NC STATE UNIVERSITY

A View From Space – Remote Sensing Applications in Water, Food, and Energy Security

Ranga Raju Vatsavai

Chancellors Faculty Excellence Associate Professor in Geospatial Analytics

Department of Computer Science, North Carolina State University (NCSU)

Associate Director, Center for Geospatial Analytics, NCSU

&

Joint Faculty, Oak Ridge National Laboratory (ORNL)

UMN-NSF Fifth Workshop on Understanding Climate Change from Data

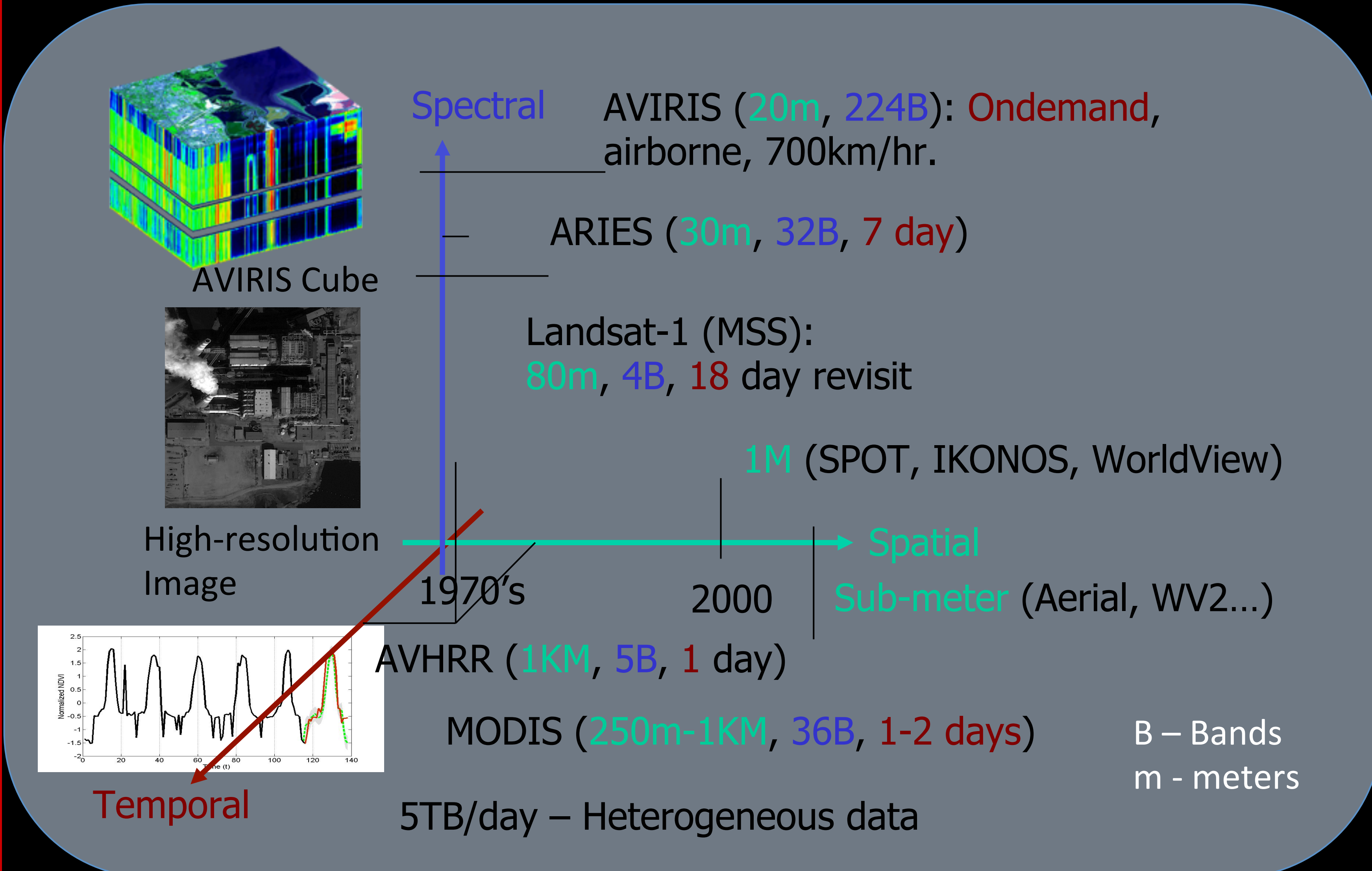
Minneapolis, MN. 8/5/2015

Outline

- Remote Sensing Overview
- Applications
  - Water
  - Food
  - Energy
- Algorithms
  - Gaussian Process (GP) Learning
  - Bi-temporal Hierarchical and Probabilistic
  - Multi-view, Semantic, and Multiple Instance Classification
- Outlook

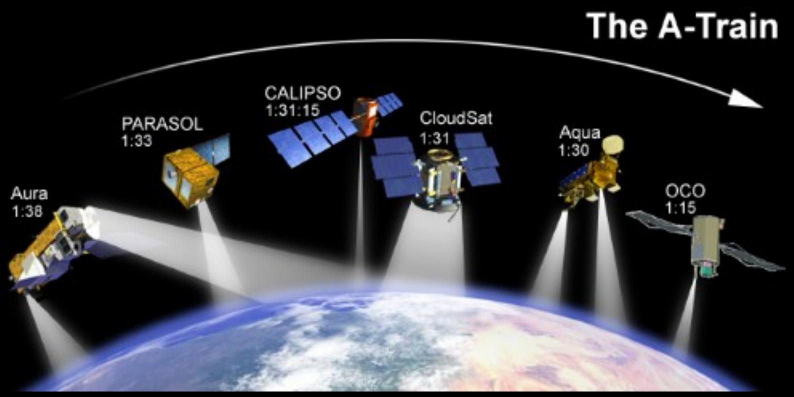
NC STATE UNIVERSITY

Big Spatiotemporal (Remote Sensing) Data



Remote Sensing

- ~130 Earth Imaging Satellites
- 33 Countries + EU



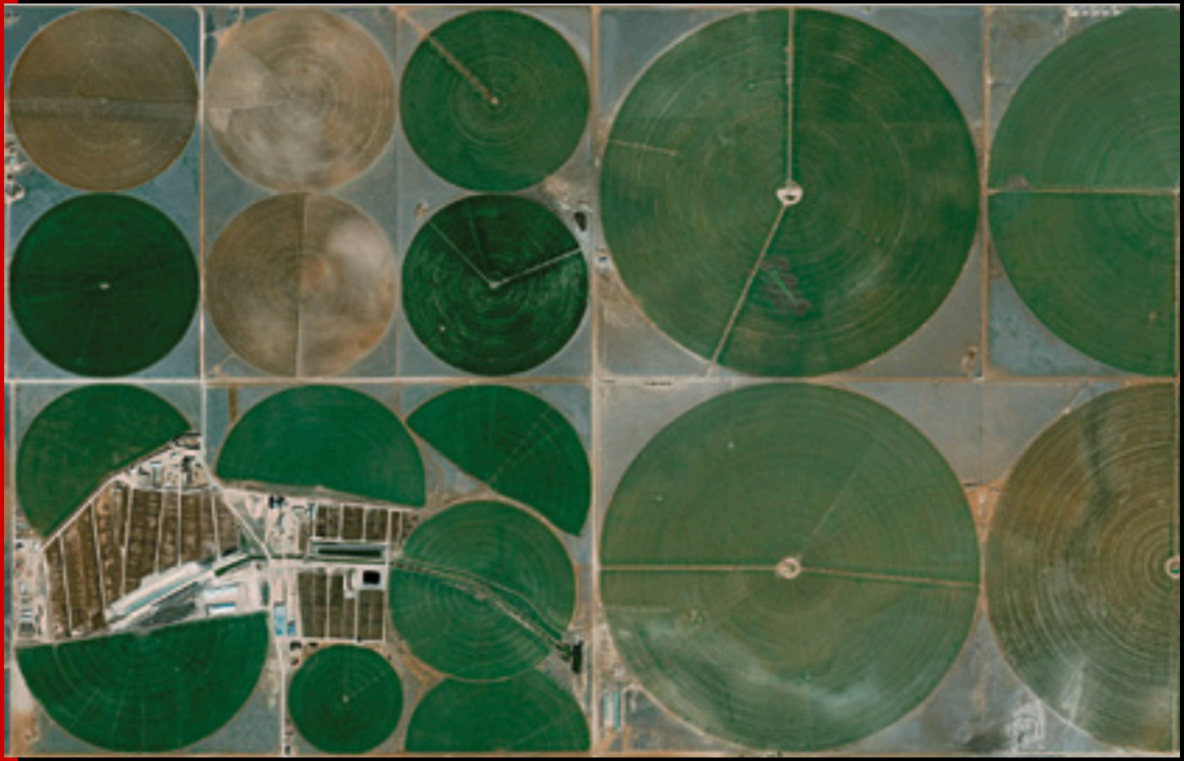
NASA

ESA

Outline

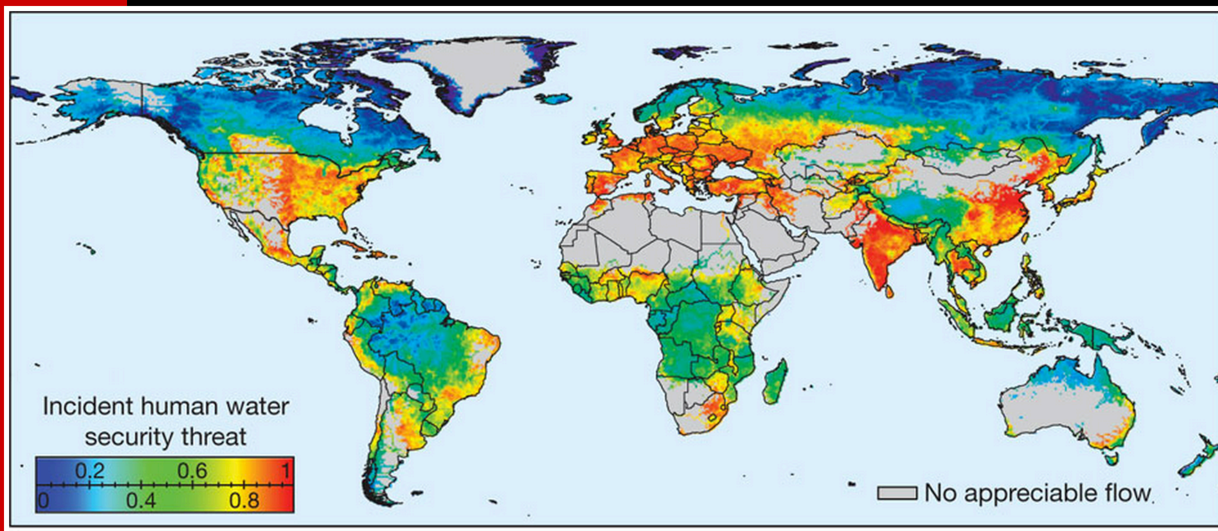
- Remote Sensing Overview
- Applications
  - Water
  - Food
  - Energy
- Algorithms
  - Gaussian Process (GP) Learning
  - Bi-temporal Hierarchical and Probabilistic
  - Multi-view, Semantic, and Multiple Instance Classification
- Outlook

Water, Food, and Energy



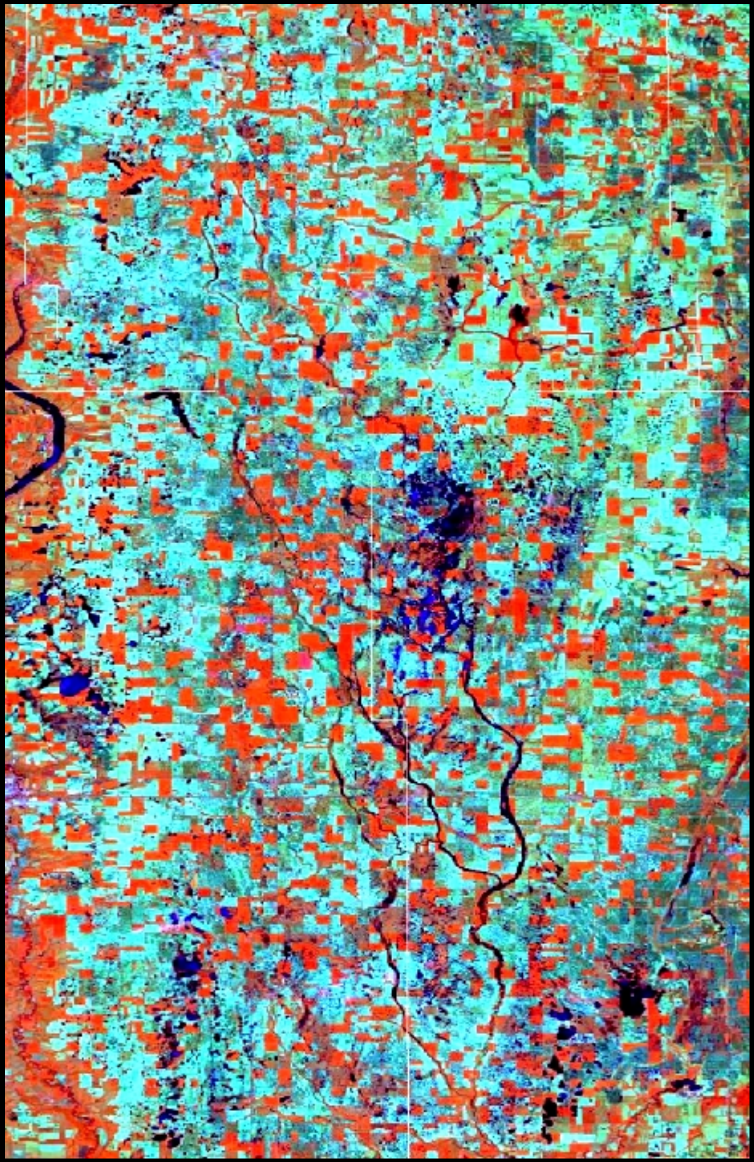
- Agriculture accounts for 70 percent of total global freshwater withdrawals
- Food production and supply chain consumes about 30 percent of total energy consumed globally
- Fast forward: 2050
    - 60% more food is needed
    - 50% energy consumption
    - 10% increase of water usage in agriculture

