GIScience Driven R&D: Interdisciplinary GIST Group at ORNL

Presented at

Purdue GIS Day

Marie Urban

November 7, 2014

www.ornl.gov/gist

Acknowledgment
Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC for the U. S. Department of Energy under contract no. DEAC05-00OR22725.
Overview of presentation

Geographic Information Science and Technology

• **About us**
  – Three scientific focus areas
  – Organization and programmatic thrusts

• **Highlights of selected applications**
  – Population assessment at scale
  – Critical infrastructure mapping
  – Energy assurance and climate impacts
  – Scalable geospatial applications
- $1.5B budget
- $750M modernization investment
- 4,400 employees
- 3,200 research guests annually
- World's most intense neutron source
- World-class research reactor
- Managing billion-dollar US ITER project
- Nation's most diverse energy portfolio
- Nation's largest materials research portfolio
- Forefront scientific computing facilities
- 50 new Purdue alum at ORNL between 2002-2014
Oak Ridge National Laboratory:
Deliver scientific discoveries and technical breakthroughs that will accelerate the development and deployment of solutions in clean energy and global security, and in doing so create economic opportunity for the nation

Signature strengths
- Computational science and engineering
- Materials science and engineering
- Neutron science and technology
- Nuclear science and technology
Education Programs to energize your career

Postdoctoral programs

Graduate degree programs/visiting faculty

Undergraduate/post-master’s programs

K–12 programs

Postdoctoral program

Fellowships

GO! Graduate Opportunities

GEM Fellows

Bredesen Center@UT

ASTRO

Undergraduate Laboratory Internship

Post-Bachelor’s/Post-Master’s Research

HERE

Lab Technology Program

Nuclear Engineering Science Lab Synthesis

ARC and ORNL summer programs

Science City Adventure

Science Saturdays

ORNL Researcher

HERE Lab Technology Program

ASTRO Undergraduate Laboratory Internship

GO! Graduate Opportunities

GEM Fellows

Bredesen Center@UT

Postdoctoral program

Fellowships
An interdisciplinary group of > 50 researchers
Fostering advances in the nexus of GIScience and computing
One of the oldest GIScience R&D groups and uniquely focused within DOE
Mission

Advance research, development, and applications of geographic information and analysis systems to support the nation’s energy, environment, and security programs, from local to global scales.

Managed by UT-Battelle for the Department of Energy
Programmatic Focus Areas

Population and Critical Infrastructure Modeling
- Population distribution in space and time
- Socioeconomic characterization
- Event specific population response

Energy Assurance
- Energy and water resources assurance and reliability
- Modeling and visualizing the “Smart Grid”
- Bioenergy and renewable energy integration

Transportation M&S
- Multimodal route optimization
- Intelligent evacuation planning

Emergency Preparedness and Response
- Time critical decision support
- Disaster risk analysis

Climate Change Science
- Climate extremes and infrastructures
- Climate change and population response
- Energy and national security

Earth Science Informatics
- Real time data integration
- Data and knowledge management
## Community leadership: Organized workshops

<table>
<thead>
<tr>
<th>Field</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GIScience</strong></td>
<td>Role of volunteered geographic information in advancing science</td>
</tr>
<tr>
<td><strong>ACM SIGSPATIAL</strong></td>
<td>Analytics for Big geospatial data</td>
</tr>
<tr>
<td><strong>IEEE ICDM</strong></td>
<td>Spatial and spatiotemporal data mining (SSTDM)</td>
</tr>
<tr>
<td></td>
<td>Knowledge discovery using cloud and distributed computing platforms</td>
</tr>
<tr>
<td><strong>ACM KDD</strong></td>
<td>Knowledge discovery from sensor data (SensorKDD)</td>
</tr>
<tr>
<td><strong>IEE/ACM Supercomputing</strong></td>
<td>Petascale (Big) data analytics: Challenges and opportunities</td>
</tr>
<tr>
<td><strong>IEEE ICCV</strong></td>
<td>Computer Vision for Converging Perspectives</td>
</tr>
</tbody>
</table>
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.
Spatial refinement of LandScan Global
Addis Ababa, Ethiopia

- 2 Xeon Quad core 2.4GHz CPUs + 4 Tesla GPUs + 48GB
- Image analyzed (0.3m)
  - 40,000x40,000 pixels (800 sq. km)
  - RGB bands
- Overall accuracy 93%
  - Settlement class 89%
  - Non-settlement class 94%
- Total processing time
  - 27 seconds
Rapid Scene Analysis
Play uavrun1output.avi
Kacha Garhi Camp, Pakistan

