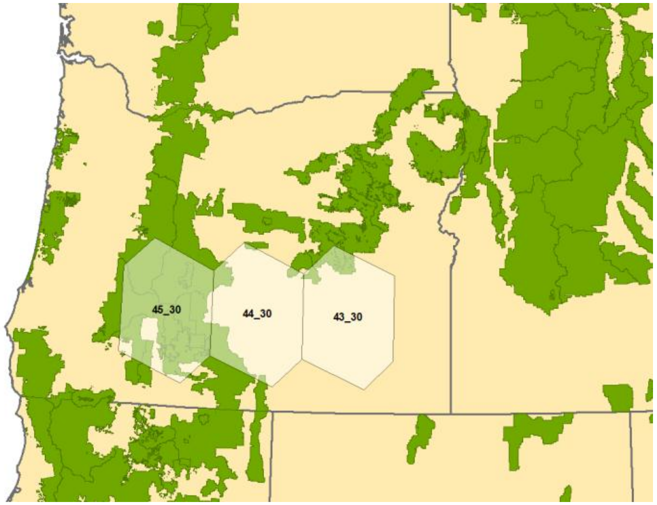
# Best Combined (L7-L8) LAI Model



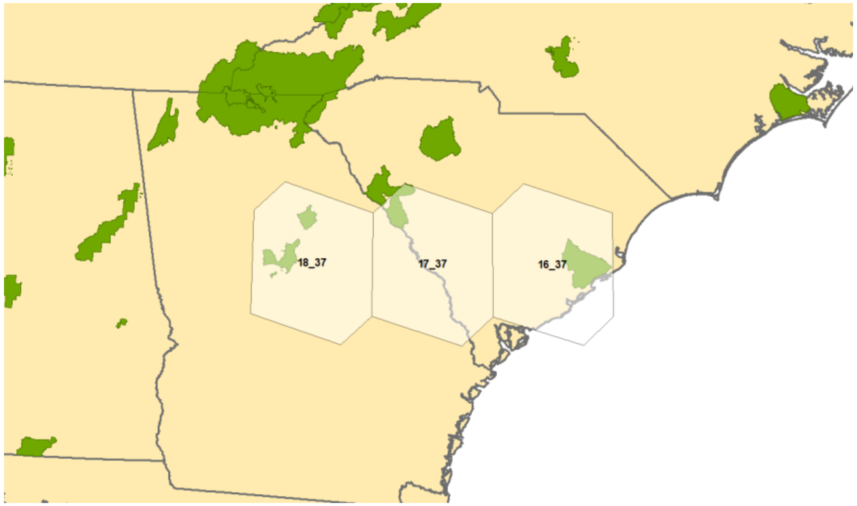

Blue circles are the minimum LAI and the red triangles are the peak LAI data. Simple ratio based on surface reflectance.

