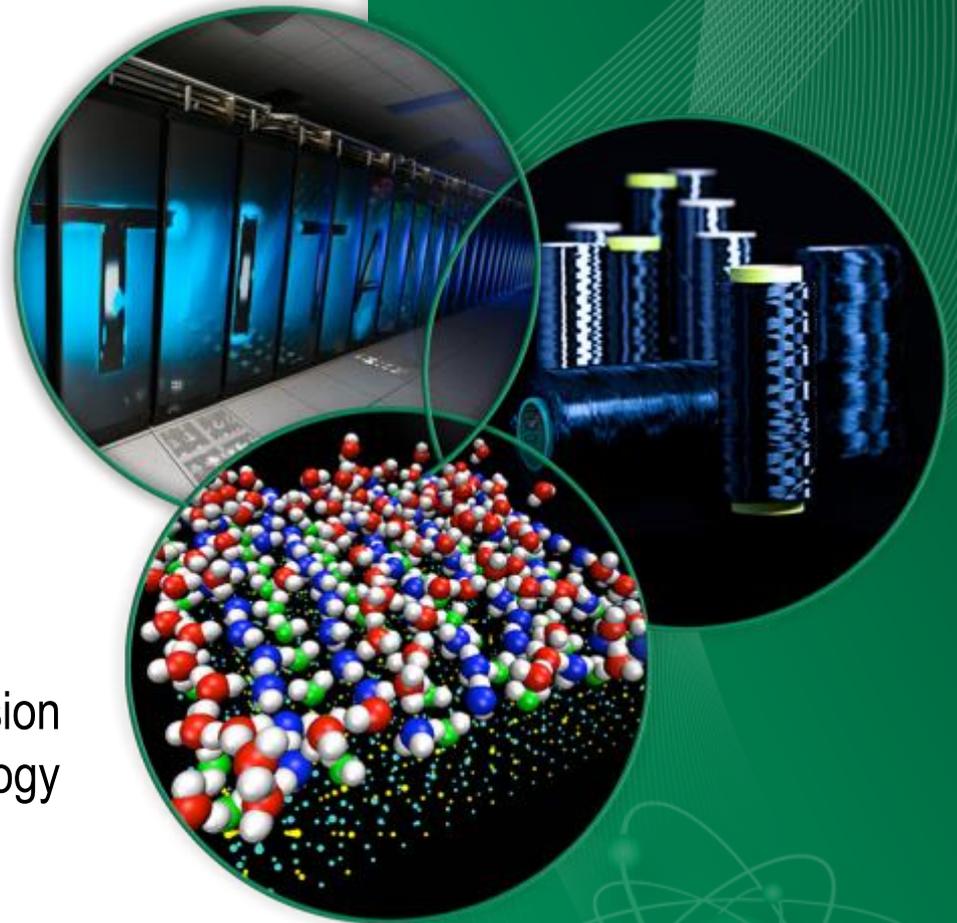


Geospatial Research Capabilities and Contributions

Nagendra Singh

singhn@ornl.gov

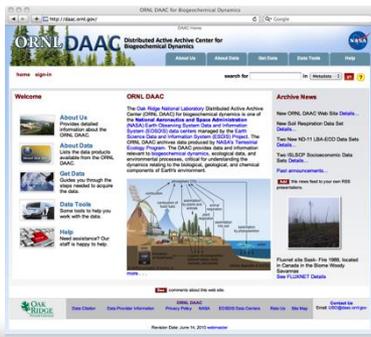
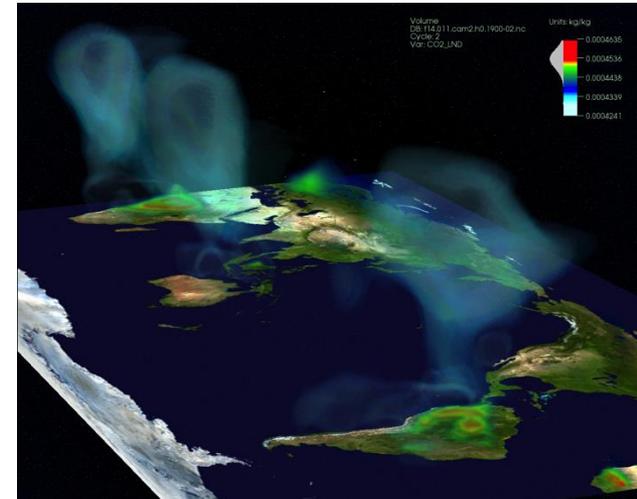
Computational Sciences & Engineering Division
Geographic Information Science & Technology



Environmental Data Science & Systems

Provide data management and analysis for large, integrated environmental databases to the nation's research community and policymakers

- Atmospheric Radiation Measurements (ARM) Archive
- Carbon Dioxide Information and Analysis Center (CDIAC)
- NASA Distributed Active Archive Center (DAAC)
- USA National Phenology Network
- National Biological Information Infrastructure

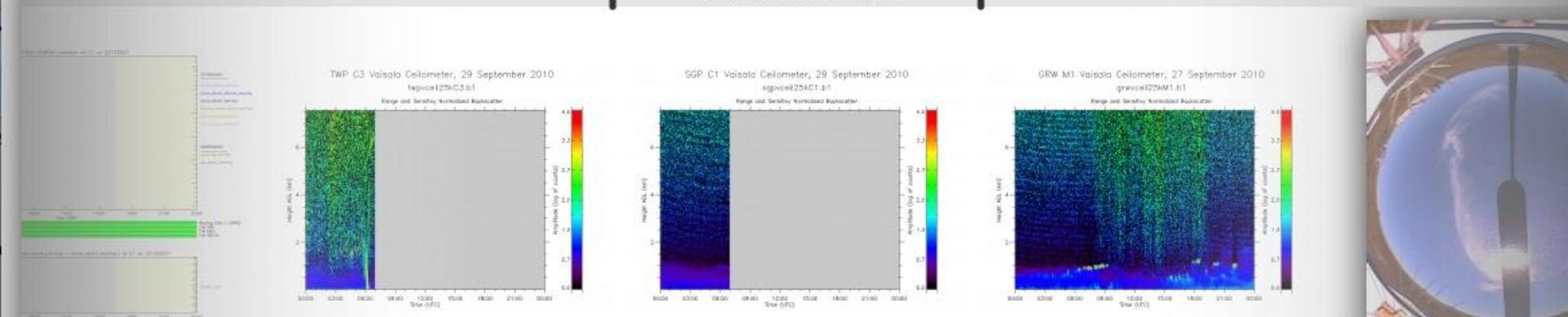


ARM Data Archive

- The **Atmospheric Radiation Measurement (ARM) Climate Research Facility** is a U.S. Department of Energy scientific user facility for the study of global climate change by the national and international research community.



