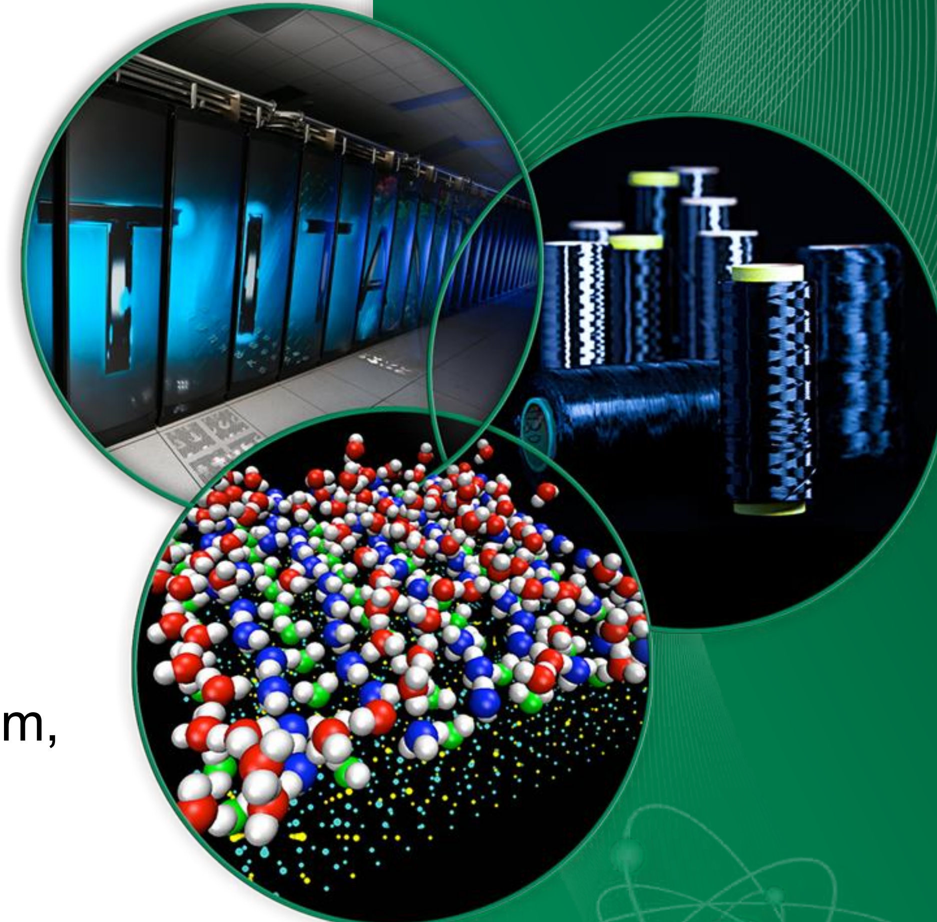
DOE Bioenergy Technologies Office (BETO) 2015 Project Peer Review



4.1.2.40 Land-Use Change Data Analysis

03/25/2015

Analysis & Sustainability

Nagendra Singh (PI)
Keith Kline, Rebecca Efroymson, Raju Vatsavai, Huina Mao, Erica Pham, Budhendra Bhaduri.

Oak Ridge National Laboratory

Goal Statement

Project Goal

Design and develop scalable tools and assessment methods to establish scientific basis for understanding and simulating effects of bioenergy policy on land cover and management.

DOE Goals

Strategic Analysis goals:

- Ensure high-quality, consistent, reproducible, peer-reviewed analyses Develop and maintain analytical tools, models, methods, and datasets to advance the understanding of bioenergy and its related impacts

Sustainability goal:

- Understand and promote the positive economic, social, and environmental effects and reduce the potential negative impacts of bioenergy production activities.

