Data Mining for Climate Change and Impacts

Karsten Steinhaeuser[1,2]
Auroop R Ganguly[1]

[1]Geographic Information Science & Technology Group (GIST)
Oak Ridge National Laboratory

[2]Computer Science & Engineering
University of Notre Dame

December 15, 2008

OAK RIDGE National Laboratory

Outline

Geographic Information Science and Technology

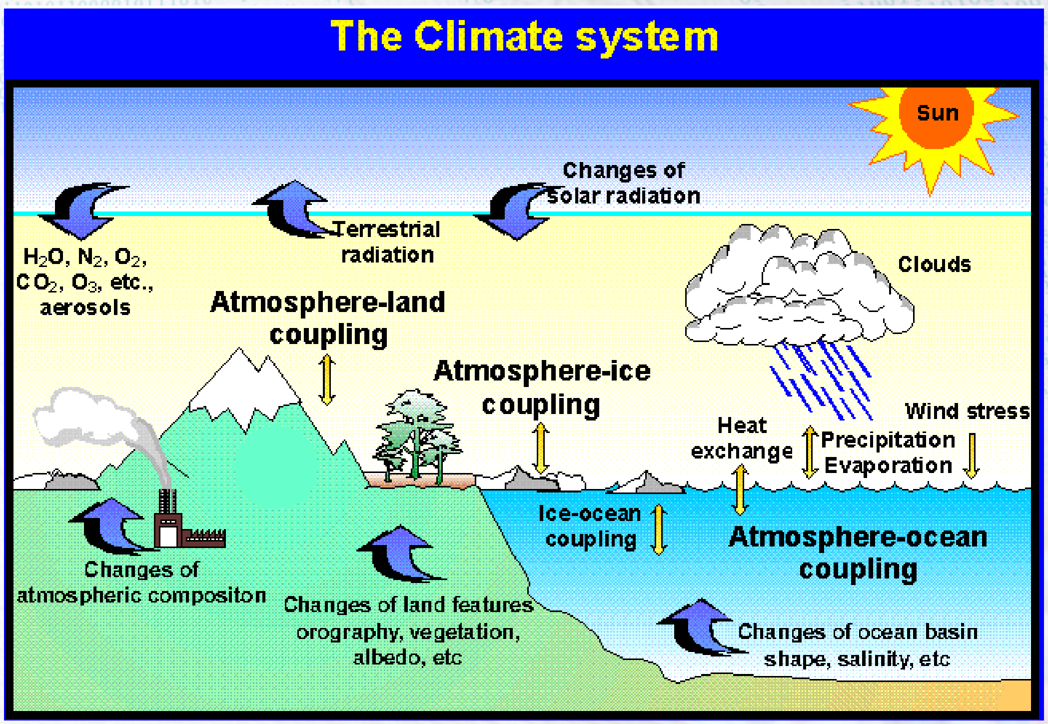
- Glossary & Introduction to Climate
- Framework for Climate Data Mining
- Relation with State-of-the-Art SSTDM
- Present Challenges
  - Computational
  - Algorithmic
- Example Applications
  - Correlation
  - Extremes
  - Uncertainty
- Take-Away Points

Glossary

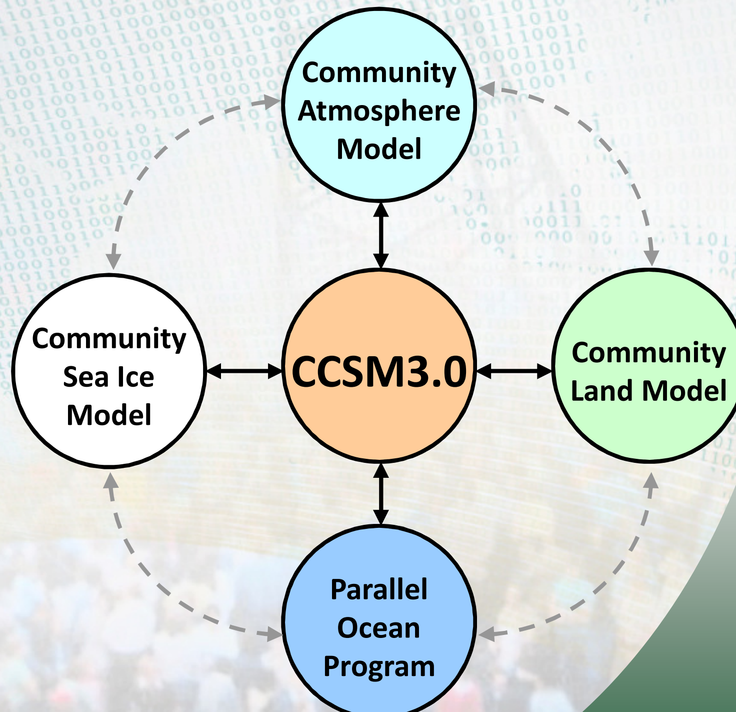
- **IPCC: Intergovernmental Panel on Climate Change**
  - 2007 Nobel Peace Prize
  - Produces "Assessment Reports", most recently AR4 (2007)
- **GCM: General Circulation Model**
  - Physical simulations of earth systems
  - Couple atmosphere, ocean, sea ice, land
- **CCSM3: Community Climate System Model**
  - Fully-coupled global climate model
  - One of primary models informing IPCC AR4
- **SRES: Special Report on Emissions Scenarios**
  - Hypothetical future states of the world
  - Define CO_2 emissions, economic development, population

How to Study Climate?

