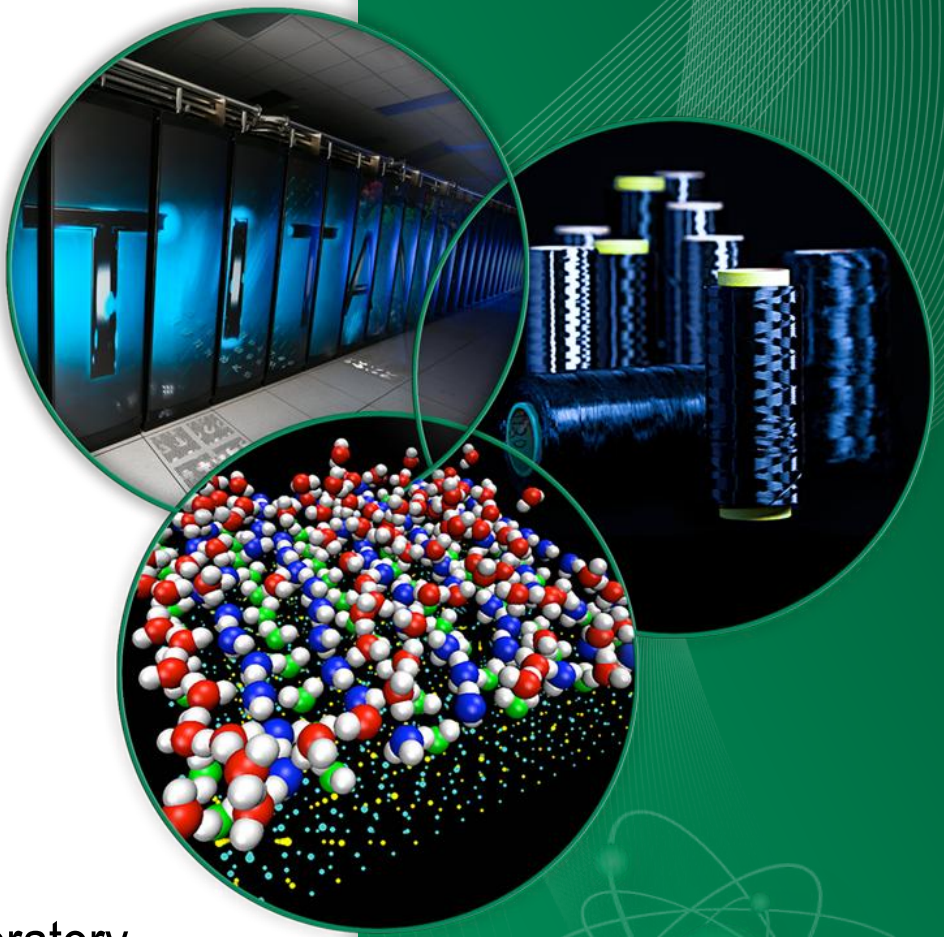
2013 DOE Bioenergy Technologies Office (BETO) Project Peer Review

Land-Use Change Data and Causal Analysis



05/21/2013

Analysis & Sustainability

Nagendra Singh (PI), Keith Kline, Rebecca Efroymson, Varun Chandola, Esther Parish, Budhendra Bhaduri

Organization: Oak Ridge National Laboratory

This presentation does not contain any proprietary, confidential, or otherwise restricted information

Goal Statement

Project Goal

- Design and develop tools and assessment methods to establish scientific basis for understanding and simulating effects of bioenergy policy on land cover and management. Two key issues to address:
  1. Consistent and reliable data
  2. Attribution of effects among causal drivers

DOE Goals

- Strategic Analysis goals:
  - Developing analytical tools, models, methods, and datasets to advance the understanding of bioenergy and its related impacts
  - Provide context and justification for decisions by establishing the basis of quantitative metrics…
- 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: FY15
- Percent complete: 60 %

Barriers

- At-A. Lack of Comparable, Transparent, and Reproducible Analysis
- At-C. Inaccessibility and Unavailability of Data
- St-C. Sustainability Data Across the Supply Chain
- St-G. Representation of Land Use

Budget

Funding for FY11: 200K
Funding for FY12: 200K
Funding for FY13: 200K
Years the project has been funded:4
Average annual funding:200K