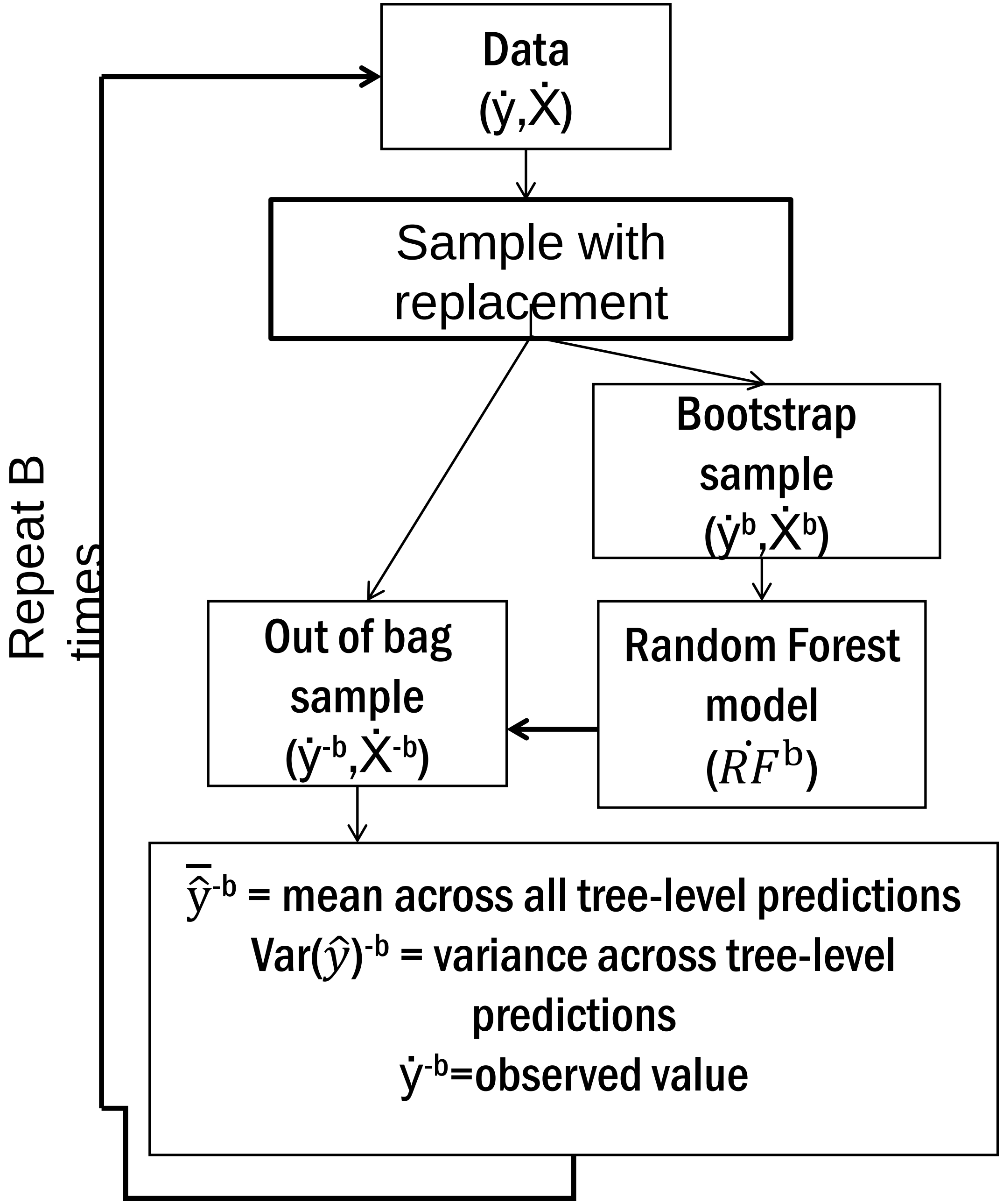
# Pilot Areas NLCD Tree Canopy Cover

**Western Study Area**



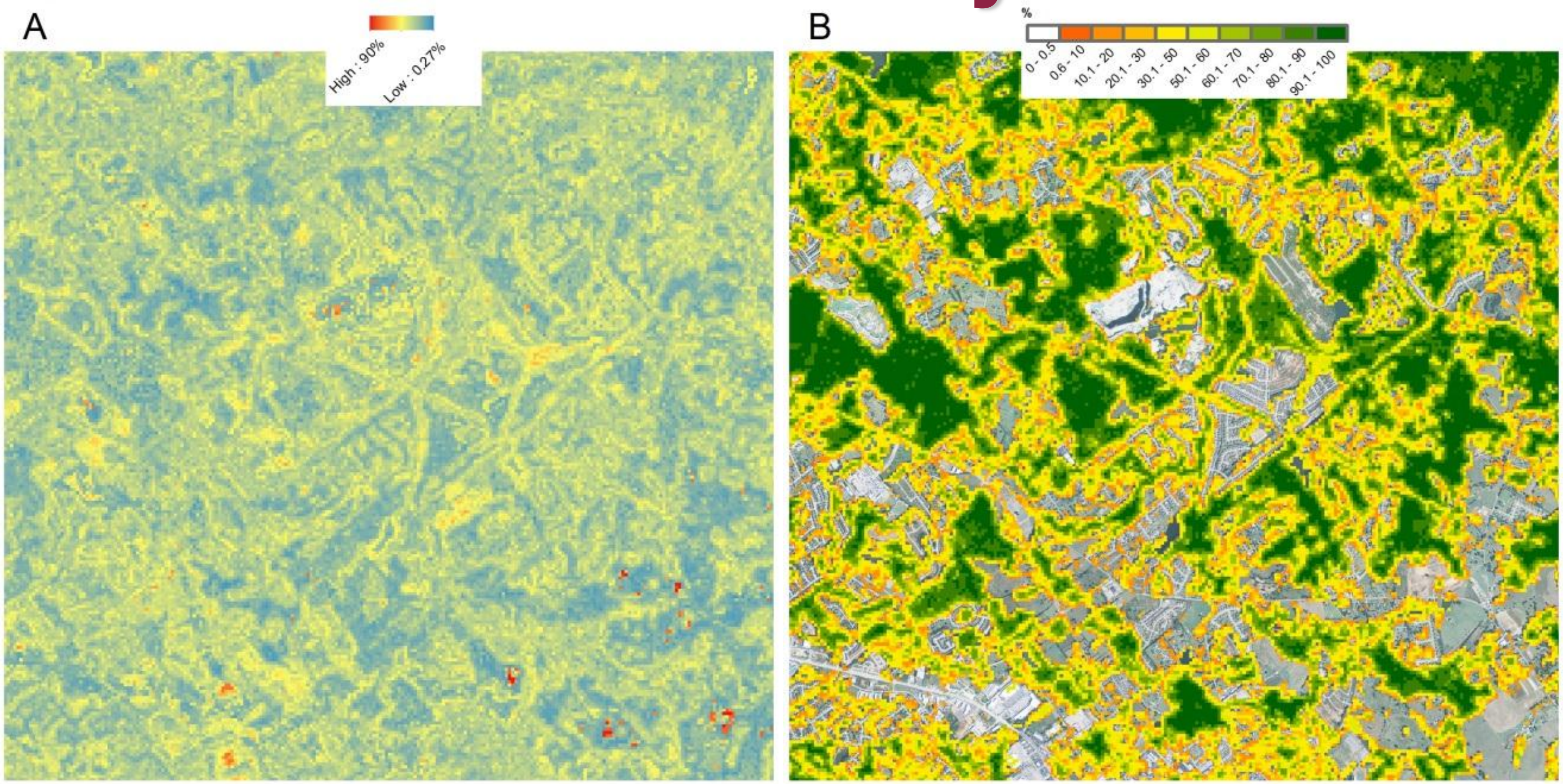

**Southern Study Area**

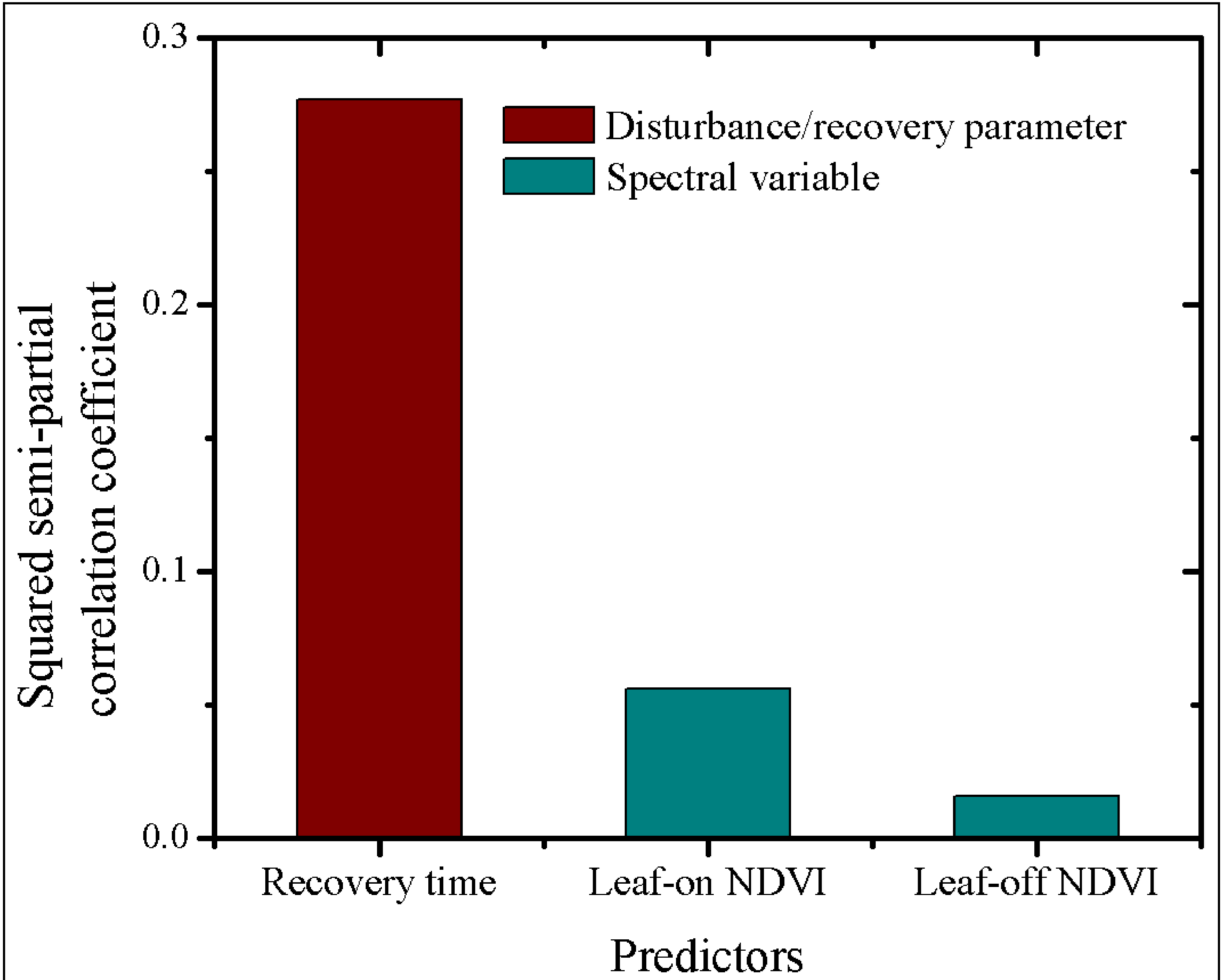




Data
(ẏ,Ẋ)
Sample with replacement
Bootstrap sample
(ẏ^b,Ẋ^b)
Repeat B times
Out of bag sample
(ẏ^-b,Ẋ^-b)
Random Forest model
(ṘF^b)
ŷ̄^-b = mean across all tree-level predictions
Var(ŷ)^-b = variance across tree-level predictions
ẏ^-b=observed value

