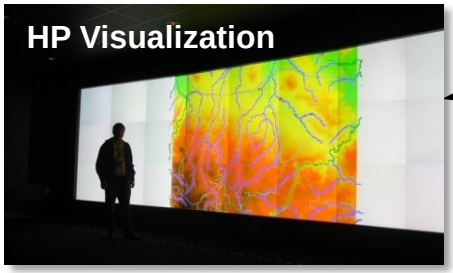
Geospatial Data Sciences

Presented by

Ranga Raju Vatsavai
Budhendra Bhaduri

Geographic Information Science and Technology
Computational Sciences and Engineering

Geospatial data sciences



Knowledge Discovery

- Data mining
- Spatial analysis
- Machine learning
- Pattern recognition
- Statistics
- Nonlinear dynamics

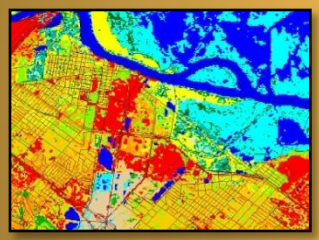
- Disaster mapping and monitoring
- Geospatial intelligence discovery
- Actionable insights/ knowledge
- Extremes
- Social networks
- National security

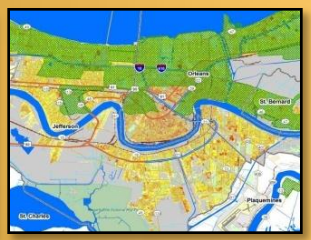
Efficient storage and retrieval methods

Terabytes of spatially referenced data per day

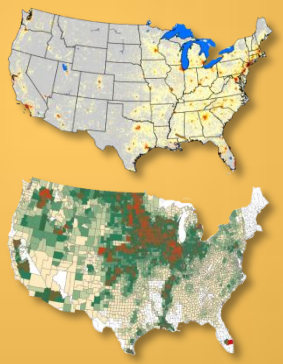
Errors and uncertainty

GeoDatabases





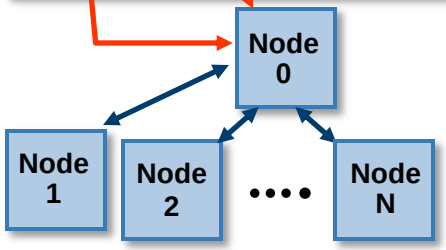
Petabytes of spatiotemporal data archives



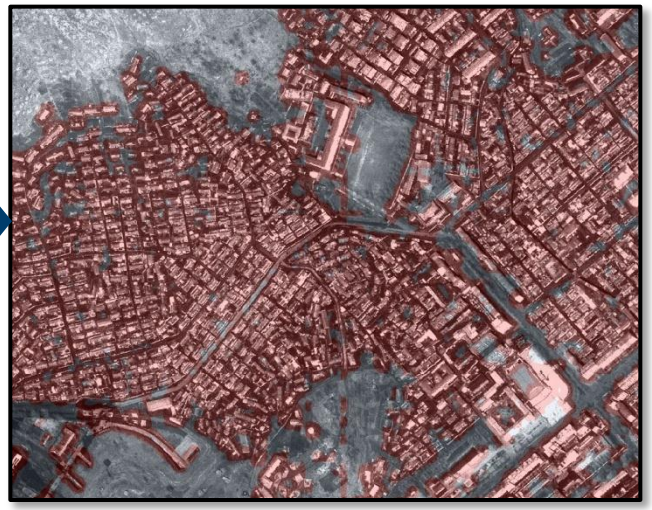
Damage	Total population	Senior
100%	791,361	90,773
75%	1,594,806	187,677
50%	1,505,196	184,372
25%	1,701,593	224,279
Total	5,592,956	687,101

Urban mapping for global population distribution

Identify urban regions from high-resolution satellite imagery by using Gabor texture analysis