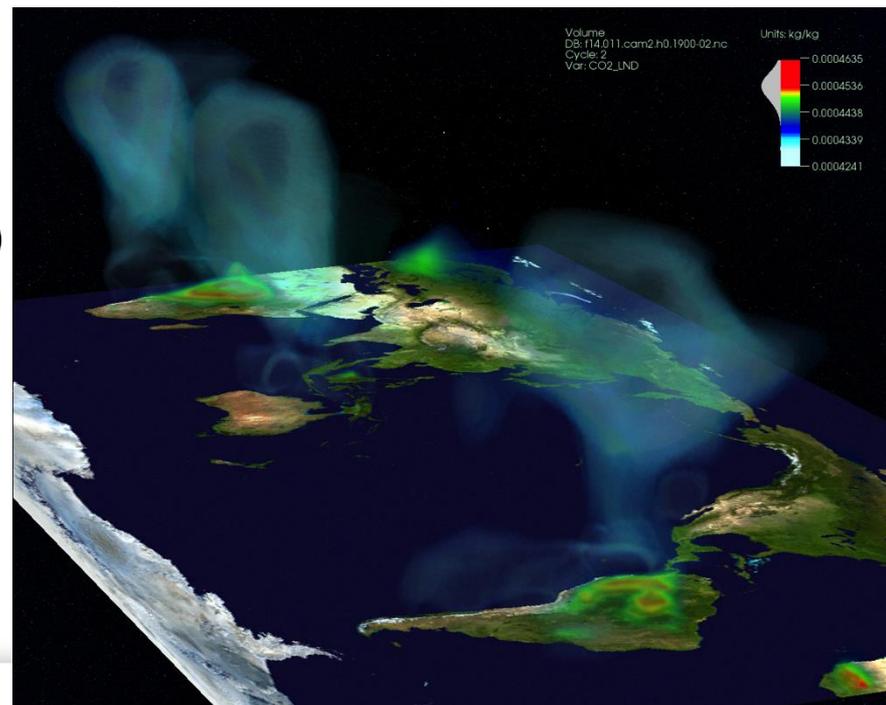
Normalized Backscatter



Vaisala Ceilometer (VCEIL)



- Provide comprehensive data, information, and research support to national and international modeling efforts, researchers and societal interests
- **300 databases include multi-disciplinary, multi-agency, multi-national data and information**
 - Carbon cycle (GHG emissions, land-use change, terrestrial fluxes)
 - Trace gases (atmospheric and oceanic)
 - Climatic data
- Satisfy ~350,000 requests for data worldwide annually

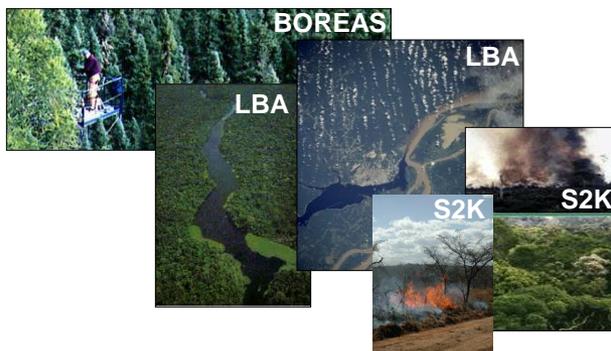


NASA Distributed Active Archive Center (DAAC): Biogeochemical dynamics, ecological data, and environmental processes

Total Data Sets = 885

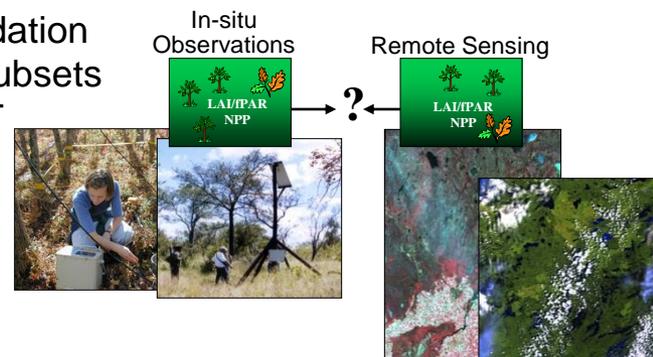
1. Field Campaigns (676)

- FIFE
- OTTER
- SNF
- BOREAS
- LBA



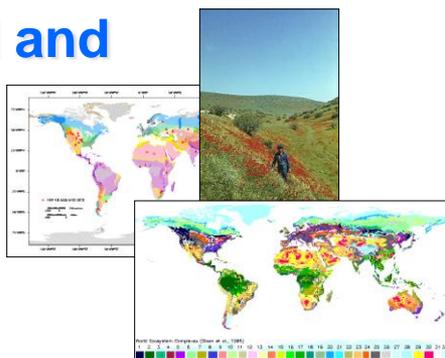
2. Validation of Land Products (21)

- Land Validation
- MODIS Subsets
- FLUXNET
- NPP
- BigFoot



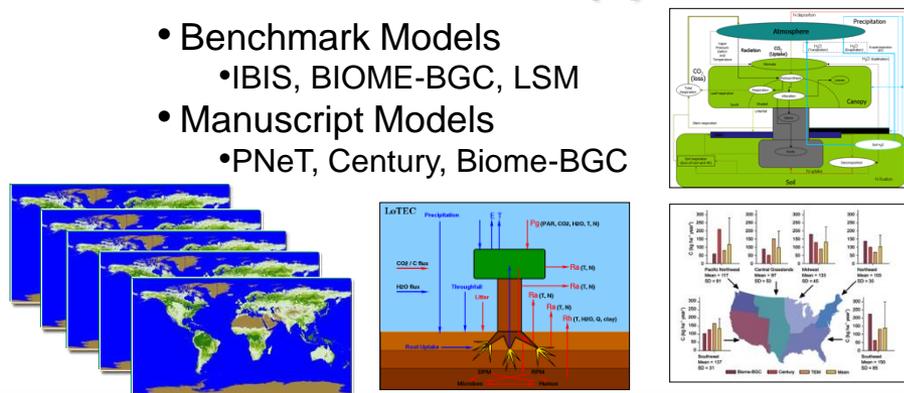
3. Regional and Global Studies (178)

- Climate
- Soils
- Vegetation
- Hydroclimatology



4. Model Products (9)

- Benchmark Models
 - IBIS, BIOME-BGC, LSM
- Manuscript Models
 - PNeT, Century, Biome-BGC



DataONE: Enabling Science Collaboration



eBird



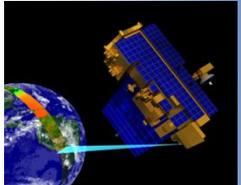
Land Cover



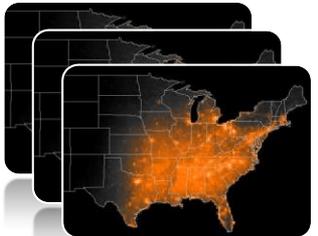
Meteorology



MODIS – Remote sensing data



Diverse bird observations and environmental data from 300,00 locations in the US integrated and analyzed using High Performance Computing Resources



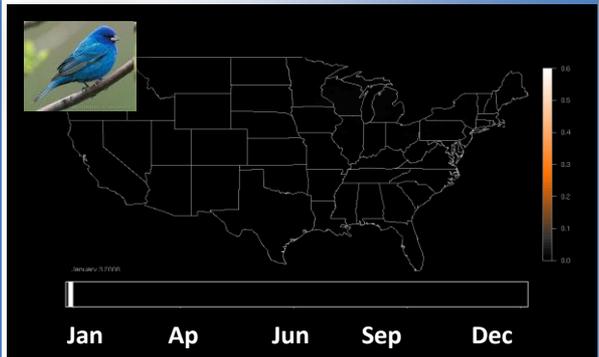
$$F(X, s, t) = \frac{1}{n(s, t)} \sum_{i=1}^m f_i(X, s, t) I(s, t \in \theta_i)$$