Quad Chart Overview

Timeline

- Project start date: FY10
- Project end date: FY17
- Percent complete: 60 %

Budget

	Total Costs FY 10 –FY 12	FY 13 Costs	FY 14 Costs	Total Planned Funding (FY 15-Project End Date
DOE Funded	274K	269K	247K	600K
Project Cost Share (Comp.)*	0	0	0	0

Barriers

- At-A. Transparent, and Reproducible Analysis
- At-C. Data Availability across the Supply Chain
- St-C. Sustainability Data Across the Supply Chain
- St-G. Land-Use and Innovative Landscape Design

Partners

Interactions/collaborations

- USGS, USDA
- Inputs from other DOE labs/ Universities

Project Overview

Context

Concern: bioenergy policy leads to effects (conversion of forestland; displacement of food production) with significant social and environmental (e.g., GHG emissions) consequences.

Objective

Develop innovative and science-based approaches to estimate changes in land use pattern using data mining and machine learning techniques on satellite data.

History- Two parts :

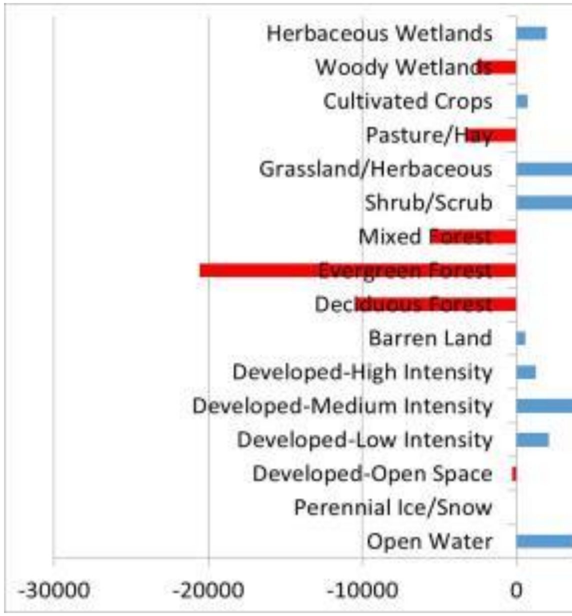
A. Consistent and reliable data
- Assess existing data & models – resolution, classes, frequency,
  - Gaps & uncertainties in data & models
- Develop tools and reliable data to accurately assess:
  - Changes and trends in vegetation patterns

B. Causal Analysis (concluded in FY14)
- Understand drivers of land use change
- Framework for causal analysis

2 – Approach (Technical)

Paradigm has to shift from interpolating land use change in the past using derived data products to monitor and measure in near-real time.

- Consistent and reliable data
- Monitor & measure
- Attribution of effects among causal drivers
- Consensus around optimal units for measuring change



Changes based on NLCD 2006 & 2011

Critical Success Factors

Availability of location-specific, on-demand, earth-observation data gives us capability to monitor changes in near real time.

Focus on providing information on trends in productivity, yield, and vegetation patterns.

Develop computationally efficient and scientifically sound algorithms to mine these massive amounts of data to extract relevant information.