- **Problem: Observations difficult (historical, future)**
- **Solution: Climate Models**
  - **Equations describing fluid dynamics, heat transfer, etc.**



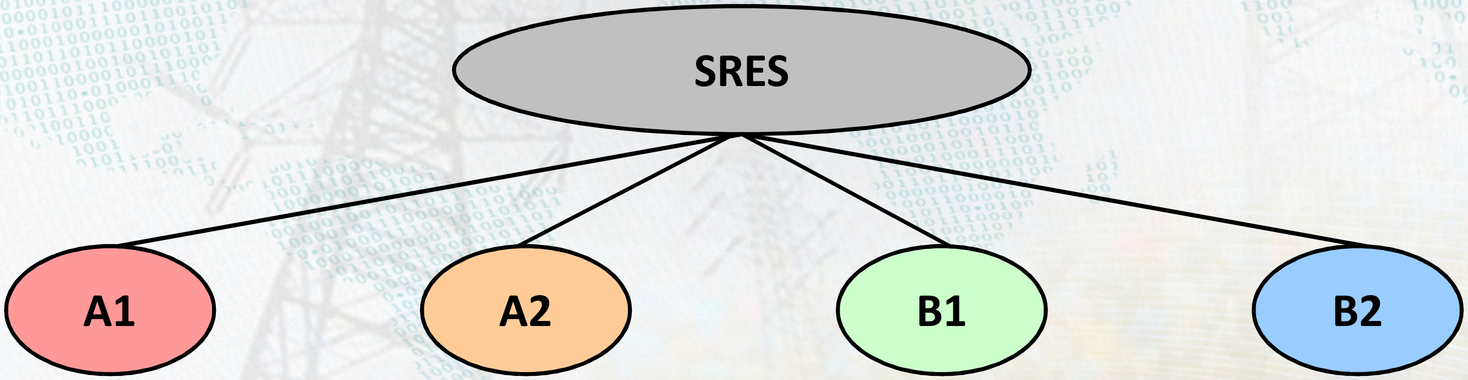
Source: CSIRO



How to Study Future Climate?

Geographic Information Science and Technology

- **Problem: What does the future hold?**
- **Solution: Test multiple hypotheses**
  - **Economic and technological development**
  - **Emissions, atmospheric composition**
  - **Population growth / decline, migration patterns**



A1
-Rapid economic growth
-High-technology
-Population peaks, declines
-Once considered extreme, but now quite realistic!

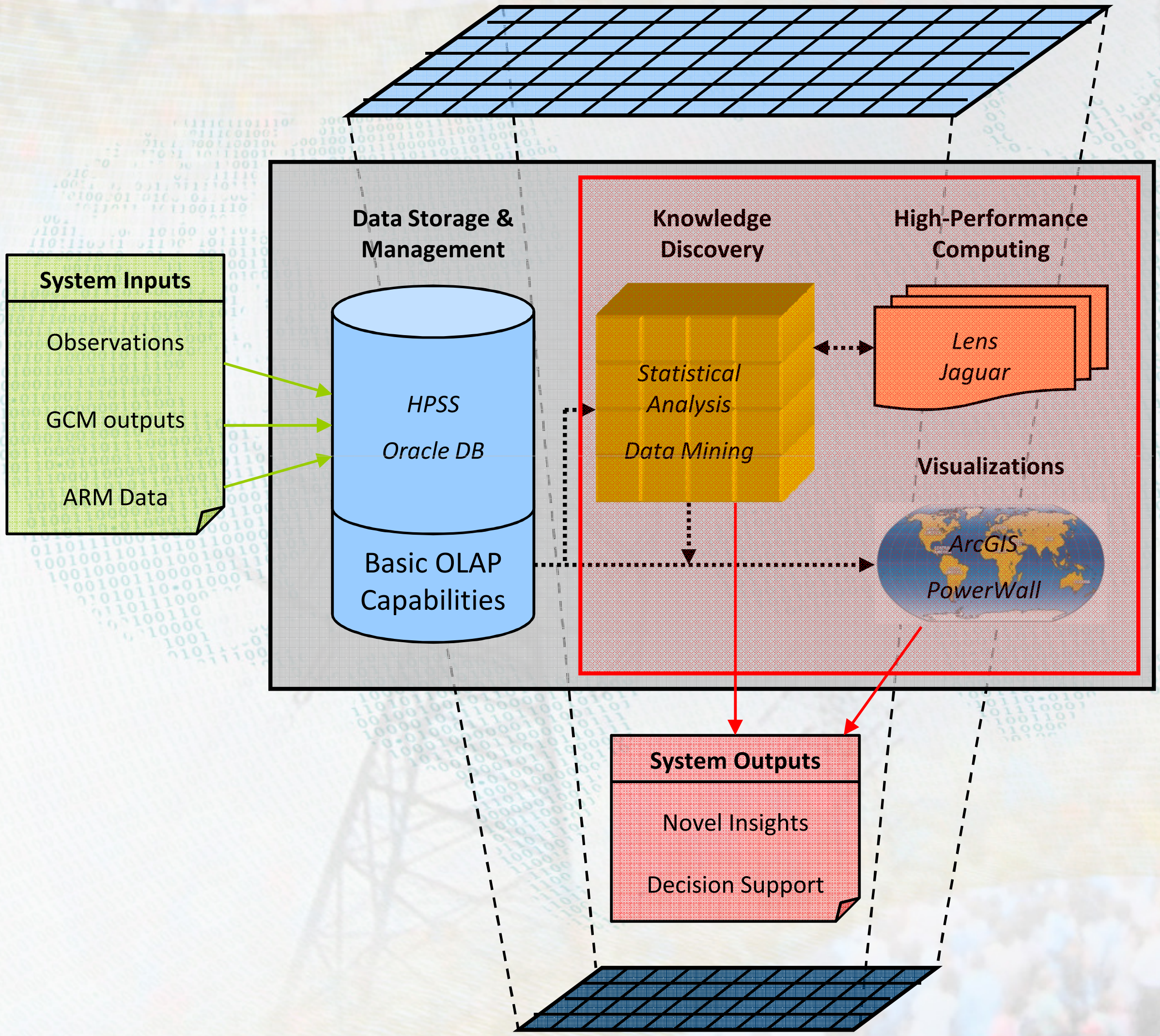
A2
-Heterogeneous world
-Self-reliant populations, regional economic grown
-Steady population increase