Spatio-Temporal Exploratory Model identifies factors affecting patterns of migration



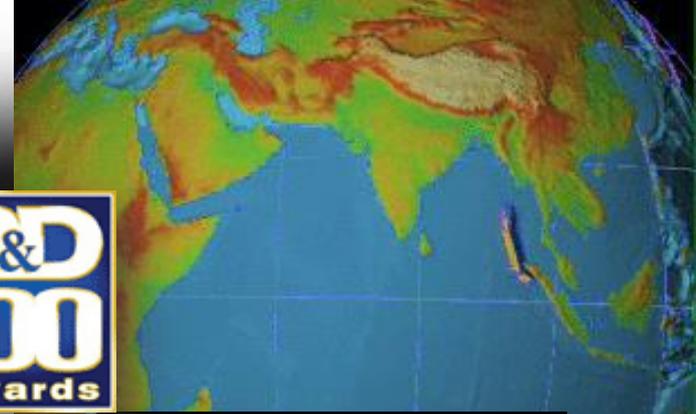
Model results

Occurrence of Indigo Bunting (2008)

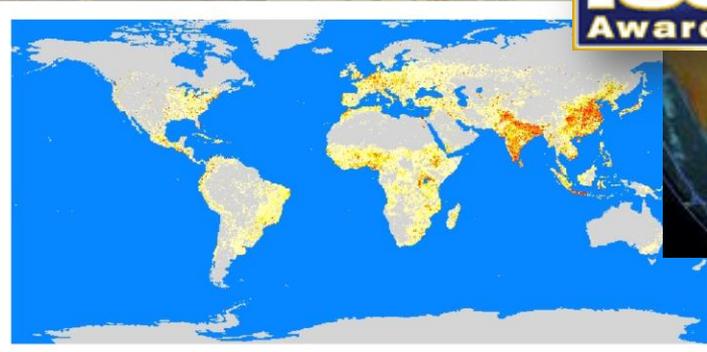
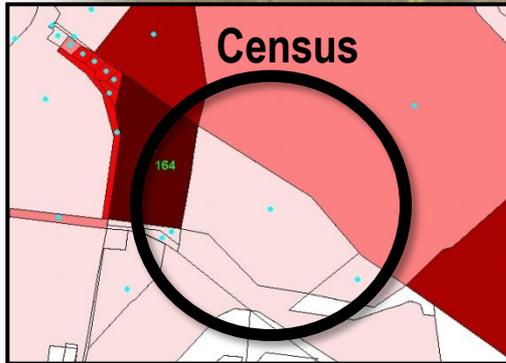


- Examine patterns of migration
- Infer how climate change may affect bird migration

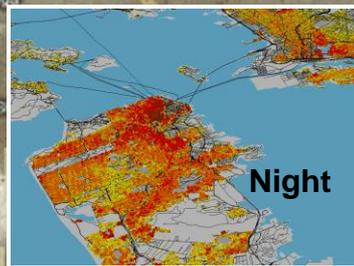
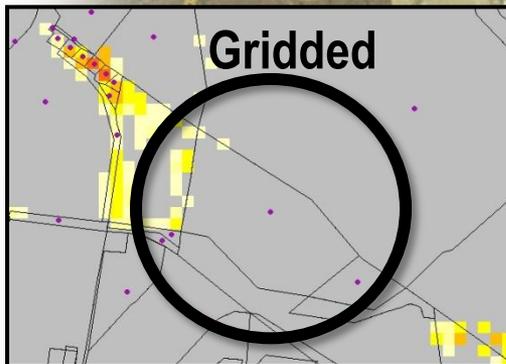
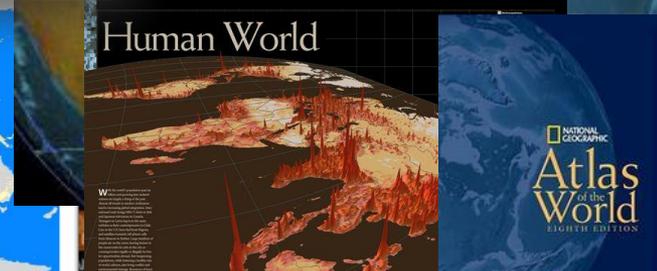
LandScan Population Distribution and Dynamics Model and Database