- Established 1980 for Afghan Refugees
- QuickBird (2004 and 2009, 4B, 2.4m)
Neighborhood mapping: From local interactions to global realizations

Damascus, Syria

- Very loosely structured
- Historical ethnic quarters/neighborhoods
- Poor residents currently being displaced in some areas with urban development/tourism

- Unstructured Settlements
- Lowest to lower middle income
- Rural migrants

- Formal Urban Planning
- Typical Urban Services
- Middle to Upper Income

Geographic Information Science and Technology
Population Density Data from Open Source

- Population/1000 ft²
- 50+ structural facility categories in 8 land use classes
- Documented data sources and methodology with traceable provenance
- Open source collection from reputed sources (> 14K)
  - Academic journals, official government statistics, corporate and university webpages, tourism brochures.
  - Utilizing other sources such as GeoCommunity, Wikipedia, Panoramia, and Wikimapia.

- Spatial resolution/extent
- Temporal resolution
  - Diurnal
  - Workweek/weekend
  - Episodic/holidays/special event
  - Seasonal
- Political and socio-economic characteristics included
  - Ethnic
  - Religious
  - Racial
  - Political affinity
  - Economic strength

https://pdt.extranet.ornl.gov/
Retaining Data Provenance is Critical

Geographic Information Science and Technology
World Spatio-Temporal Analysis Mapping Project

- Illuminate patterns of geopolitical and strategic importance from over 30+ data sources like the CIA World Fact book, World Bank, UN, and Correlates of War.
  - ~250 countries and geographic entities
  - ~2100 socioeconomic and demographic attributes
  - 25 years

- Develop geovisualization, geopolitical ontologies, and geonarratives

- Enable analysts to ask questions from data and just not look for answers
Critical infrastructure data development

Geographic Information Science and Technology

We develop and maintain spatially enabled, foundation level data for a number of critical infrastructures for research and operational communities.
Readiness, Response, and Recovery

Geographic Information Science and Technology

Integration of crowdsourcing and social media data

Population mobility and advanced evacuation modeling and simulations

Global security infrastructure protection support

Augmentation of EARSS mobile and web services

Special events population model

High Performance SAR data processing for disaster response

Managed by UT-Battelle for the Department of Energy
Climate Change Science

- Feedback among climate, land use, and population distribution
- Climate induced hazards and infrastructure impacts
- Spatiotemporal translation of regional climate impacts on local decisions
Population mobility modeling framework

Geographic Information Science and Technology

- Quick response emergency evacuation system to analyze evacuation scenarios any place in the world.
- Ability to perform time variant modeling such as daily commuter traffic (engineering) modeling.

Flexible and global scope

- High resolution population (LandScan Global and LandScan USA) and street networks (NAVTEQ and OSM)
- Mesoscale (link based) to microscale simulations (TRANSIMS and MITSIM)

Scalable data integration and modeling

- Web application using WebGL and data streaming techniques to display vehicular movement across the respective networks.
- Ongoing integration with mobile platforms and cloud computing.

Dynamic visualization and access

Managed by UT-Battelle for the Department of Energy
Climate influences the spatiotemporal distribution of human populations

- Understanding human exposure to climatic extremes and the local, regional and global distribution of resources provides insights into potential migration and resettlement patterns.
- Bangladesh: 162 Millions population, 27 millions household.
- Full scale Bangladesh simulation is estimated to require 20,000 CPU hours or 400 GPU hours for a single simulation.

Runtime for Different Numbers of Agents on GPU Based Parallel Computing Platforms (ORMAC ABM Platform)
• Provides a mechanism for understanding how people make decisions related to travel.

• The foundation is a least cost analysis – people make decisions to minimize costs.

• People behave fairly predictably in this regard, especially in aggregate, but may change course in certain situations.
Energy Assurance

Geographic Information Science and Technology

- Cyberphysical systems in energy
- Spatiotemporal assessment of renewable energy potential
- Early warning systems for climate extremes
- Bioresource monitoring for energy security
- Geographically scalable spatiotemporal optimization for energy supply chain
  - PRISM
Design and develop a robust and scalable spatiotemporal data mining framework utilizing high resolution spatial and temporal data streams (MODIS and AWiFS)

Geocomputation based strategy

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
- Greenup Onset
- Dormancy Onset
- Peak
- Length of growing season
Wide area biomass monitoring in near real time is becoming a reality

- MODIS Tile (4800x4800 pixels)
  - ~23 million locations/time series
  - 161 time steps (bi-weekly over 7 years)