B1
-Rapid move towards service industries, more efficient technologies
-Population peaks, declines

B2
-Local solutions to economic, social, environmental issues
-More diverse innovation
-Slow but steady population increase

Framework for Climate Data Mining

Geographic Information Science and Technology



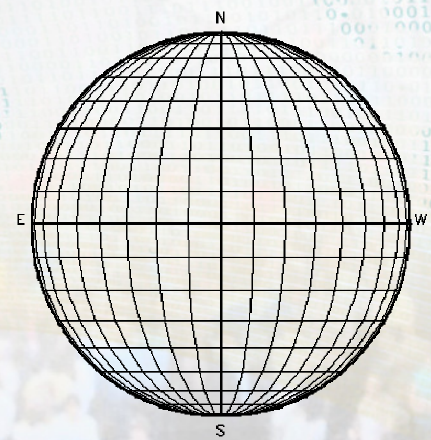
Global Climate Data
(100x100 km scale)

Large-Scale Climate Analysis Framework
(leverage existing Methods, Resources and Expertise)

Local Decision Making
(1-10 km scale)

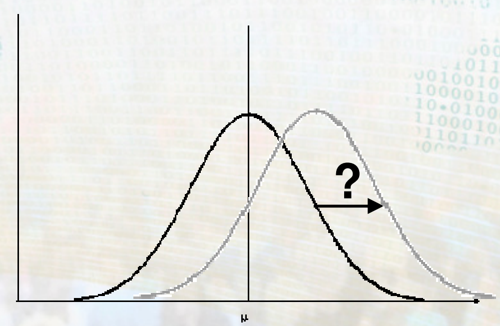
Relation with SSTDM

- **Complex Spatio-Temporal Data**
  - "First Law of Geography" – everything is related to everything, but near things more than distant
  - But also more complex relations
    - Long-range spatial dependence (teleconnection)
    - Non-linear dependence structure
    - Long-memory temporal processes
- **Measures of Correlation**
  - Mutual Information
  - k-Nearest Neighbors
  - Kernel Density Estimators



Relation with SSTDM

- **Classification and Prediction**
  - Datasets of climate variables $(x_{i1}, x_{i2}, \ldots, x_{id}, y_i)$ can be classified or regressed
  - But predictions often complicated by concept drift
    - Change in distribution of x_i's over time
    - Change in distribution of y over time
  - Non-linear dynamics produce chaotic behavior
- **Methods**
  - Spatial Autoregressive Models
  - Support Vector Machines
  - Neural Networks
  - LLE/ISOMAP for dimensionality reduction