LandScan Global



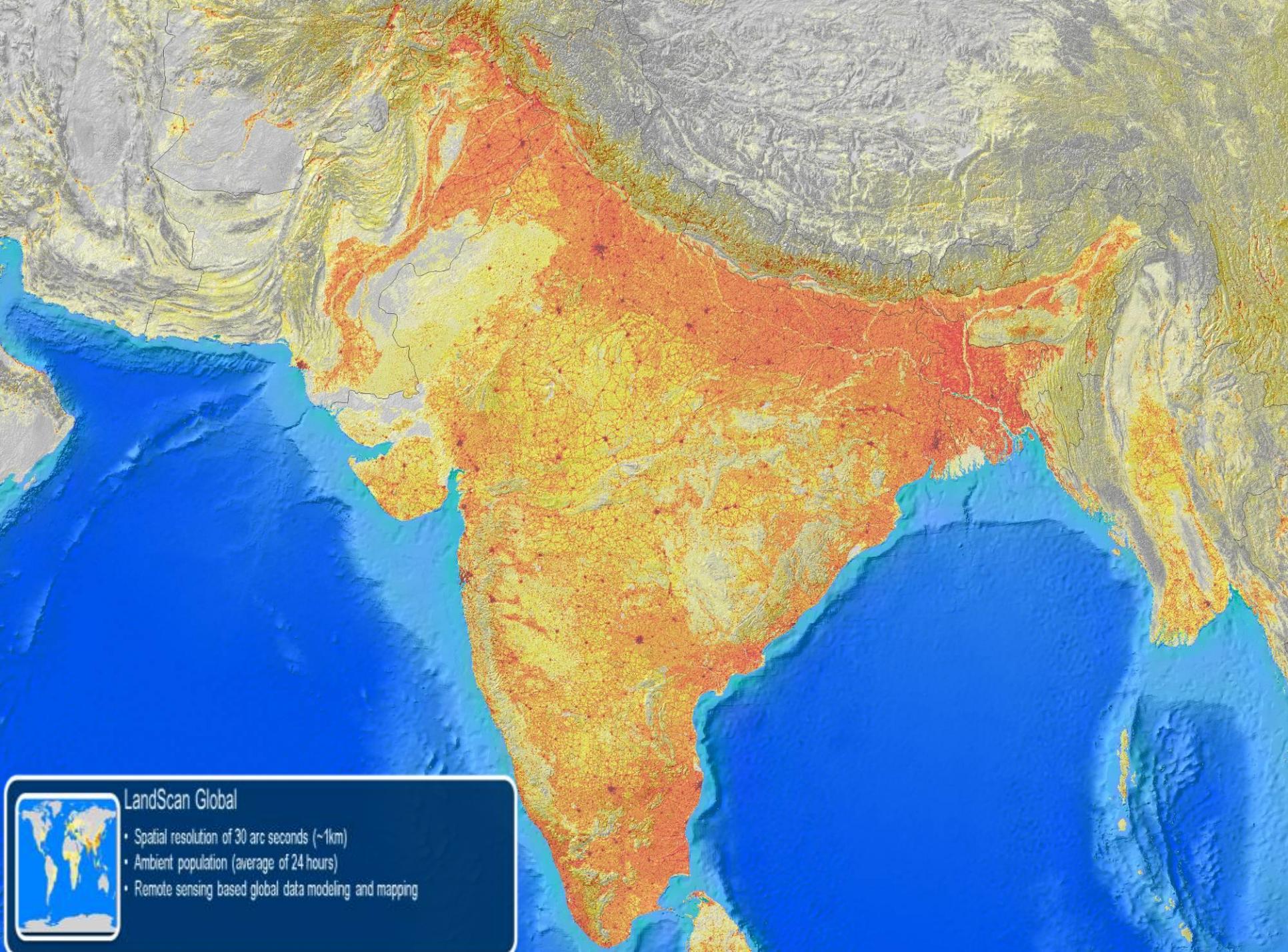
Human World



LandScan USA



As the finest population distribution data ever produced for the world and the US, LandScan Global and LandScan USA are the community standard for estimating population at risk