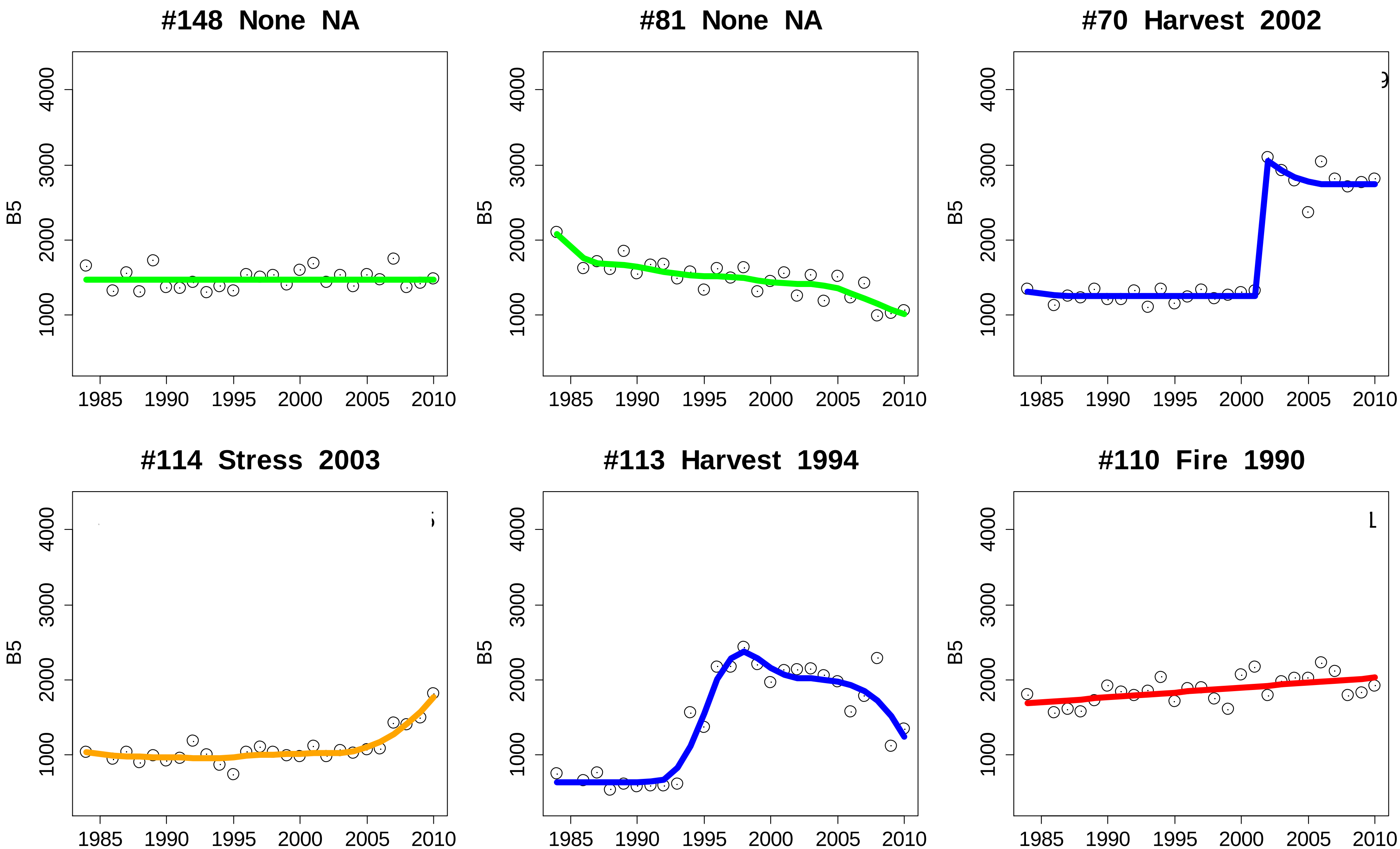
# Percent TCC Uncertainty



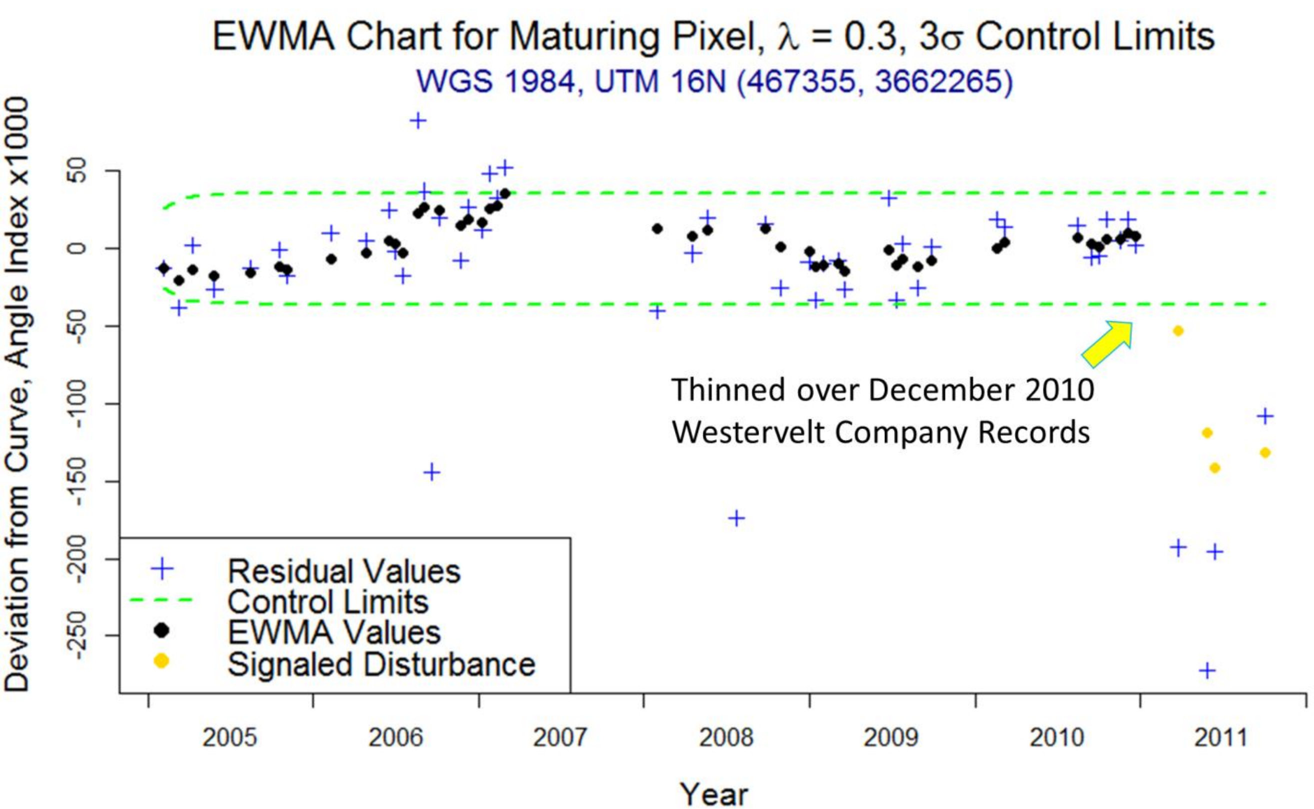


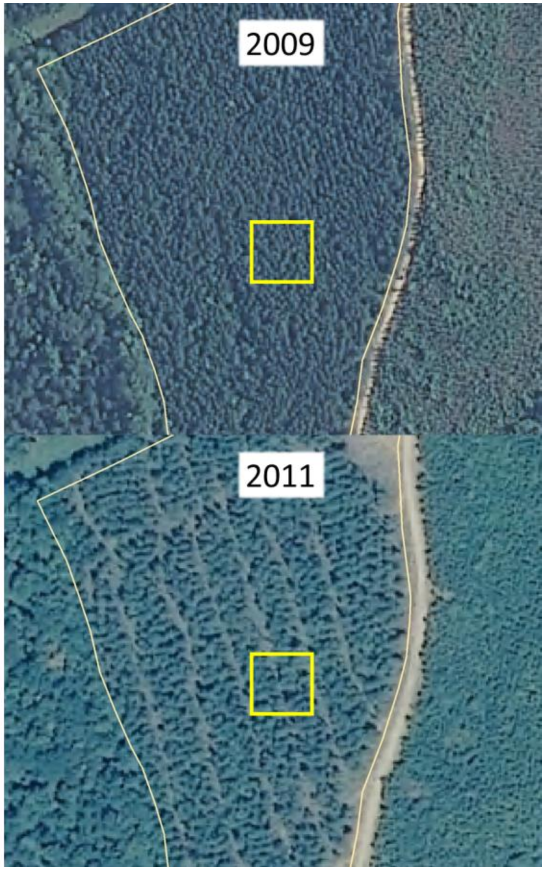
0.3
0.2
0.1
0.0
Squared semi-partial correlation coefficient
Disturbance/recovery parameter
Spectral variable
Recovery time
Leaf-on NDVI
Leaf-off NDVI
Predictors