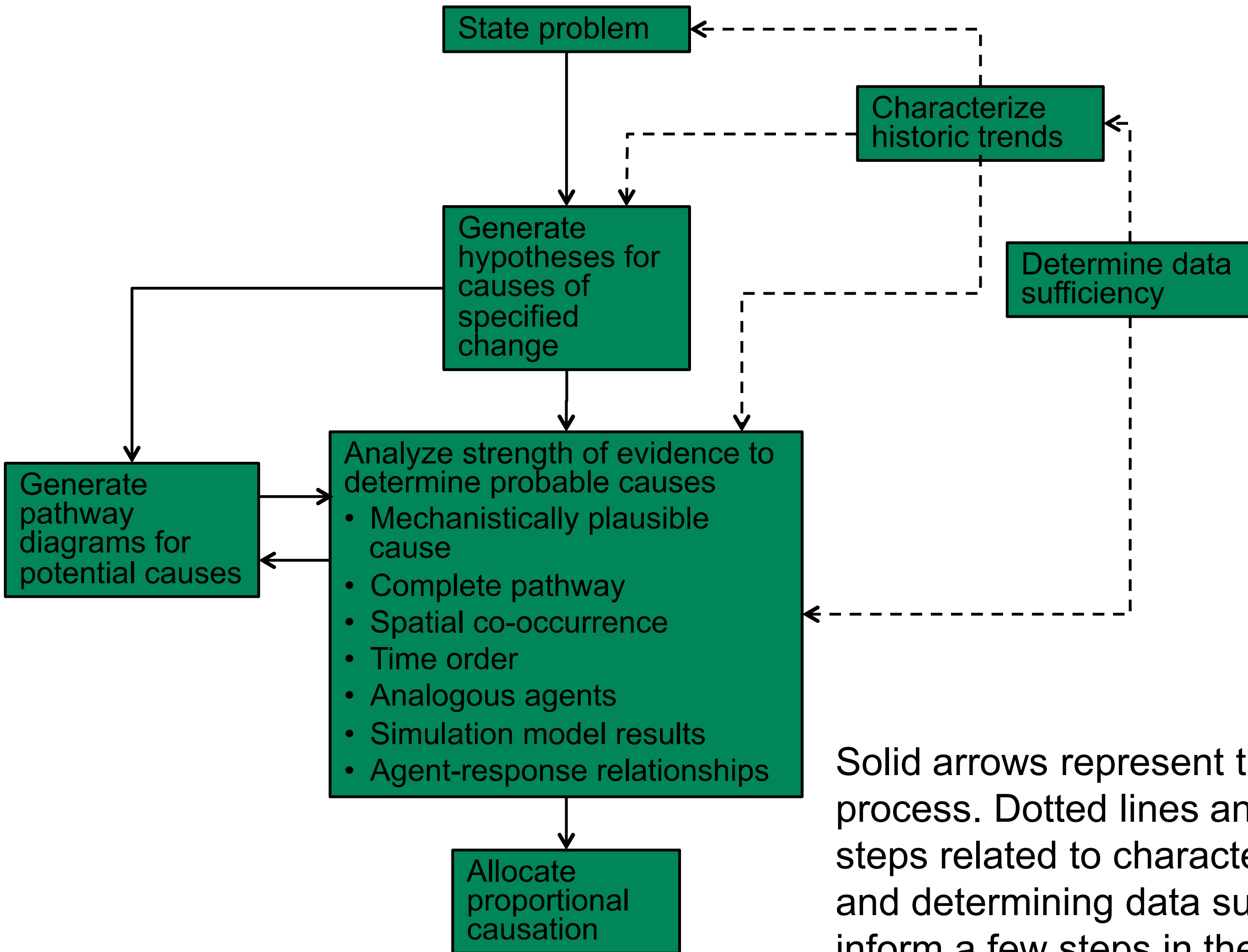
Disseminate information to stakeholders in easily accessible and timely manner.

2 – Approach (Management)

- Biweekly project meetings with team members
  - Assess project progress, discuss issues, analyze results
  - Plan future steps and tasks
  - Discuss issues (computing, data) and deviations from desired outcome
    - Challenge on balancing between speed versus accuracy
    - Tracking the state of art in computing, data analytics, machine learning
- Integration between data and causal analysis task.
- Collaboration with international land-use change community for inputs and coauthor publications.
- Time management can be tricky with quarterly deliverables
  - Balance time between deliverables and overall objectives.
- Conversation with other researchers for collaborations.
- Quarterly progress reports and discussions with BETO managers; monthly written updates.

3 – Technical Accomplishments/ Progress/Results

Development of Causal Analysis Framework



Solid arrows represent the order of steps in the process. Dotted lines and arrows indicate that steps related to characterizing historic trends and determining data sufficiency are required to inform a few steps in the process.

