Geospatial Data Sciences

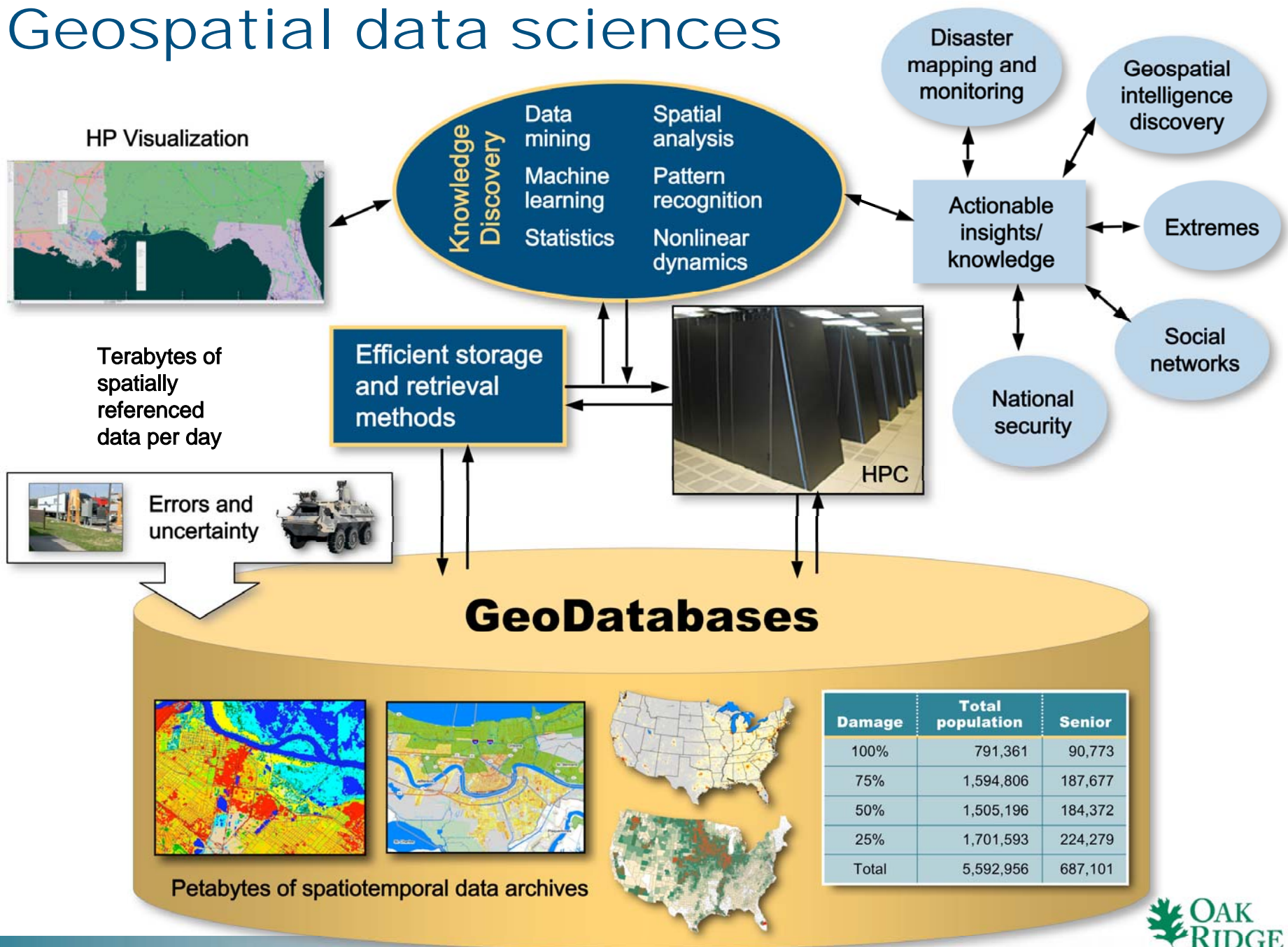Presented by

Ranga Raju Vatsavai
Auroop Ganguly
Budhendra Bhaduri