# Time Series Disturbance Metrics: Shapes



# Example EWMA Flagged Pixels



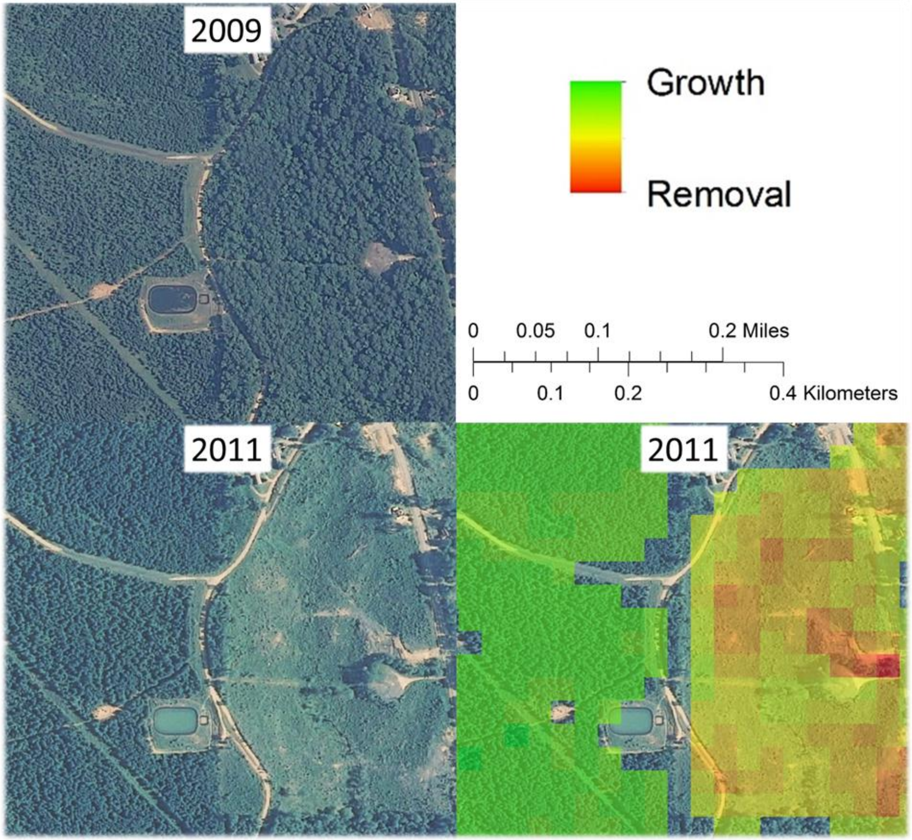

EWMA value, $z_n = (1 - \lambda)\, z_{n-1} + \lambda x_n$

where $\lambda$ = tuning parameter between 0 and 1

$x_n$ = pixel value at time n and $z_{n-1}$ = EWMA value at time n-1

# EWMACD/LCMS

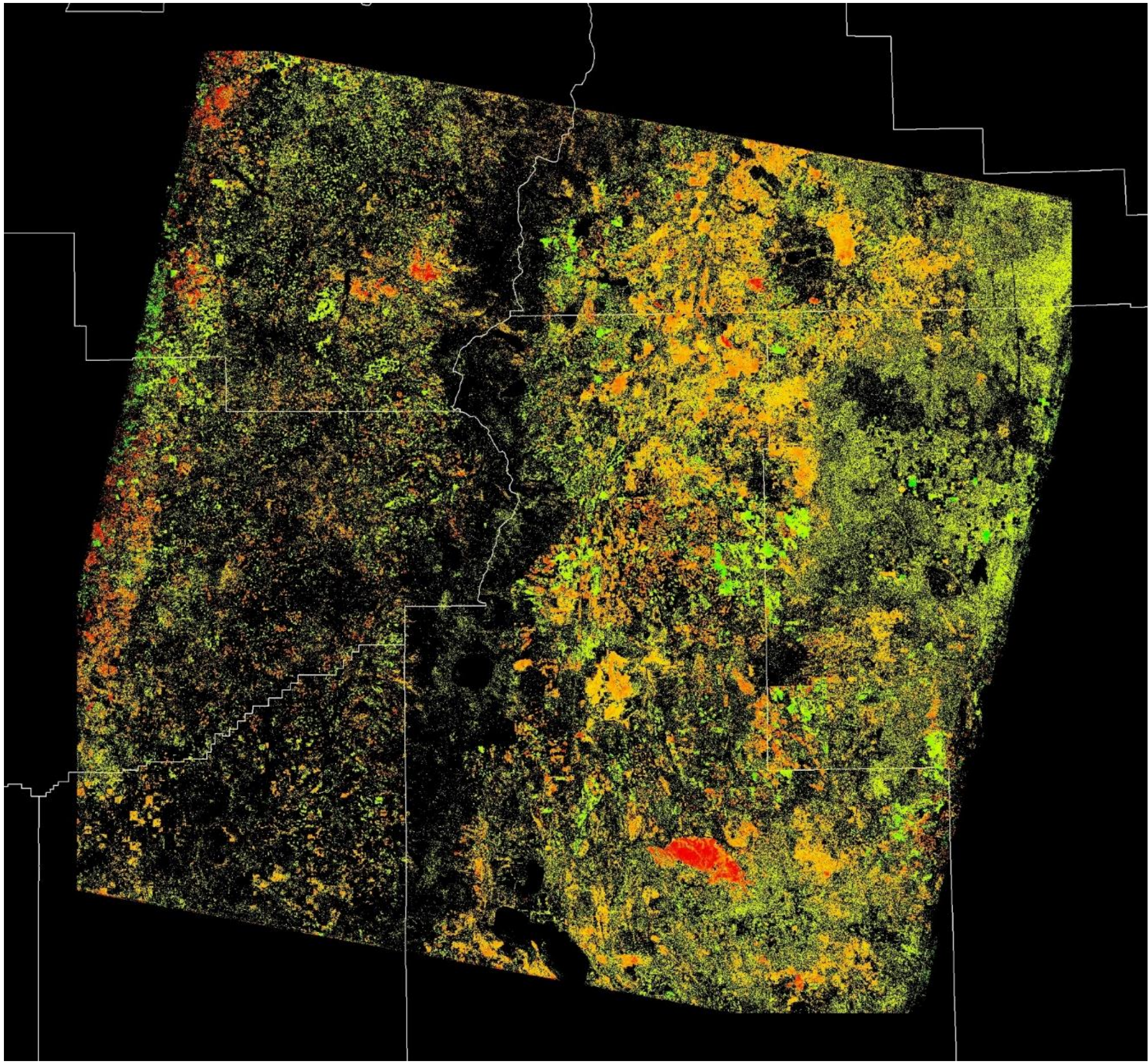
- Dynamic retraining functionality added
- Code actively being developed and distributed (Creative Commons)
  - doi: 10.7294/W4WD3XHK




Source: **Brooks, E. B.**, Thomas, V. A., Wynne, R. H., Blinn, C. E., and Coulston, J. W. (2014) "On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data." *IEEE Transactions on Geosciences and Remote Sensing*, 52(6), 3316-3332.

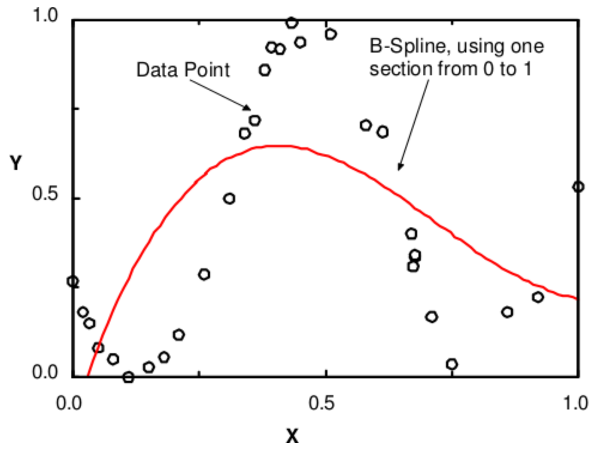
# EWMACD/LCMS

- Completed processing for pilot study (six scenes, full TM history)
  - 45/30 shown
- Accuracy comparable to other LCMS base learners



# Shapes

- B-splines are dependent on user-defined knots and degrees



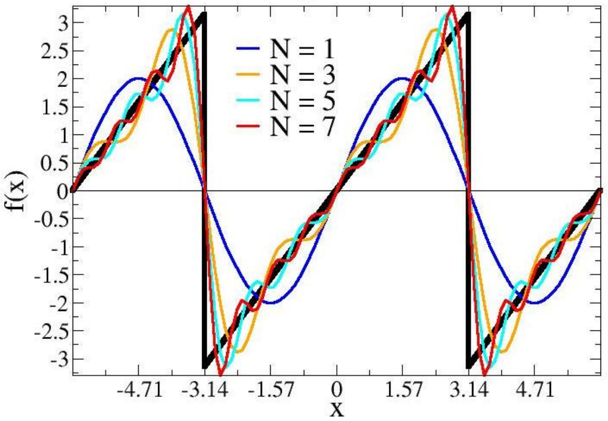

B-spline with not enough sections.