* Causal Analysis Task

3 – Technical Accomplishments/ Progress/Results

Types of evidence in strength-of-analysis approach

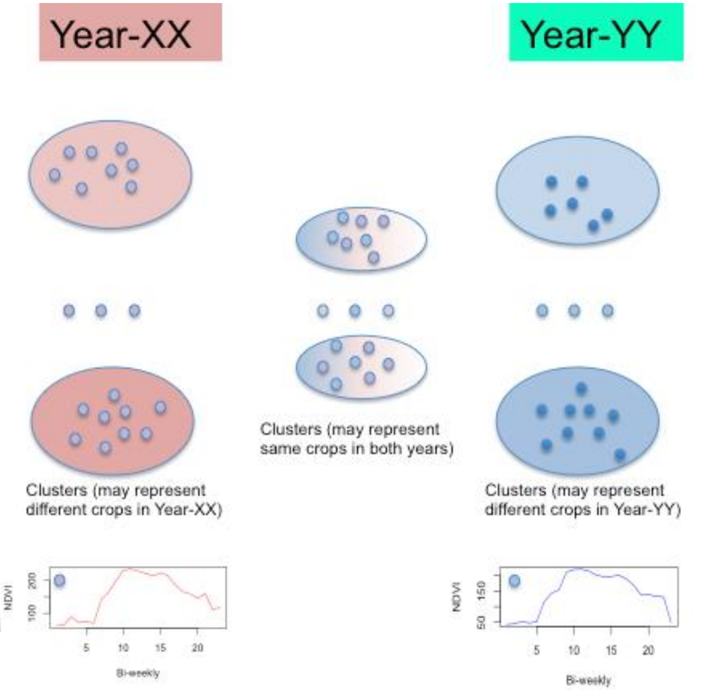
Evidence	Question
Plausibility	Is there a reasonable mechanism (based on science) to relationship between bioenergy policy and deforestation?
Complete pathway	Is there a complete pathway from bioenergy policy to deforestation, or is part of pathway blocked? (A→B→C)
Spatial co-occurrence	Is deforestation occurring where bioenergy policy was implemented or biofuel crops were grown?
Time order	When did deforestation (or change in rate) occur, relative to bioenergy policy?
Analogous agents	Is the hypothesized relationship similar to other cases involving bioenergy or related drivers?
Simulation model results	Do simulation model results support or contradict the hypothesized relationship? Was model validated/verified?
Agent-response relationship	Is there a quantitative relationship between cause or intermediary agent in causal chain and deforestation?

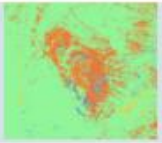
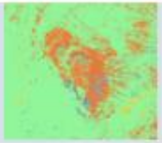
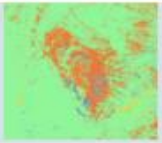
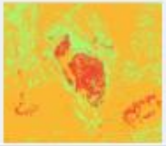
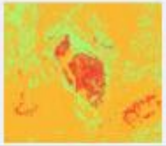
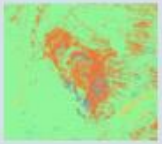
* Causal Analysis Task

3 – Technical Accomplishments/ Progress/Results

Development of new Time Series Clustering method

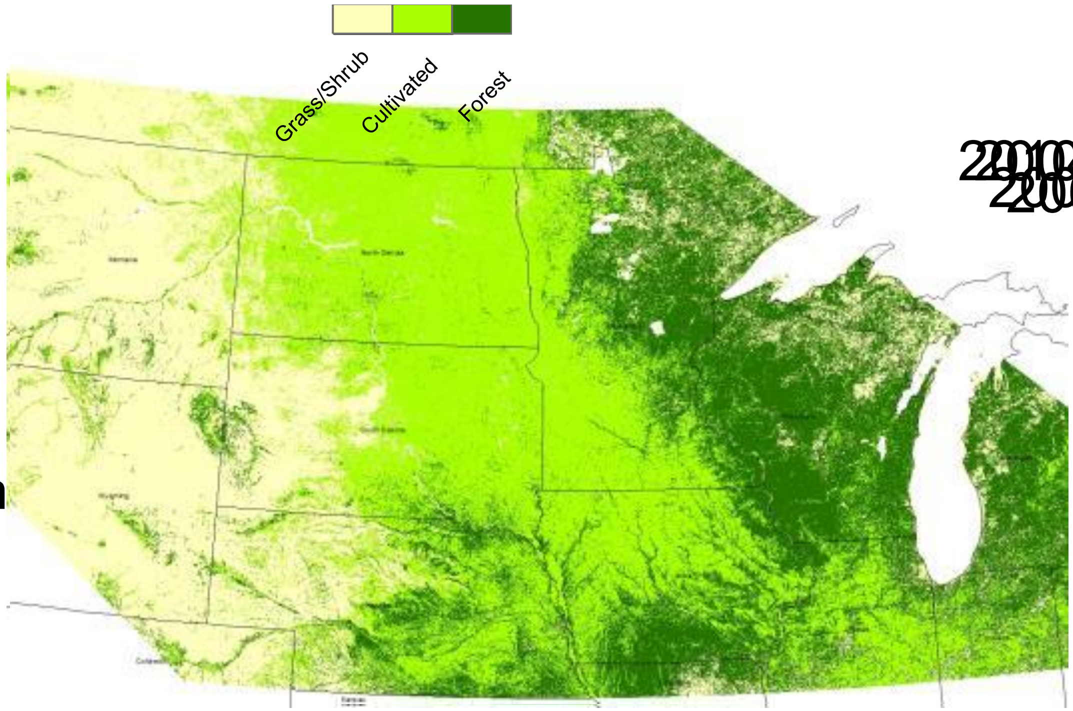
- Existing methods > Cluster Each Year>Typically K-Means
- New Method
  - Given two time series (Y1 and Y2): How similar they are?
  - Group time series into clusters in a hierarchical manner
  - Use the composite cluster model to predict labels for both images
  - Computationally efficient, can be run on a desktop (tera-bytes of data)