Partners

Interactions/collaborations

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

Project Overview

Context

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

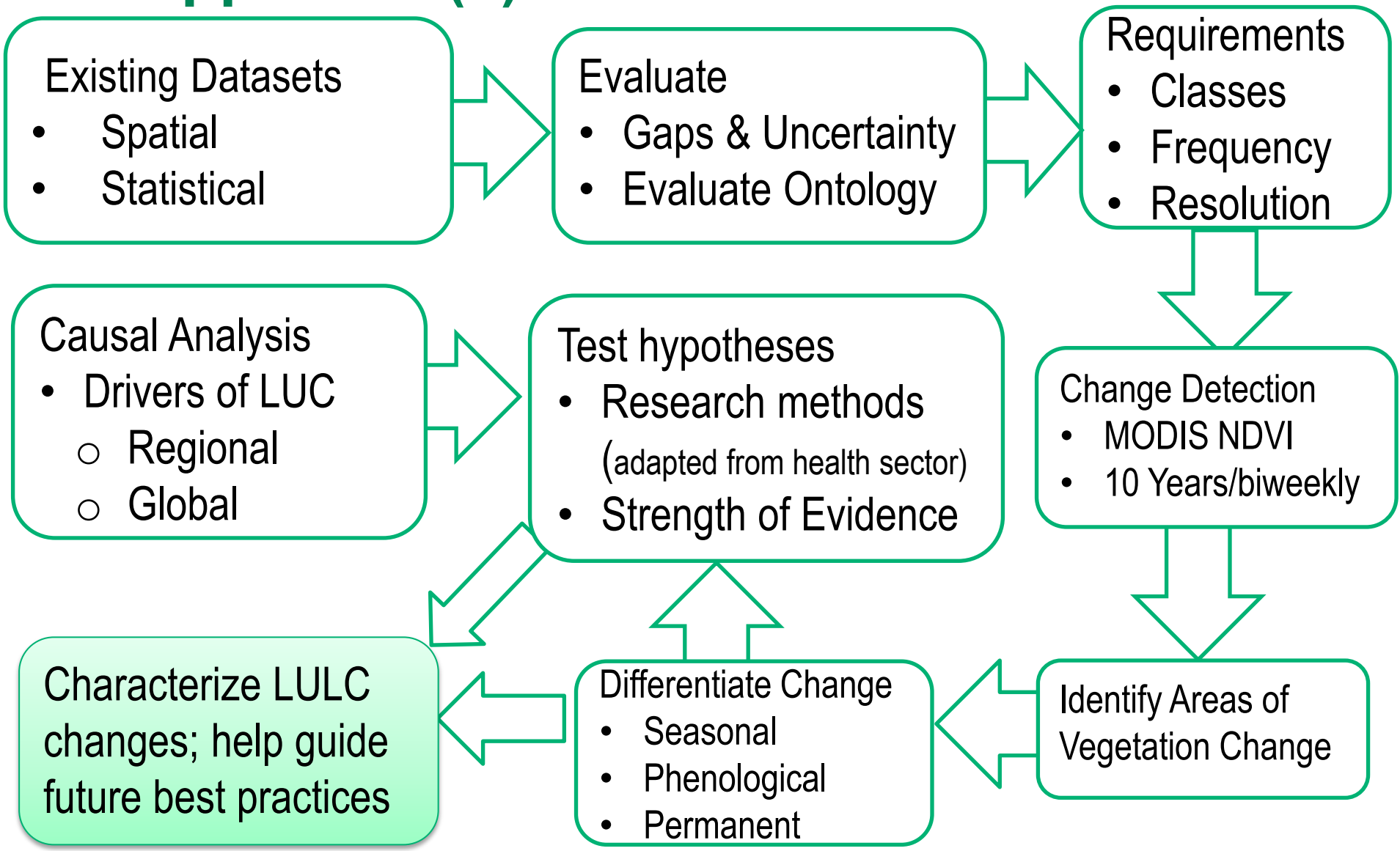
History

- Land-Use Change and Bioenergy workshop (2009)
  - 50 international experts
  - Recommendations to address priority research issues and uncertainty
- Project Development Steps:
  - Pin-point problems related to current land-cover data sets used to assess bioenergy effects
  - Identify key improvements and missing components: reliable time-series data, causal analysis
  - Now moving toward solutions