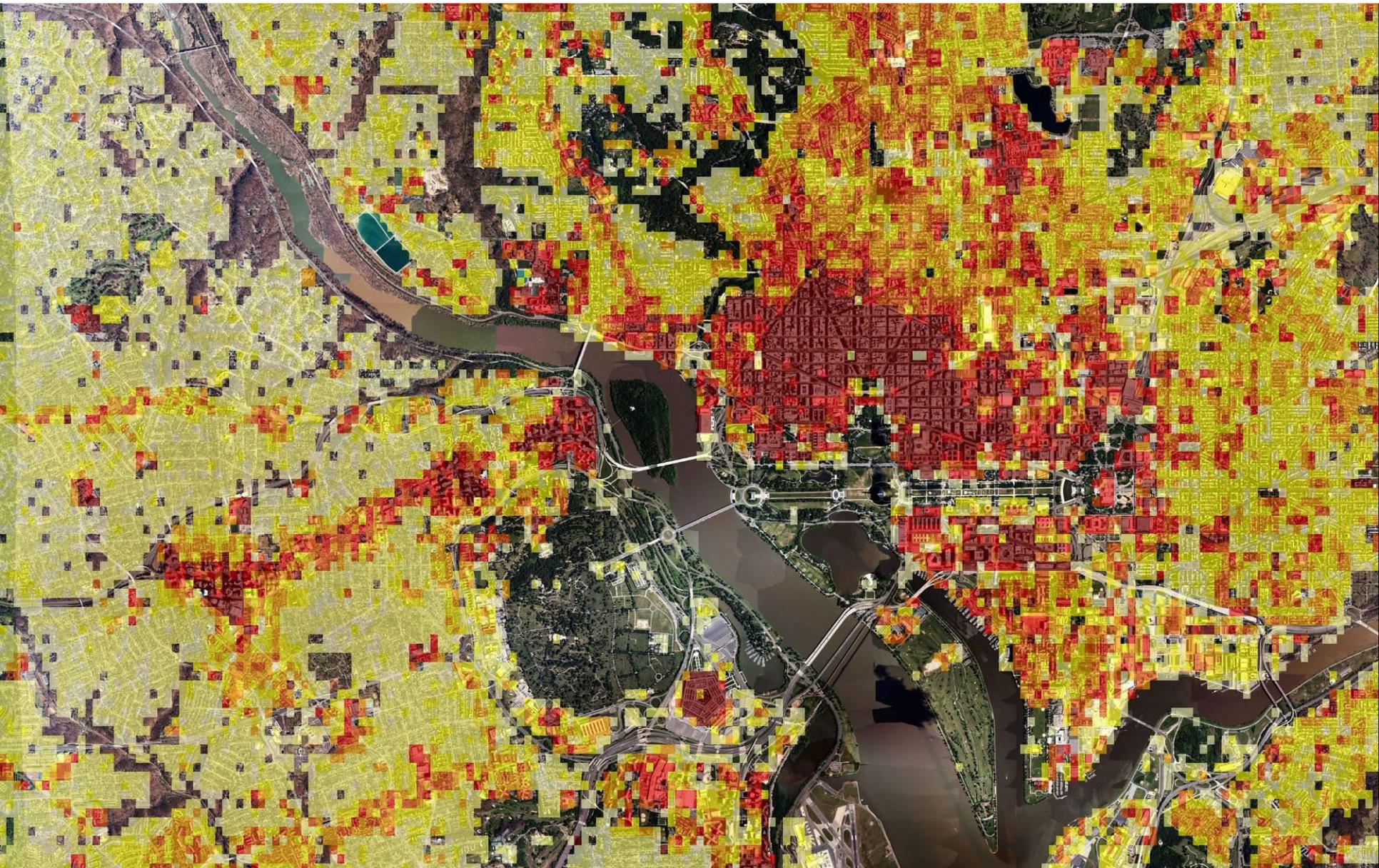


LandScan Global

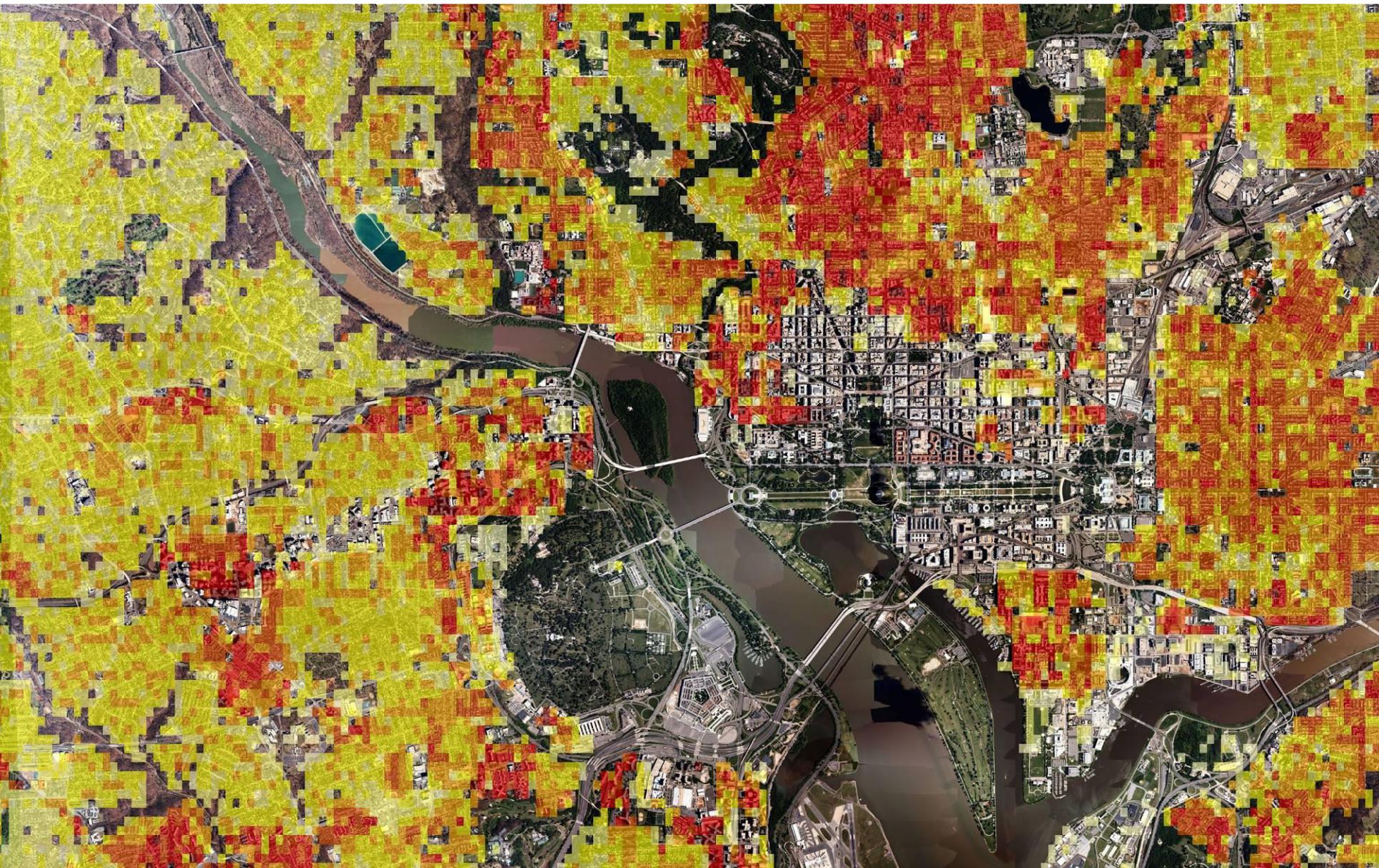
- Spatial resolution of 30 arc seconds (~1km)
- Ambient population (average of 24 hours)
- Remote sensing based global data modeling and mapping