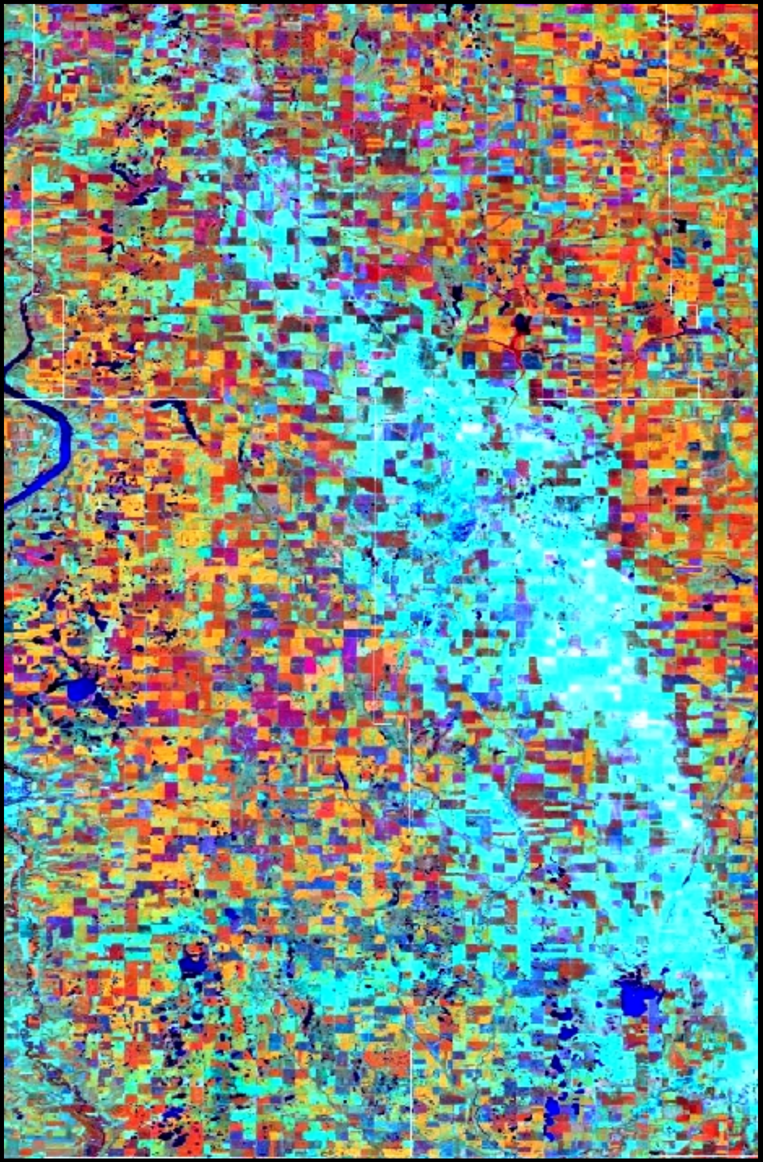
Global Threats to Human Water Security



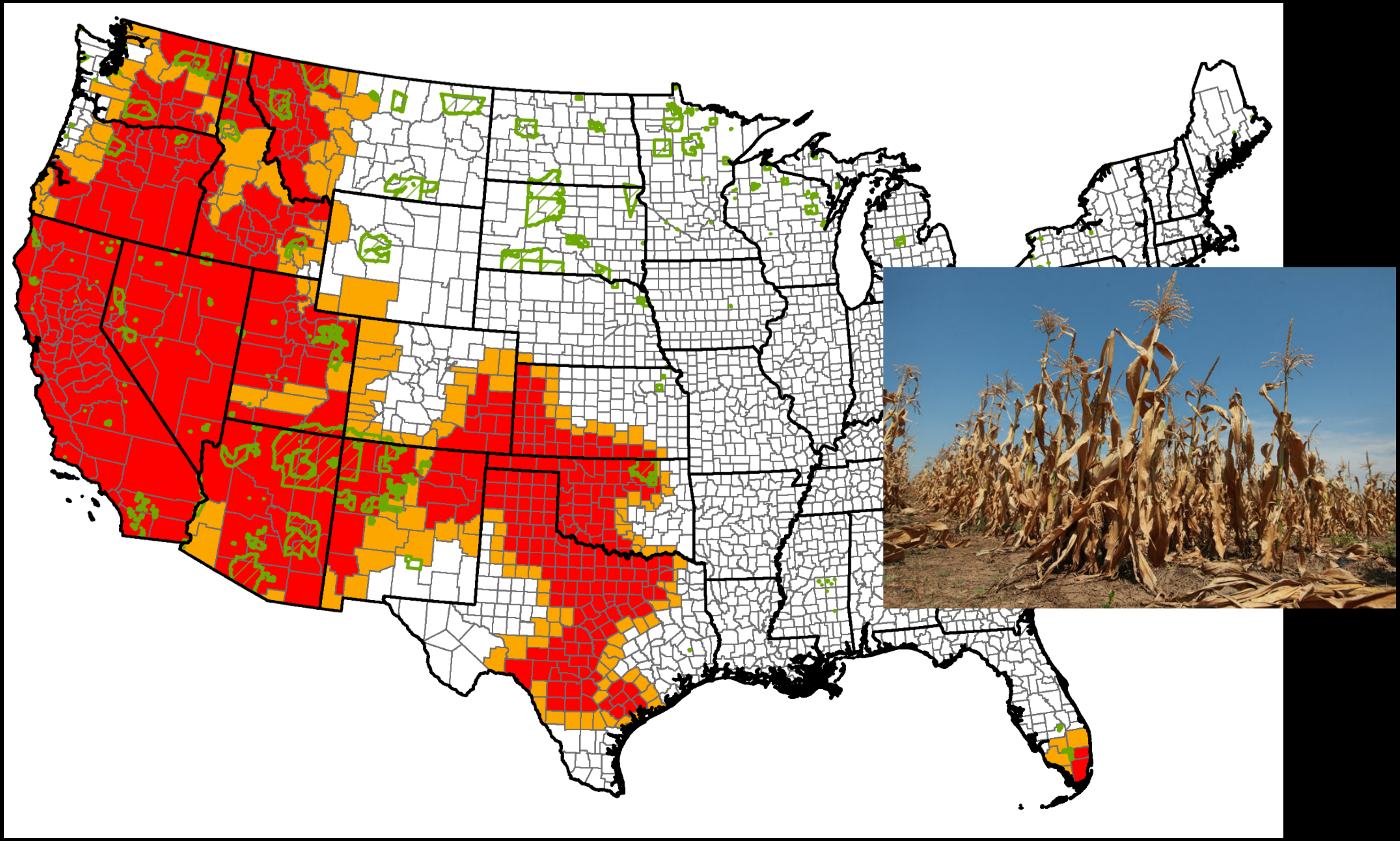
- C. J. Vörösmarty, et. al. Global threats to human water security and river biodiversity. Nature 467, 555–561 (30 September 2010).
- C. J. Vörösmarty. Global water resources: Vulnerability from climate change and population growth. *Science* 289: 284-288. 2000.

Food – Weather Threats





Droughts



USDA-2015

Human Impacts

Pollution

Biomass Monitoring

- Changes are dynamic and multifaceted
  - Population pressure (Present: ~7B; 2050: ~9B)
  - Bioenergy demands/policies
    - Strategic goals: Reduce gasoline use by 20% by 2017 and 30% by 2030.
    - 2007: 6.8 billion gallons
    - 2030: 60 billion gallons
  - Increasing emphasis on Feedstocks (DOE/OBP, "Biomass: Multi-Year Program Plan," March 2008).
  - Emphasis on growing energy crops (Cellulosic ethanol)
  - Diseases
  - Natural disasters
- This will lead to significant land use changes in US and other countries

NC STATE UNIVERSITY

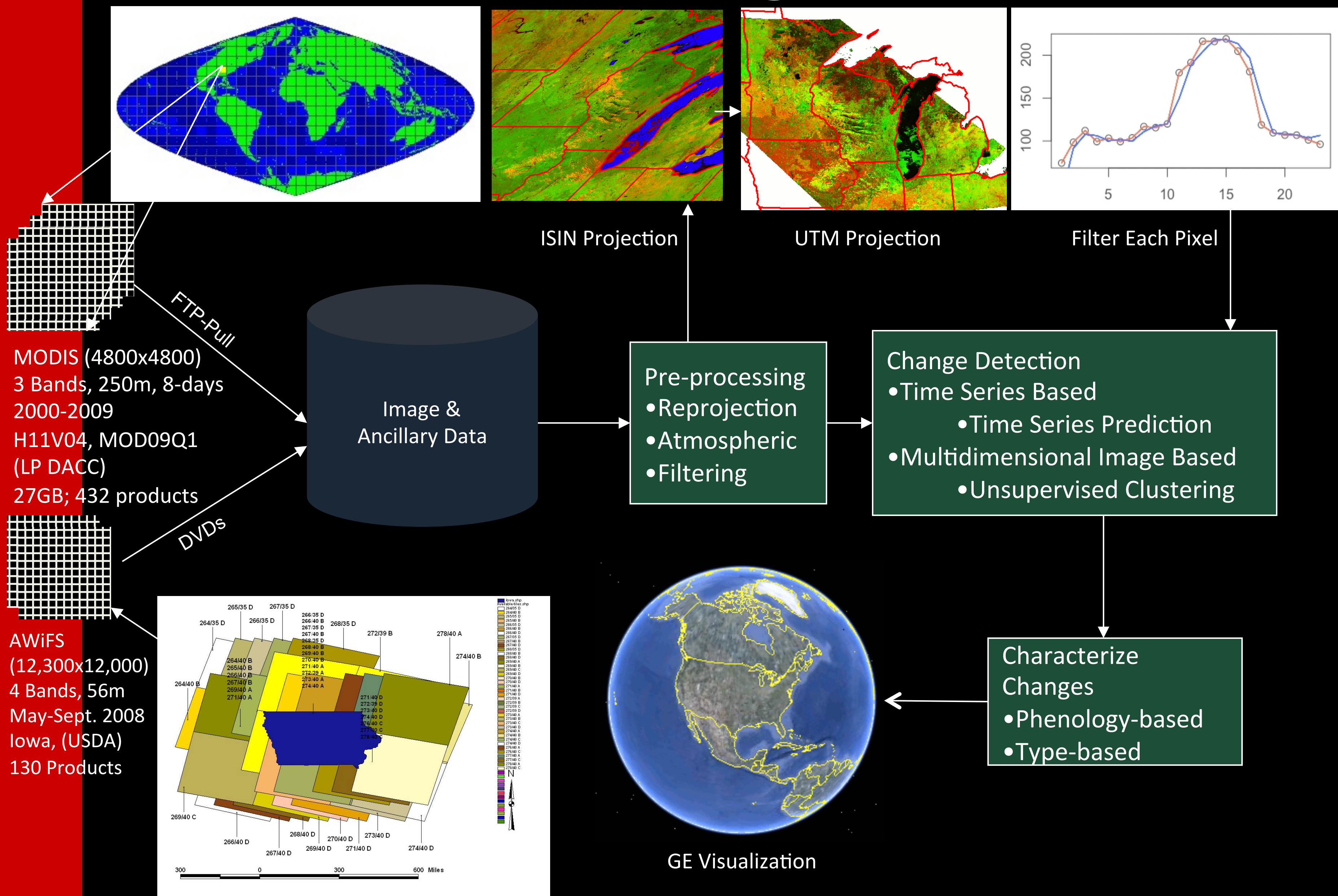
Biomass Monitoring

- Supporting the national bioenergy infrastructure will demand moving to operational mode
  - Existing federal mapping efforts are slow, for example NLCD (Started: 1992, Released:2000; 2nd Ver. Started: 2001, Released: 2007) and Cropland Data Layer (CDL): Annual (not wall-to-wall)
  - Dynamic assessment of "State of Biomass"
- Timely and accurate biomass monitoring is extremely important for both economic and energy security
  - Crops are susceptible to diseases, natural disasters, droughts, early frost, etc.
    - 1970: Naturally occurring leaf blight disease destroyed crops ~ $ 1B
    - 2008: Iowa flood damages to croplands ~ $3B

Seasonal Changes

AVHRR NDVI 1KM (1981-2000)

Biomass Monitoring Framework

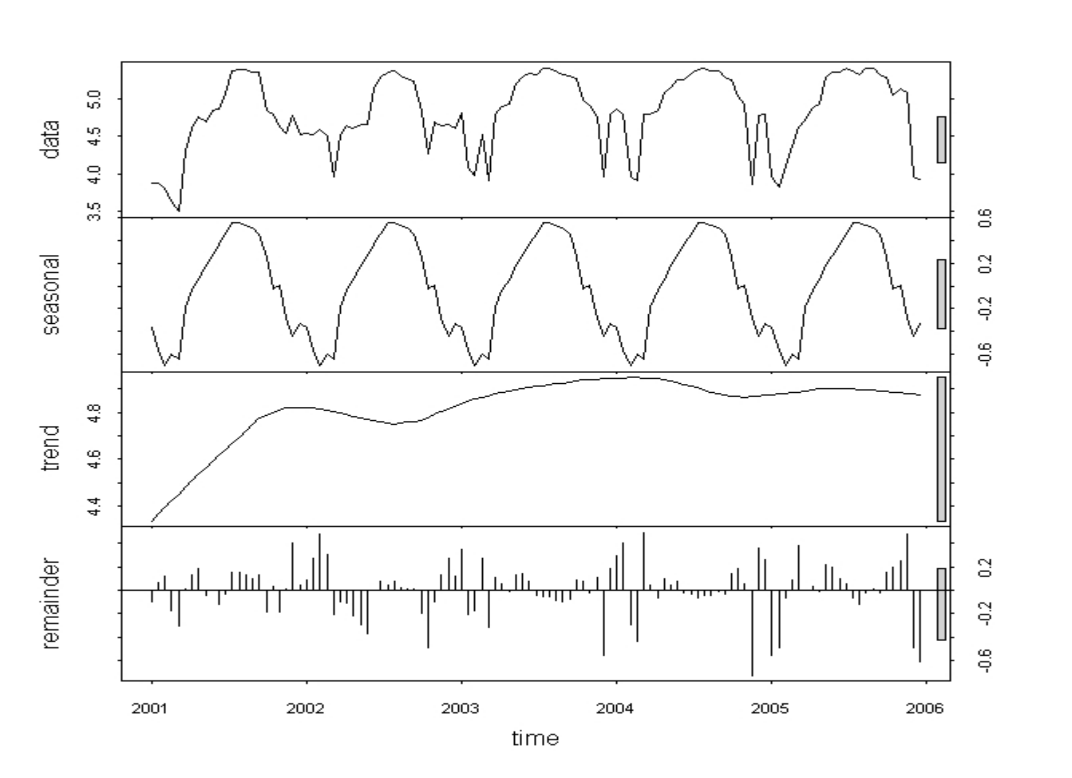


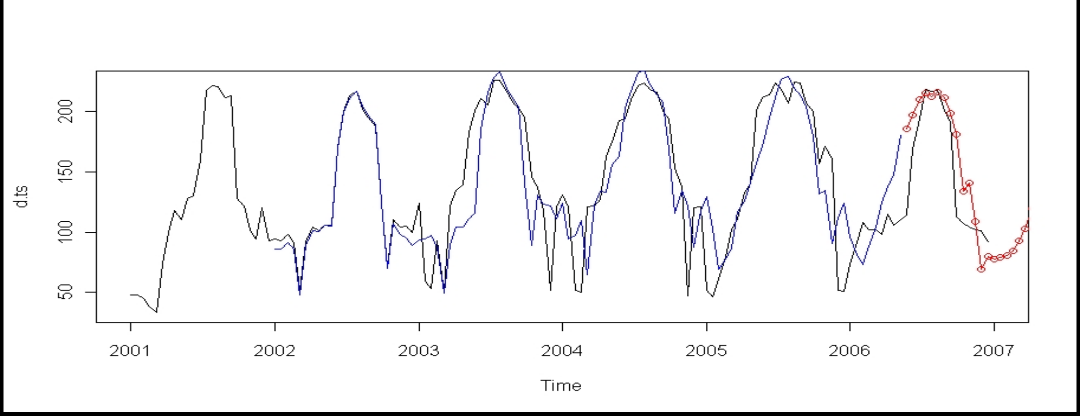
NC STATE UNIVERSITY

Time Series Based Change Detection

Basic algorithm

- Learn from past observations, that is, build a model that fits to all previous observations (NDVI time series)
- Using the model
  - Predict NDVI at next time step
- Determine if there is a change
  - Compare predicted value with observed (current NDVI image) value
  - If the difference is within a threshold, no change, else "possible change"
- Challenges
  - Which model
  - What is the appropriate threshold