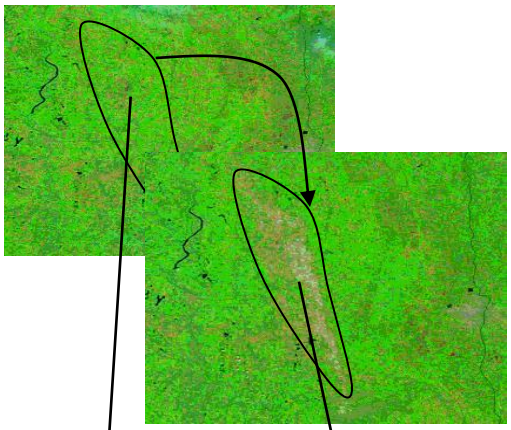
Tasks

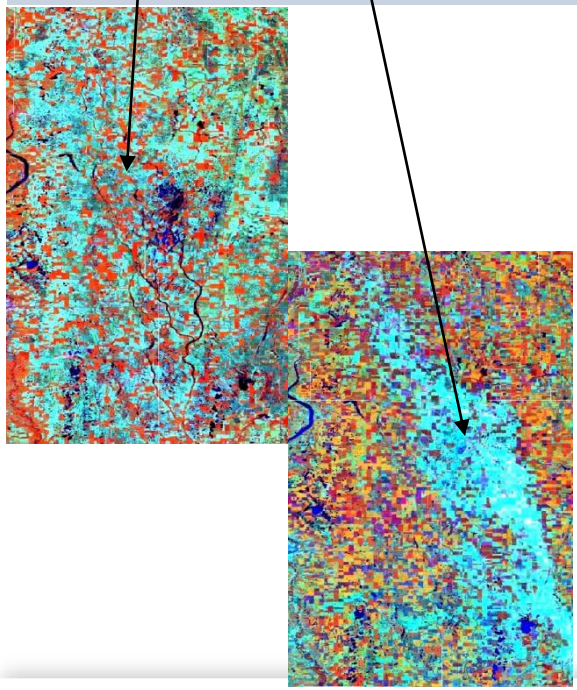
- Analyze and test existing land use/land cover (LULC) data for their suitability to compute impacts of bioenergy policy
- Develop and apply change detection techniques to raw MODIS NDVI data to understand changes before and after biofuel policies
- Devise a causal analysis framework for LULC change

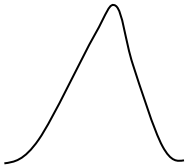
1. Approach (1) – Overview



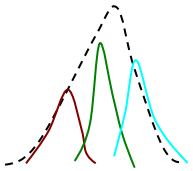
1. Approach (2) – Change Detection







Aggregate Class (Agriculture)



Fine (Sub-)Classes (Corn, Soy, Wheat)

Biomass Monitoring

MODIS (250m, NDVI, 16 day)

- Coarse-spatial resolution
- High-temporal resolution
- Good for regional, global monitoring, but not ideal for crop identification

Change Vs. No Change

Fine-resolution Information Extraction

AWiFS (56 m, 4B, 5d)
NAIP (1m, 4b, 1y)

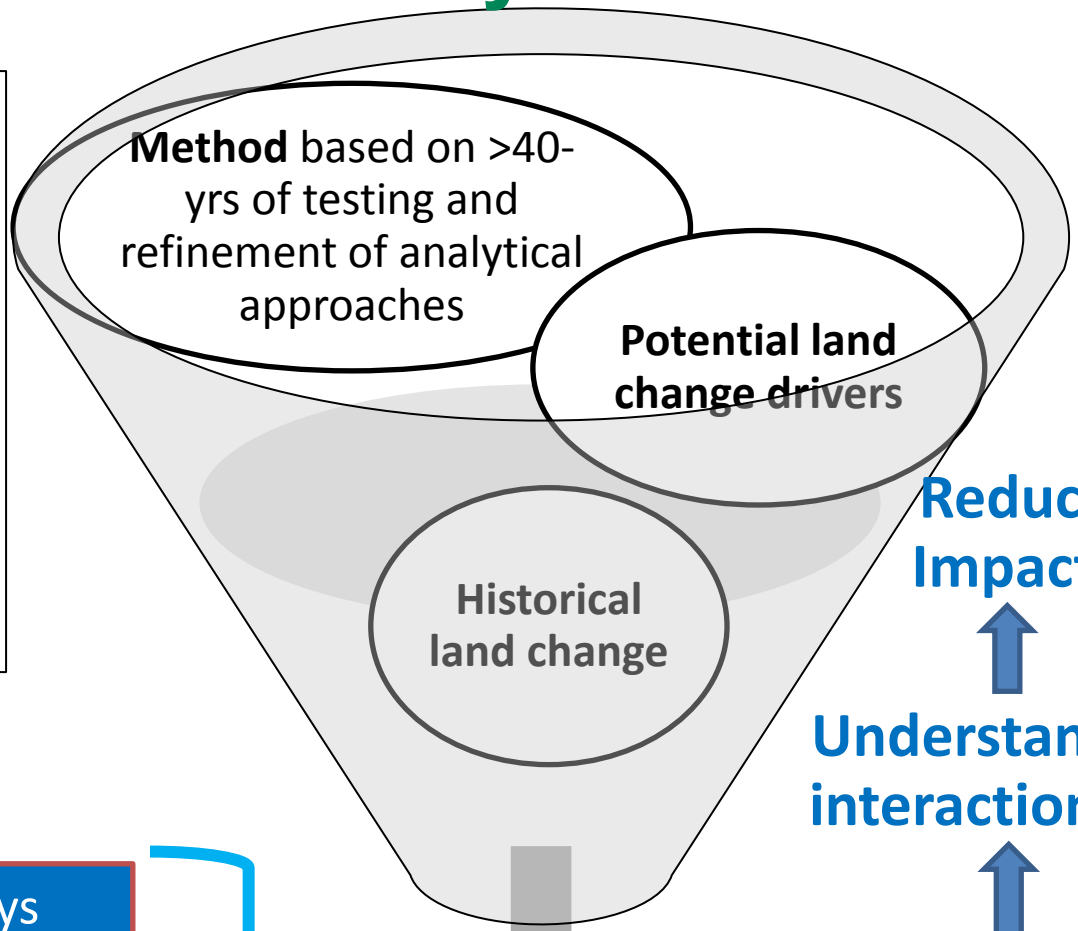
- Moderate-to-high spatial
- Moderate temporal
- Used for crop type and condition extraction
- High training and computational requirements