- Bayes Information Criterion has limitations
  - Sample size must be much larger than the number of parameters.

- <u>Check</u> validity of the chosen knots, degrees etc.

# Harmonic Regression

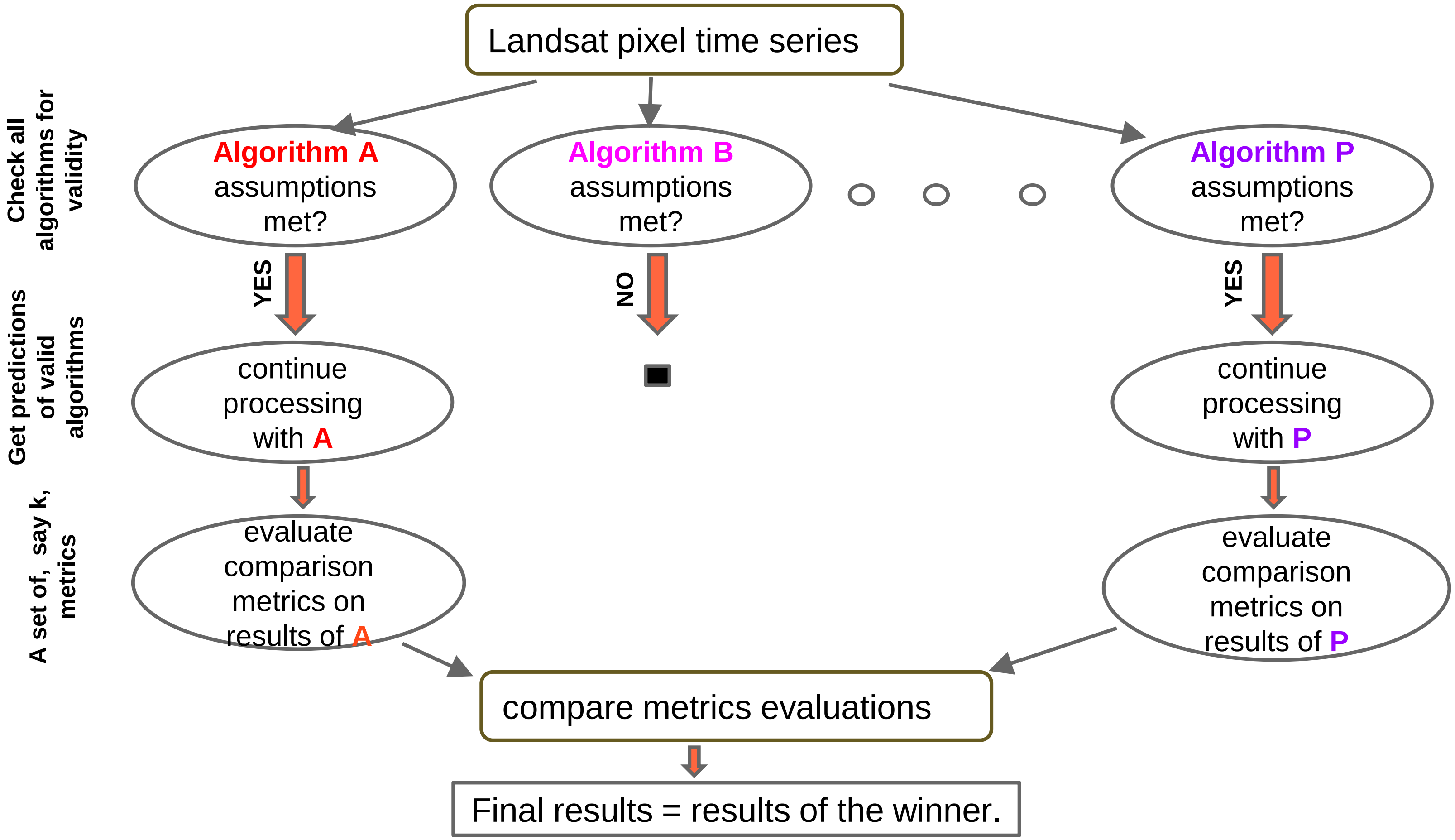
- Assumes smoothness in the initial (training) data



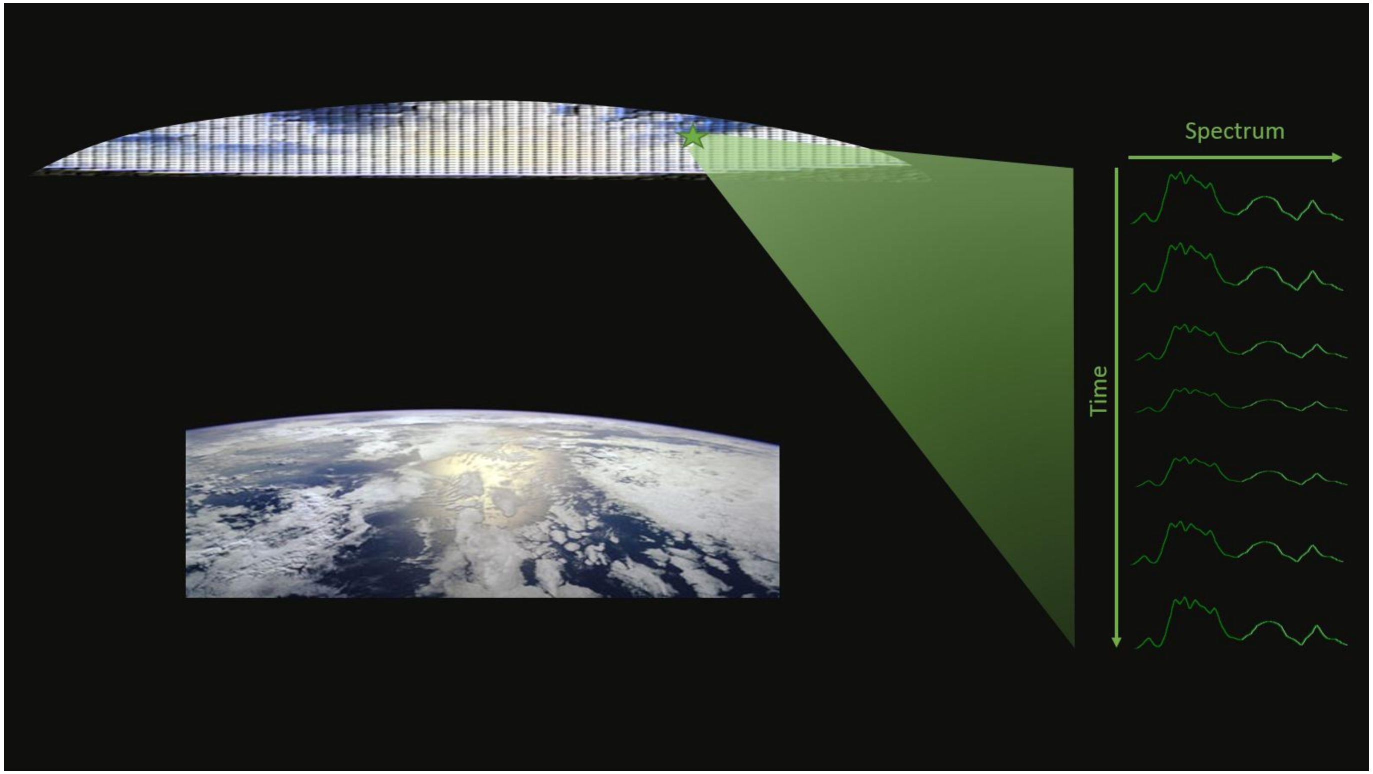

N=2 harmonics not enough to capture the dynamics of this sawtooth data.