Geographic Information Science and Technology
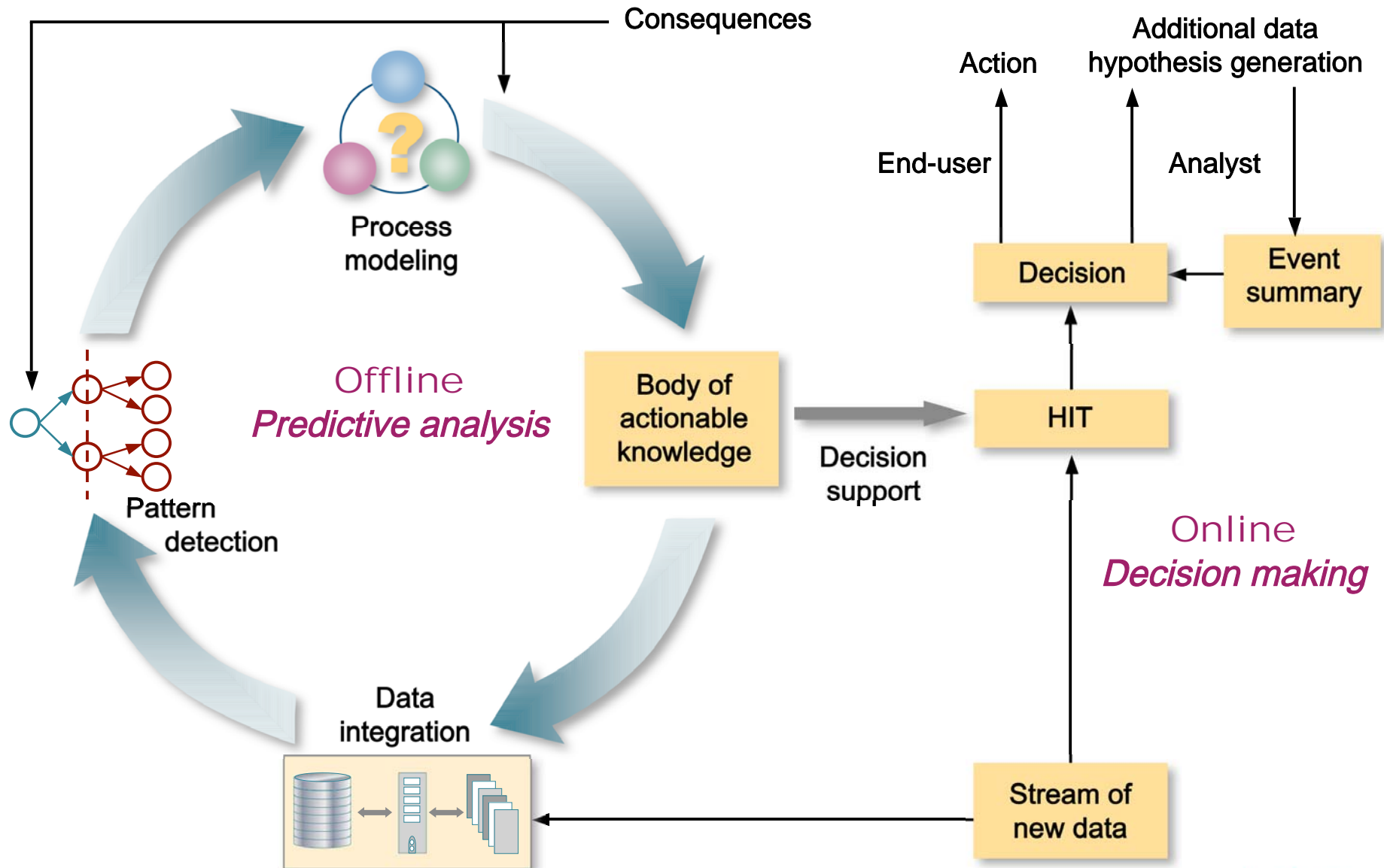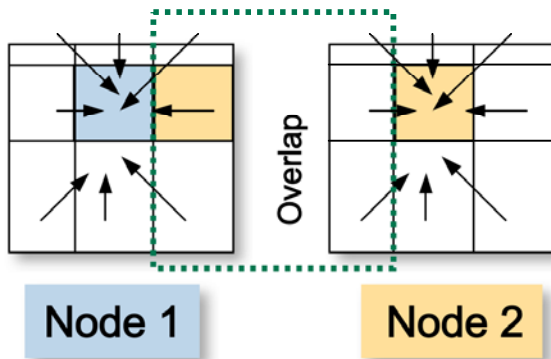Computational Sciences and Engineering