Characterize Changes

1. Approach (3) – Causal Analysis

Epidemiology (Webster's) "sum of factors controlling the presence or absence of a disease or pathogen"

Our goal: Determine the sum of factors controlling the presence or absence of specific changes in land cover or land management.



Hypothesized pathways

Counterfactual scenarios

Statistical analyses

Causal analysis framework

Reduce Impacts

Understand interactions

Location-specific analyses

1. Approach (4) – Management

- Bimonthly meetings that bring together the data and causal analysis halves of the projects to make sure that they are integrated
- Meetings and email correspondence as required to handle other issues
- A lead team member for each milestone who is responsible for ensuring that the right deadlines are met, the right collaborators are involved, and that related projects mutually benefit from the analyses
- Conversations with other laboratories and researchers that have led to collaborations
- Monthly discussions with BETO and LRM regarding the status of the project and quarterly reports

2. Technical Accomplishments/Results (1)

Geographic Data and Analysis Methods

- Gaps and Uncertainties in existing LULC datasets.
  - Ontology for LULC datasets ✓
  - Manuscript for DOE review ✓
- Change Detection Algorithm ✓
- Processing MODIS data ✓
  - Computational process, debugging ✓
  - Interpretation to distinguish Land cover before and after change (forthcoming - due 3rd quarter FY13)

Causal Analysis

- Framework and report for causal analysis related to LULC change (2012) ✓
- Manuscript on the importance of causal analysis when assessing LULC change effects of bioenergy policy (due 4th quarter FY13)

Iowa Case Study

- Changes in land cover in mid-western USA and testing of possible correlation between bioenergy growth and changes (4th quarter FY13)

2. Technical Accomplishments/Results (2)

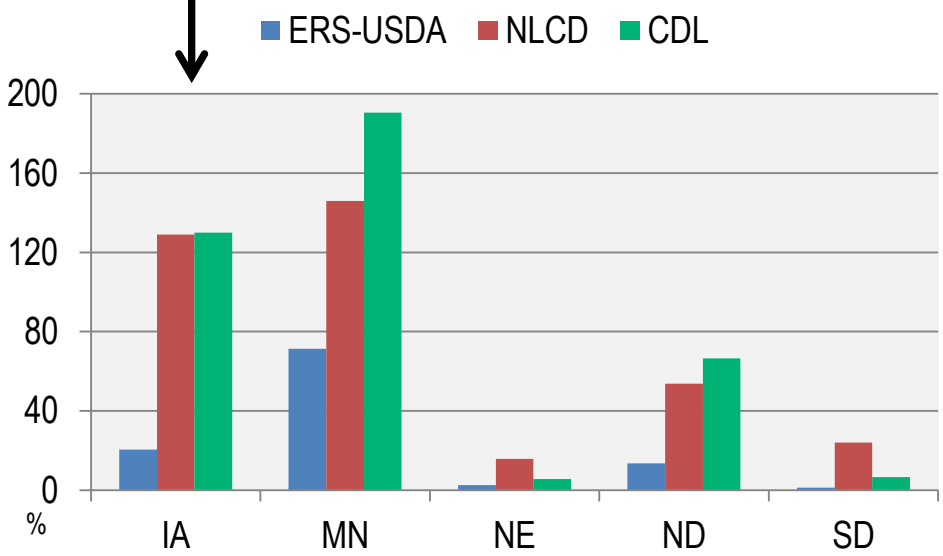
Data Analysis

- Major sources of uncertainty stemmed from area aggregation techniques, pixel-level mathematics, outdated data, and incorrect use of land-cover classes and aggregation of classes
- Data-centric approach taken to compute significant land-use changes in mid-western US using bi-weekly MODIS NDVI data for 10 years; effort will identify areas in mid-western US which underwent clearly specified land-cover change
- Spatial classification algorithm developed which will use annual MODIS NDVI profiles as features to classify a pixel into one of several land cover types
- Algorithm incorporates semi-supervised learning methodologies to efficiently learn a classifier using limited set of ground truth examples and large number of unlabeled examples.