	Y1	Y2	Y3	Y4
Y1				
Y2				
Y3				

3 – Technical Accomplishments/ Progress/Results

- Produces realistic changes by taking into account the crop growing patterns.
- Provides fast estimates of regional changes in vegetation pattern (fire, climate, disease etc.)
- High temporal resolution (upto16 days) and continuous time series
- Computational framework to quickly analyze massive amounts of data



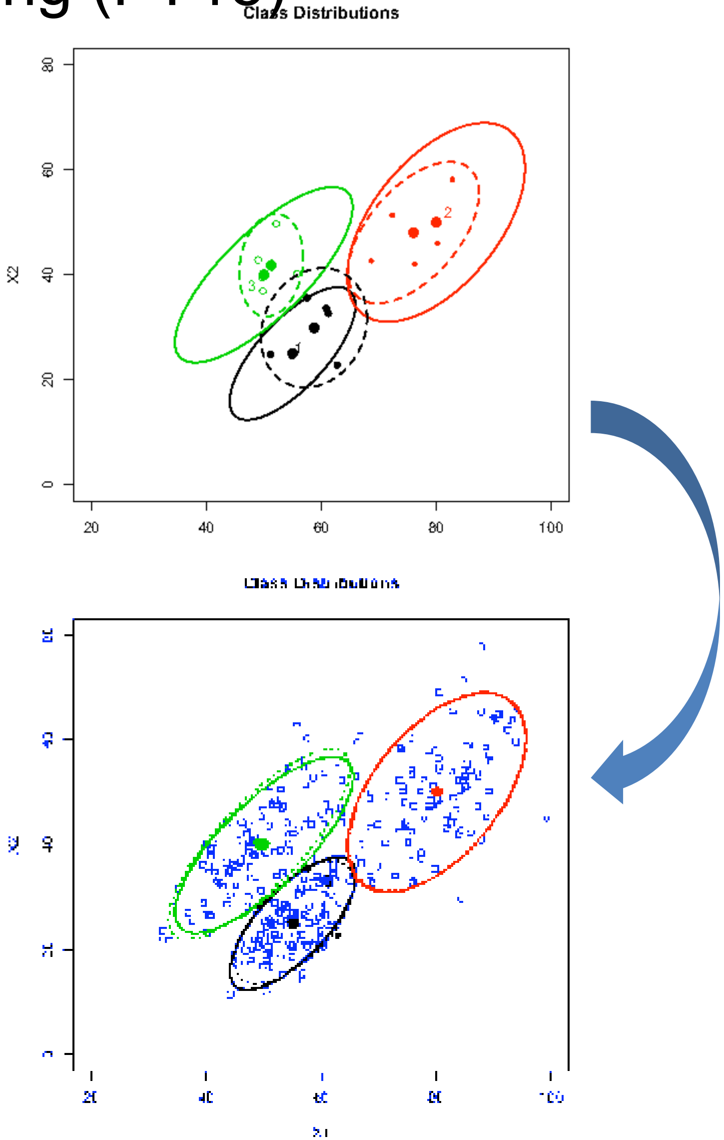
Challenges

- Significant post-processing
- Clusters have to be manually labelled
- Getting “sufficient” Ground-truth is difficult