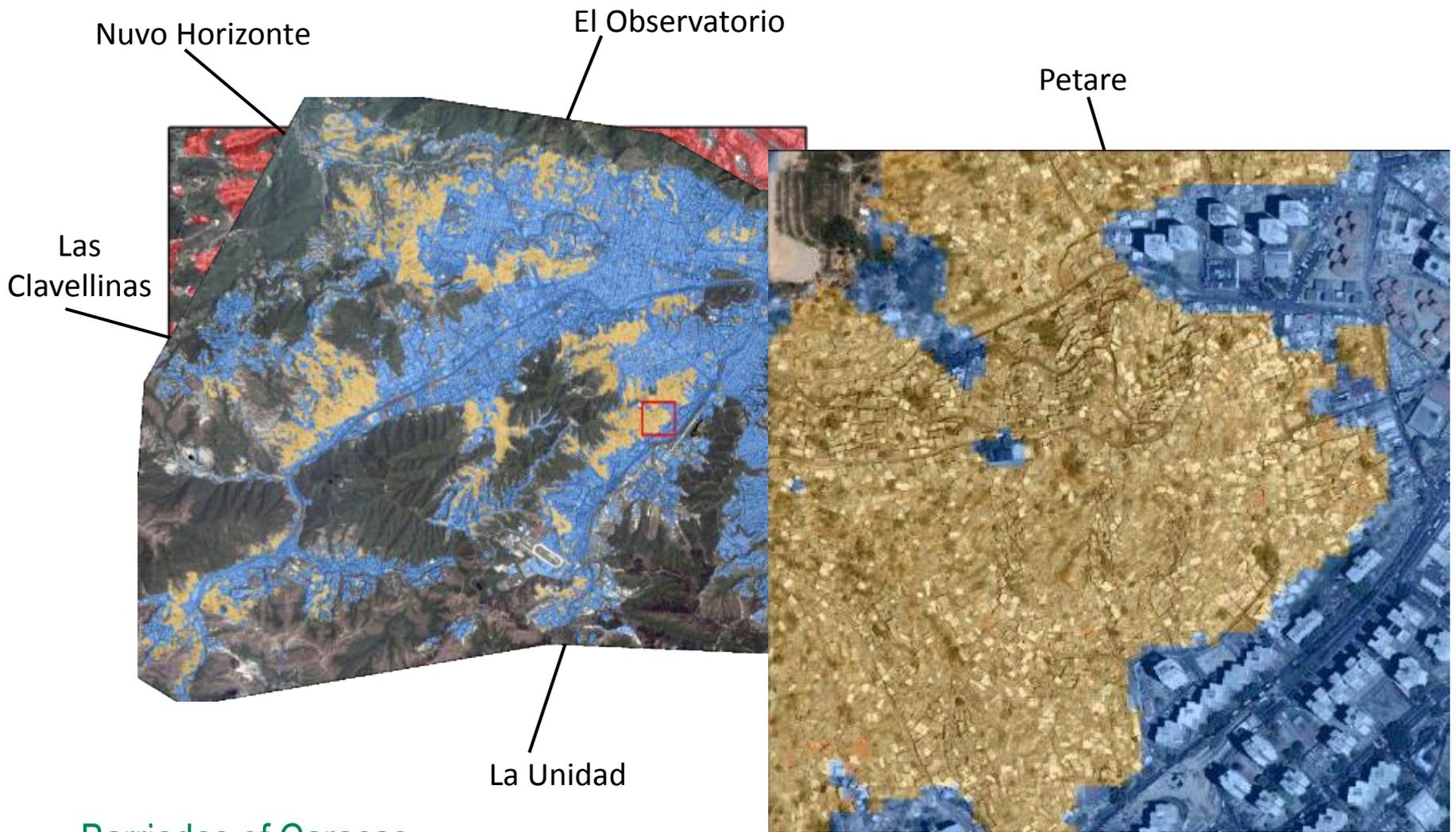
Day Time Distribution



Night Time Distribution

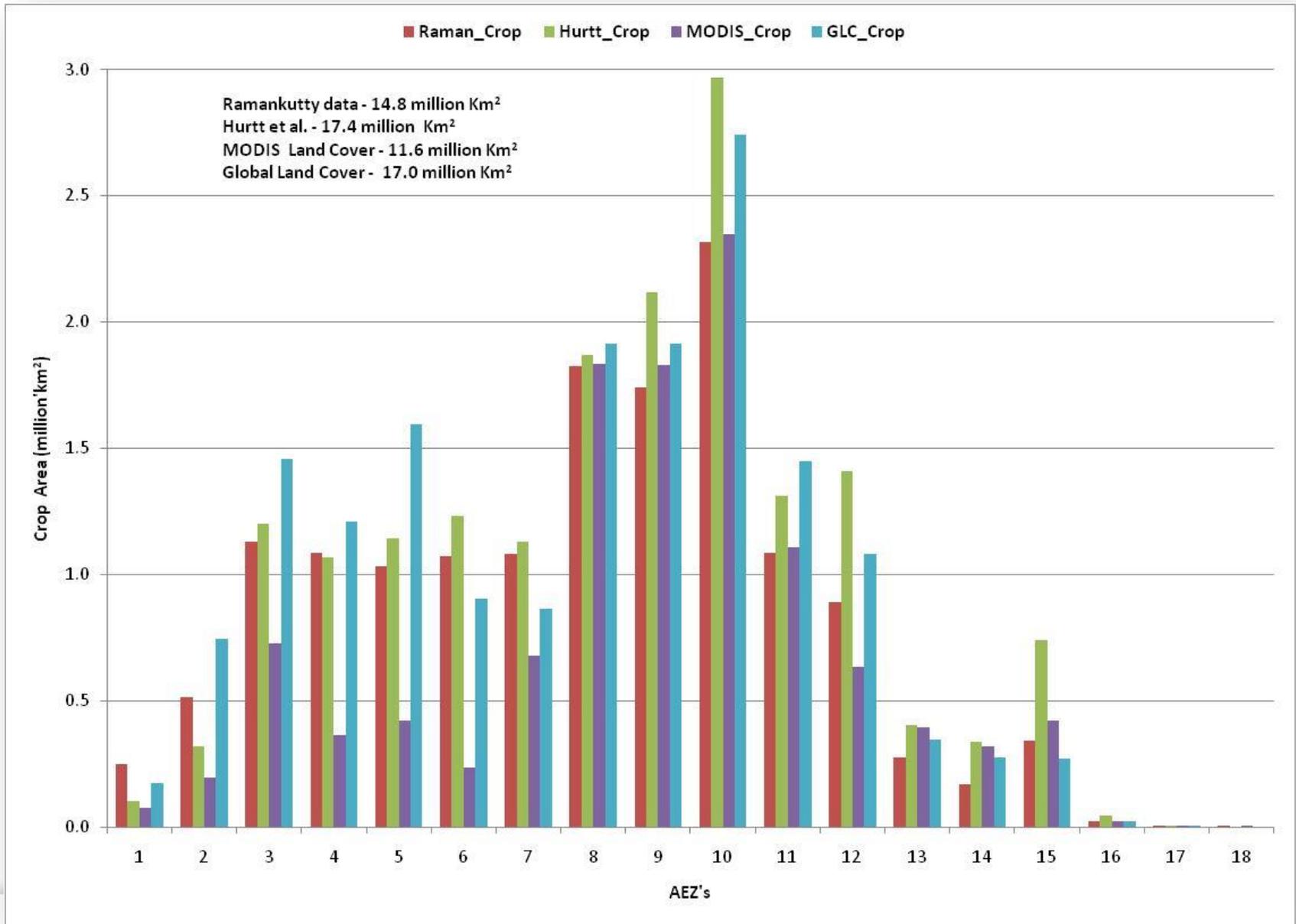


Settlement Mapping



Barriadas of Caracas

Challenges in Global Land Cover Data



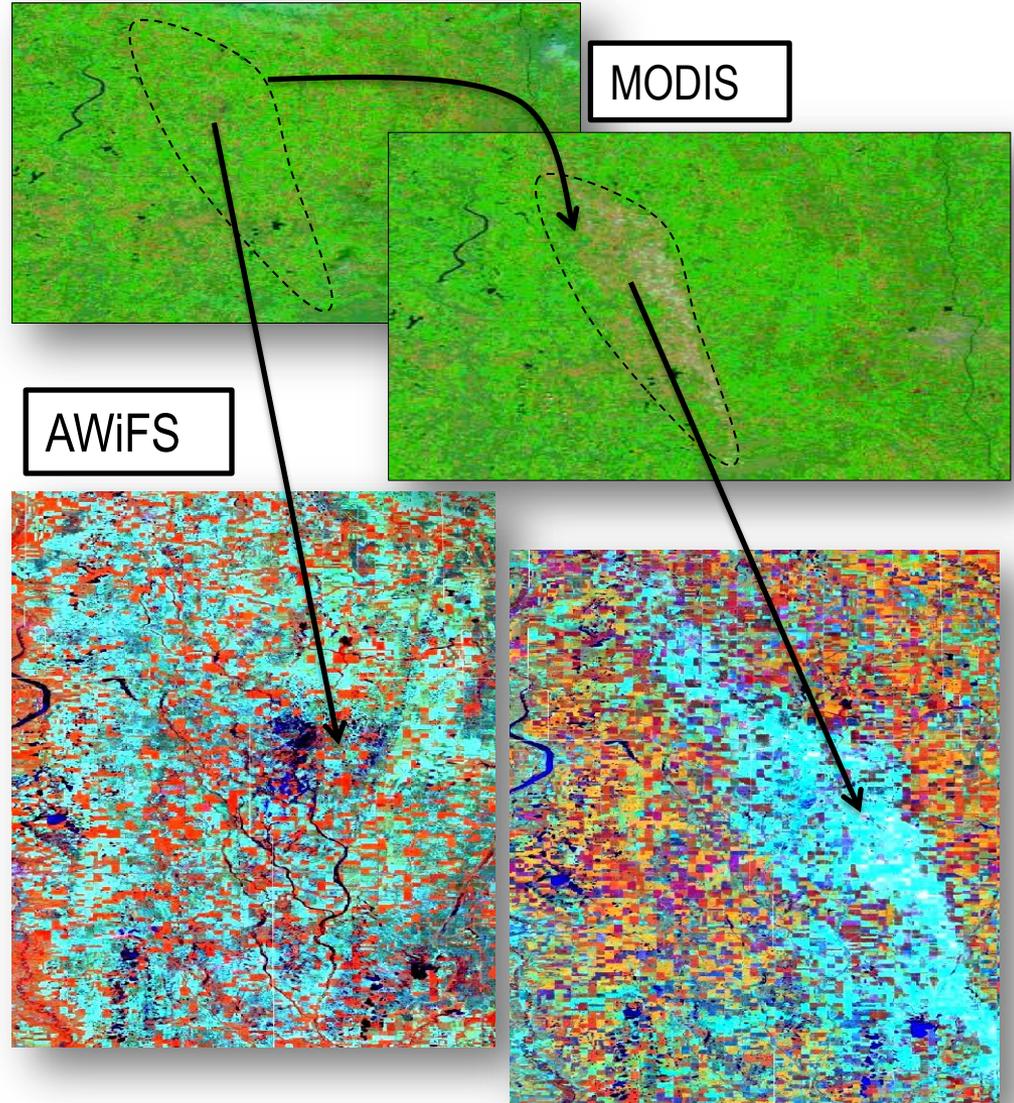
Challenges in Global Land Cover Data

- Different spatial and temporal resolutions
- Variable Coverages
 - Global, National, Regional
- Different classification methodology
- Different class definitions
- Variable update frequency
- GlobCover
 - 300m, global, LCCS,2009
 - Preceded by GLobCover (2006)
- MODIS-MCD12Q1
 - 500m, global, IGBP, 2004
 - Preceded by MOD12Q1
- Regional
 - Corine, AfriCover
- NLCD
 - 30m, US, 2006, 1 crop class
 - Preceded by NLCD 2001
- Local
 - CDL, GAP