NC STATE UNIVERSITY

Gaussian Process (GP) Regression

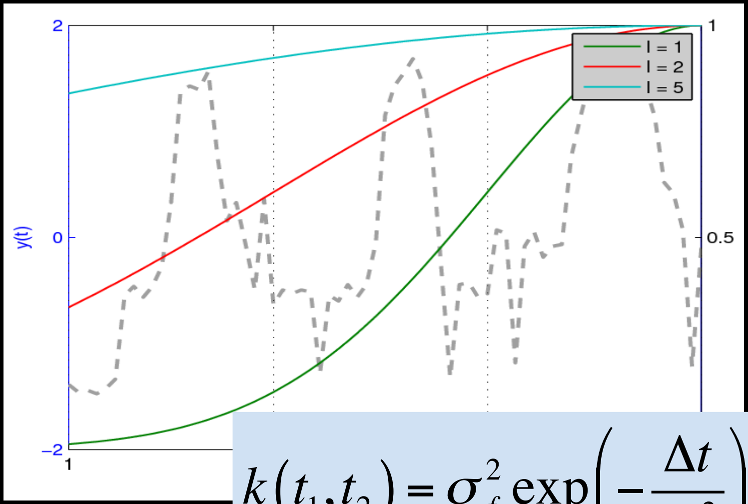
$$y_i = f(x_i) + \varepsilon$$

- GP Prior

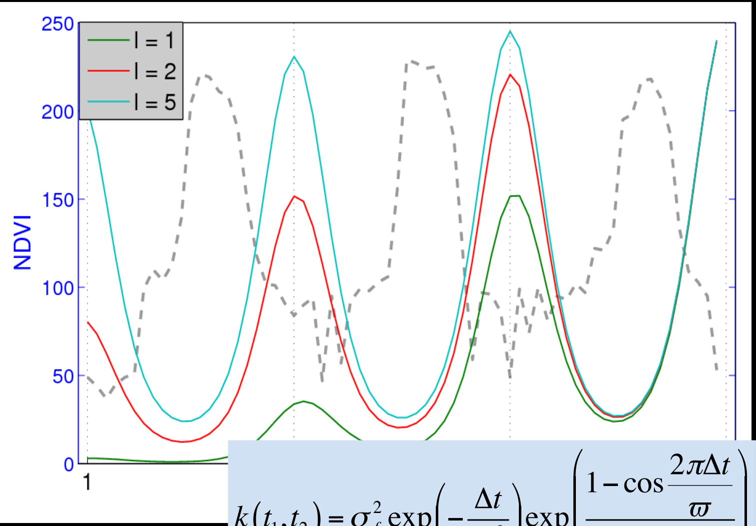
$$f(x_1), f(x_2), \ldots, f(x_n) \sim N(m(x), K)$$

$$K[k][j] = k(x_i, x_j)$$

- Covariance
  - Closer time instances should have similar values
  - Can capture seasonality via sinusoid covariance function

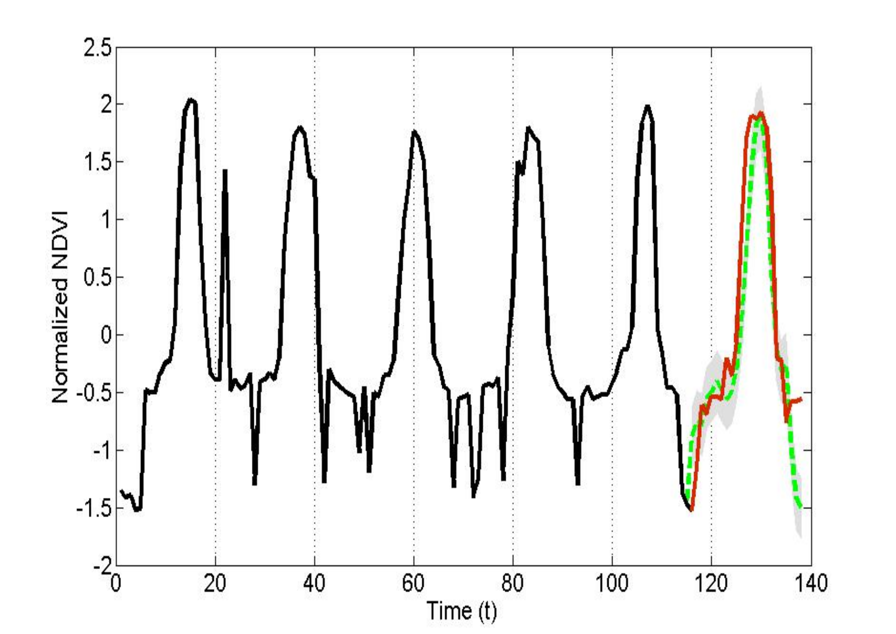


$$k(t_1, t_2) = \sigma_f^2 \exp\left(-\frac{\Delta t}{2l^2}\right)$$



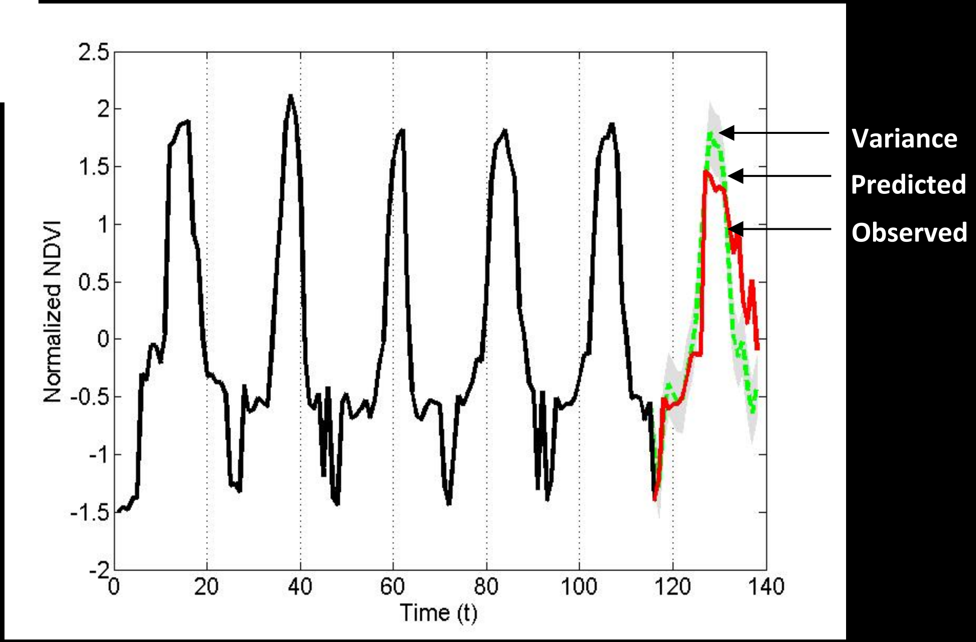
$$k(t_1, t_2) = \sigma_f^2 \exp\left(-\frac{\Delta t}{2l^2}\right) \exp\left(\frac{1 - \cos\frac{2\pi\Delta t}{\varpi}}{a}\right)$$

GP Based Change Detection



No Change

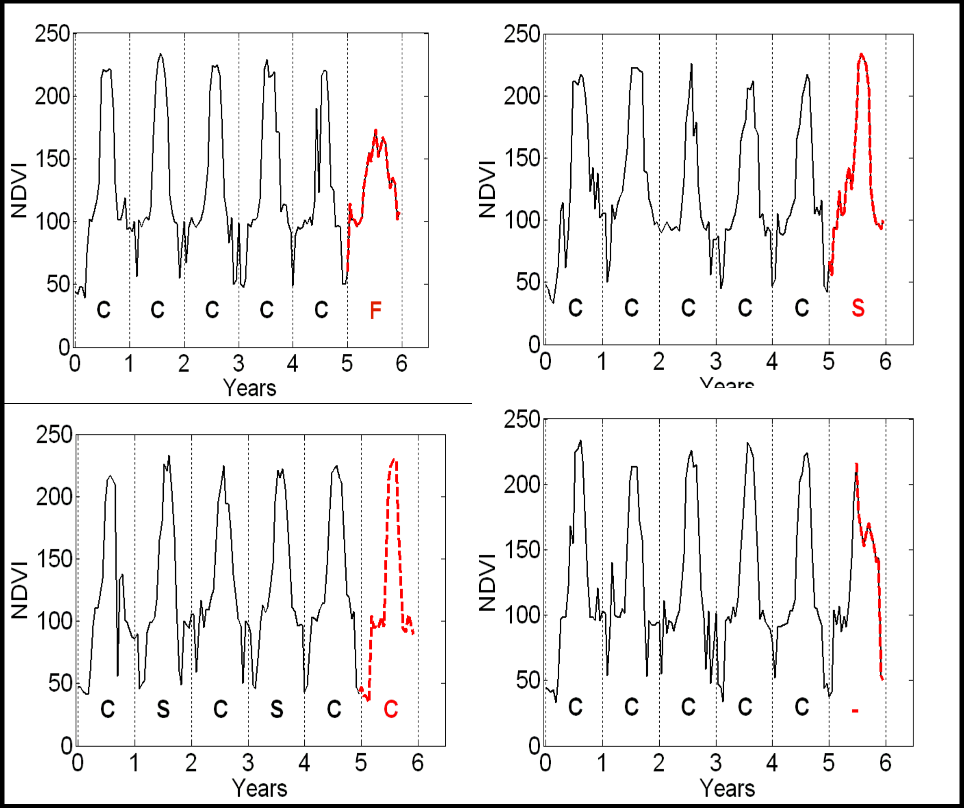
Change



NC STATE UNIVERSITY

Results

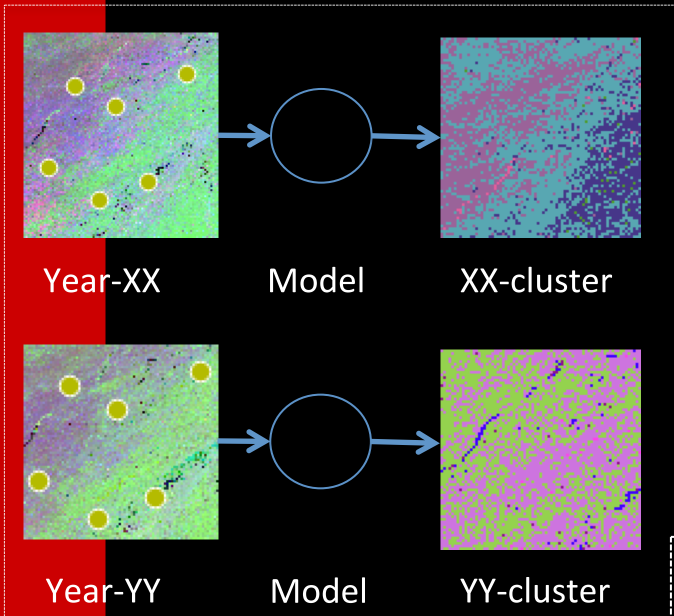
- MODIS Time Series From Iowa
  - 6 years (2001-2006)
  - 23 Observations/ year
- Labeled data: 97
- Accuracy: 88%



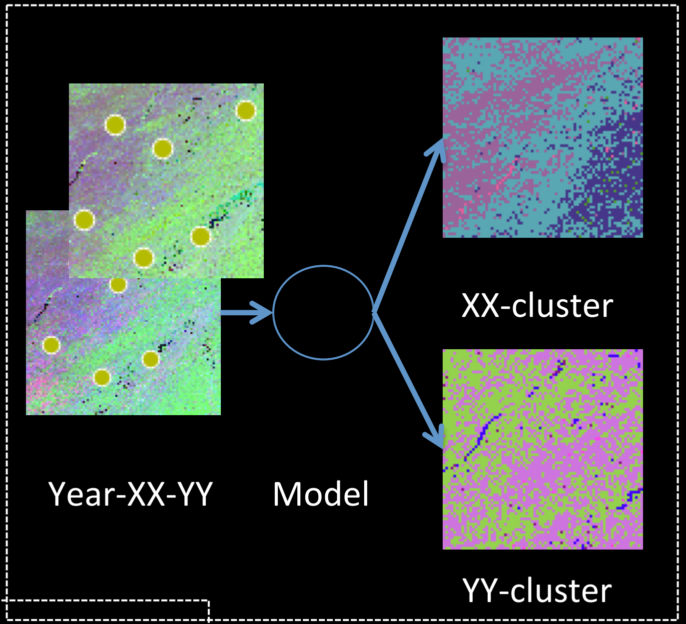
C-Corn; S-Soy; F-Fallow

Varun Chandola, Ranga Raju Vatsavai: A scalable gaussian process analysis algorithm for biomass monitoring. Statistical Analysis and Data Mining 4(4): 430-445 (2011)

Biannual Changes

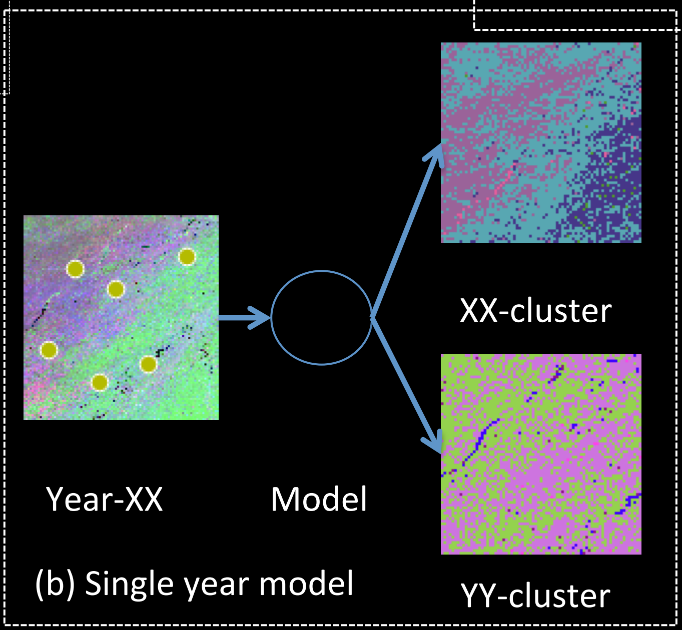


(a) Year-wise independent cluster model



(c) Combined year model

Unsupervised Methods

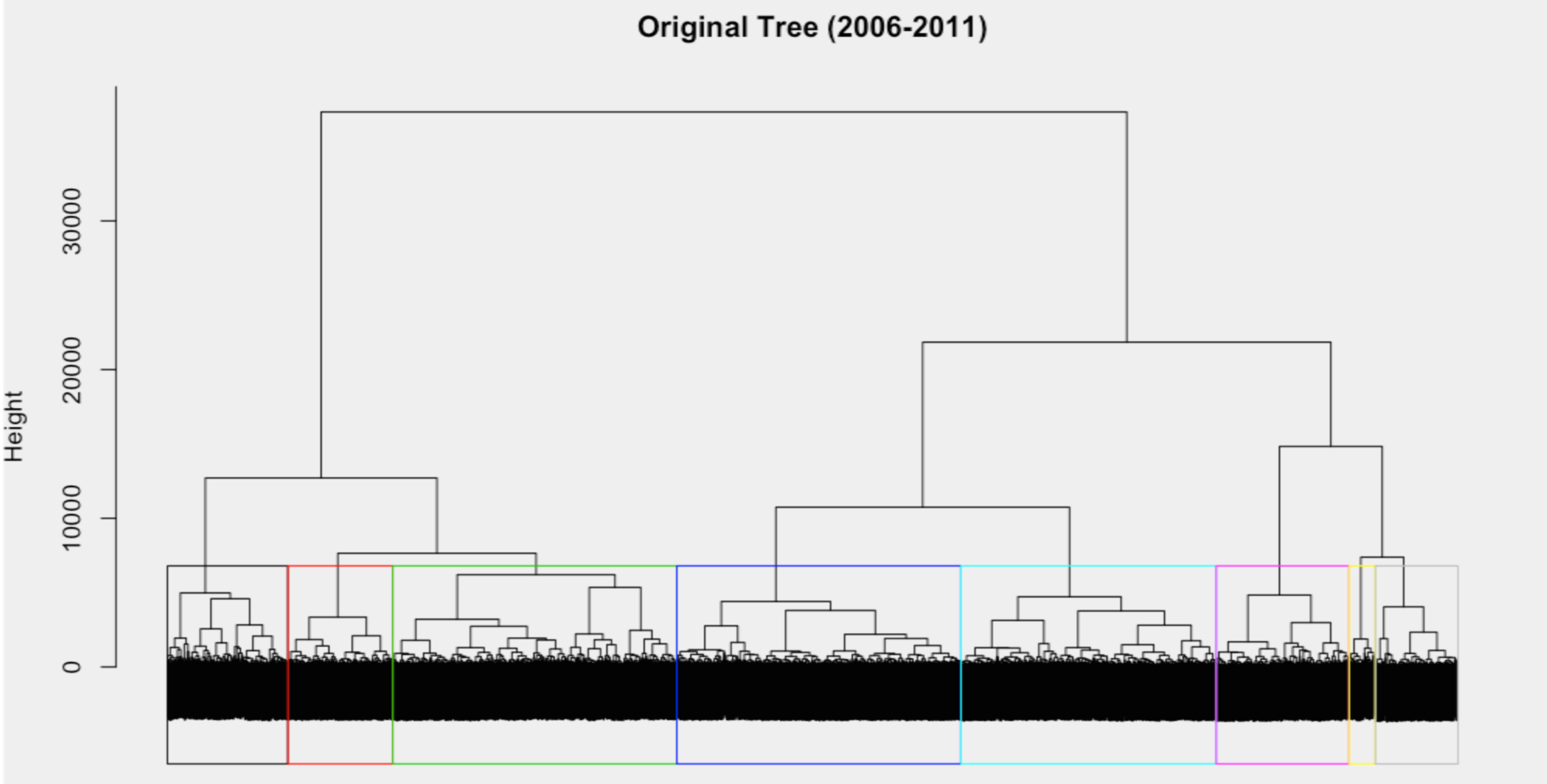


(b) Single year model

Hierarchical Change Detection

- Hierarchical clustering
  - Grouping NDVI time-series by similarity
- Extract change relationships
- Generate change image