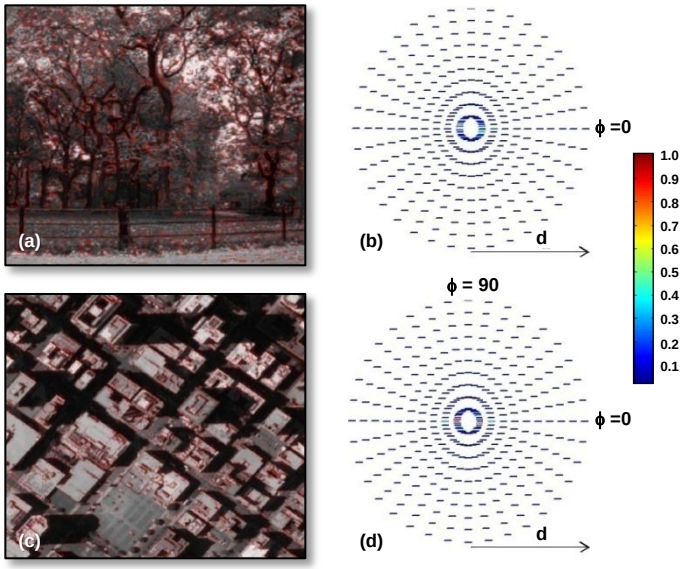
- Convert to Gabor space
- Compute energy variance inside sliding window size W
- Perform iterative data clustering

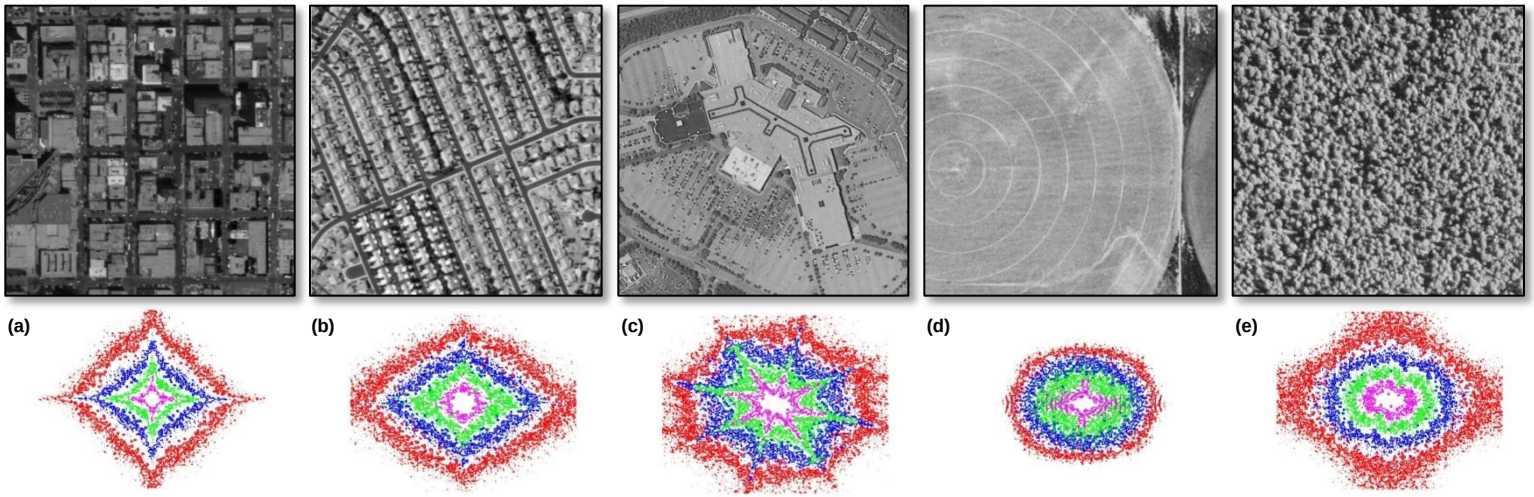


	Area (sq km)	Imagery data	Intermediate data created	Time 10 nodes
Sample area	114	219 MB	~3.9 GB	~30 min
Projected computing requirements				
Global populated land area	56,863,754	~103 TB	~1.8 PB	~30 years

Statistical analysis for satellite image characterization

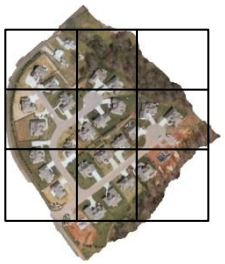
- Power spectrum analysis
- Image gradient distribution
- Wavelet analysis





Challenge is to extend automated processing to fine-resolution images over large geographic regions