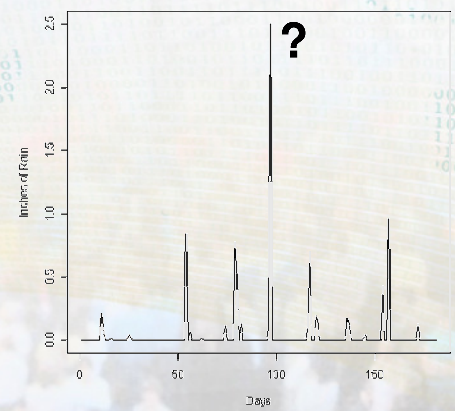


OAK RIDGE National Laboratory

Relation with SSTDM

Geographic Information Science and Technology

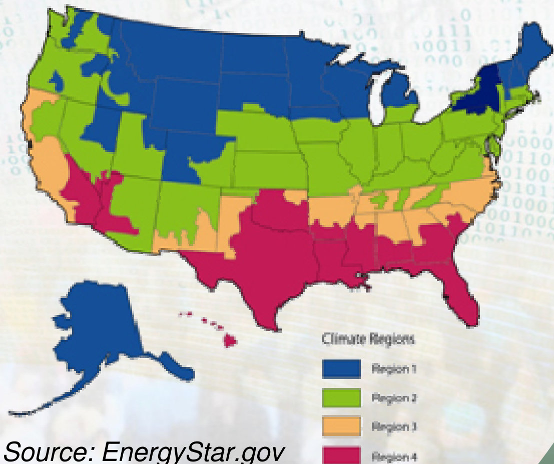
- Outlier Detection
  - Data point can be anomalous in space and/or time
  - But candidate outliers could be ambiguous
    - Legitimate extreme value caused by climate variability
    - Measurement error
    - Recurrence vs. non-repeatable patterns
  - Can we distinguish between the two?
- Methods
  - Extreme Value Theory
  - Time-Series Clustering
  - Change Detection



OAK RIDGE National Laboratory

Relation with SSTDM

- **Co-Location and Clustering**
  - Multivariate clustering to identify climate regions
  - But clusters may not be intuitive
    - Cover geographically disparate locations
    - Extend over both space and time
- **Methods**
  - k-Means Clustering
  - Principal Component Analysis
  - Singular Value Decomposition



Source: EnergyStar.gov

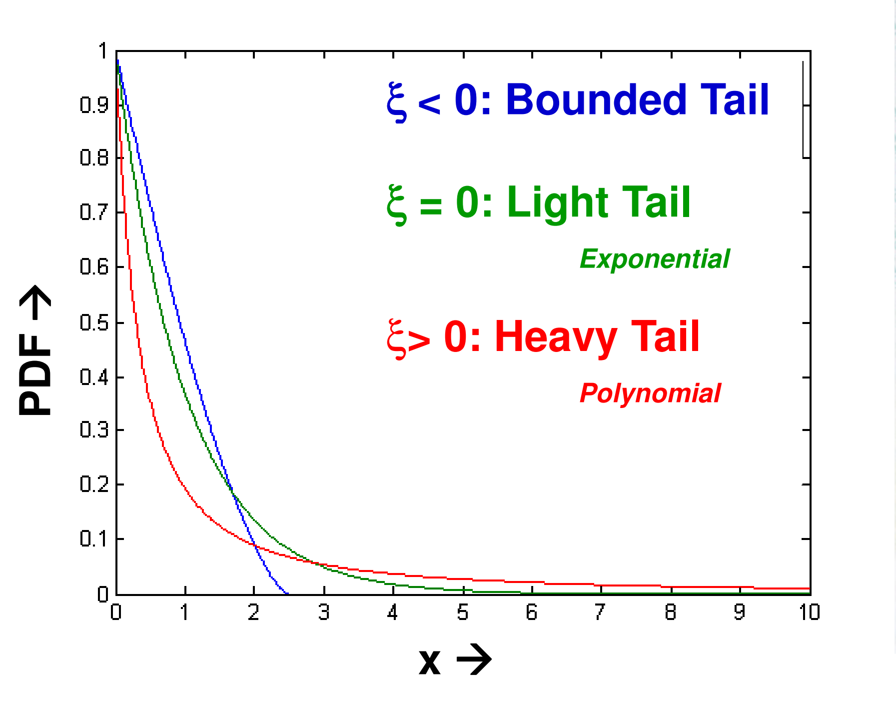
Present Challenges

- **Computational: Lots of Data**
  - Observation: NCEP/NCAR Reanalysis
    - 50+ climate variables
    - Daily/6-hourly intervals (1948-present), global (2.5ºx2.5º grid)
  - Climate Model: CCSM3.0
    - 100+ climate variables
    - Daily/6-hours intervals (1870-2100), global (1.4ºx1.4º grid)
    - 40 scenarios, multiple initial-condition ensembles
    - Example: 100 years daily outputs ~850GB (compressed)
  - IPCC AR4 incorporates 20+ climate models
- **Algorithmic: Spatial Dependence**
  - Data dependencies make parallelization difficult
- **Computational/Algorithmic: Non-Linear Processes**
  - Many current methods assume linear correlations
  - Non-linear analogs often computationally expensive

Example: Correlation & Extremes

Extreme Value Theory

$$F_{\sigma,\xi}(y) = \begin{cases} 1-[1+(\xi y/\sigma)]^{-1/\xi}, & 1+(\xi y/\sigma)>0, \xi \neq 0 \\ 1-e^{-y/\sigma}, & \xi = 0 \end{cases}$$



Plots courtesy of "The Mathworks"