Geospatial data sciences

HP Visualization

Terabytes of spatially referenced data per day

Errors and uncertainty

Knowledge Discovery

Data mining | Spatial analysis
Machine learning | Pattern recognition
Statistics | Nonlinear dynamics

Efficient storage and retrieval methods

HPC

Disaster mapping and monitoring

Geospatial intelligence discovery

Actionable insights/ knowledge

Extremes

Social networks

National security

GeoDatabases

Petabytes of spatiotemporal data archives

Damage	Total population	Senior
100%	791,361	90,773
75%	1,594,806	187,677
50%	1,505,196	184,372
25%	1,701,593	224,279
Total	5,592,956	687,101

OAK RIDGE National Laboratory

Pattern detection and decision support

Consequences

Process modeling

Offline
Predictive analysis

Pattern detection

Body of actionable knowledge

Decision support

Data integration

Action

End-user

Additional data hypothesis generation

Analyst

Decision

Event summary

HIT

Online
Decision making

Stream of new data

OAK RIDGE
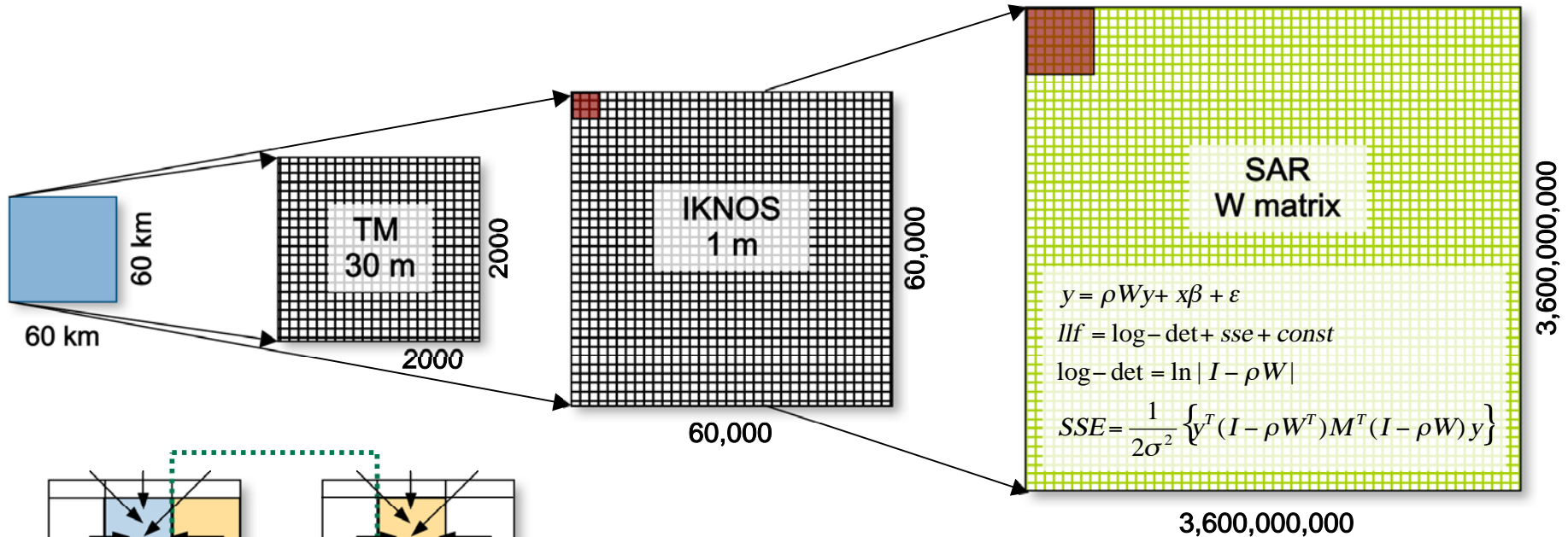National Laboratory

Spatial classification and prediction

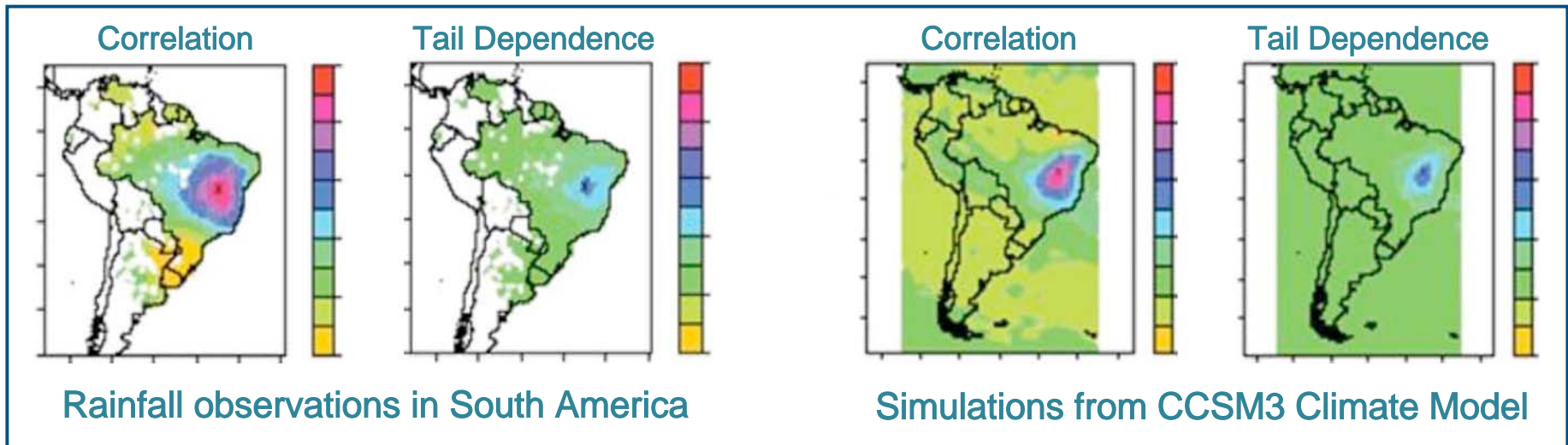- Spatial autoregressive regression (SAR)
- Markov random fields (MRF)

60 km × 60 km

TM 30 m — 2000 × 2000

IKNOS 1 m — 60,000 × 60,000

SAR W matrix — 3,600,000,000 × 3,600,000,000

$$y = \rho W y + x \beta + \varepsilon$$
$$llf = \log{-}\det + sse + const$$
$$\log{-}\det = \ln | I - \rho W |$$
$$SSE = \frac{1}{2\sigma^2}\left\{ y^T (I - \rho W^T) M^T (I - \rho W)\, y \right\}$$