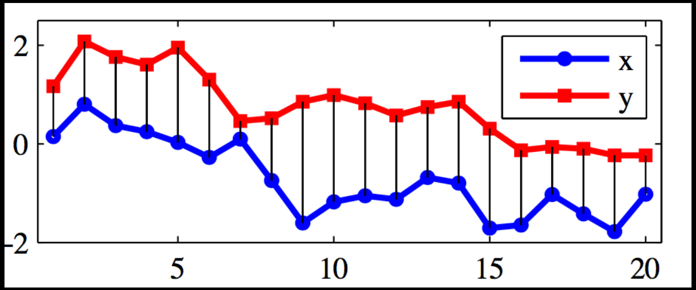
Hierarchical Model

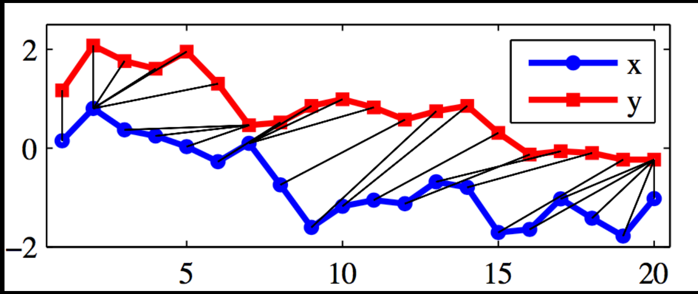


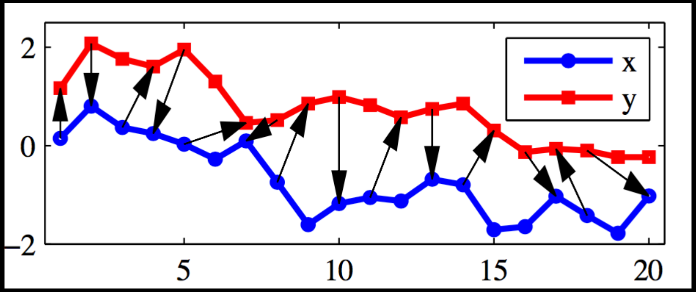
1	2	3	4	5	6	7	8
759	221	759	322	682	355	280	70

Similarity Measures

- Dynamic Time Warping (DTW; Berndt and Clifford, 1994)
- Edit Distance on Real Sequences (EDR; Chen et al., 2005)
- Minimum jump costs (MJC; Serra and Arcos, 2012)





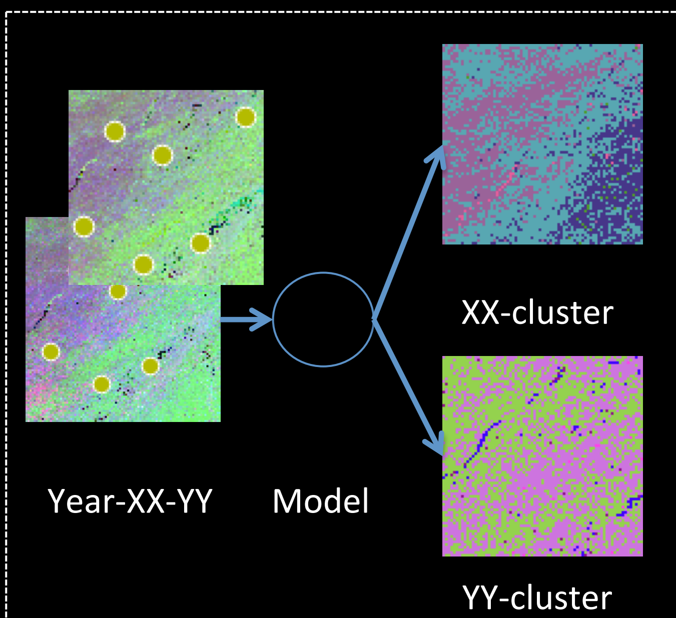


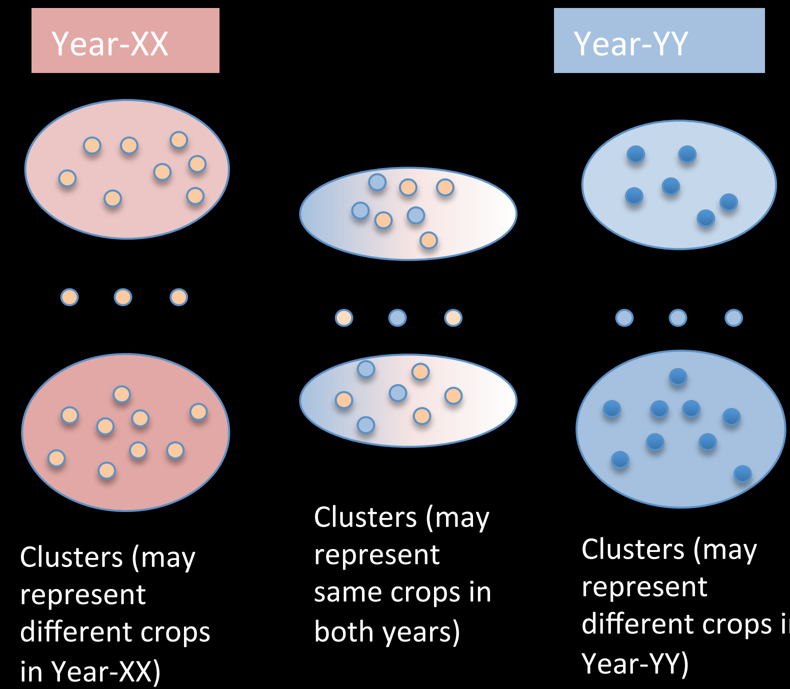
Source: Joan Serra, Josep Ll. Arcos. An Empirical Evaluation of Similarity Measures for Time Series Classification

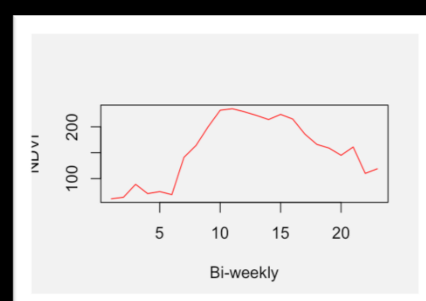
NC STATE UNIVERSITY

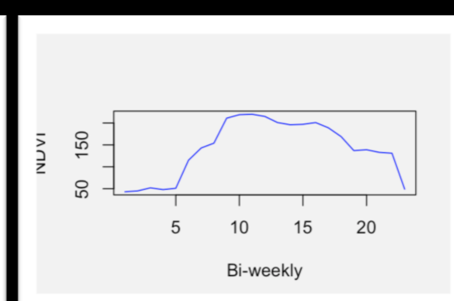
Combined Model

- Build model on samples from Y1 and Y2 (Y12.HM)
- Use Y12.HM to predict labels for Y1 and Y2



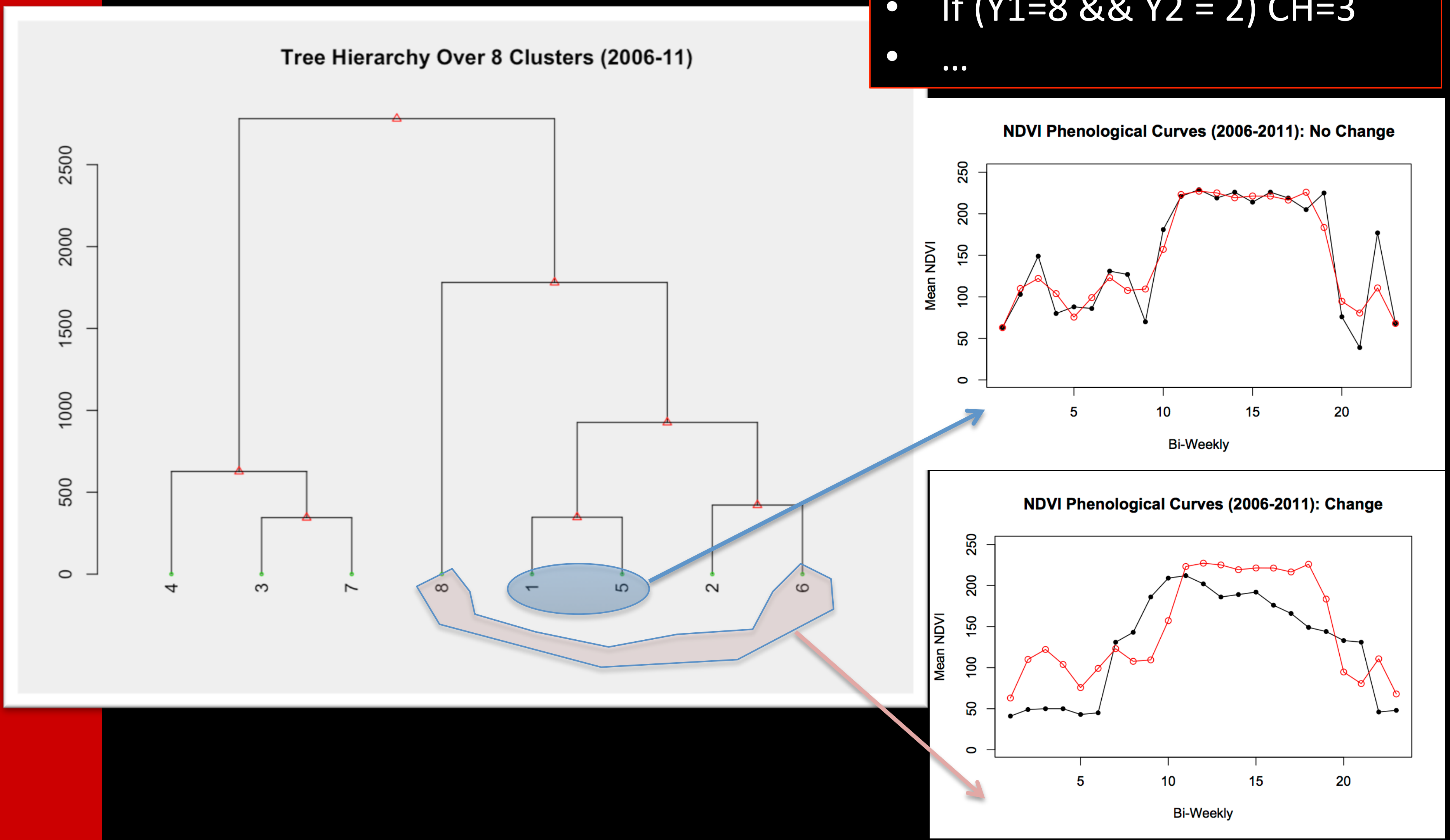




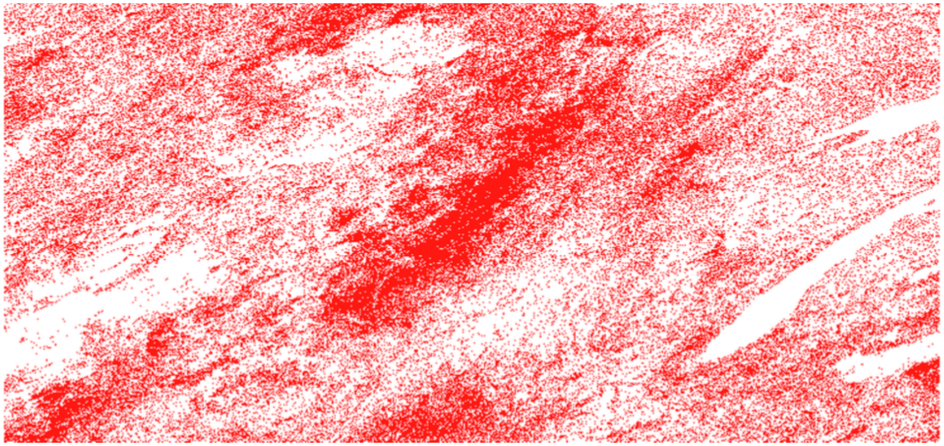


Extract Hierarchical Changes

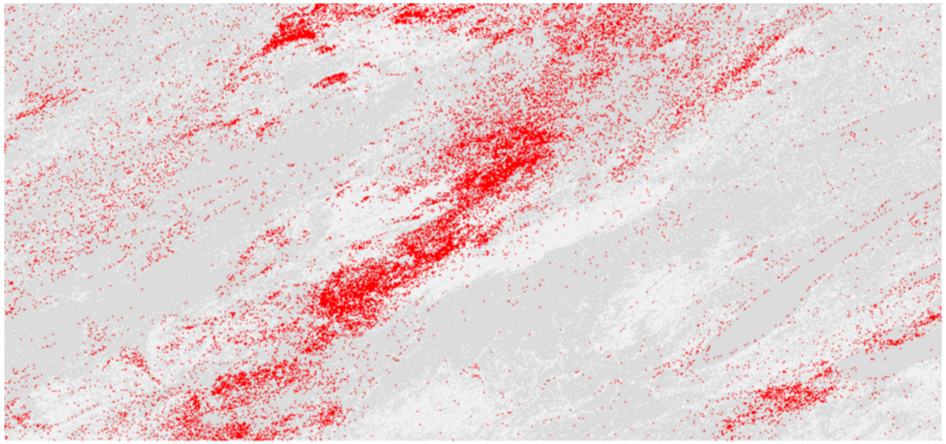
- If (Y1=1 && Y2 = 6) CH=2
- If (Y1=8 && Y2 = 2) CH=3
- …



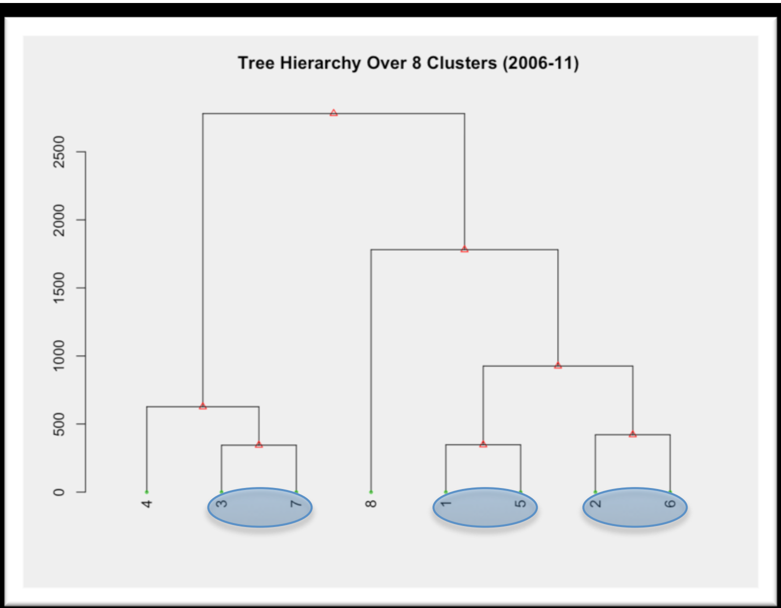
Results



K-Means



Hierarchical



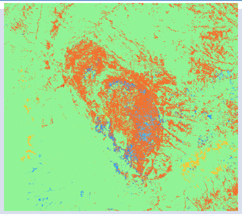
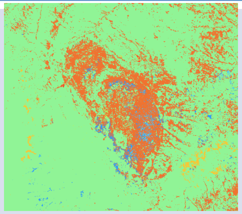
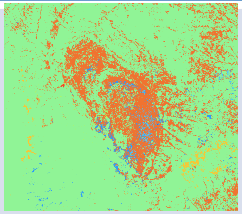
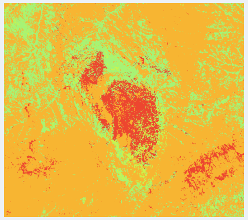
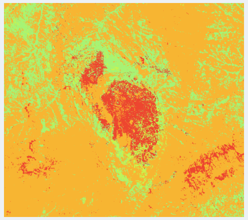
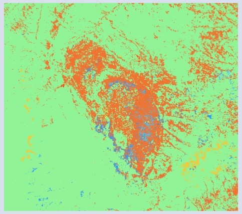
K-Means over predicts changes (3-7;1-5;2-6)