EXCEEDENCES OVER THRESHOLD

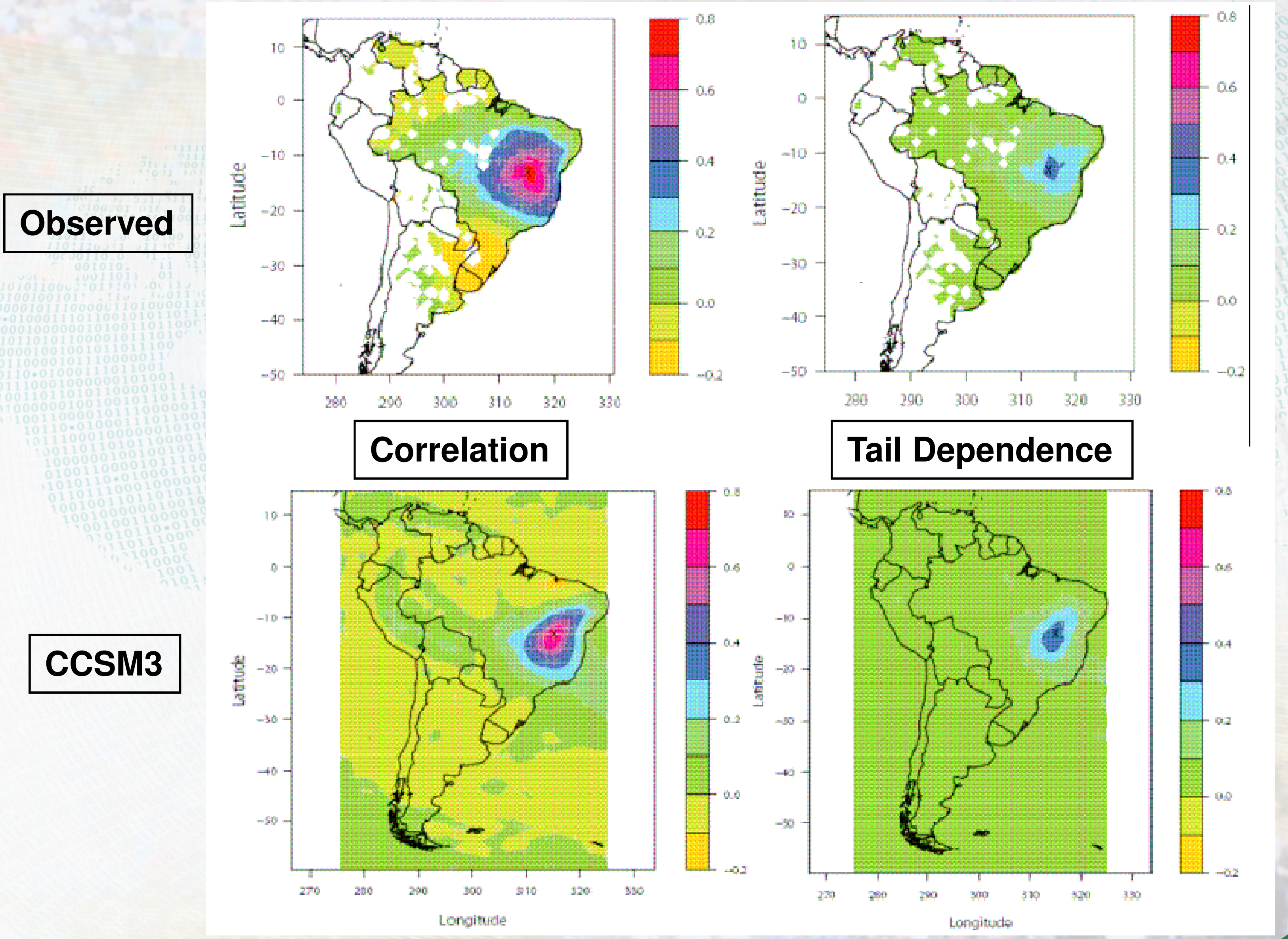
Prob. (X – u | X > u)