2. Technical Accomplishments/Results (3)

- Subjective aggregation will lead to divergent conclusions
- Definition of LULC varies considerably across datasets
- High variation in estimates from different datasets

From	To	1000 Acre	Change %
Grass	Corn/Soy	5,488	1.34
Corn/Soy	Grass	4,287	1.32
Difference		**1,201**	
Grass + Idle	Corn/Soy	6,598	1.51
Corn/Soy	Grass + Idle	4,450	1.38
Difference		**2,148**	
Grass	Corn/Soy + Other Crop	7,966	1.94
Corn/Soy + Other Crop	Grass	8,641	1.96
Difference		**-675**	
Grass+Idle	Corn/Soy + Other Crop	9,496	2.15
Corn/Soy + Other Crop	Grass+Idle	10,409	2.36
Difference		**-913**	



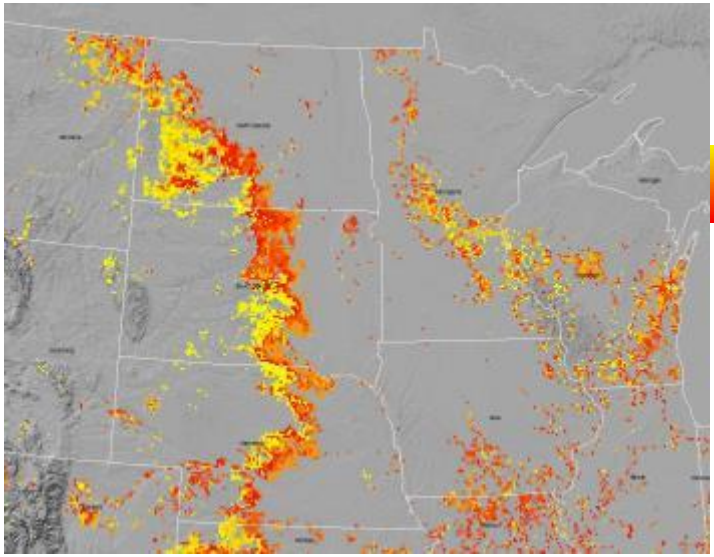
"Grassland" area 2006: % variation with respect to the USDA Ag Census for 2007

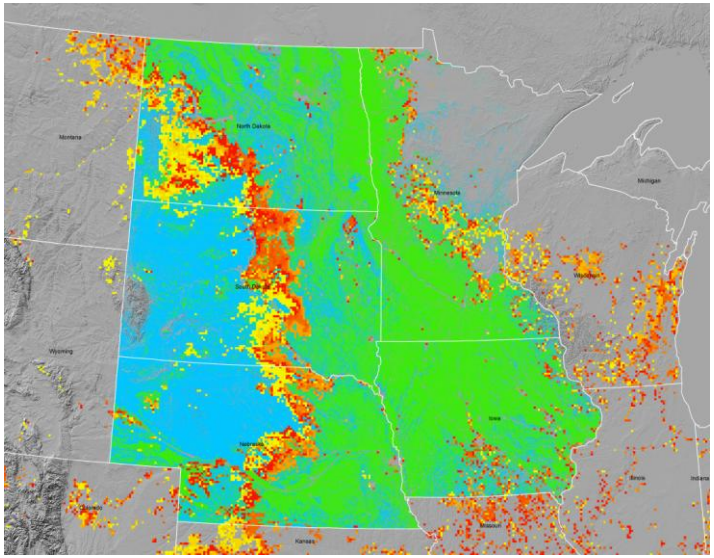
TAG Cloud For Pasture in Various Datasets

2. Technical Accomplishments/Results (4)

We implemented two solutions to convert large scale bi-weekly MODIS NDVI data for 10 years into a change statistic for every MODIS pixel.

1. A Message Passing Interface (MPI) based solution for deployment on a traditional multi-node high performance computing architecture.
2. A Map-reduce based solution for deployment on a Hadoop based cloud architecture.