- <u>Check</u> presence of discontinuities in the training data (and, of course, periodicity)

# Saxena Thesis



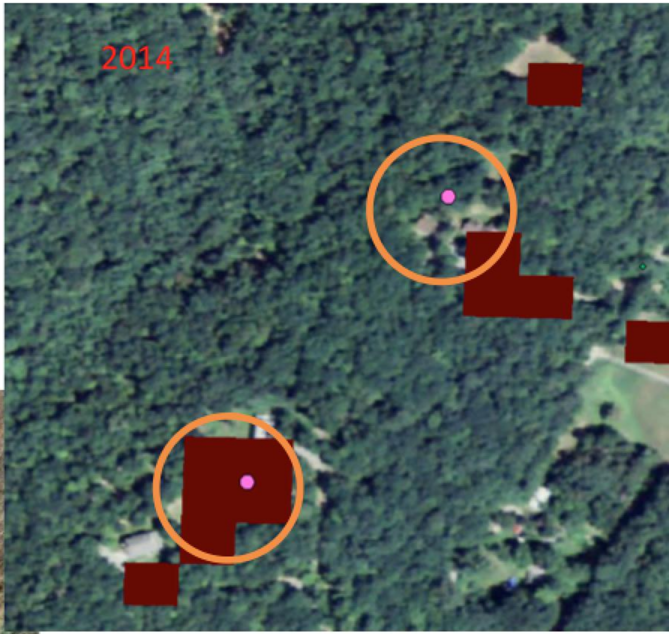
# Hadoop/Starchives

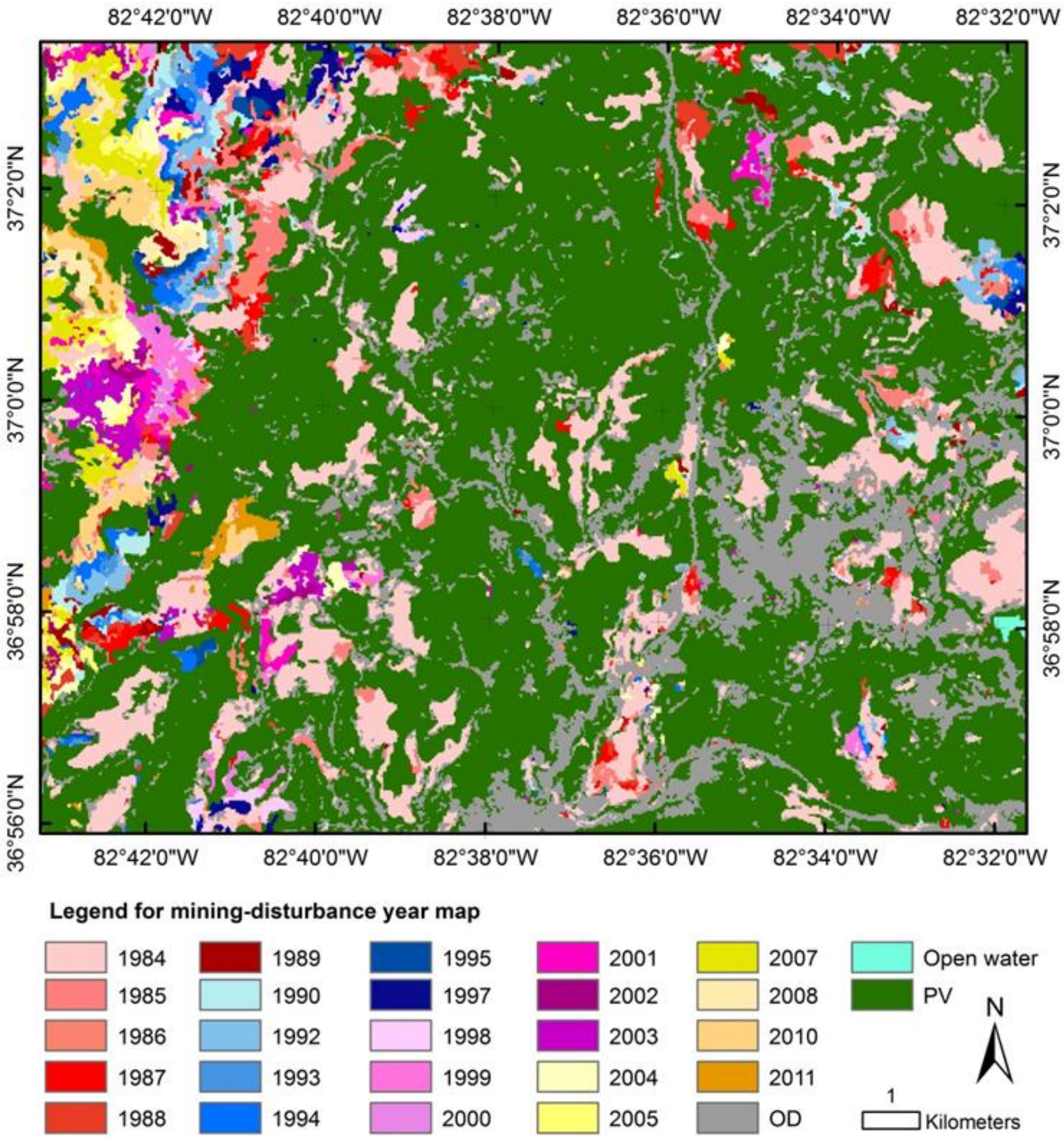


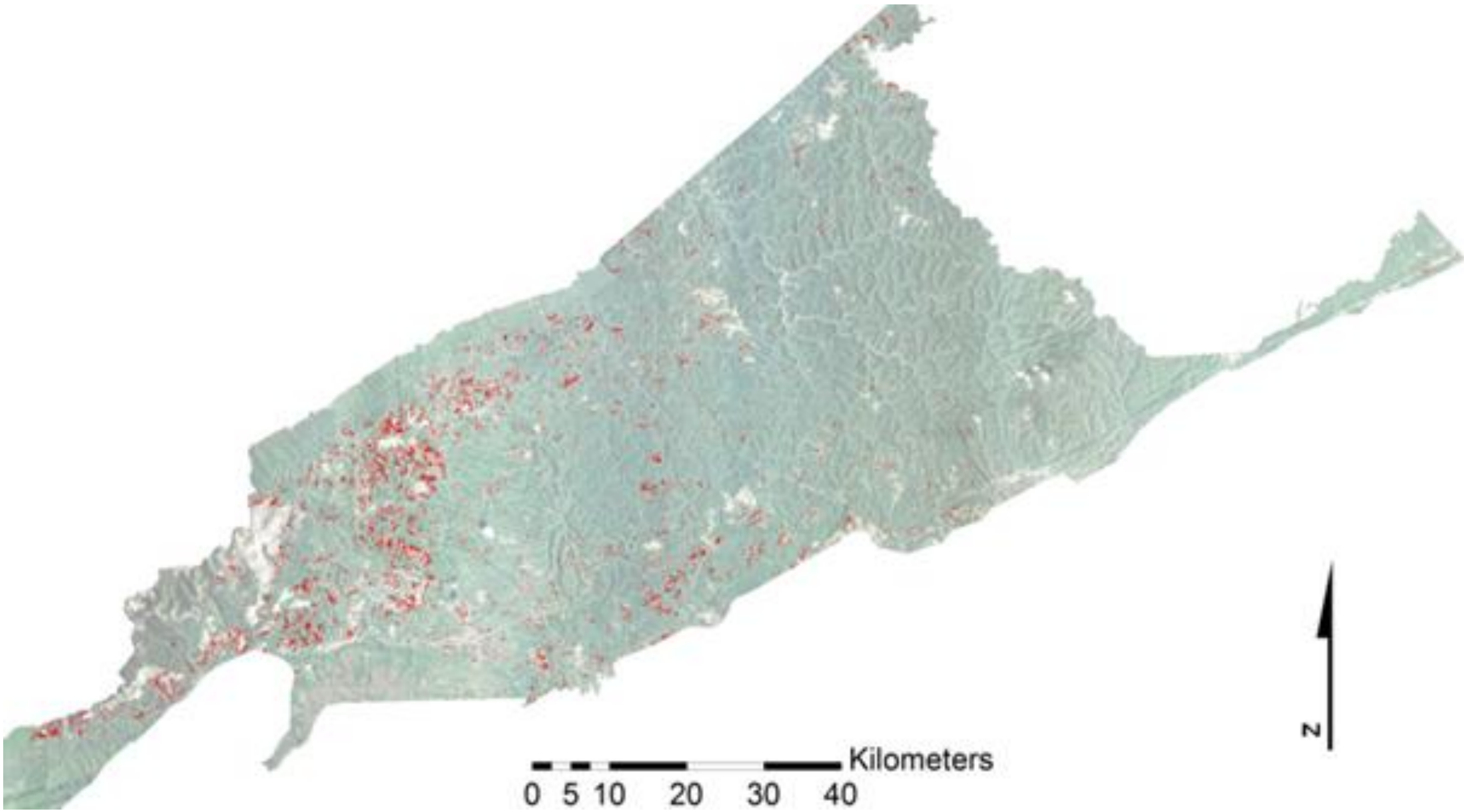

- EWMACD implemented successfully on Hadoop via starchive architecture

# Hotspot Conversion Tracking Through Landsat


Using housing permit data we are able to verify that the program can accurately detect subtle disturbances to the forest ecosystem.