PARAMETERS ESTIMATED BY MAXIMUM LIKELIHOOD

Scale Parameter: σ

Shape Parameter: ξ

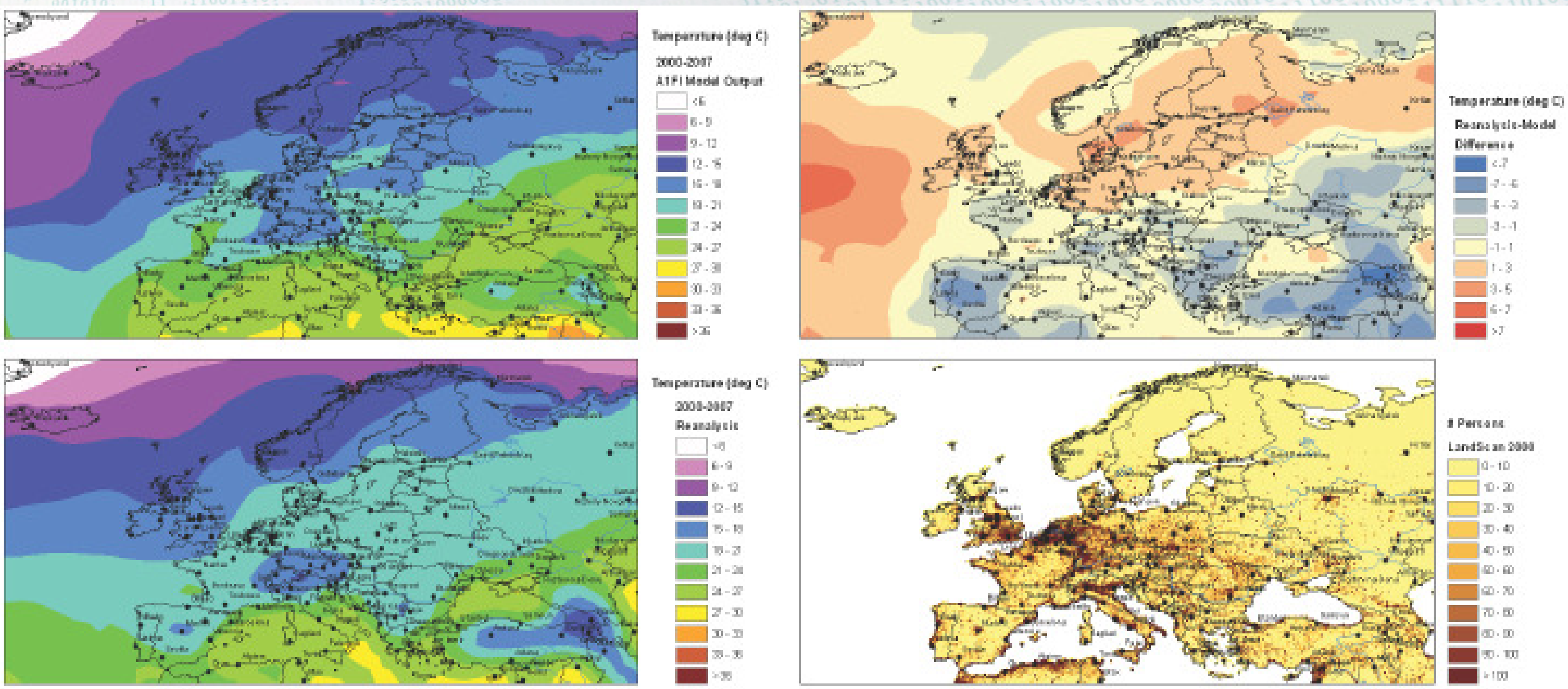
Example: Correlation & Extremes



Example: Extremes & Uncertainty

Heat Waves over Europe

- Annual 3-day nighttime minima event
- Compare model outputs with observations (bias)