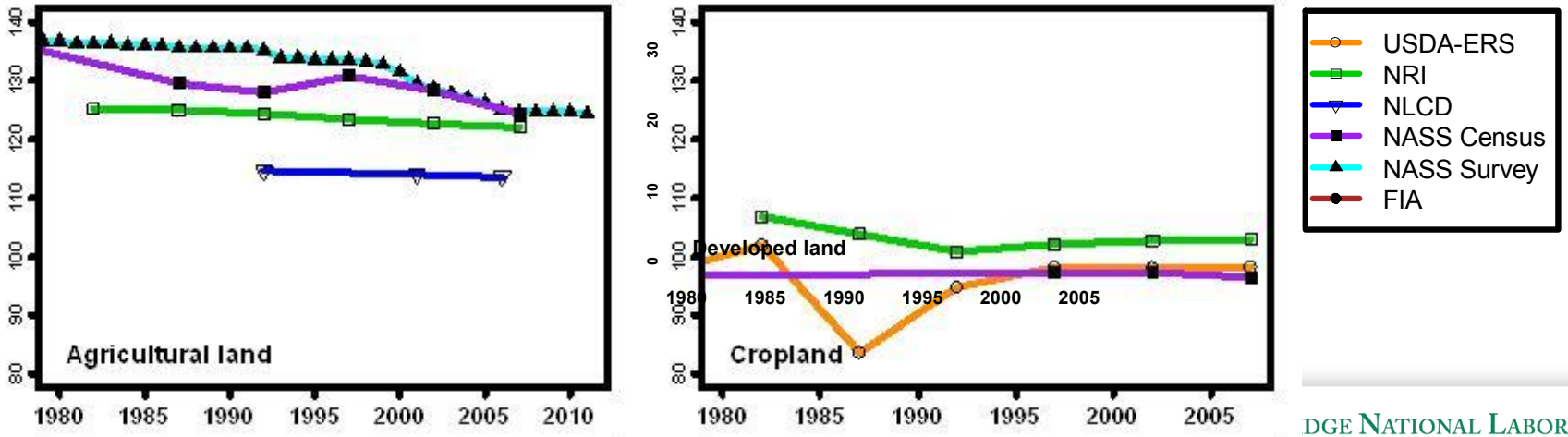




2. Technical Accomplishments/Results (5)

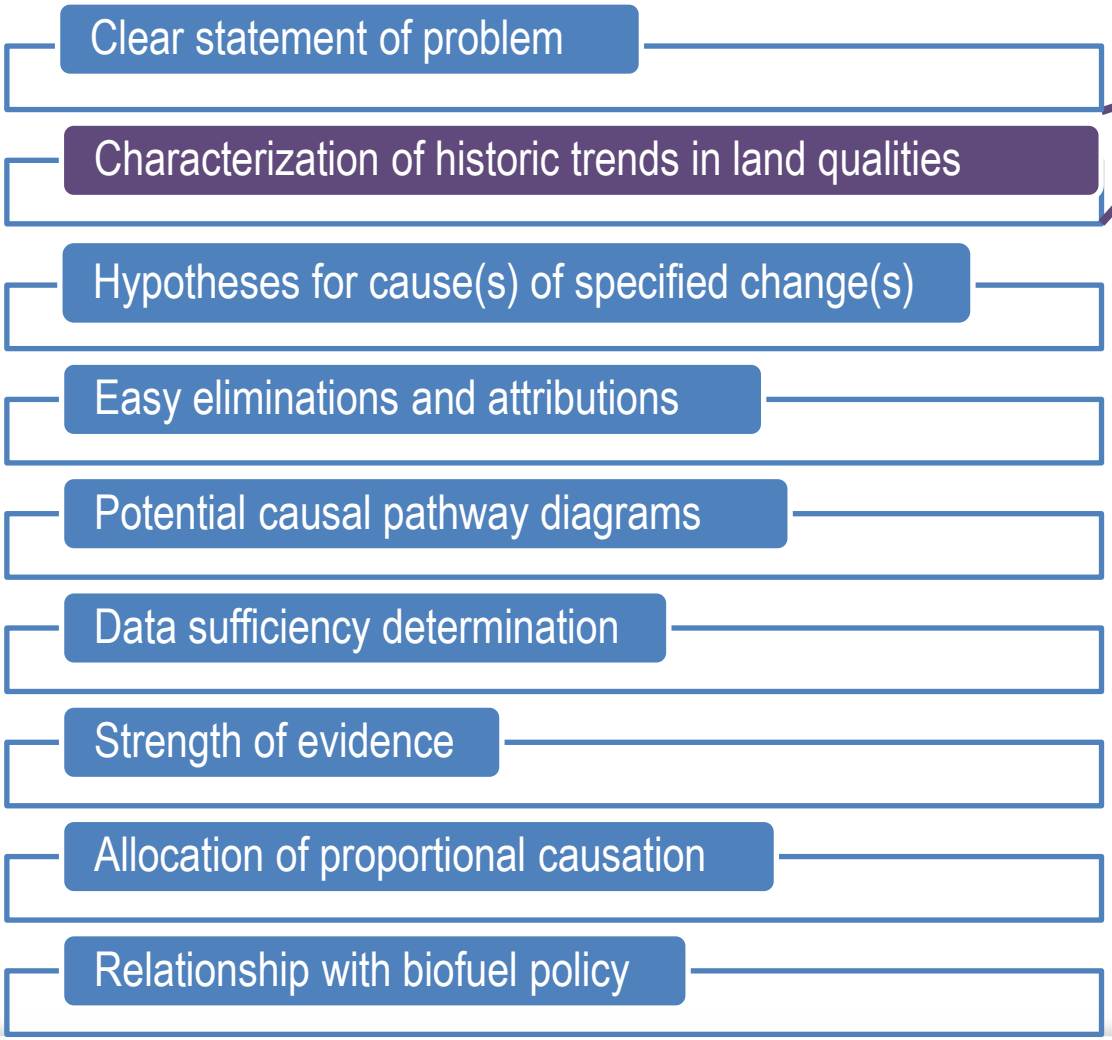
Preparation of report, “Investigation of Iowa land-use change for bioenergy feedstock production”