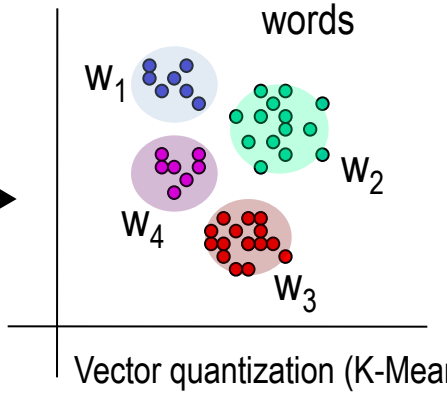
Semantic classification



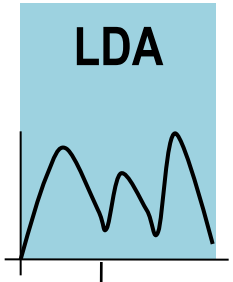
Examples for each semantic category

Features

Visual words

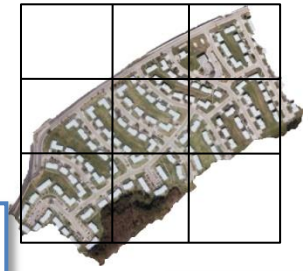


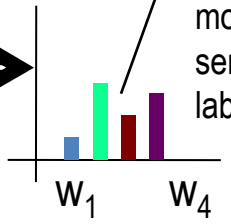
Vector quantization (K-Means)

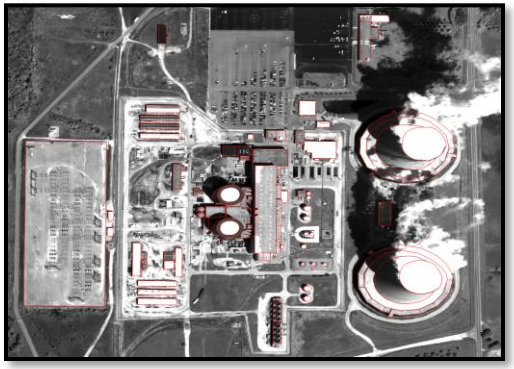


Predict most likely semantic label

New sample







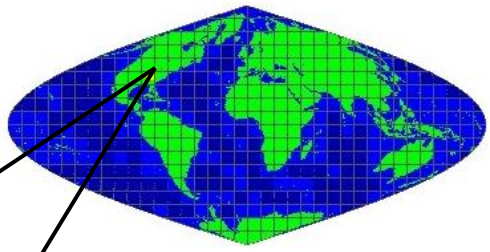
Challenges

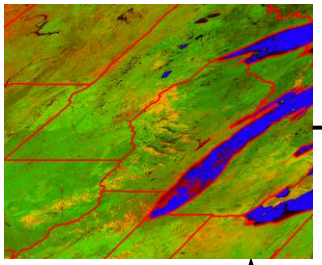
- What are the right features?
- What are right objects?
- Which spatial relationships?
- Bottom-up vs. top-down?

Why Semantics?

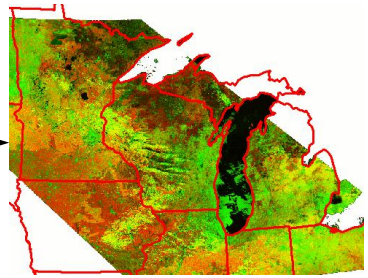
Set of objects like "switch yard," "containment building," "turbine generator," "cooling towers" AND their spatial arrangement may imply a semantic label like "nuclear power plant"

Biomass monitoring framework

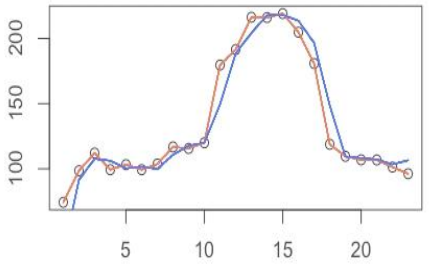




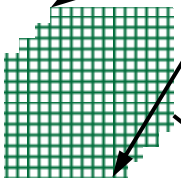
ISIN Projection



UTM Projection



Filter Each Pixel



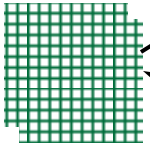
MODIS (4800x4800)
3 bands, 250 m, 8 days
2000–2009
H11V04, MOD09Q1
(LP DACC)
27 GB; 432 products