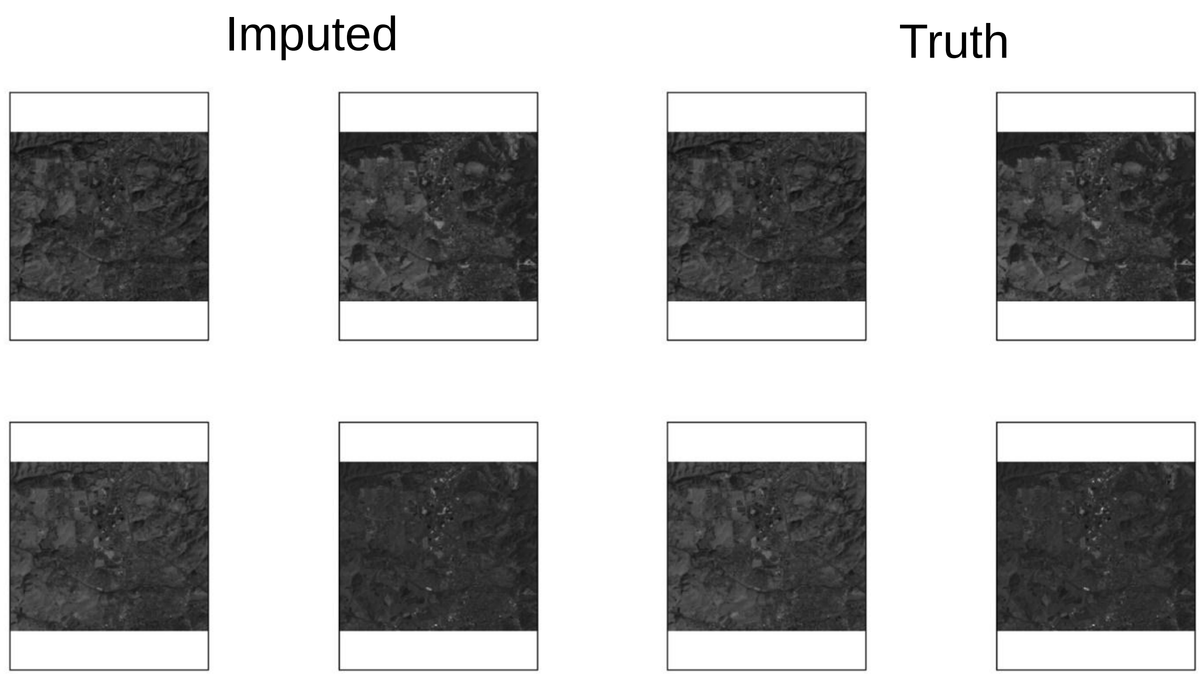

82°42'0"W
82°40'0"W
82°38'0"W
82°36'0"W
82°34'0"W
82°32'0"W
37°2'0"N
37°0'0"N
36°58'0"N
36°56'0"N
Legend for mining-disturbance year map
1984
1985
1986
1987
1988
1989
1990
1992
1993
1994
1995
1997
1998
1999
2000
2001
2002
2003
2004
2005
2007
2008
2010
2011
OD
Open water
PV
N
1
Kilometers



Kilometers
0
5
10
20
30
40
N

# Window Regression

- **Testing the applicability of WR in destriping ETM+ data and filling in cloud-masking gaps**

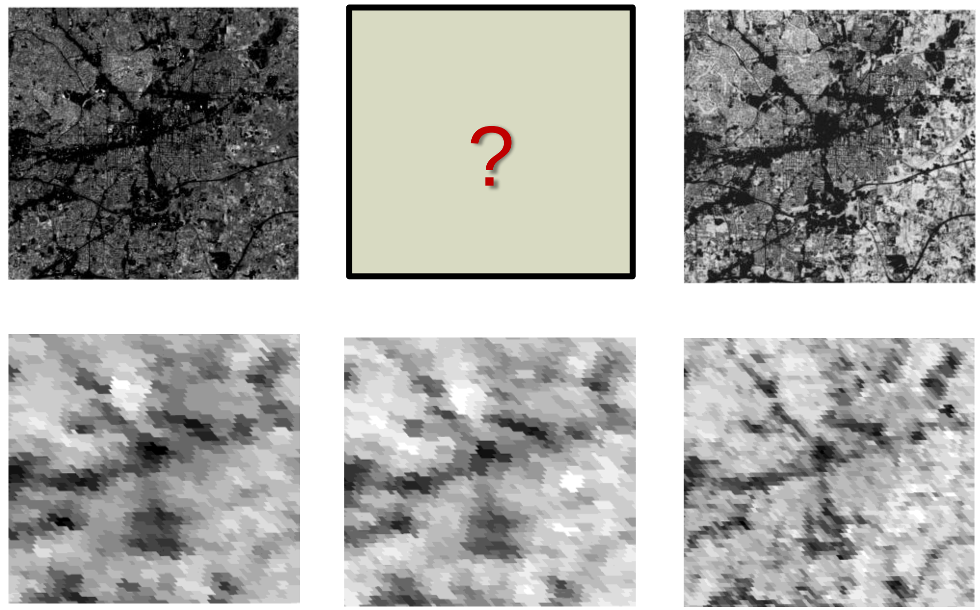



**Source:** Brooks, E. B., Wynne, R. H., and Thomas, V. A. "The applicability of window regression for repairing Landsat ETM+ data."

# Problem Statement: Predict L($t_p$) from available data

High Spatial Resolution Data . . .
e.g., Landsat

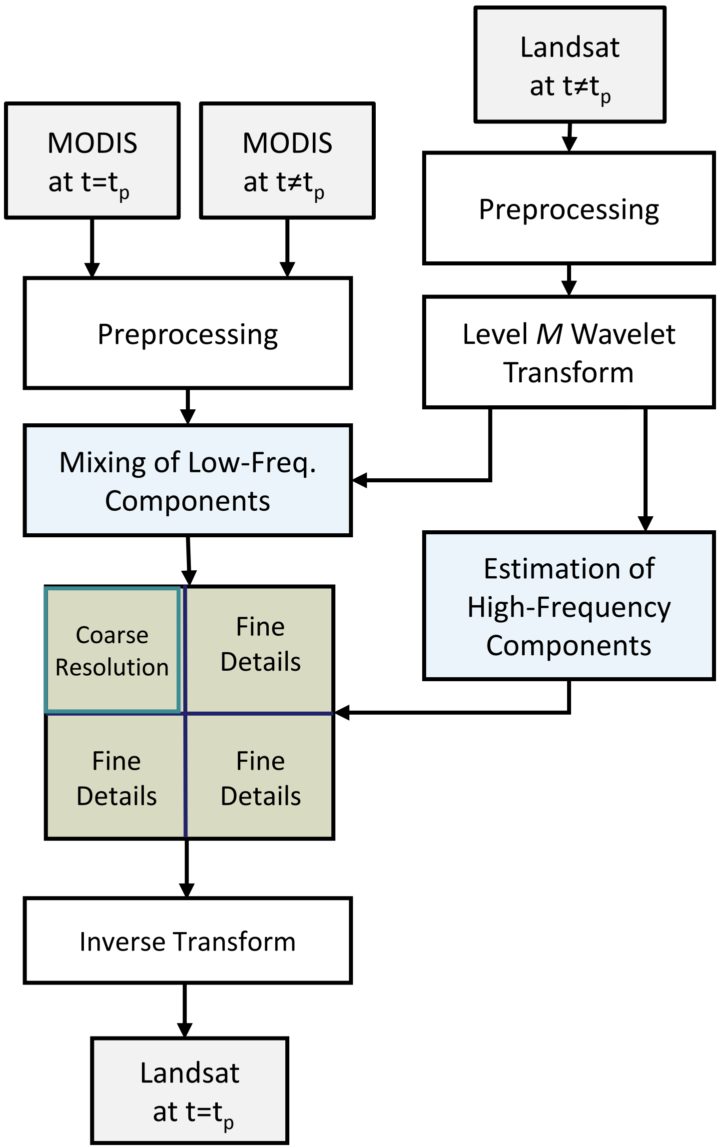
Coarse Spatial Resolution Data . . .
e.g., MODIS