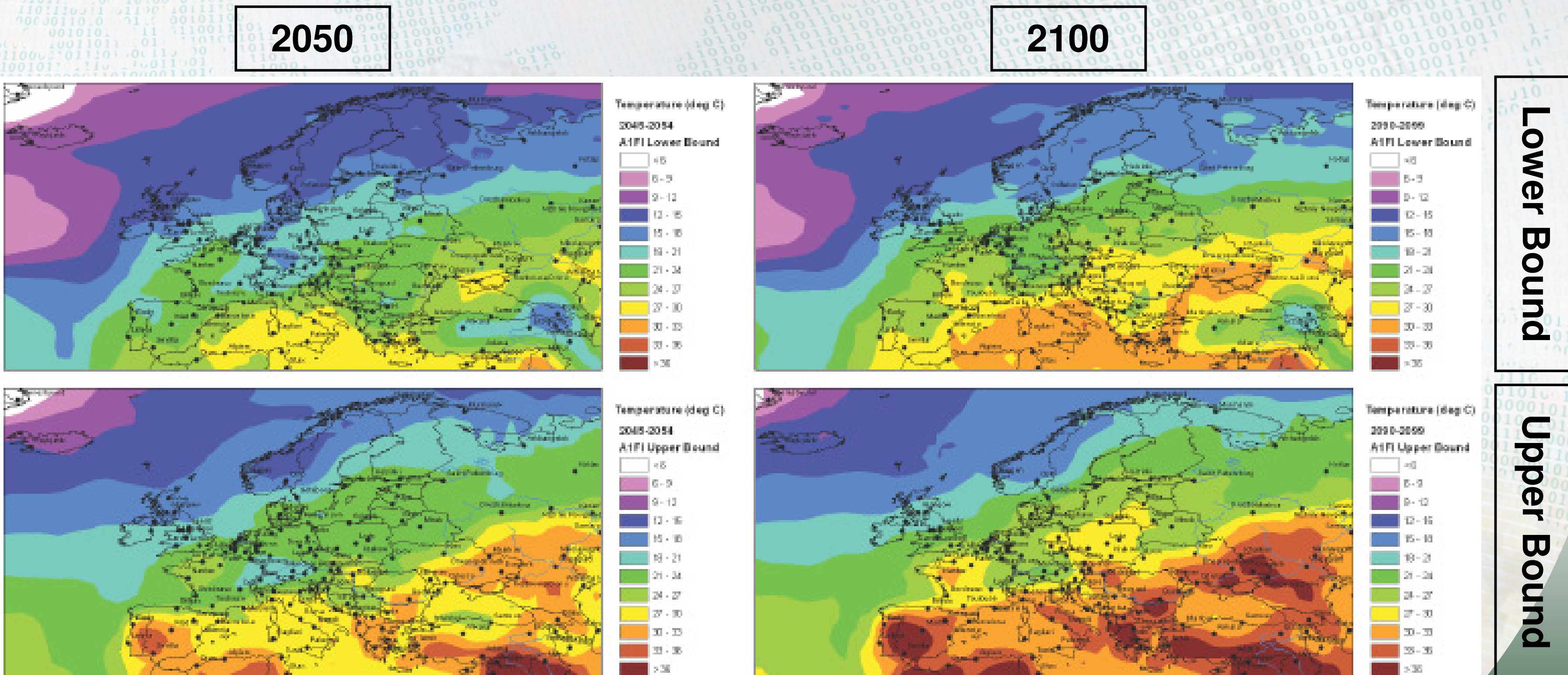
OAK RIDGE National Laboratory

Example: Extremes & Uncertainty

Geographic Information Science and Technology

Heat Waves over Europe

- Compare model outputs with observations
- Correct projections for bias, assign uncertainty bounds