3 – Technical Accomplishments/ Progress/Results

Development of Semi-Supervised Learning (FY15)

- Approach
  - Combine labeled (small ground-truth samples) and unlabeled (large samples from image) to build model.
  - Improve model parameter estimation (similar to as if all the samples are labeled).
- Utility
  - Can deal with small samples and uncertainties.
  - Improves accuracy in classification tasks.



3 – Technical Accomplishments/ Progress/Results

Semi-supervised method that admits both labeled (ground truth) and unlabeled (from image) time series for constructing the model (FY15).

4. Relevance

Project consistent with DOE- BETO Objectives

- ✓ Consistent and reliable data ensures that sustainability and analysis results are not biased by data selection.
- ✓ Algorithms and tools based on sound scientific principle which can mine large volumes of data to understand impacts of bioenergy on environment.
- ✓ Better land use management through real time monitoring and visualization and stakeholder input.

Project addresses challenges & barriers

- ✓ Methodology developed ensures ‘comparable transparent and reproducible analysis’ through use of consistent and unbiased data.
- ✓ Data produced at required temporal and spatial resolution for measuring ‘sustainability at landscape scale’. Approach is scalable over time and space
- ✓ The approach ‘captures the dynamic state of land use and management’ by analyzing data at two week intervals over long time periods.

5 – Future Work

Refine, calibrate and rigorously test the classification algorithm to accurately estimate changes in vegetation pattern over a variety of regions.

Develop a system for monitoring and characterization of changes in vegetation using satellite data in near-real time to:

- Visualize changes over any time period over any region of the globe.
- Understand real time impact of natural/anthropogenic events (fire, disease, hurricane etc.) on bioenergy resources.
- Understand impact of climate change on bioenergy by integrating with future climate scenarios.

Integrate monitoring system with KDF for online usage and analysis.

Summary

Approach

– Apply innovative machine learning techniques to satellite data to understand changes in biomass in near real time.
– User defined model to monitor and track changes.

Technical accomplishments

– Causal Analysis framework for land use change.
– Development of change detection system for analyzing massive amounts of data.
– Development of semi-supervised model to minimize errors and noise.

Relevance

– Approach removes bias and uncertainty associated with derived data products.
– Enables analysis over any geographic area over any range of time.
– Allows to ‘measure’ and ‘monitor’ trends in vegetation pattern at scale.

Future Work

– Refine , optimize, calibrate and test the semi-supervised modelling framework.
– Develop a system for monitoring and characterization of changes in near real time.
– Understand impact of climate-change, population growth on bioenergy.

Additional Slides

(Not a template slide – for information purposes only)

- *The following slides are to be included in your submission for Peer Evaluation purposes, but will **not** be part of your oral presentation –*
- *You may refer to them during the Q&A period if they are helpful to you in explaining certain points.*

Responses to Previous Reviewers' Comments

Because of the challenges and the potentially small spatial scale of the bioenergy effect, reviewer commented that "it is unclear whether current methods will be able to yield useful results." – We have successfully demonstrated the utility of machine learning and data mining techniques to understand changes in vegetation patterns. Based on suggestions we have moved to a semi-supervised algorithm where the model maps changes based on user can provide input

The relevance of the causal analyses approach: Given that current policy in the USA (EPA and CARB) and EU policy (affecting US exports), as well as general public opinion, coincide in assigning significant penalties to biofuels for indirect land-use change (ILUC), this work is essential. If the scientific and political communities agreed with reviewer's comments that ILUC caused by biofuels is "too small" to be of significance, then we would not be doing this research. However the causal analysis task ended in FY14 due to lack of continued funding.