Year	K-Means	HC
2001-02	33	08
2001-03	29	08
2001-04	30	06
2001-05	31	06
2001-06	34	08
2001-07	31	06
2001-08	33	06
2001-09	30	06
2001-10	36	08
2001-11	35	09

Ranga Raju Vatsavai. Hierarchical Change Detection. (Under Review)

NC STATE UNIVERSITY

Other Applications

	Y1	Y2	Y3	Y4
Y1				
Y2				
Y3				

Online Change Browser

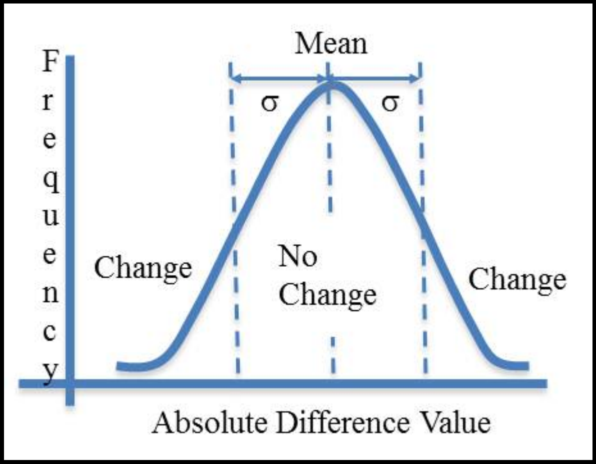
NC STATE UNIVERSITY

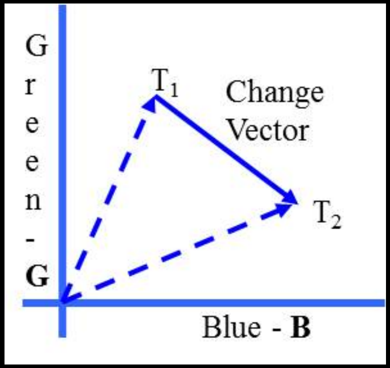
Damage Assessments

- Settlement Dynamics
  - Damages to existing structures
  - New construction
- Biomass
  - Forest fires
  - Floods and Hail Storms
  - Disease

Bi-temporal Change Detection

- Image Differencing
  - $I_{Diff}(i,j) = I_2(i,j) - I_1(i,j)$
  - Thresholding, Sensitive to noise
- Ratio of Means
  - $I_{Ratio}(i,j) = I_2(i,j) / I_1(i,j)$
  - Robust to multiplicative noise
- Inner Product and Spectral Correlation
- Multivariate Alteration Detection (MAD)
- L. Bruzzone, F. Bovolo, 2013



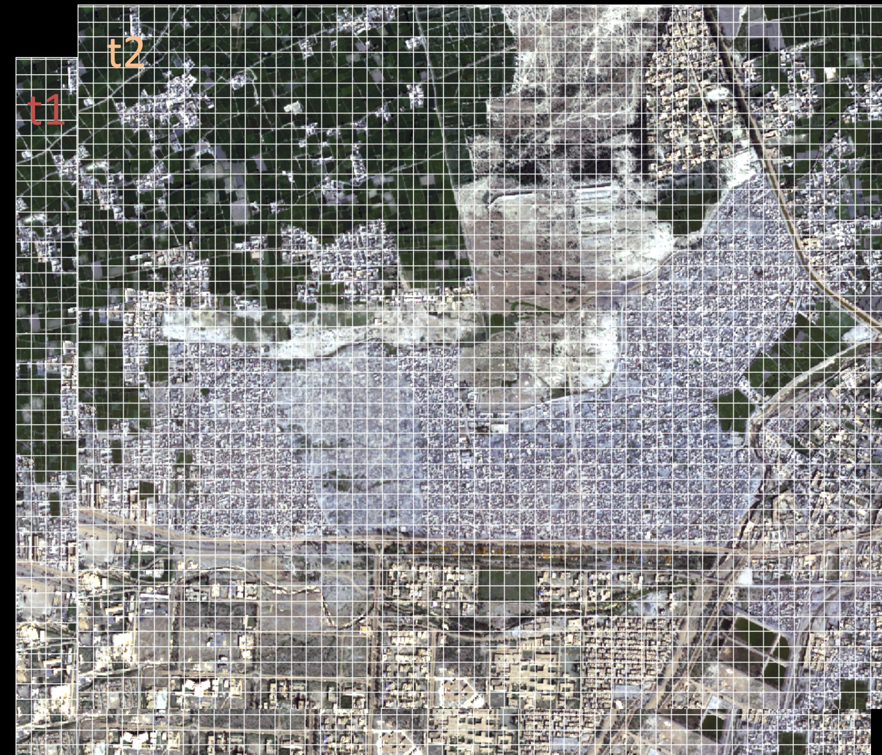


Limitations

- Point based – at individual pixel (or small neighborhood)
- Mostly Univariate
- Multivariate (e.g., MAD) techniques produce multi-band change maps
- Mostly the output is continuous (requires thresholding)

Probabilistic Approach

- Divide image into fixed grids
- Model that data in a grid is generated by probability distribution
- Estimate the overlap between two grids (distributions)
  - No change: distributions should be highly overlapping
  - Change: less overlap between distributions
- Clustering to find change and possible change groups (GMM)



Highly overlapping to No overlap

NC STATE UNIVERSITY

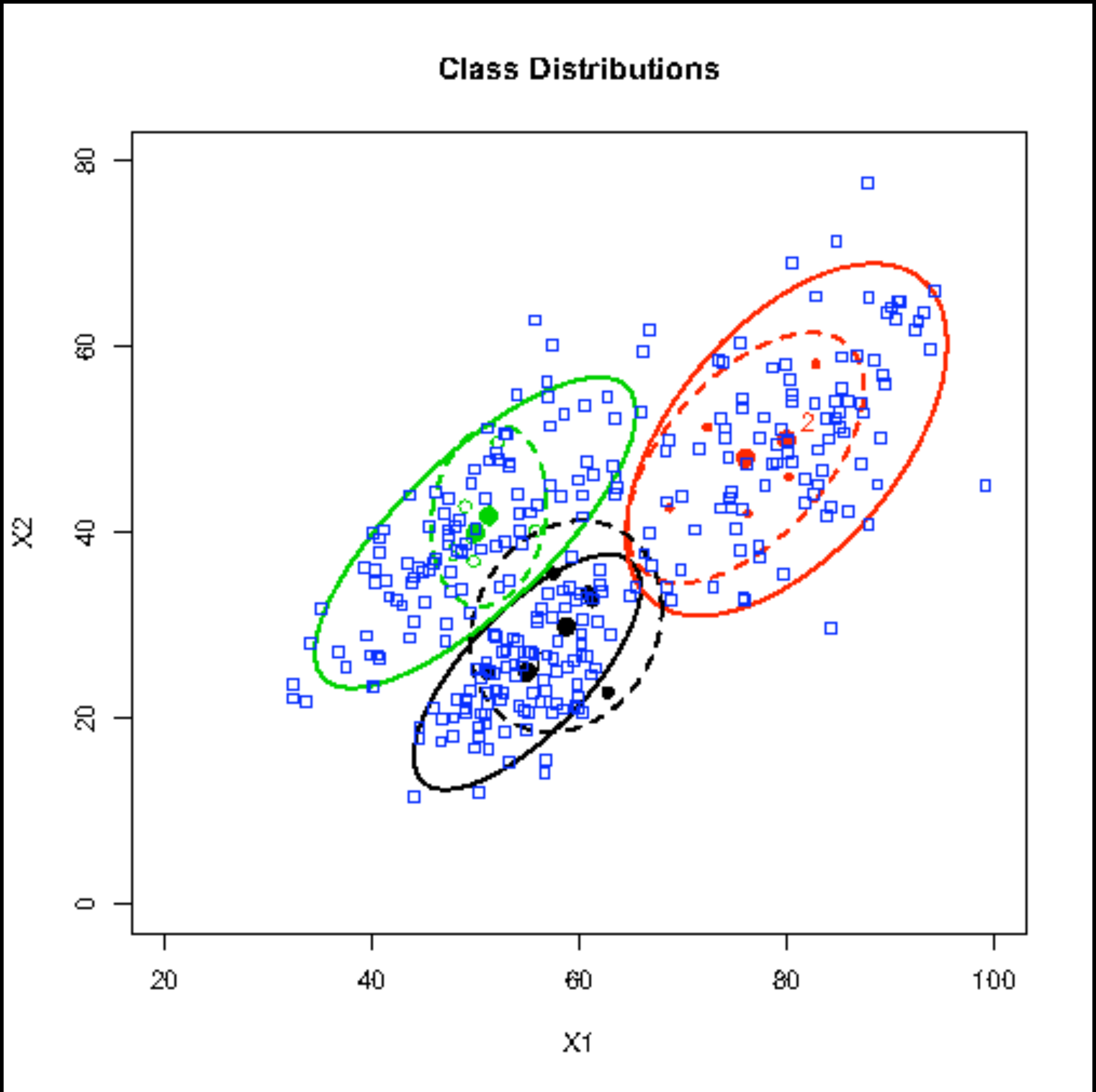
Clustering

- Distribution over grid-pair distances
- Gaussian Mixture Model (GMM)

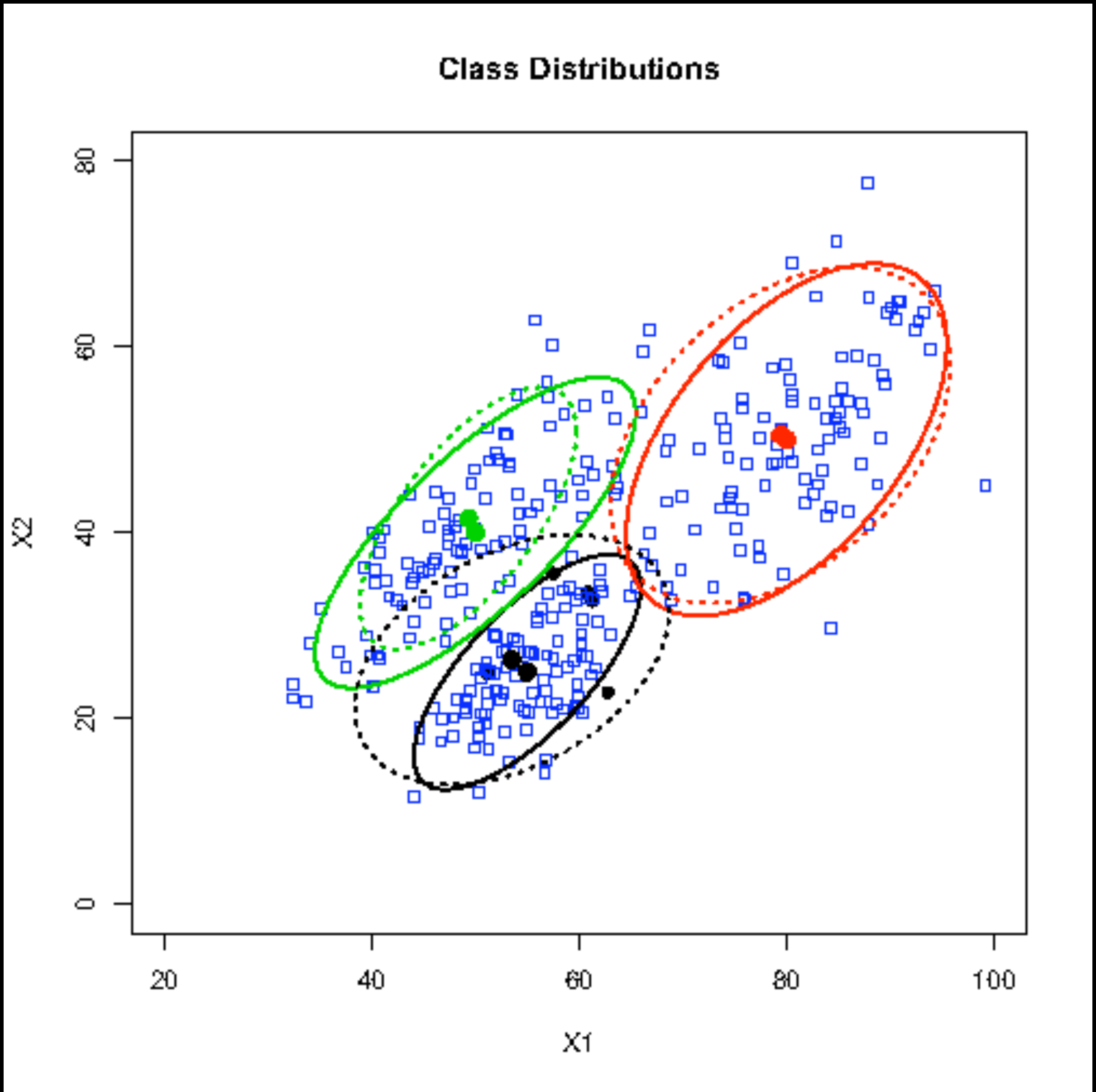
$$P(x_i \mid \Theta) = \sum_{j=1}^{K} \alpha_j P_j\left(x_i \mid \theta_j\right)$$

- Compute Model Parameters Using Expectation Maximization (EM)