- Among forest/ grassland/ cropland/ developed-other classes:
  - Little or no alteration in change trends before and after biofuel policies
  - Possible reduction in rate of net loss of agriculture land to developed-other classes
  - Most changes involve rotations at cropland/grassland interface
- Conclusion: reliable LULC time series datasets are needed to test causal hypotheses and to develop and validate improved LULC models that incorporate regional mechanisms



2. Technical Accomplishments/Results (6)

Causal Analysis Framework



Essential to clearly define effect. E.g., what is meant by "land-use change"?

- land management
- land cover
- carbon stocks
- nutrient cycling

Defined in measurable terms

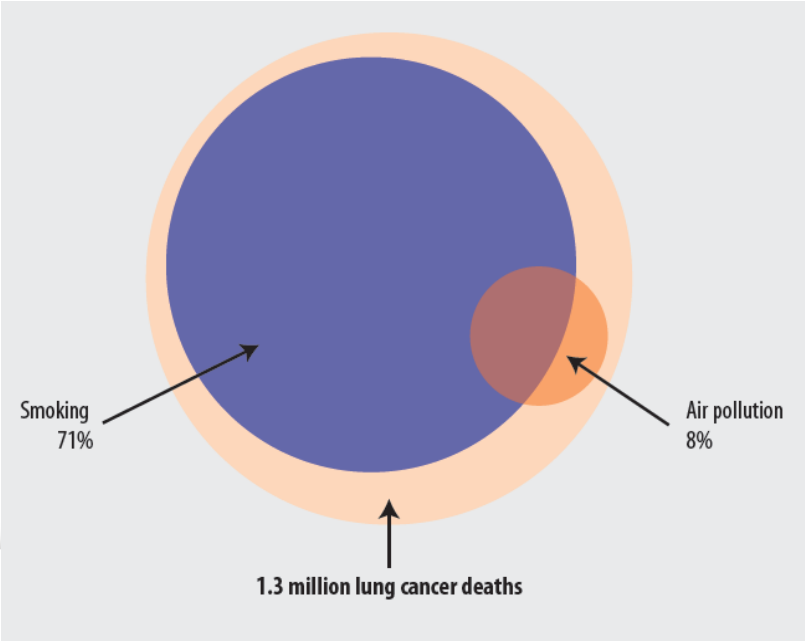
2. Technical Accomplishments/Results (7)

- Clear statement of problem
- Characterization of historic trends in land qualities
- Hypotheses for cause(s) of specified change(s)
- Easy eliminations and attributions
- Potential causal pathway diagrams
- Data sufficiency determination
- Strength of evidence
- Allocation of proportional causation
- Relationship with biofuel policy

Evidence

- Plausible cause and pathway
- Spatial co-occurrence
- Time order
- Analogous drivers
- Simulation model results
- Driver-response relationships

World Health Organization 2009 example of allocation



3. Relevance

- **Project consistent with DOE BETO Objectives**
  - ✓ **Strategic Analysis –** "supports efforts to better understand and ***characterize the complex drivers of land-use change and gather more accurate land-use data*** to feed into these analyses"
  - ✓ **Environmental Sustainability –** **"…**land use—should be monitored along the entire bioenergy supply chain"
- **Project will help meet DOE BETO Targets**
  - ✓ Evaluate and compare the sustainability of biofuels produced from agricultural residues, energy crops, forest resources, and algae
- **Project supports Strategic Analysis Platform Activities**
  - ✓ Ensure high-quality, consistent, reproducible… analyses
  - ✓ Develop analytical tools, methods, and datasets to advance understanding…
  - ✓ Convey results of analytical activities to wide audience