Overlap

Node 1   Node 2

- Challenge
  - Large neighborhood matrices
  - Overlapping computation

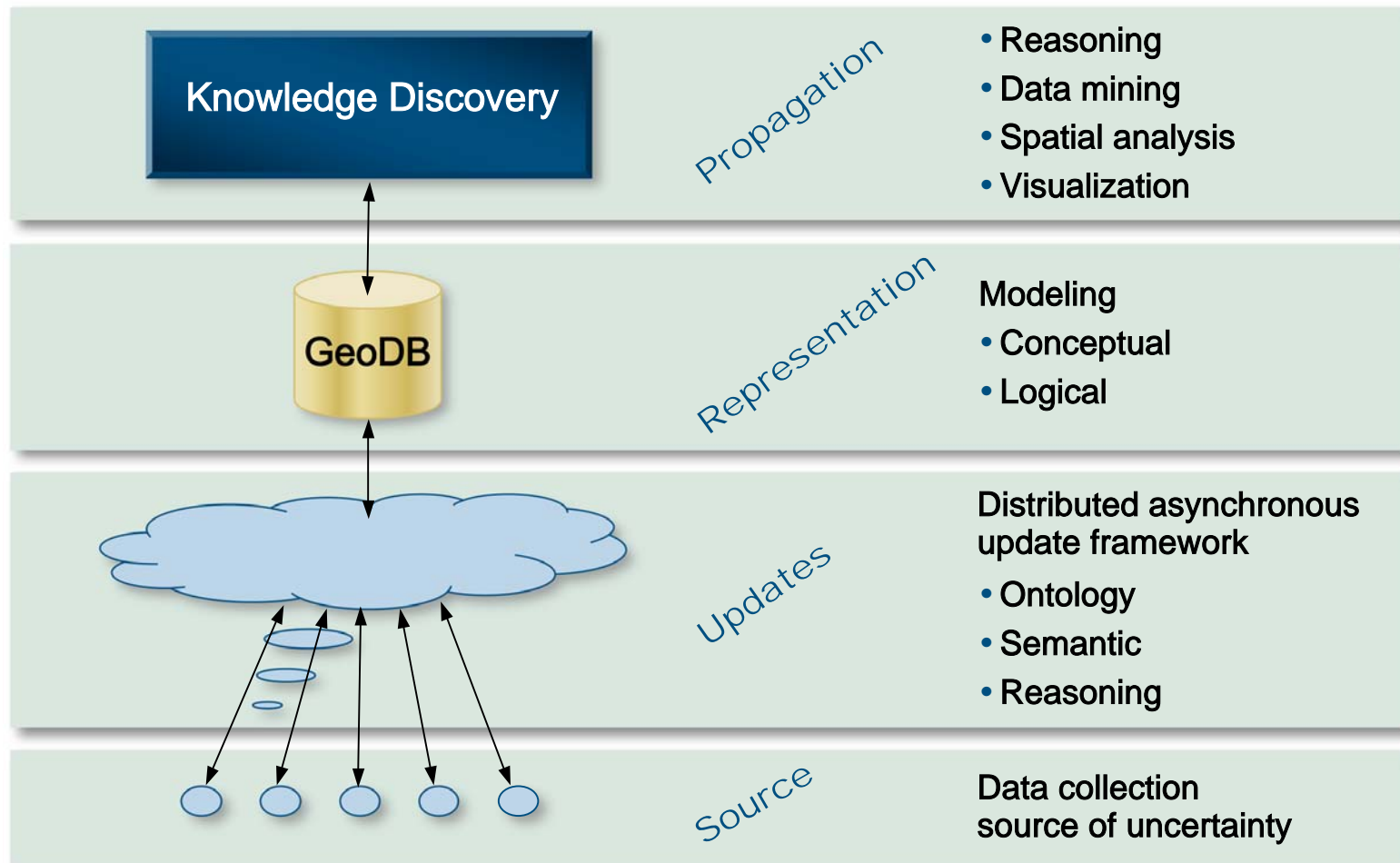OAK RIDGE National Laboratory

Extremes

- A new measure to describe the relationship among extremes in space and time:
  - Simultaneous occurrence of 100-year return levels:
    - Perfect dependence → 100 year event.
    - No dependence → 10000 year event.

Correlation   Tail Dependence   Correlation   Tail Dependence

Rainfall observations in South America   Simulations from CCSM3 Climate Model

- Challenge
  Spatiotemporal dependence estimation is computationally expensive
  (1000 grid cells results in scanning more than half a million pairs).

OAK RIDGE
National Laboratory

Managing uncertainty in spatiotemporal environments (MUSE)

- **Propagation**
  - Reasoning
  - Data mining
  - Spatial analysis
  - Visualization

- **Representation**
  - Modeling
    - Conceptual
    - Logical

- **Updates**
  - Distributed asynchronous update framework
    - Ontology
    - Semantic
    - Reasoning

- **Source**
  - Data collection source of uncertainty

Knowledge Discovery

GeoDB

- Challenge
  - Expensive query processing and spatial analysis

Contacts

Ranga Raju Vatsavai

Geographic Information Science and Technology
Computational Sciences and Engineering
(865) 576-3569
vatsavairr@ornl.gov

Auroop Ganguly

Geographic Information Science and Technology
Computational Sciences and Engineering
(865) 241-1305
gangulyar@ornl.gov

Budhendra Bhaduri

Geographic Information Science and Technology
Computational Sciences and Engineering
(865) 241-9272
bhaduribl@ornl.gov

OAK RIDGE
National Laboratory