FTP-Pull

Image and Ancillary Data

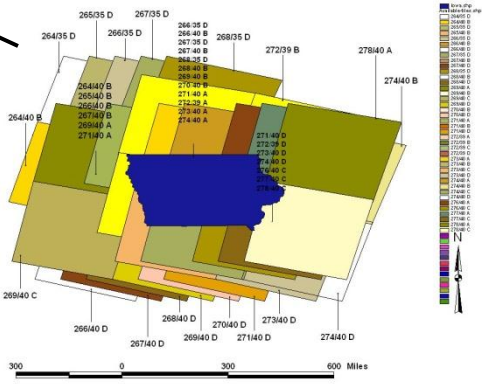
Pre-processing
- Reprojection
- Atmospheric
- Filtering

Change detection
- Time series based
  - Time series prediction
- Multidimensional image based
  - Unsupervised clustering



DVDs

AWiFS
(12,300x12,000)
4 bands, 56 m
May–Sep 2008
Iowa (USDA)
130 products



GE Visualization

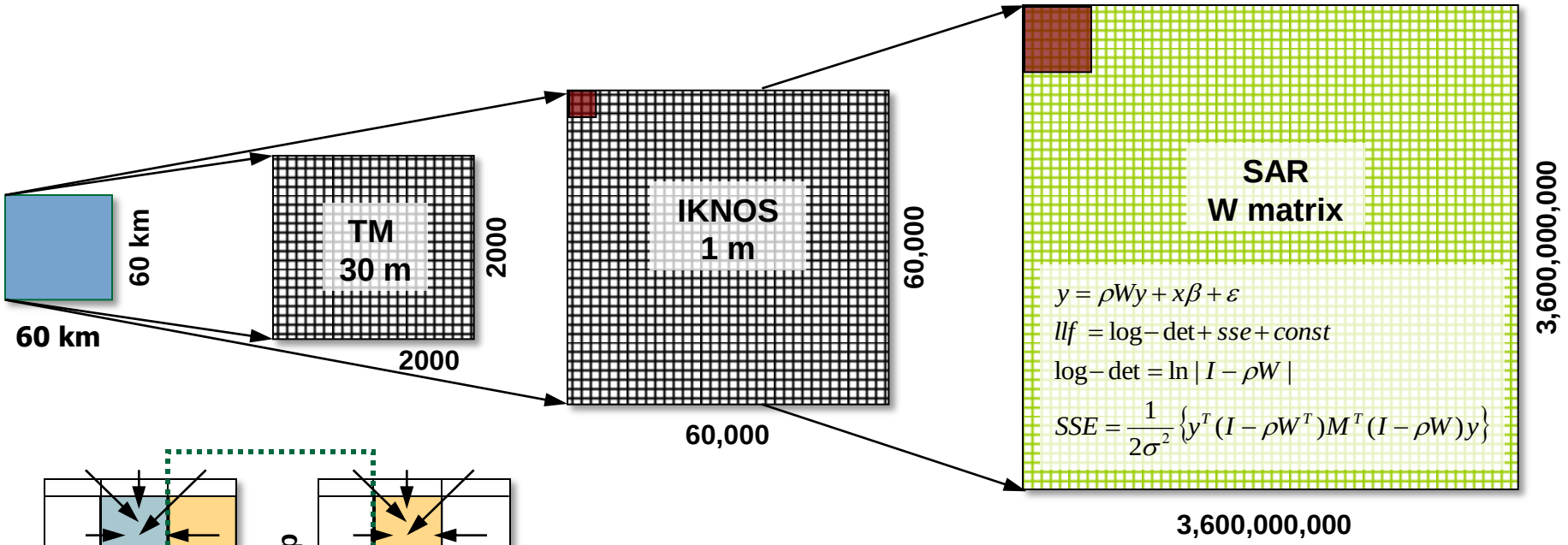
Characterize changes
- Phenology based
- Type based

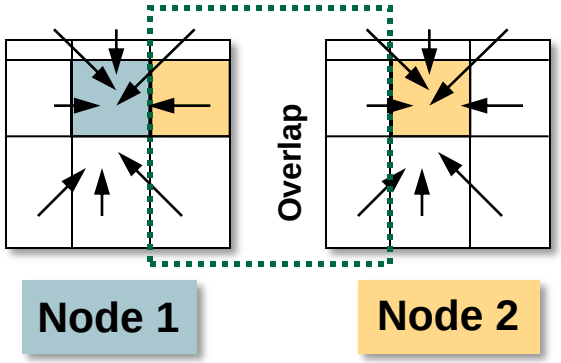
Challenges

- Daily Data, needs to be processed before new images arrive
- Computationally expensive algorithms