3. Relevance (2)

- Improving understanding of “land-use change” through
  - Definitions and clarifying semantics (land management, land cover, stocks and flows, other measurable characteristics)
  - Explaining effects of different aggregation techniques and pixel-level mathematics
  - Developing improved, time-series data sets
- Developing evidence-based methods for attributing changes among bioenergy and other causes
- Revealing deficiencies in land-cover trend analysis and assumptions regarding links between bioenergy policy and LULC change
- Developing spatial classification algorithm to accurately characterize changes in LULC patterns
- Provide inputs (data, definitions, plausible mechanisms) to support improved modeling with GTAP, BioLUC and other platforms
- Enabling corrective actions (e.g. best practices), to reduce negative effects and optimize positive effects of biofuel development

4. Critical Success Factors

Challenges for achieving results include

- Data to support causal analysis and land-use change models require sufficient
  - categorization
  - resolution
  - extent
  - duration and
  - frequency
- Ability to consider potential drivers of land-use change and the evidence associated with each

Project success will positively impact commercial viability of bioenergy

- Uncertain estimates of land-use change and GHG emissions present challenges to investment and development of sustainable bioenergy
- Reducing uncertainty will speed social acceptance of bioenergy

5. Future Work

- Use HPC and cloud computing to establish an integrated monitoring framework that allows correlating changes obtained from MODIS time series imagery with existing LULC databases like USDA crop data layer.
- Develop a consensus benchmark dataset for use in global analysis of LULC from energy policy by involving stakeholders to review options for best available data, classification, and ontology issues
- Identify, collect data sets, and test hypotheses for regional causal factors of LULC in priority change zones around the world
- Define and develop science-community consensus around optimal units for measuring and modeling changes in land management, land quality, land services and functions

Summary

- Approach
  - Apply robust data and causal analysis techniques to quantify and understand linkages between LUC & bioenergy.
- Technical accomplishments
  - A change detection system to accurately characterize changes in land cover
  - An analytical framework to identify causes of LULC change and assess the role of bioenergy as one potential driver
- Relevance
  - Accurate data and robust causal analysis are essential to improve analysis and projection of land changes associated with energy policies
  - Better understanding of “real drivers” at appropriate scales enables achievement of strategic goals to enhance positive effects and minimize negative effects of biofuel policy
- Critical success factors and challenges
  - Significant time and effort required to develop consensus on consistent definitions, approaches and complete data collection at global scale
  - Accurate LULC change information is critical to achieve sustainable production pathways
- Future Work
  - Integrate data and causal analysis to clarify relationships among bioenergy and drivers of LUC and to identify best practices.

Additional Slides

Acronyms

AwiFS - Advanced Wide Field Sensor

CDL - Cropland Data Layer

DOE- Department of Energy

ERS - Economic Research Service -USDA

FIA - Forest Inventory and Analysis

FY- Financial Year

GTAP - Global Trade Analysis Project

LRM – Laboratory Relationship Manager

LULC – Land Use Land Cover

MODIS - Moderate Resolution Imaging Spectroradiometer

NAIP - National Agriculture Imagery Program

NASS - National Agricultural Statistics Service

NDVI - Normalized Difference Vegetation Index

NLCD - National Land Cover Database

NREL – National Renewable Energy Laboratory

ORNL – Oak Ridge National Laboratory

USDA - United States Department of Agriculture

USGS - United States Geological Survey

Responses to Previous Reviewers' Comments

Sig. Weakness: The project approach to project management including milestones and deliverables was not clearly delineated or discussed in the presentation.

- These are described in this review – see Slide 8, 9.

It is not clear how forestlands data will be compiled and interfaced with other land uses. Also, it is not clear how cause-and-effects will be modeled

- Good question given divergent global data sets and definitions for "forest" land cover and land use along with many forest degradation issues. We have proposed moving toward new ontologies that provide measurable, verifiable data on metrics including carbon stocks, carbon and nitrogen flux rates, etc. This is a long-term objective and will take time to do properly, as noted above under "challenges."
  Regarding cause and effects, we plan to apply the methods developed for epidemiology to identify most probable drivers of specific observed (measured) changes and then apply the causal analysis framework which stresses a strength-of-evidence approach.

This is a critically-important project that is being handled by very capable people. The major weakness I note is the failure to take into account the greenhouse gas emissions associated with management approaches adopted during land use change.

- Good point. This is an example of specific data sets and measurements that are needed (standard classifications for land management interventions and their intensities, along with corresponding GHG emissions associated with each). At this stage, we are not trying to do the GHG accounting. Better estimation of land-use change and attributions are essential prerequisites to assessing the net GHG emissions associated with biofuels. This project is of modest scope/funding and is therefore focusing on just two priority needs: improved data and causal analysis.

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., Singh.N, & Dale,V. 2013, Cultivated hay and fallow/idle cropland confound analysis of grassland conversion in the Western Corn Belt. (accepted *PNAS*)