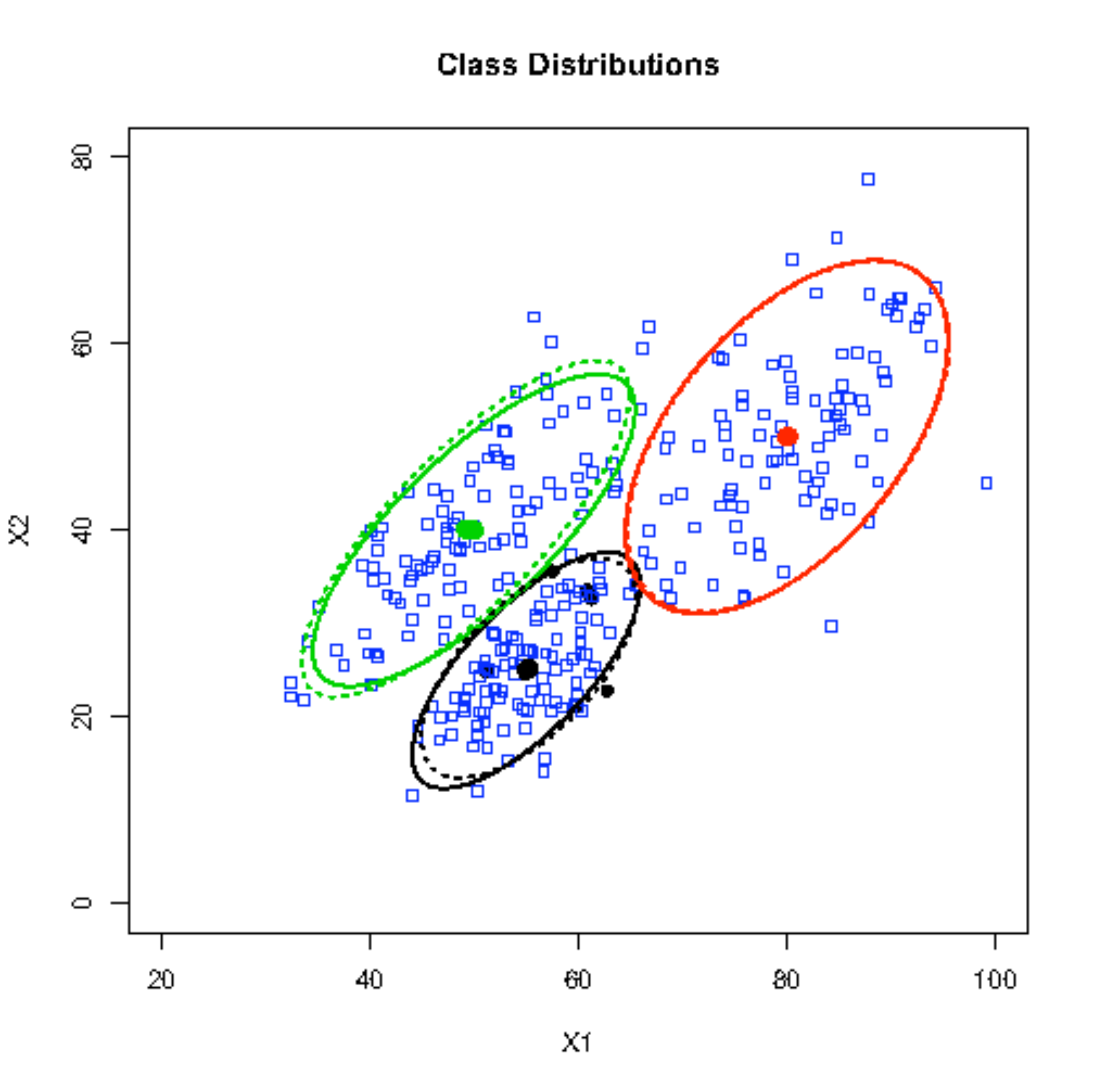
GMM Execution Trace



GMM Execution Trace



GMM Execution Trace



NC STATE UNIVERSITY

GMM Execution Trace

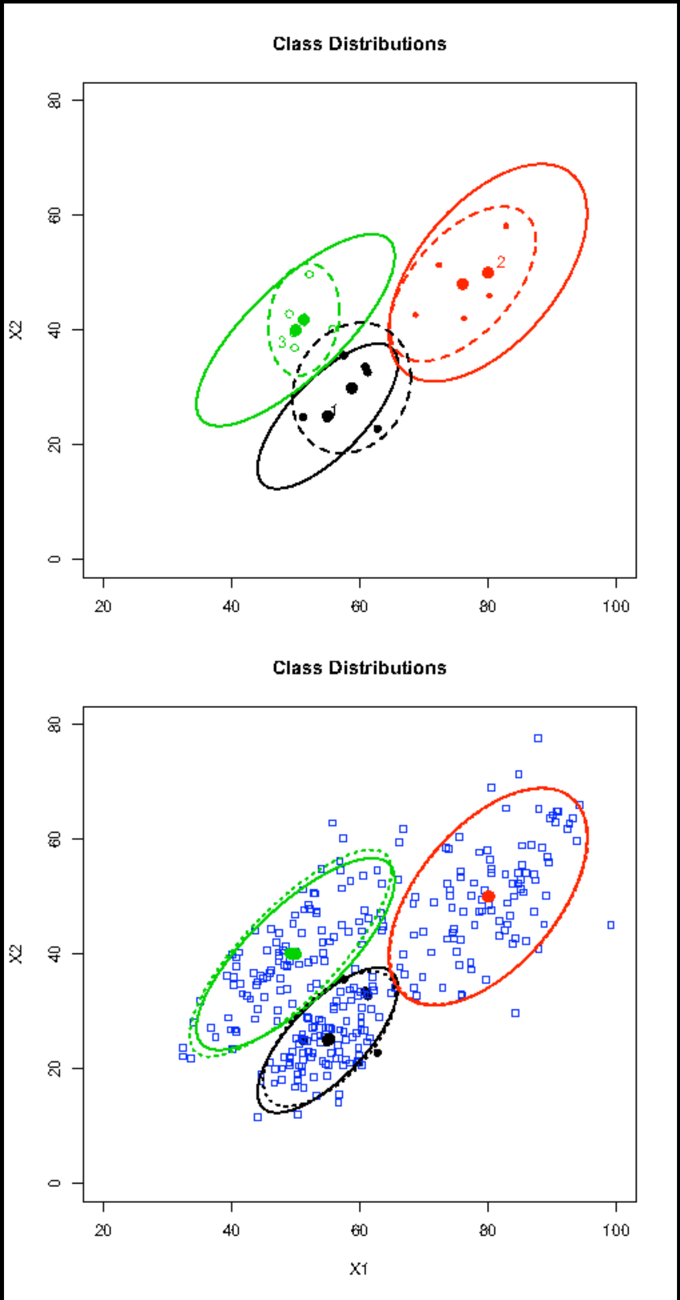
- Expectation Maximization (EM)
- E-Step

$$e_{ij} = \frac{\left|\hat{\Sigma}_j^k\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(x_i - \hat{\mu}_j^k\right)^T \hat{\Sigma}_j^{-1,k}\left(x_i - \hat{\mu}_j^k\right)\right\}}{\sum_{l=1}^{M}\left|\hat{\Sigma}_l^k\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(x_i - \hat{\mu}_l^k\right)^T \hat{\Sigma}_l^{-1,k}\left(x_i - \hat{\mu}_l^k\right)\right\}}$$

- M-Step

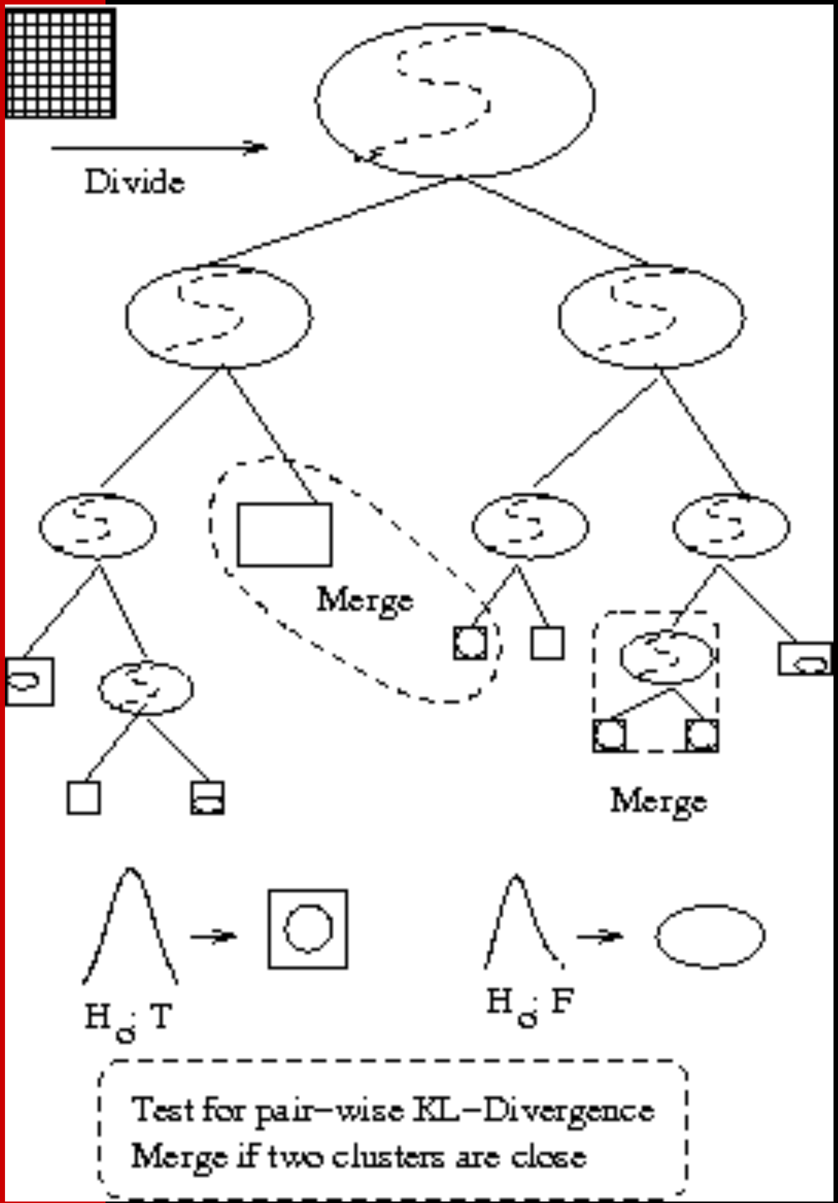
$$\alpha_j = \frac{\sum_{i=1}^{N} e_{ij}}{N}, \qquad \hat{\mu}_j^{k+1} = \frac{\sum_{i=1}^{N} e_{ij} x_i}{\sum_{i=1}^{N} e_{ij}},$$

$$\text{and} \quad \hat{\Sigma}_j^{k+1} = \frac{\sum_{i=1}^{N} e_{ij}\left(x_i - \hat{\mu}_j^{k+1}\right)\left(x_i - \hat{\mu}_j^{k+1}\right)^T}{\sum_{i=1}^{N} e_{ij}}$$



NC STATE UNIVERSITY

Challenge: How Many Clusters?



Inputs: D, sample dataset; significance (default p-value = 0.05), initial K (default = 2), nClusters = K

Loop 1: WHILE (TRUE):

Loop 2: FOR 1:nClusters

Statistical test: Shapiro-Wilk test.

Check: **IF** a cluster fails statistical test,

THEN split that cluster into two

clusters using GMM-Clustering;
increment nClusters and K;

ELSE accept cluster,

decrement nClusters

Clustering: GMM-Clustering(failed-cluster data-samples, new K)

Merge: Compute KL-Divergence,

IF two-clusters are closer than threshold,

THEN decrement K

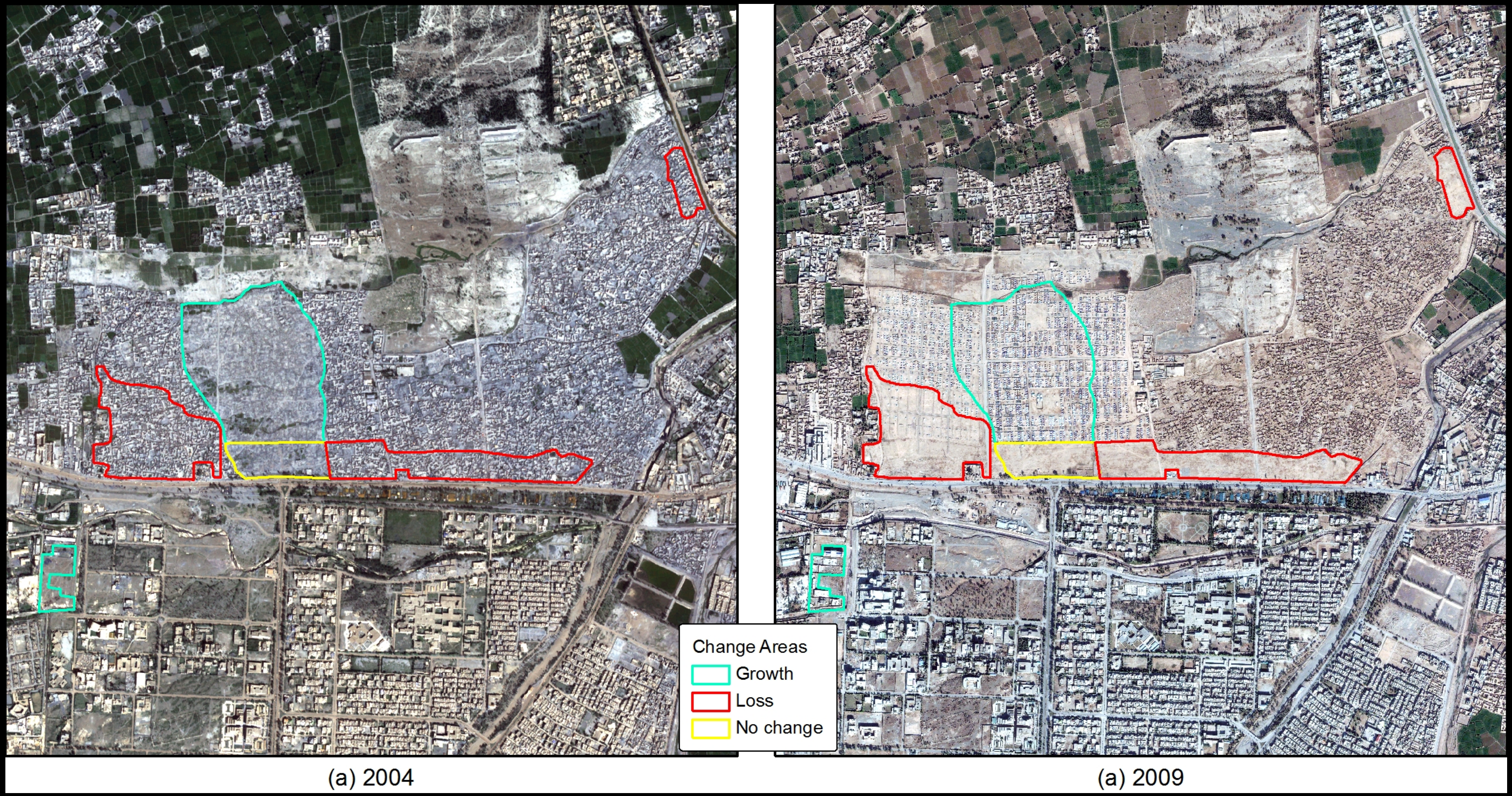
continue (Loop 2)

Check: **IF** nClusters = 0 (break, Loop 1)

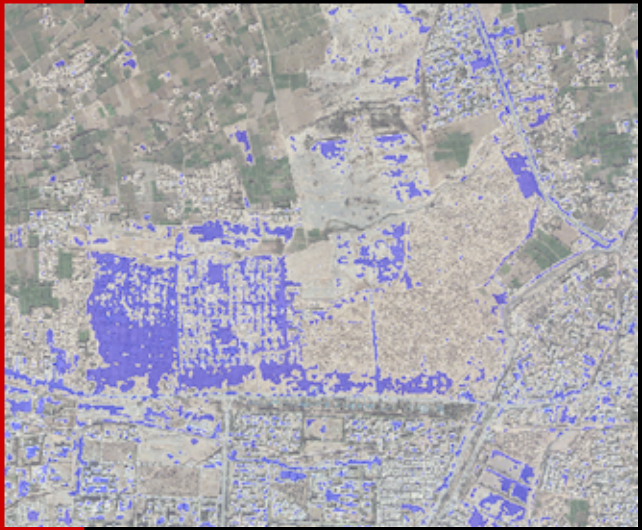
Output: Parameter vector Θ.