Spatial classification and prediction

- Spatial autoregressive regression (SAR)
- Markov random fields (MRF)





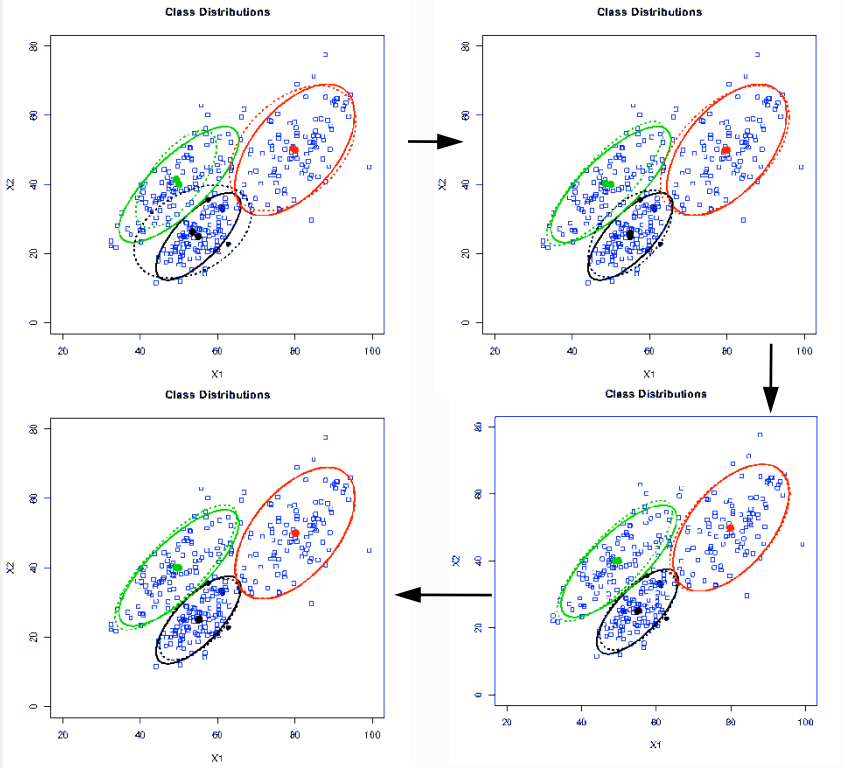
Challenges

- Spatial homogeneity
- Large neighborhood matrices
- Overlapping computation

Large geographic area classification

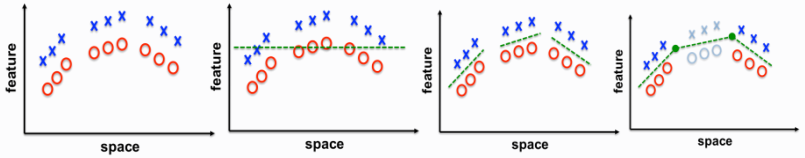
Semi-supervised Learning

Insufficient Ground-truth



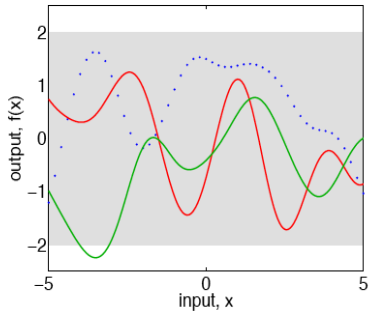
Gaussian Process (GP) Learning

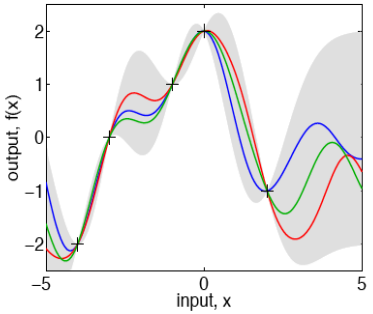
Spatial Heterogeneity

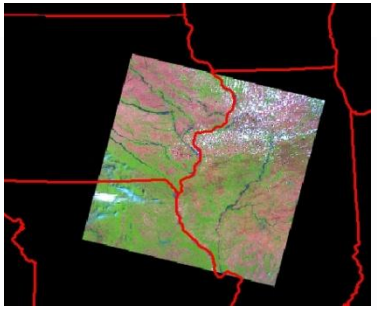


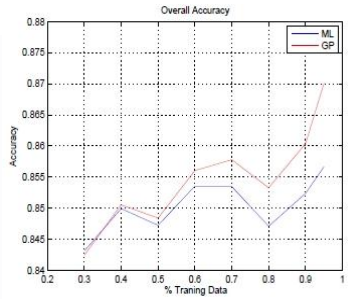
Change of distribution of over space is modeled by

$$p(x \mid y) \sim N(\mu, \Sigma) \qquad p(x(s) \mid y) \sim N(\mu(s), \Sigma(s))$$