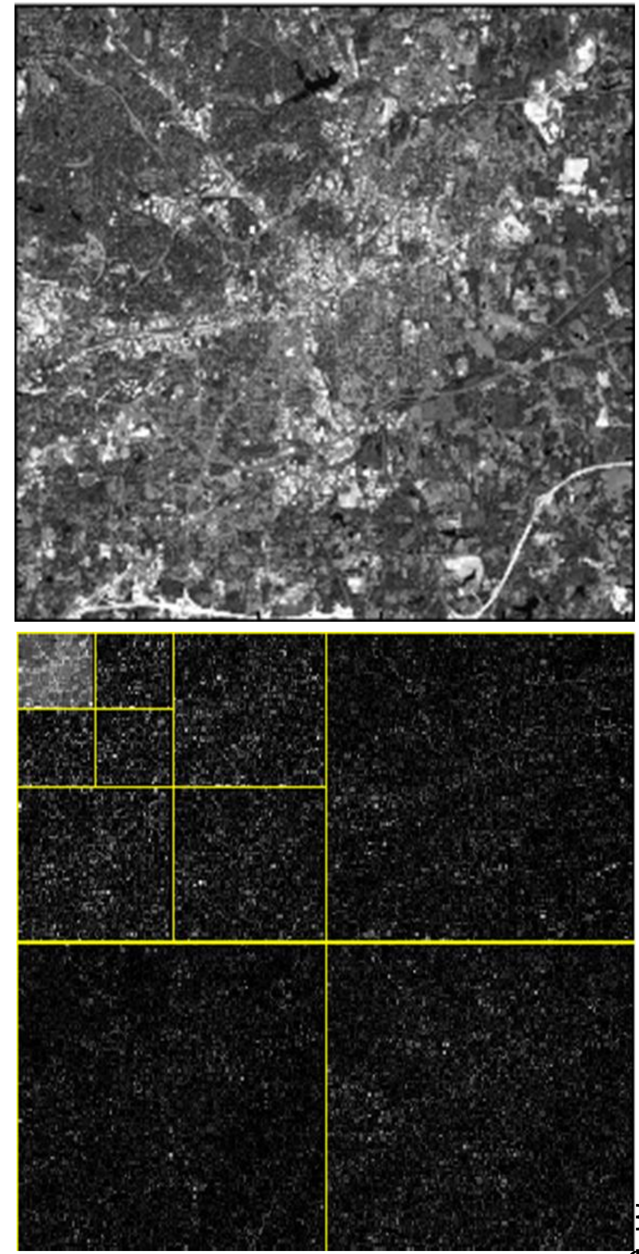



. . .

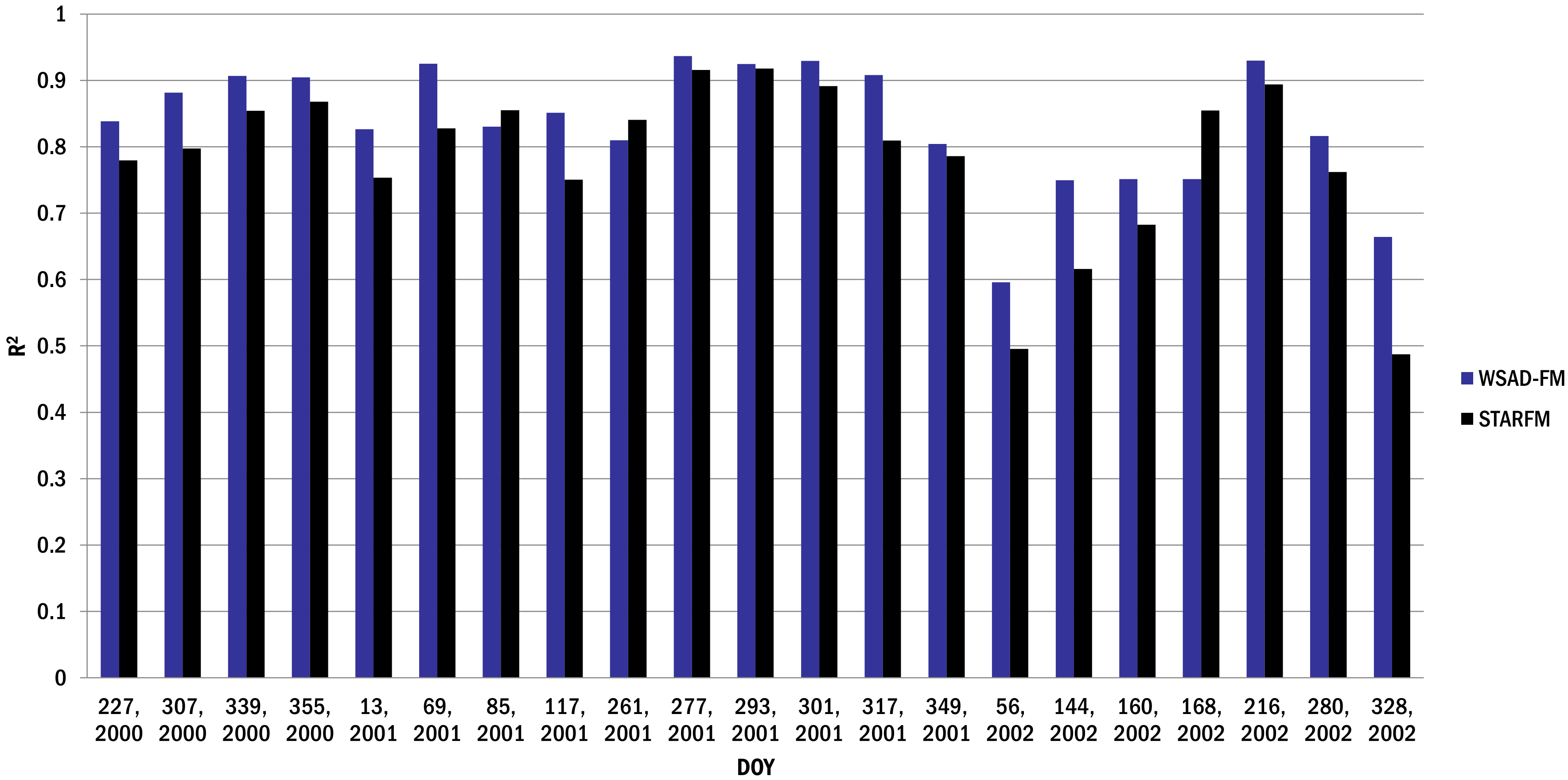
$$L(t_p) = f\left(M(t_p), L(t_k), M(t_k)\right)$$

# Wavelet-based Spatiotemporal Adaptive Data Fusion Model (WSAD-FM)



# Red Band (STAR-FM vs. WSAD-FM)



# WSAD-FM: Example

- At $t_p = 147$

Actual Landsat Red Band Image

Predicted Landsat Red Band Image

**$R^2$=0.8593**

# "Change" easy to find…meaningful change less so

# Are we…

- Crowd sourcing when it makes sense to do so?
- Maximizing the potential utility of the cirrus band?
- Really ready to bring algorithms to data? (algorithms, data & structures, architecture, …)
- Estimating spatially-specific uncertainty in our higher order products (static and change)?
- Maturing from our early “kids in the candy shop” multitemporal analyses?