Results

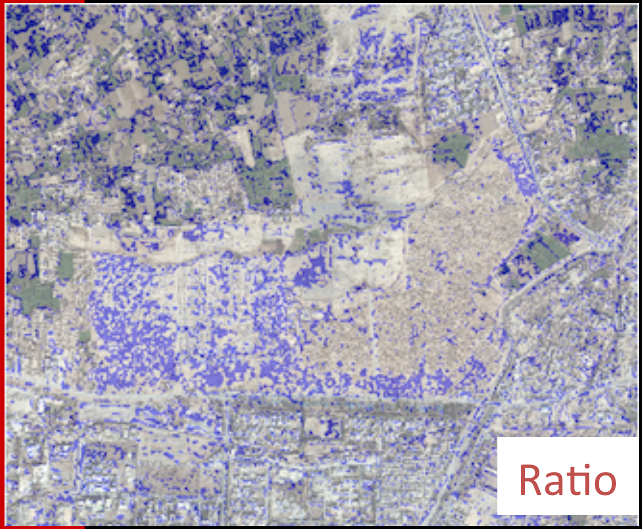
- Kacha Garhi Camp, Pakistan
- Established 1980 for Afghan Refugees
- QuickBird (2004 and 2009, 4B, 2.4m)

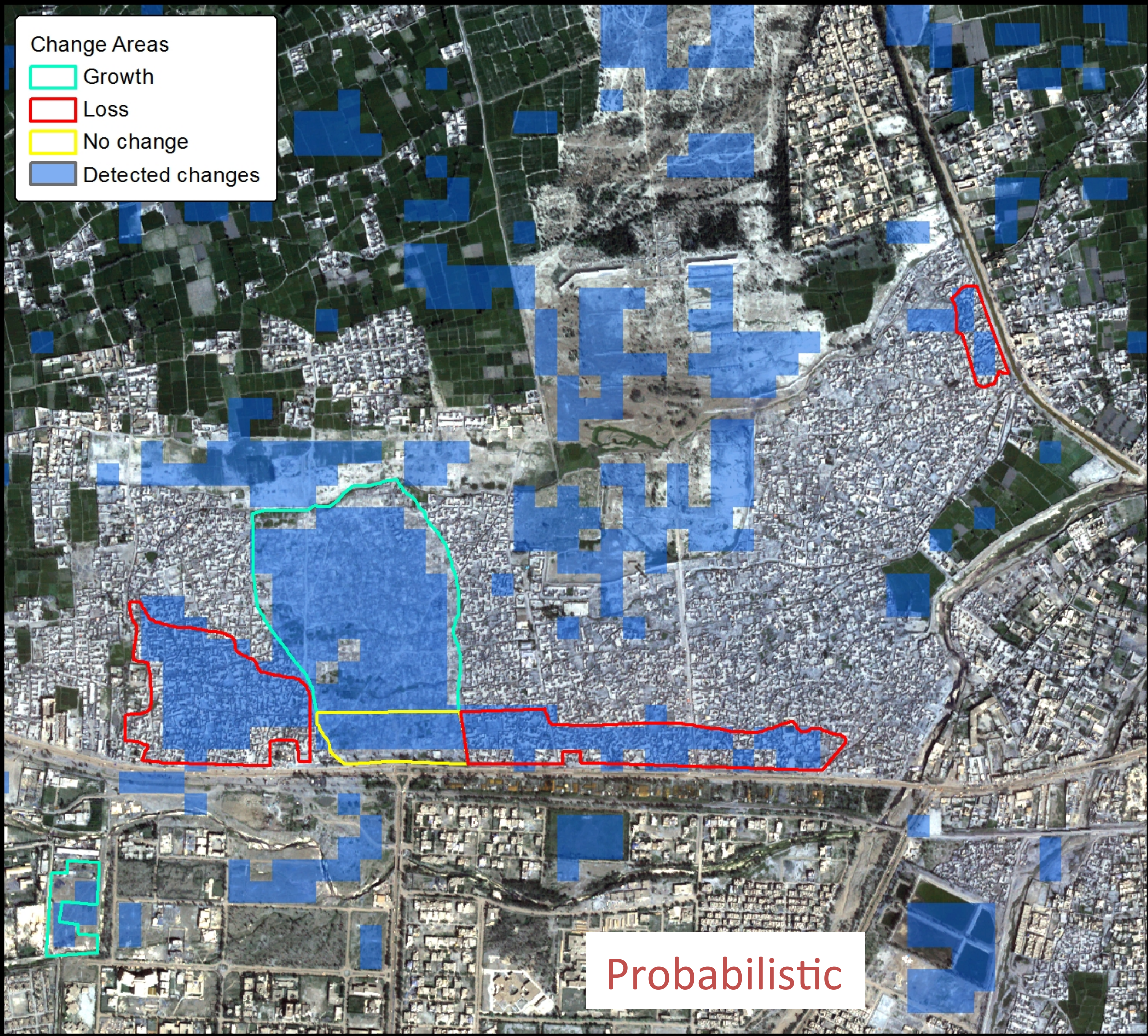


Results



Difference

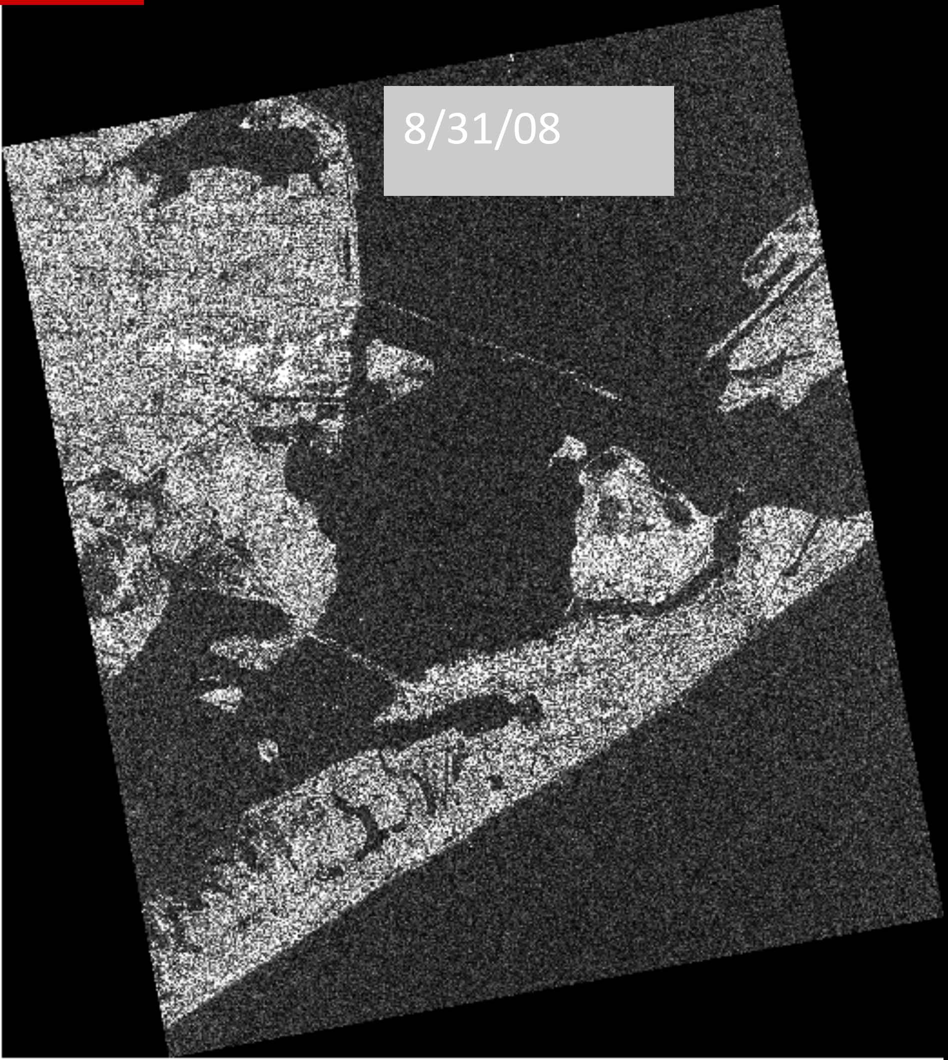


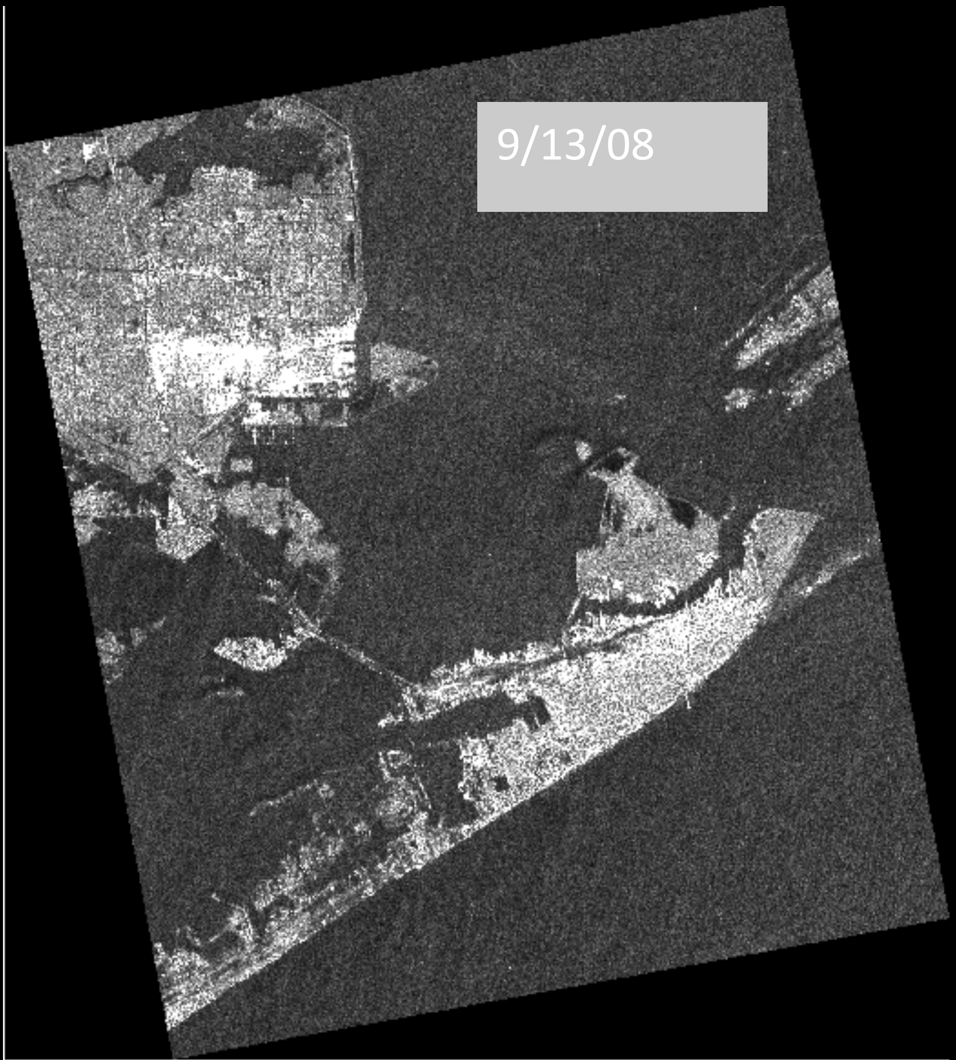


Ranga Raju Vatsavai, Jordan Graesser: Probabilistic Change Detection Framework for Analyzing Settlement Dynamics Using Very High-resolution Satellite Imagery. ICCS 2012: 907-916

Results

- SAR Imagery during Ike – noise, spatial resolution (1.56m vs. 12.5m)