A reviewer asked about the other side of the coin, "food security." -While the scope of this project was not focused on food security, the results are relevant. Understanding the effects of bioenergy policy on land use and productivity is critical for improved understanding of relationships to food security. The tools and algorithms we are developing can map what land covers are changing, when and where, and we can easily integrate this information with population mapping (which we work on extensively (both global 7 national) to answer questions about food security if mandated to do so.

Publications, Presentations, and Commercialization

Singh, N. & Bhaduri, B.L. *Suitability of Land Cover datasets for Feedstock Estimation.* Association of American Geographers (AAG) annual meeting, Las Vegas, NV.,2009

Singh,N. & Bhaduri B.L. - *The Effect Of Biofuel on Land Cover Change Using Multi-Year MODIS Land Cover Data*. 2010 IEEE International Geoscience and Remote Sensing Symposium. Honolulu, Hawaii.

Chaudhuri G, Bhaduri B.L., Singh N, & Clark K. 2011 *Expanding Bioenergy and Land Use Change: A Spatially Explicit Modeling Approach,* Association of American Geographers (AAG) annual meeting, Seattle, WA,2011

Chandola V., & Vatsavai., R.R., 2011, Scalable Gaussian Process Analysis Algorithm for Biomass Monitoring., Statistical Analysis and Data Mining . 4(4):430-445

Kline, Keith L., Oladosu, G.A., Dale, V.H., McBride, A., & Singh, N. Perspectives on Land-Use Change Analyses. CRC Workshop on Life Cycle Analysis Of Biofuels, ANL, Oct 2011.

Kline, Keith L., Dale, V.H., McBride, A., & Singh, N. "Top Ten Steps to Improve Quantification of Land-Use Change Effects of Bioenergy Systems." IEA Joint Workshop--Quantifying and managing land use effects of bioenergy, Campinas, Brazil, Sept 2011.

Singh, N. et al., Estimating Land Use Land Cover Change in Iowa due to Bioenergy, Association of American Geographers (AAG) annual meeting, New York, 2012

Kline, Keith L., et al. LUC dynamics and improving sustainability assessments: models, science and causal analysis. , Ecological Society of America Annual Meeting, Portland,2012.

Kline, Keith L., et al. ORNL Research: land-use change, global bioenergy crop models and indicators of sustainability. Brazilian Bioethanol Science and Technology Laboratory, Campinas, Brazil, Oct 2012.

Publications, Presentations, and Commercialization

Vatsavai R.R., Chandola, V.,, and Bhaduri, B. , *Large Scale Remote Sensing Data Mining for Biomass Monitoring: Recent Advances and Future Challenges*, Proceedings of 7th International Conference on Geographic Information Science (GIScience), 2012.

Kline, Keith L., et al. "“Sustainable” Development, Energy, Assessment, Land-Use Change." Research Collaboration Network Workshop on Sustainable Bioenergy Systems, Merida, Mexico, May 2012.

Chandola V. Large Scale Machine Learning for Massive Remote Sensing Data — A Case Study in Biomass Monitoring. ASPRS Annual Conference, Baltimore, MD, 2013

Singh, N. et al., 2013. Spatial Pattern of Cropland Changes in the Great Plain regions . IEEE International Geoscience and Remote Sensing Symposium. Melbourne, Australia (accepted for presentation)

Kline, K. L., Singh, N., & Dale, V. H. (2013). Cultivated hay and fallow/idle cropland confound analysis of grassland conversion in the Western Corn Belt. Proceedings of the National Academy of Sciences, 110(31), E2863-E2863.

Vatsavai R.R. & Bhaduri B.,2013 Crop production estimation . Remote Sensing–Beyond Images, 18-21

Vatsavai R.R. & Singh N., Unsupervised Phenological Clustering Technique for Biomass Monitoring. International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM-14).

Singh N. & Vatsavai R.R..Unsupervised Land Cover Change Detection Using Biweekly MODIS NDVI Data. AGU Fall Meeting 2014.