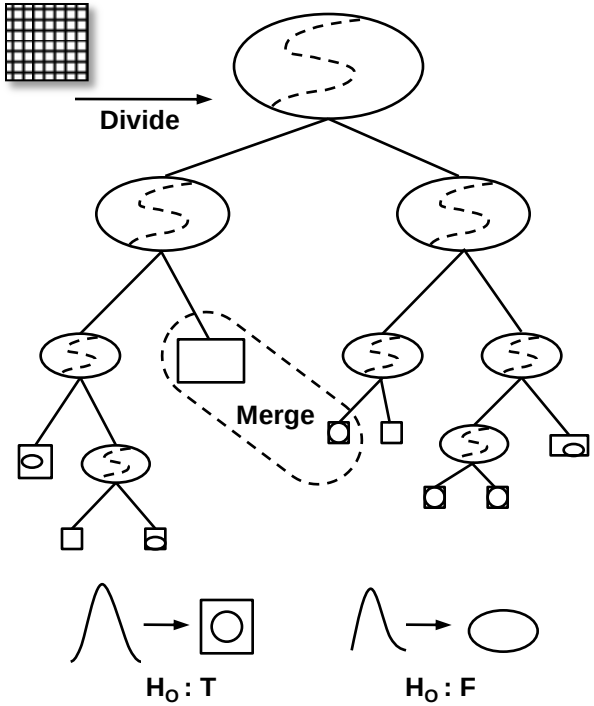




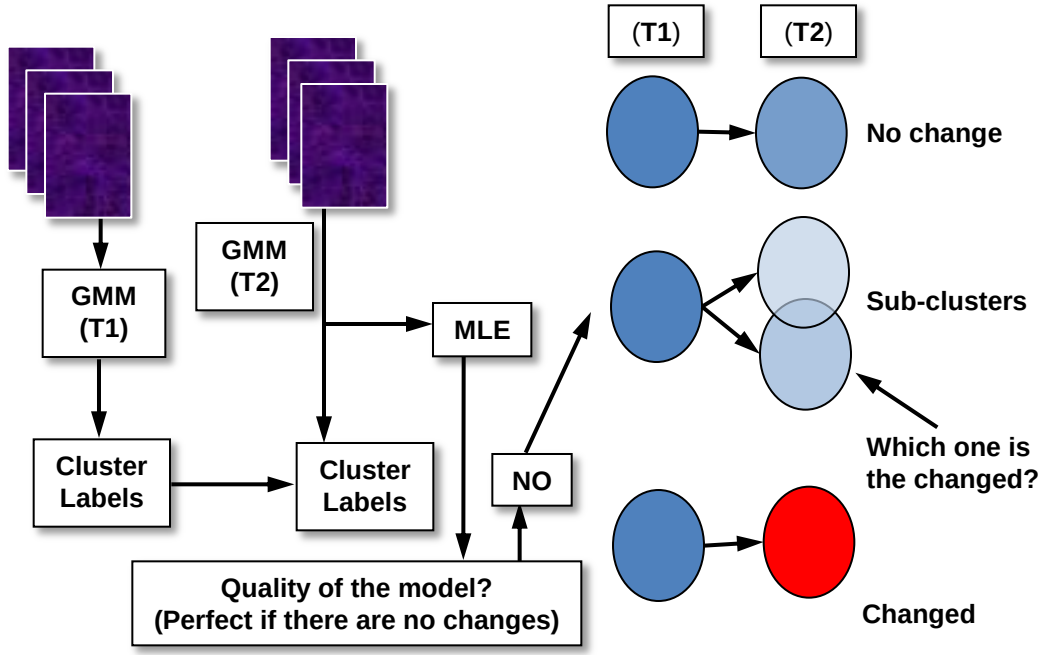
Challenges

- **Large extents, insufficient ground-truth**
- **Spatial heterogeneity**

Unsupervised change detection

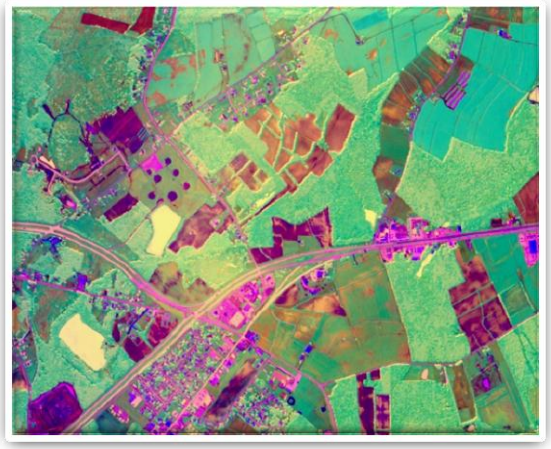


Test for pair-wise KL-Divergence