Dynamic monitoring of biomass

- Identify and characterize changes
 - Phenological (crop growth)
 - Land use (crop type)
- Develop analytical capabilities to monitor crop biomass from regional to global scales
- Provide operational support



Technical Challenges

- **Large spatial and temporal extents**

- High volume data throughput
- Inconsistent model parameters resulting from geographic heterogeneity

- **Diversity of data**

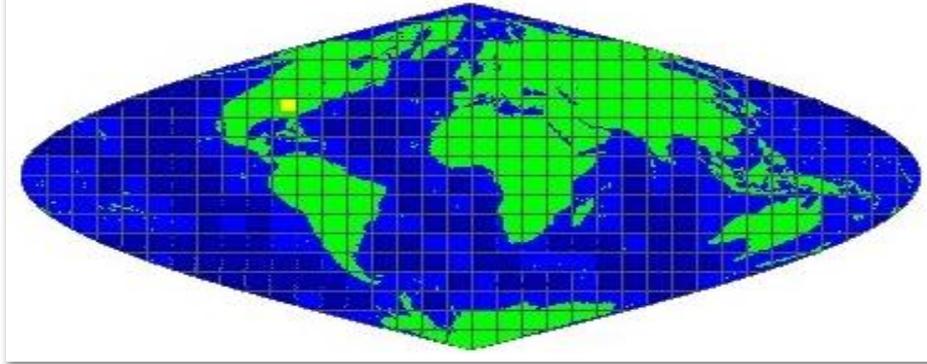
- Multi-resolution (spatial and temporal)
- Overlapping data products
- Disparate attributes (projections)
- Noise (such as cloud cover)

- **Change is natural**

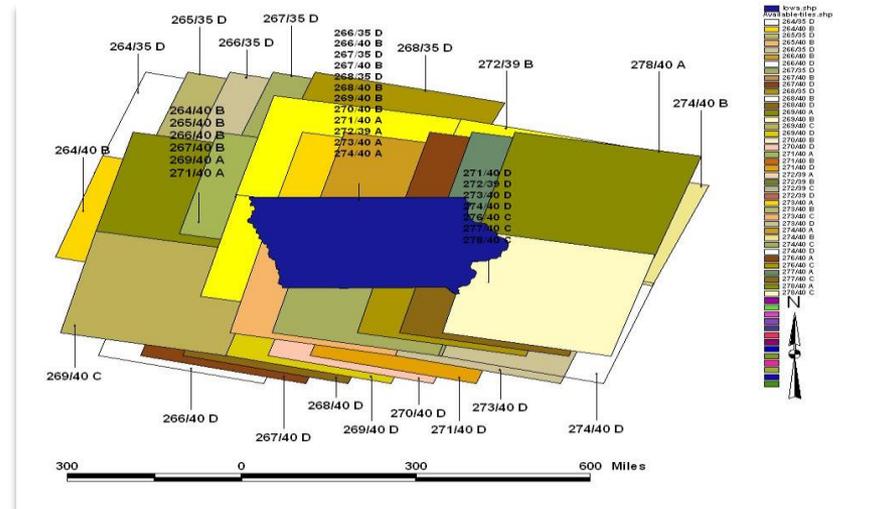
- Crop phenology (changes in NDVI are expected due growing nature of crop)
- Changes temporary (rain, snow, dry)

- **Defining agricultural subclasses (such as row crop to corn)**

- Insufficient and noisy ground truth data



MODIS (4800x4800), 3 Bands,
250m, 16-days, 2001-2009, 27GB;
432 products



AWiFS (12,300x12,000), 4 Bands,
56m, May-Sept. 2008, Iowa, (USDA),
130 Products

Geocomputation based strategy

Design and develop a robust and scalable spatiotemporal data mining framework utilizing high resolution spatial and temporal data streams (MODIS and AWiFS)

Preprocessing

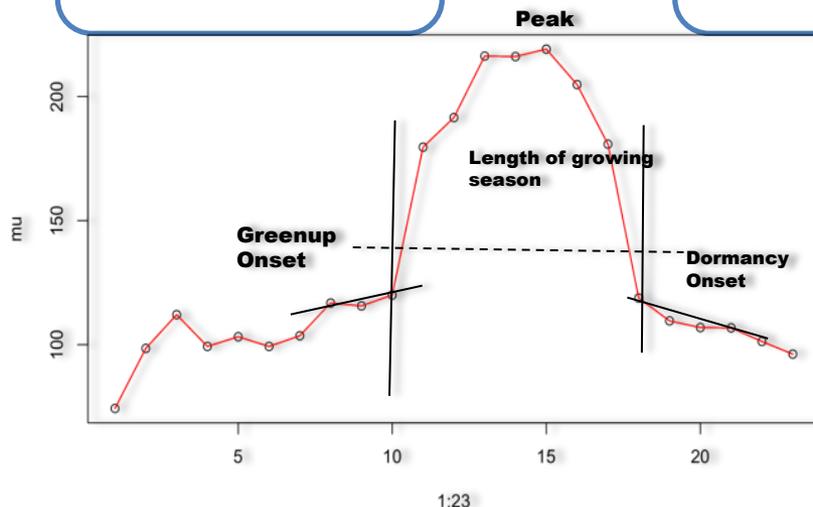
- Reprojection
- Atmospheric corrections
- Time series filtering

Change detection

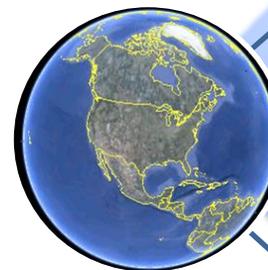
- Time series prediction
- Unsupervised multidimensional geospatial image clustering

Change characterization

- Classification
- Phenology-based
- Crop Type-based



Key features of crop phenology



Google Earth

NASA World Wind

Other thin clients

Wide area biomass monitoring in near real time is becoming a reality

- MODIS Tile (4800x4800 pixels)
 - 42,76,383 time series
- FROST: An SGI Altrix ICE 8200 Cluster at ORNL
 - 128 compute nodes each with 16 virtual cores and 24 GB of RAM
- Multicore (multithreaded) and Distributed (message passing) computing strategy

Serial

- 41,105 seconds (11.4 hours)

Threads (16)

- 5,872 seconds (1.6 hours)

MPI (96 nodes)