- 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Time Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serial</td>
<td>41,105 seconds (11.4 hours)</td>
</tr>
<tr>
<td>Threads (16)</td>
<td>5,872 seconds (1.6 hours)</td>
</tr>
<tr>
<td>MPI (96 nodes)</td>
<td>604 seconds (10 minutes)</td>
</tr>
<tr>
<td>MPI + Threads (1536 cores)</td>
<td>34 seconds</td>
</tr>
</tbody>
</table>
EARSS: Hurricane Irene (8/26, 1625)
Spatial Analysis and Decision Assistance

SADA provides a spatially enabled modeling environment that integrates:

- advanced methods in sample design,
- risk assessment,
- remedial design, and cost-benefit analysis

Risk communication, uncertainty assessment and geovisualization are a central focus of the work.

18K+ users

www.sadaproject.net
We support a global user community

**LandScan Global**
- Spatial resolution of 30 arc seconds (~1km); Ambient population.
- ~1350 registered organizations worldwide since 2005; 115 commercial licensees. Enterprise standard for NATO.
- Google search returns 225,000 web pages.

**LandScan USA**
- Spatial resolution of 3 arc seconds (~90m); Nighttime and daytime population.
- 836 organizational requests, to support over 110,734 users, for the HSIP Gold 2010.

**Bioenergy KDF**
- Bioenergy infrastructure analysis with 1600 Datasets; 1450 Map Services.
- 13,000 Visits from 134 Countries; Avg. Time on Site: 6:00. (Since Feb. 2013)
- 1156 registered users since January 2011.
We support a global user community

Geographic Information Science and Technology

WebTRAGIS
- Multi-Modal Routing Analysis and Visualization Tool
- 220 Visits Since Oct. 2013, Avg. Time on Site: 4:10
- 55 Registered users since September 2012.

Population Density Tables (PDT)
- Open Source Data Collection for Population Density Modeling
- 55 Registered users since November 2011.

Centralized Used Fuel Resource for Information Exchange (CURIE)
- 2,235 Used Fuel related Documents, Data, and Images.
- 4,800 Visits from 65 Countries; Avg. Time on Site: 7:24. (Since Feb. 2013)
- 112 registered users since January 2013.
Managed by UT-Battelle
for the Department of Energy

We support a global user community

Geographic Information Science and Technology

Energy Awareness Resiliency Standardized Services (EARSS)

- Infrastructure Disruption Visualization Tool
- 105 unique users during Hurricanes Irene and Sandy
- Re-publication from FEMA Geoportal to State Emergency Operation Centers

Homeland Security Extreme Weather Event Anticipation Tool (HEAT)

- Interfaces Climate Impacts with Critical Infrastructure Disruption Models
- 25 Evaluation Accounts with DHS, EPA and Local planners as part of Regional Risk Assessment Program

Production and Research Informatics for Strategic Materials (PRISM)

- Collaboration, Data Management, Analytics, and Visualization Platform supporting Strategic Materials Supply Chain Analysis community
- Organizational Experimental Accounts with DLA, DOE, and NGA
Oak Ridge Urban Dynamics Institute
Delivering transformational science and technology capabilities

- Science and informatics for energy and urban infrastructures
  - Data from individual components (sensors) of infrastructure networks (energy, water, transportation, telecommunication,...)
  - Data from users of infrastructure (human network)
- Characterization of the interaction between the human dynamics and integrated infrastructures
  - Discovering emerging behavior of urban systems over large spatial and temporal scales (at unprecedented resolution)
- Efficient data management, analysis, creation, and visualization of meaningful information within useful timeframe
- Developing interdisciplinary bridge between foundational R&D, operational communities, and industry
Data and computing for urban science

Big Spatiotemporal Databases

- Social Media
- LiDAR
- Text
- Images
- Sensors
- Video

Challenges (Increased I/O: 150X, Computation: $O(n^2)$-$O(n^3)$)

Smart Cities
- Sustainable mobility
- Energy efficiency
- Resilient infrastructure

Energy
- Transportation
- Electricity
- Water
- Buildings

Environment
- Pollution
- Air
- Water
- Noise
- Public health

Security
- Cyber
- Communication
- Disaster management

Feature Extraction
- SIFT/SURF
- Node degree
- ...

Data Fusion
- Multi-resolution
- Multi-sensor
- Multi-modal

Feature Selection
- PCA
- Compression
- ...

Knowledge Discovery
- Data Mining
- Machine Learning