Merge if two clusters are close

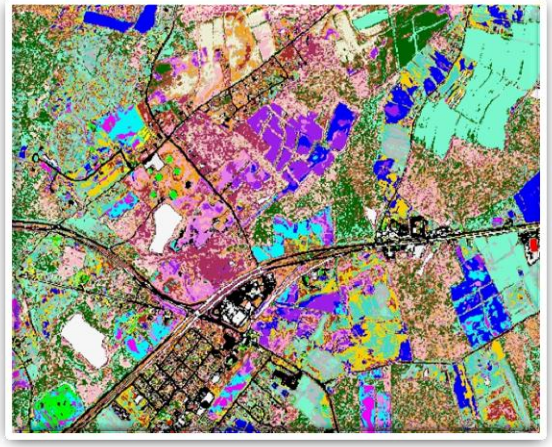


Challenges

- How many clusters?
- Which clusters changed?
- Iterative and computationally expensive algorithms

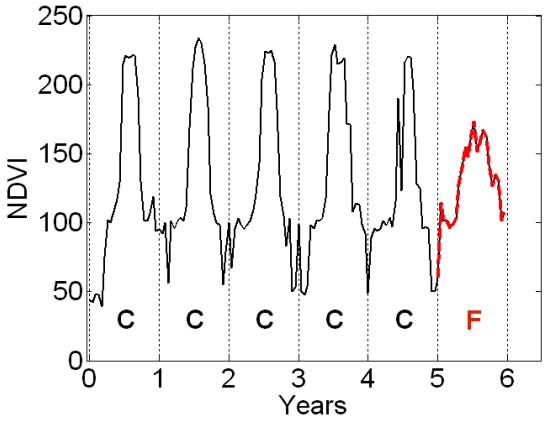


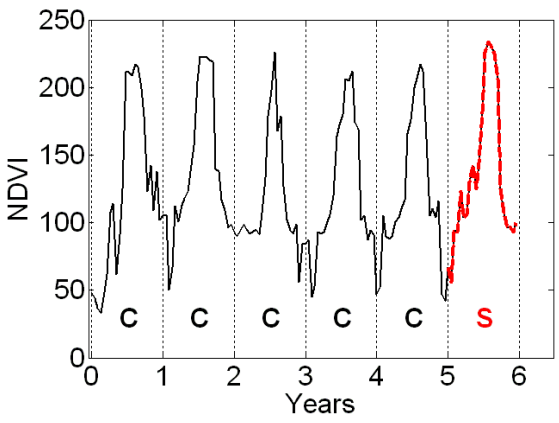
AVRISS (224 Spectral Bands)