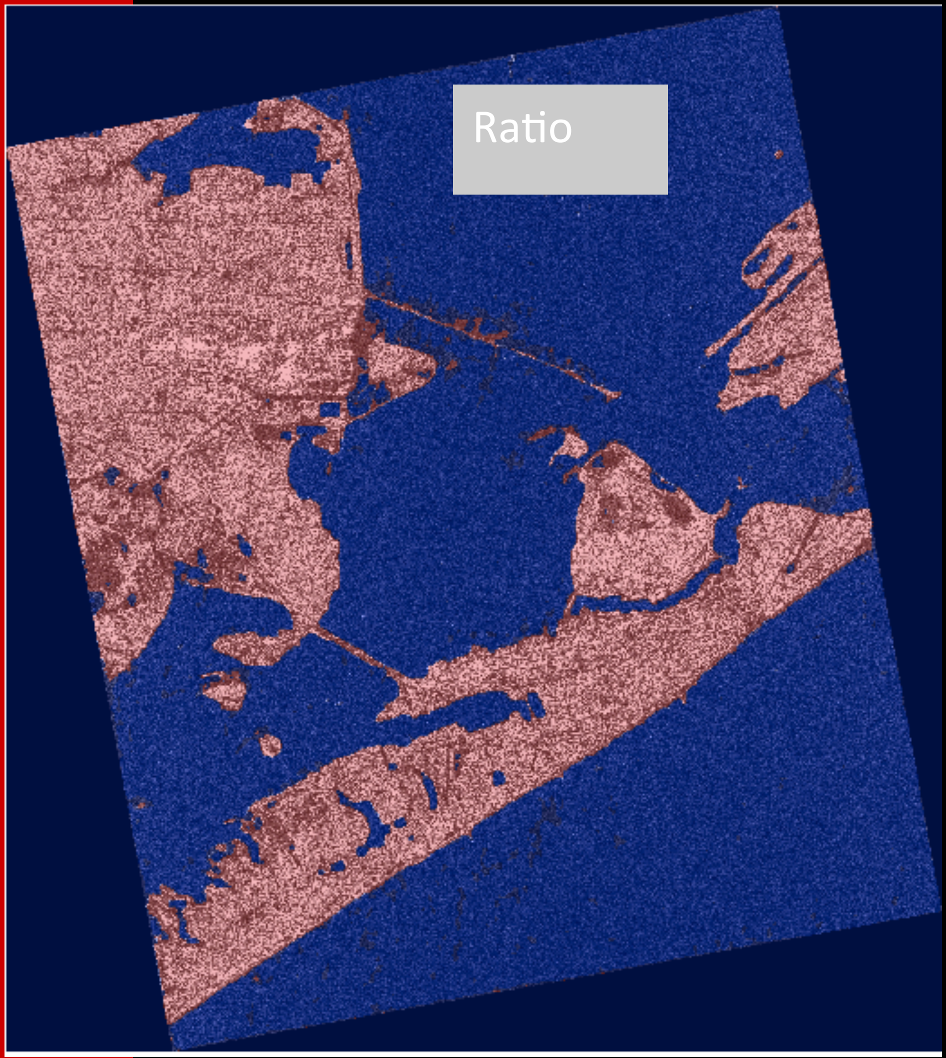


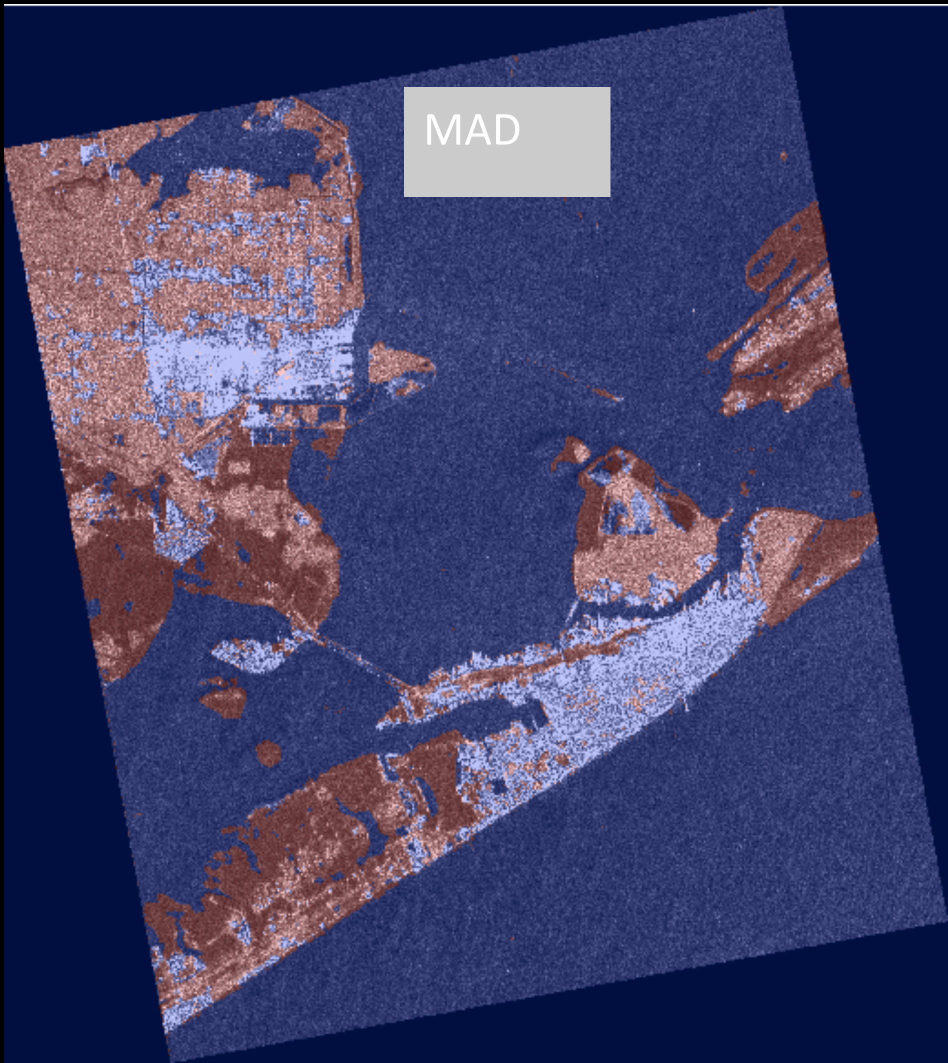


NC STATE UNIVERSITY

Results

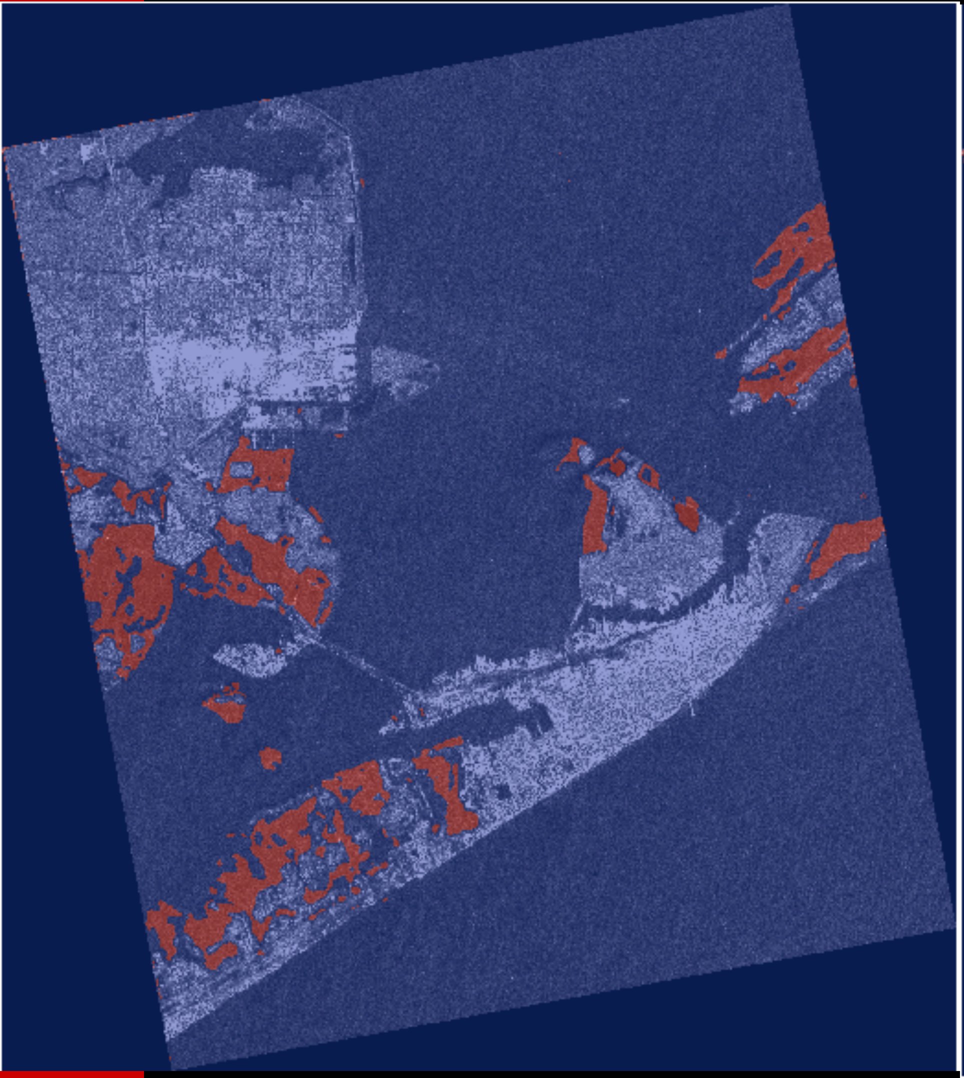
- Off-the-shelf techniques predict almost every pixel as change

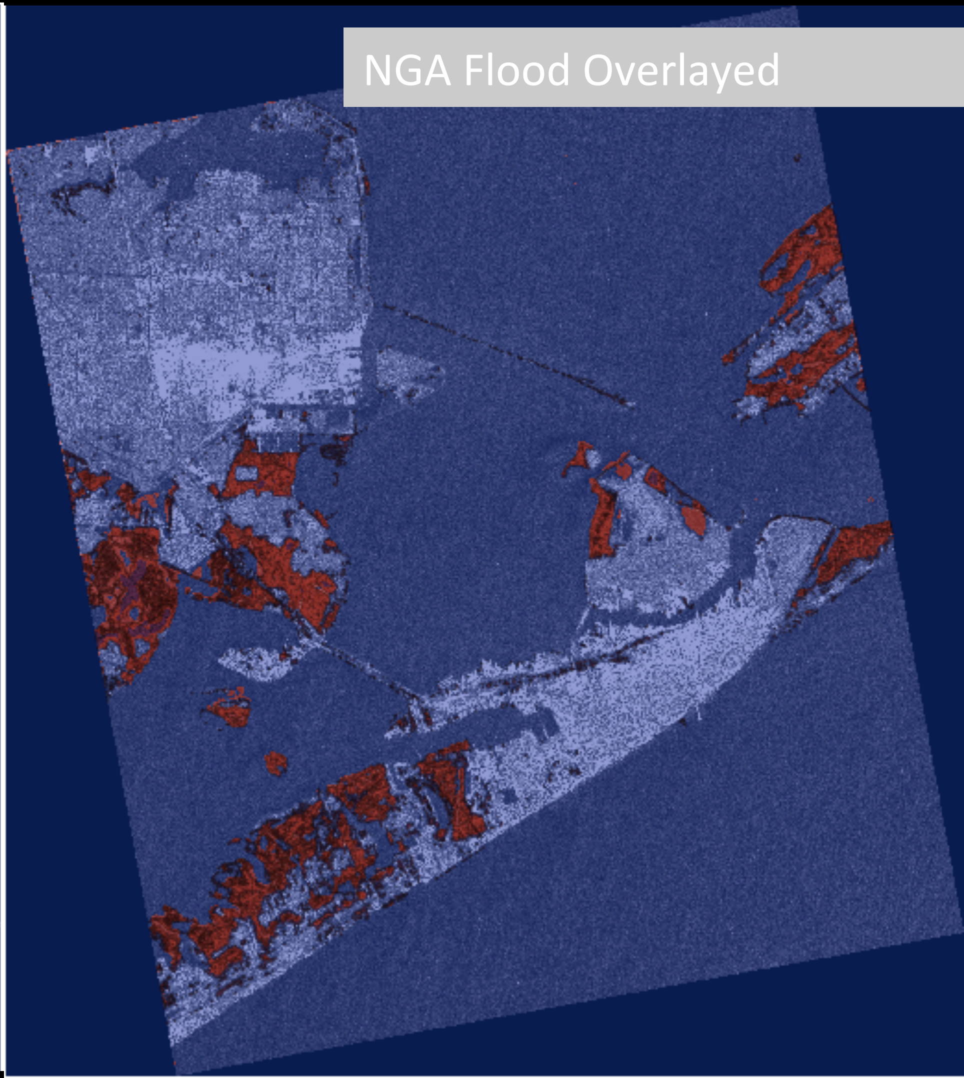




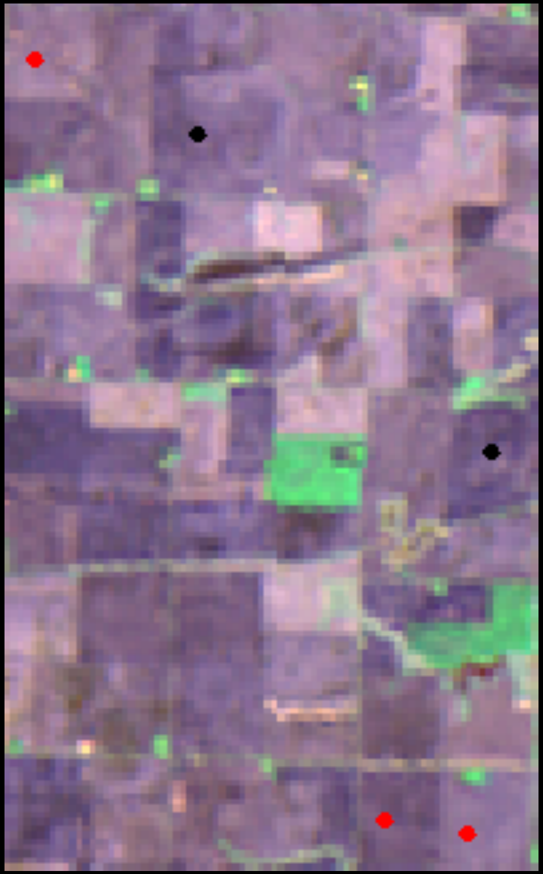
Results

- Probabilistic Approach

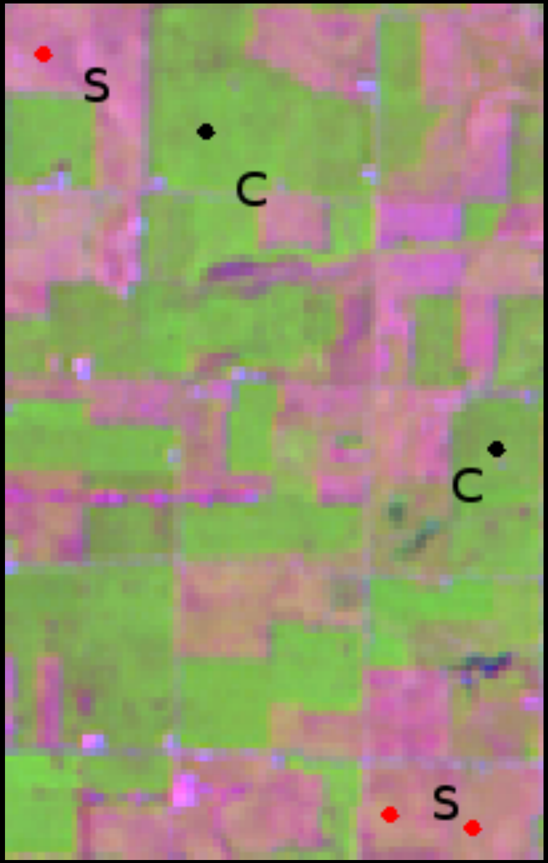




Multi-temporal Classification



AWiFS (May 3, 2008;
FCC (4,3,2))



AWiFS (July 14, 2008;
FCC (4,3,2))

Thematic Classes: C-Corn, S-Soy

Multi-temporal Classification

	corn	soy	alfa	grass	water	dvlpd	forest	wetlnd
corn	0.00	957.98	2000.00	1999.98	2000	1999.45	1859.75	2000
soy	957.98	0.00	2000.00	2000.00	2000	2000.00	1999.11	2000
alfa	2000.00	2000.00	0.00	2000.00	2000	1998.70	1999.89	2000
grass	1999.98	2000.00	2000.00	0.00	2000	1790.64	1973.95	2000
water	2000.00	2000.00	2000.00	2000.00	0.00	2000.00	2000.00	2000
dvlpd	1999.45	2000.00	1998.70	1790.64	2000	0.00	1817.02	2000
forest	1859.75	1999.11	1999.89	1973.95	2000	1817.02	0.00	2000
wetlnd	2000.00	2000.00	2000.00	2000.00	2000	2000.00	2000.00	0.00

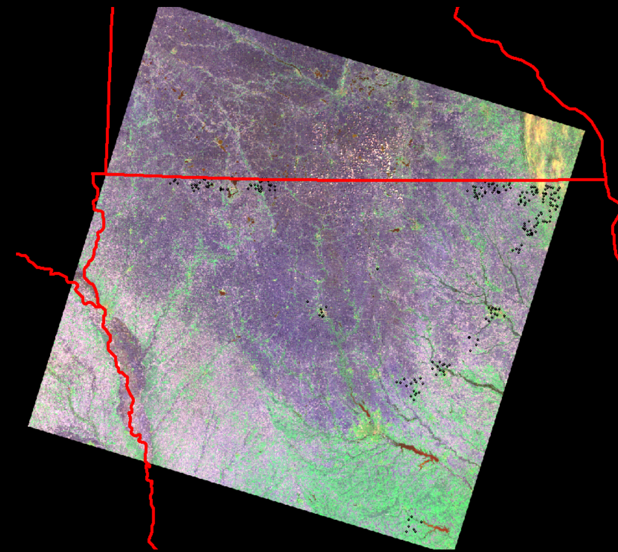
TABLE 6. Transformed Divergence Between Classes from May Image

	corn	soy	alfa	grass	water	dvlpd	forest	wetlnd
corn	0.00	1610.59	2000	927.95	2000	2000.00	1993.94	1999.65
soy	1610.59	0.00	2000	1252.87	2000	1997.30	2000.00	2000.00
alfa	2000.00	2000.00	0.00	2000.00	2000	2000.00	2000.00	2000.00
grass	927.95	1252.87	2000	0.00	2000	1992.04	1999.50	1999.76
water	2000.00	2000.00	2000	2000.00	0.00	2000.00	2000.00	2000.00
dvlpd	2000.00	1997.30	2000	1992.04	2000	0.00	2000.00	1999.31
forest	1993.94	2000.00	2000	1999.50	2000	2000.00	0.00	1734.34
wetlnd	1999.65	2000.00	2000	1999.76	2000	1999.31	1734.34	0.00

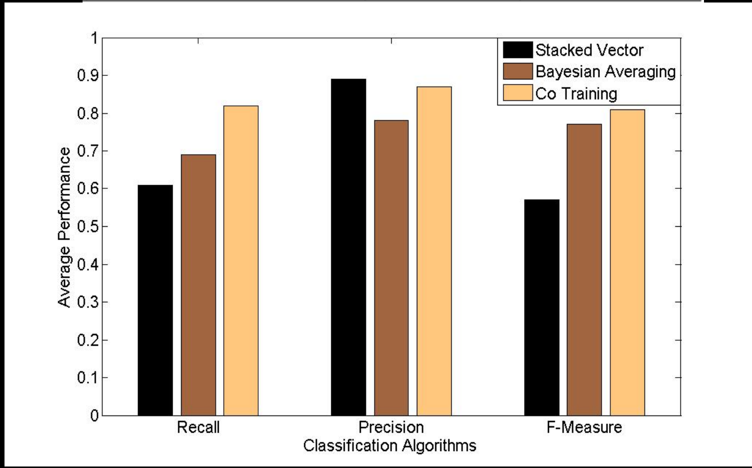
TABLE 7. Transformed Divergence Between Classes from July Image

Multi-view Approach

- Multi-temporal images are different views of same phenomena
  - Learn single classifier on different views, chose the best one through empirical evaluation
  - Combine different views into a single view, train classifier on single combined view – stacked vector approach
  - Learn classifier on single view and combine predictions of individual classifiers – multiple classifier systems
    - Bayesian Model Averaging
  - Co-training
    - Learn a classifier independently on each view
    - Use predictions of each classifier on unlabeled data instances to augment training dataset for other classifier



Class	Training	Validation
Corn	261	261
Soybean	225	225
Alfa alfa	27	27
Grass	189	180
Water	18	18
Developed	90	99
Deciduous Forest	117	117
Wetlands Forest	18	36
Total:	945	963



Varun Chandola, Ranga Raju Vatsavai: Multi-temporal remote sensing image classification - A multi-view approach. CIDU 2010: 258-270

Conclusions and Outlook

- Remote Sensing is a Key Resource in Understanding Food, Water, and Energy Resource Monitoring
- Continuous Monitoring
  - Full automation is still a challenge
  - Multi-*: sensor, resolution, temporal
- Mining for Interesting Patterns
  - Automated Event Generation
- Modeling Spatial and Temporal Relationships
- Computational Challenges
  - $O(N^3)$
  - Approximate solutions
  - Exploitation of true heterogeneity of modern compute node

NC STATE UNIVERSITY

Acknowledgements

- Initial work was carried out at the Oak Ridge National Laboratory
- Collaborators: V. Chandola, B. Bhaduri, J. Grasser, A. Cheriyadat, and E. Bright