# *Miner: A Spatial and Spatiotemporal Data Mining System

Ranga Raju Vatsavai+ and Budhendra L. Bhaduri
Geographic Information Science & Technology Group
Computational Sciences and Engineering Division
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Shashi Shekhar[1] and Thomas E. Burk[2]
[1]Department of Computer Science and Engineering
[2]Remote Sensing & Geospatial Analysis Laboratory
University of Minnesota, MN, USA

## SSTDM Overview

Spatial and Spatiotemporal Data Idiosyncrasies

- Huge Amounts of spatial and spatiotemporal data
- Multimodal, multi-source, multi-resolution
- Complex data types (points, lines, polygons, raster, networks, graphs, ...)
- Spatial and spatiotemporal autocorrelation
- Spatial constraints (non-Euclidian distance, obstacles, ...)

- Limitations of conventional techniques
  - Independent and identically distributed (iid)
  - Hypothesis driven analysis may not be practical in many situations
  - Spatial analysis and spatial statistics do not scale well

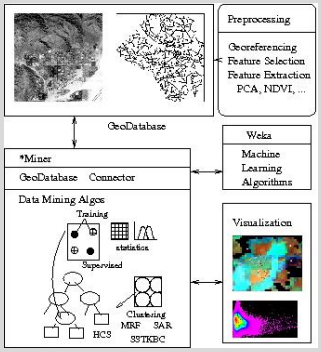
## Definition

- Spatial Data Mining (SDM) is the process of discovering interesting, useful, and non-trivial patterns from large spatial databases [1]

## Spatial Pattern Families

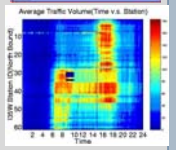
- Spatial outliers, discontinuities, extremes
- Spatial Clustering
- Spatial Classification
- Spatial Prediction
- Spatial Co-location

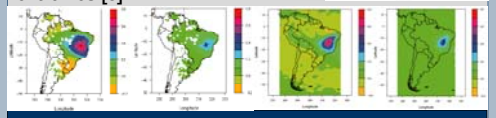
## Architecture




## Spatial Outliers

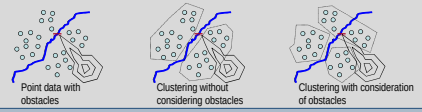
- Spatial outlier is a spatially referenced object whose non-spatial attribute is significantly different in its neighborhood though it may not be significantly different in the entire population [2].
- Multivariate extreme value theory for estimating multivariate dependence among climate extremes [3]



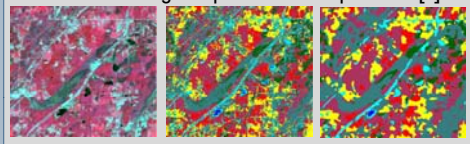
## Spatial Clustering

- Spatial Clustering is the process of grouping a set of spatial objects into clusters such that intra-class similarity is high and inter-class similarity is low.
  - Spatial similarity
  - Spatial constraints (physical obstacles, network distance)




## Spatial Classification

- Given a set of input-output pairs (training data), the objective of supervised learning is to find a function that learns from these i-o pairs, and predicts an output for any unseen input data (new samples).
  - Spatial Autocorrelation (near-by pixels may have same labels) – Markov Random Fields
  - Fewer training samples – Semi-supervised [4]

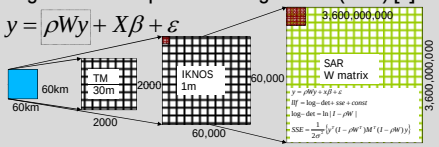


Landsat ETM – FCC Image | Classification with iid assumption | Spatial classification

## Spatial Prediction

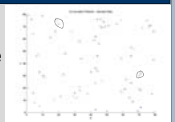
- Regression vs. Spatial Autoregression (SAR) [5]

$$y = \rho W y + X\beta + \varepsilon$$



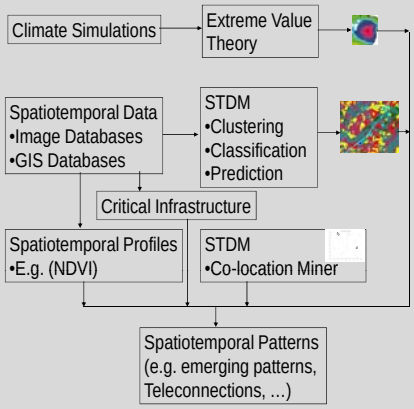

## Spatial Co-location Mining

- Co-location patterns represent subsets of Boolean spatial features whose instances are often located in close geographic proximity (e.g., Nile crocodile and Egyptian plover) [6].



| | | Spatial Data Mining (SDM) | Spatiotemporal Data Mining (STDM) |
|---|---|---|---|
| Input Data | | Often implicit Relationships, complex types | Another dimension-Time<br>Implicit relationships<br>Changing over time |
| Statistical Foundation | | Spatial autocorrelation | Spatial autocorrelation and Temporal correlation |
| Output | Association | Co-location | Spatiotemporal association<br>Mixed-Drove pattern<br>Sustained Emerging pattern |
| | Clusters | Hot-spots | Flock pattern<br>Moving Clusters |
| | Outliers | Spatial outlier | Spatiotemporal outlier |
| | Prediction | Location prediction | Future Location prediction |

## Application




## References

1. S. Shekhar, R. R. Vatsavai, M. Celik. "Spatial and Spatiotemporal Data Mining: Recent Advances." In the Next Generation Data Mining, Hillol Kargupta. et. al. (Eds). CRC Press. 2008.
2. Chang-Tien Lu, Dechang Chen, Yufeng Kou, "Algorithms for Spatial Outlier Detection," 3rd IEEE International Conference on Data Mining (ICDM'03), 2003.
3. Gabriel Kuhn, Shiraj Khan, Auroop R. Ganguly, Marcia L. Branstetter, Geospatial-temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America, Advances in Water Resources Volume 30, Issue 12, , December 2007, Pages 2401-2423.
4. Ranga Raju Vatsavai, Shashi Shekhar, Thomas E. Burk: An efficient spatial semi-supervised learning algorithm. Parallel Algorithms Applications. 22(6): 427-437 (2007)
5. Baris M. Kazar, Shashi Shekhar, David J. Lilja, Ranga Raju Vatsavai, R. Kelley Pace: Comparing Exact and Approximate Spatial Auto-regression Model Solutions for Spatial Data Analysis. GIScience 2004: 140-161
6. Yan Huang, Shashi Shekhar, Hui Xiong: Discovering Colocation Patterns from Spatial Data Sets: A General Approach. IEEE Trans. Knowledge and Data Engineering 16(12): 1472-1485 (2004)

## Collaborators

Auroop Ganguly, Eddie Bright, Chris Symons @ ORNL

## Contact Information

+Ranga Raju Vatsavai
Computational Sciences and Engineering
Oak Ridge National Laboratory
1, Bethel Valley Road
Oak Ridge, TN 37831-6017

Tel.: +1 (865) 576-3569
Fax: +1 (865) 241-6261
Email: vatsavairr@ornl.gov

*This poster has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains, and the publisher by accepting the article for publication, acknowledges that the United States Government retains, a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.*