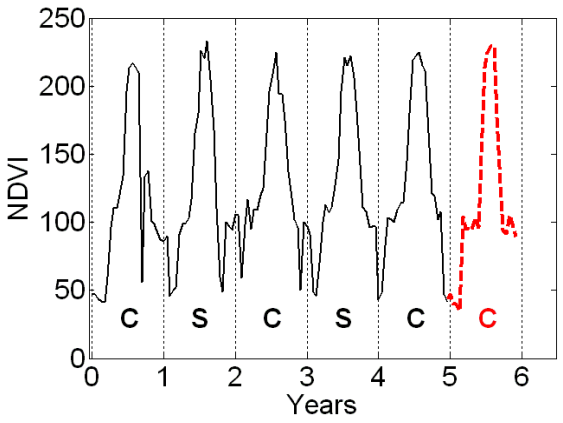


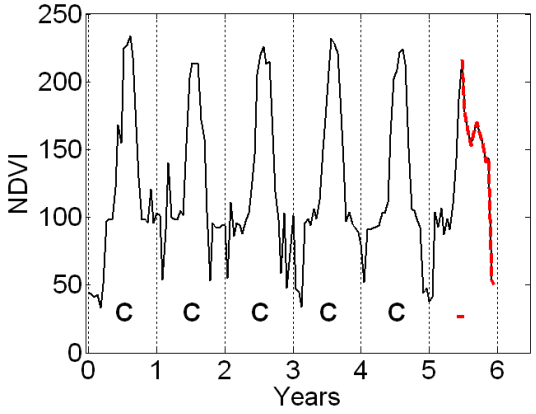
GX-Means Clustering

Online change detection using multitemporal remote sensing images









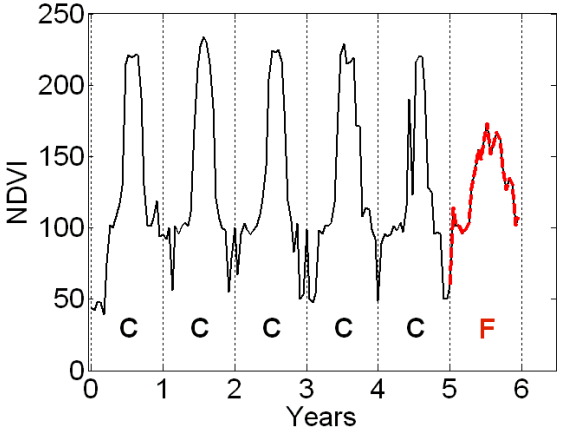
C - Corn
S - Soy
F - Fallow

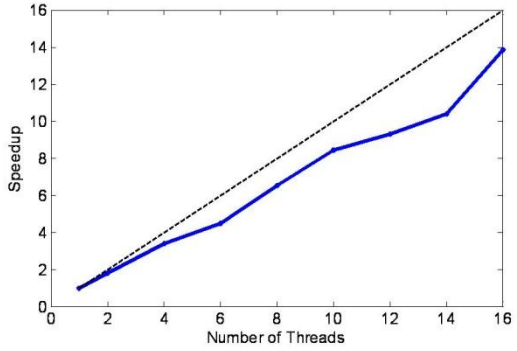
Challenges

- Existing approaches cannot deal with all types of changes
- Existing approaches are mostly applicable to static data
- Missing and noisy data
- Not scalable to massive streaming data

Parallel approaches

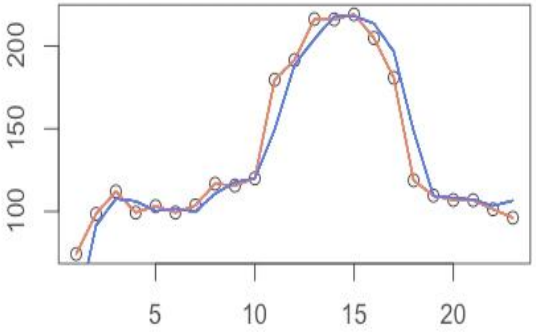
Change detection on multi-core machines





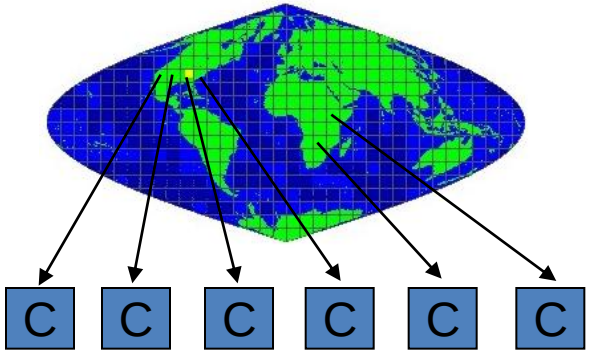
14x speedups on 16 virtual cores

Filtering and NDVI on many-core machines



~20x speedups

Knowledge discovery on Clouds



Highly scalable data intensive computing

Challenges

- **Distributed and heterogeneous data sources**
- **Heterogeneous computing architectures**
  - **Multi-core**
  - **Many-core**
  - **Distributed**

Contacts

Ranga Raju Vatsavai

Geographic Information Science and Technology
Computational Sciences and Engineering
(865) 576-3569
vatsavairr@ornl.gov

Budhendra Bhaduri

Geographic Information Science and Technology
Computational Sciences and Engineering
(865) 241-9272
bhaduribl@ornl.gov