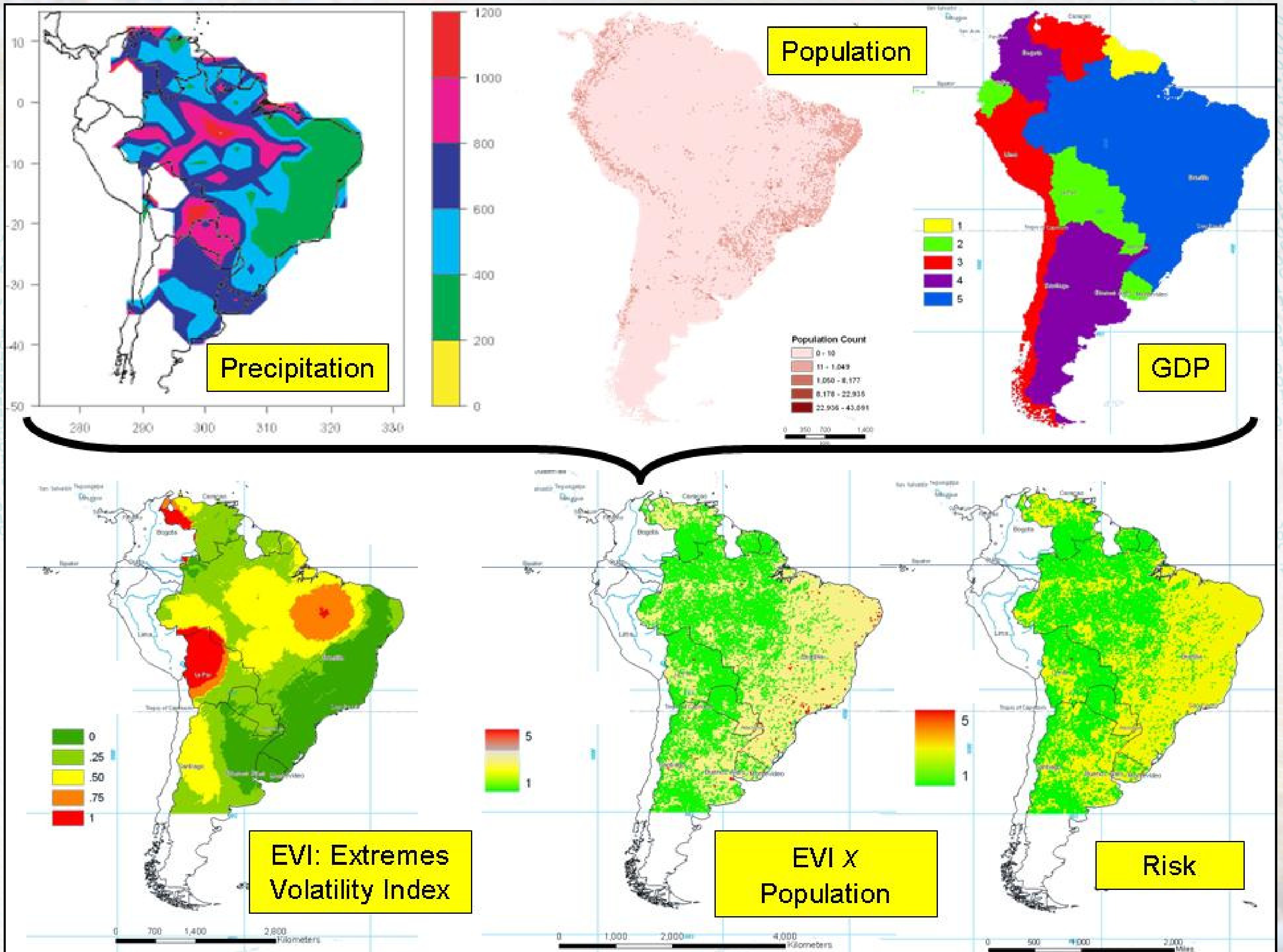


OAK RIDGE National Laboratory

Example: Climate Change Impacts

Example: Climate Change Impacts

Geographic Information Science and Technology



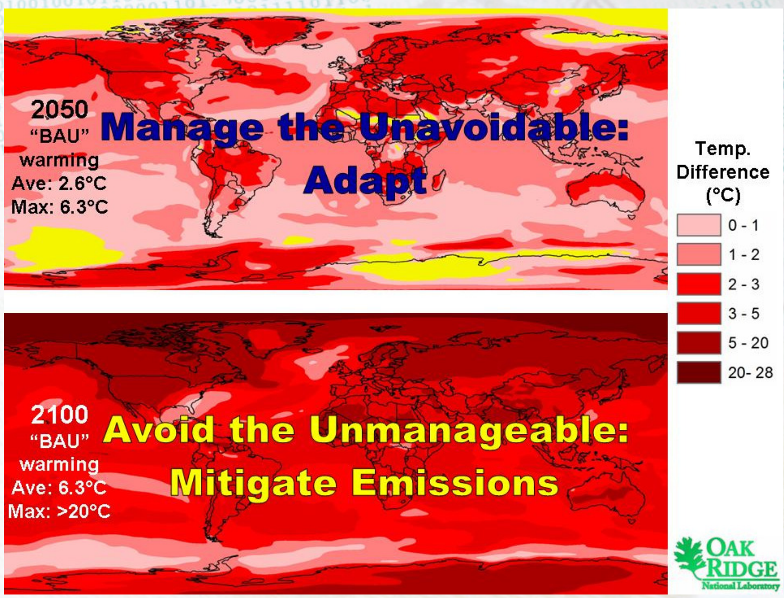
 Managed by UT-Battelle for the Department of Energy

Oak Ridge National Laboratory

Example: Climate Change Impacts

Geographic Information Science and Technology

- First ever Climate Change Wargame
  (Center for New American Security, July 2008)
- Role playing exercise involving world leaders, climate scientists
- Supporting materials produced by the GIST Group at ORNL
- Generated significant attention, publicity



Source: Jeff Tollefson's Nature Blog


CLIMATE CHANGE WARGAME CONSORTIUM

Center for a New American Security

Published online 5 August 2008 | Nature | doi:10.1038/454673a

naturenews

News

Climate war games

Role-play negotiations test the outcomes of global warming.

Jeff Tollefson (/news/author/Jeff+Tollefson/index.html)

More than 40 negotiators from Asia, Europe and the United States converged on Washington DC last week for what was billed as the first major war game involving global warming.

The Center for a New American Security, a Washington-based national-security think tank, gathered together climate scientists and experts in security, environmental policy and business for the role-playing exercise. Each was assigned to one of four teams, representing Europe, the United States, China and India, and charged with negotiating the best deal for their team.

Under the scenario, set in 2015, greenhouse-gas emissions are rising, and the latest climate models paint a grim picture of the future if business were to continue as usual. Extreme weather, including droughts, storms and floods, is on the rise. The United Nations is calling for international cooperation on emissions reductions, adaptation, disaster relief and shortages of crucial resources such as food and water.

Climate policy expert Eileen Claussen joined scientists and business executives, including Jim Owens (left), chief executive of construction tech firm Caterpillar, in role-playing negotiations.

M. NGAN/AFP/Getty

Conference organizers worked with scientists at Oak Ridge National Laboratory in Tennessee who provided a climate simulation up to 2100 based on the worst-case scenario proposed by the Intergovernmental Panel on Climate Change. This scenario mixes rapid growth with continued reliance on fossil fuels — something the organizers say is reasonable, given that emissions are currently trending higher than projected.

OAK RIDGE National Laboratory

Take-Away Points

- Climate research is a “hot” area (pun intended)
- Computer science has contributed computation, but little in terms of analysis
- Massive quantities of data posing novel challenges
  - Spatio-temporal dependence structures
  - Uncertainty and impacts analysis
- Data mining can make significant contributions
  - Apply existing methods in new context
  - Develop/implement new algorithms
  - High-performance data analysis
- Even simple techniques can lead to novel insights
- Window of opportunity to contribute to IPCC AR5

References

CCSM3 Model Output http://www.earthsystemgrid.org/

NCAP/NOAA Reanalysis Data http://www.cdc.noaa.gov/

Population Data http://www.ornl.gov/sci/landscan/

CNAS Climate Change Wargame http://www.cnas.org/node/149
and Materials http://www.ornl.gov/sci/knowledgediscovery/WarGaming/

Thanks to the Geographic Information Science & Technology (GIST) Group and the entire Computer Science & Engineering Division (CSED) at ORNL, especially M. Branstetter, D. Erickson, E. Parish, and N. Singh.

Acknowledgment This research was funded by the Oak Ridge National Laboratory (ORNL), managed by UT Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.