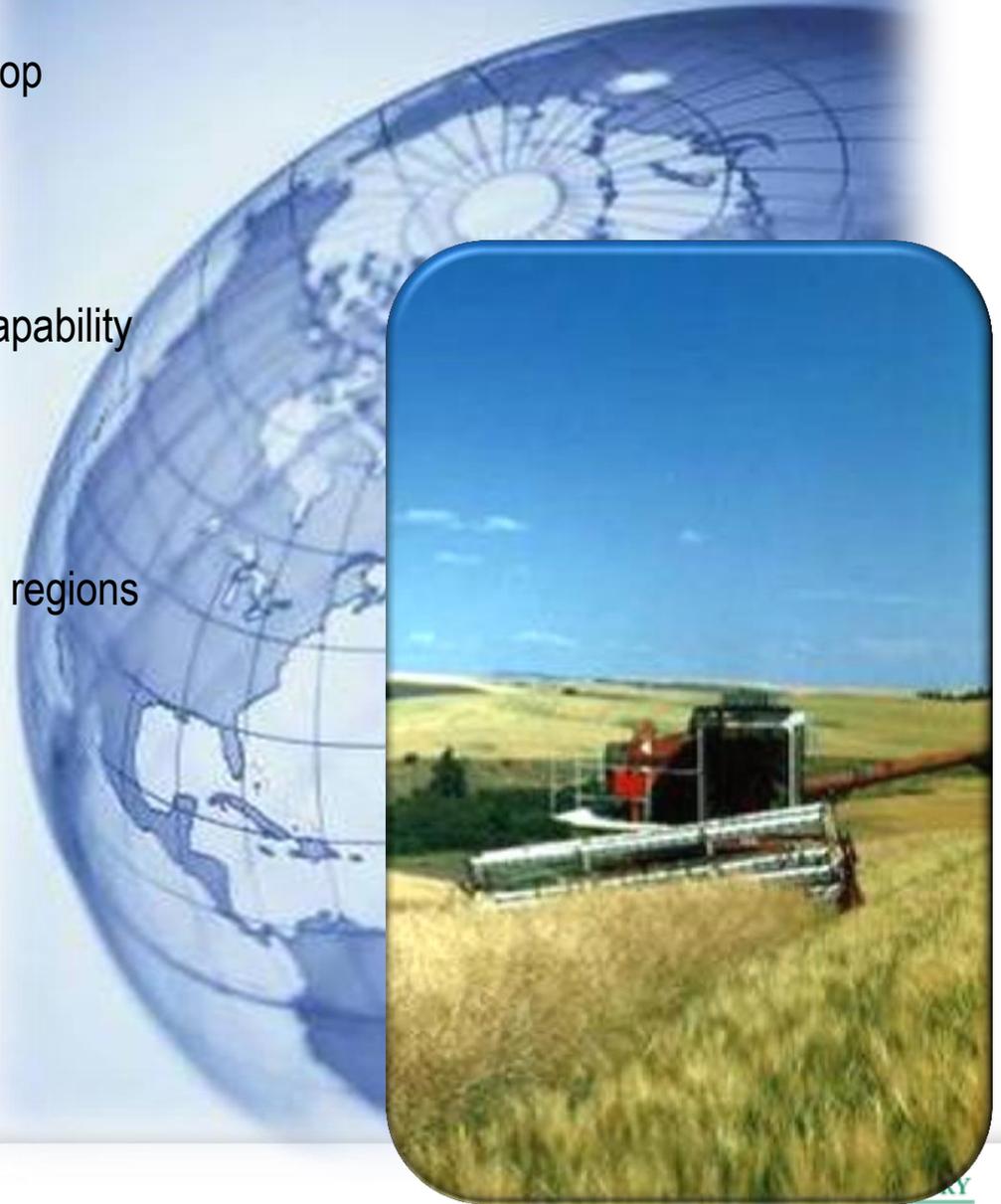
- 604 seconds (10 minutes)

MPI + Threads (1536 cores)

- 34 seconds

Benefits

- Provides continuous monitoring system for crop biomass to address
 - Natural changes
 - Disruptive changes
- Advanced characterization and forecasting capability
 - Trends in land use and crop patterns
 - Yield estimation
 - Early warning systems
- Computationally scalable to large geographic regions
- Flexible and extendable to
 - Crop stress
 - Evapotranspiration
 - Forest Fire Monitoring
 - Natural Hazards



Bioenergy Knowledge Discovery Framework

Databases

- ~1600 curated spatial data sources
- 1500 downloadable data, 1450 Map Services, 257 Field trial datasets
- Billion Ton update Data Explorer and Download tools

Knowledge Bases

- ~ 300 curated resources describing models and important journal articles
- 263 Web resources

Models

- Resources links to 38 domain models
- Commodity routing model
- Infrastructure logistics planning model

Tools

- Geospatial and Graphical Visualization
- Spatial Analysis and Querying
- Faceted Search and News Feeds

What can one do with Bioenergy KDF?

Search

- Data, publications, documents, and models
- Subject matter experts*

Contribute

- Data, publications, documents, and models
- Provide feedback and requirements

Associate

- Data, knowledge, and people (publications with data; documents with documents)

Analyze

- Spatial analysis with geographic data
- Scenarios with domain specific models

Share

- Data or analysis results with everyone, selected users (groups), or individuals based on contributor's preference

Visualize

- Spatial overlays and geographic visualization
- Conventional visualization (Tables, graphs, and charts)

Collaborate

- Organize special interest groups
- Communicate on a forum*

Bioenergy Knowledge Discovery Framework

- Brings together all relevant data
 - Helps move towards standardized datasets, assumptions, and guidelines
- Increases accessibility of data to researchers
- Provides access to analytic tools/models
- Provides secure collaboration communities to share research, data, and in-progress publications

<https://bioenergykdf.net/>

The screenshot shows the homepage of the Bioenergy Knowledge Discovery Framework (KDF). At the top, there is a navigation bar with the KDF logo, the text "BIOENERGY KNOWLEDGE DISCOVERY FRAMEWORK", and "U.S. DEPARTMENT OF ENERGY". To the right of the logo is a search bar and links for "Register", "Contact Us", and "Sign-In". Below the navigation bar are links for "OVERVIEW", "TOOLS & APPS", "MAP", and "BIOENERGY LIBRARY".

The main content area features a green banner with the text: "The Bioenergy KDF provides access to collaboration, data management, analysis, and visualization tools designed to support bioenergy infrastructure research. By integrating bioenergy spatial data with socioeconomic and industrial factors, the KDF helps facilitate informed decision making by providing researchers, policy makers, industry, and consumers. To get started, access or add your bioenergy data and use the KDF's tools to synthesize, analyze, and visualize information in a spatially integrated manner." To the right of this text is a small video player with the caption "Feed more and watch a short webthrough video".

Below the banner is a section titled "What Would You Like To Do?" with four interactive buttons: "ADD DATA", "FIND DATA", "VISUALIZE DATA", and "FIND TOOLS & APPS".

The "Featured Content" section includes a card for "Land-Use Change and Bioenergy Workshop" submitted by "Biodiesel Refinery Capacity" and "Optimal Biorefinery Locations and".

The "News" section lists several articles with dates: "Apr 9, 2013: US Transportation, Agriculture Secretaries to sign 'Farm to Fly' agreement at ABLC.", "Apr 12, 2013: BioAmber sets price range for \$128M IPO.", and "Apr 11, 2013: Dinnien blasts RFS Reform Bill: 'Nothing'".

The "Events" section lists: "Monday, May 20, 2013 - Thursday, May 23, 2013: 2013 Project Peer Review" and "Friday, July 19, 2013: Bioenergy KDF Release".

The "Popular Data Sets" section lists: "Algae/Terrestrial Feedstock Land-Use Competition Acres of Pasture", "Algae/Terrestrial Feedstock Land-Use Competition from 2007 Census of Agriculture", and "Algae/Terrestrial Feedstock Land-Use Competition from 2010 CDL".

At the bottom right, there is a blue box with the text "WHY SHOULD YOU JOIN?" and an image of a man in a cap.