Predictive Insights on Complex Nonlinear Processes and Unusual Behavior from Heterogeneous Data

Auroop R. Ganguly, Ph.D.
Computational Sciences and Engineering Division
Oak Ridge National Laboratory

Complex and Unusual Patterns in Data

- **Weather, Hydrology and Climate**
  - Origin of the Lorenz Equation
  - Searching for Chaos from Real Data
  - Extreme Values, Anomalies and Teleconnections

Natural Disasters and Impacts

- **Transportation and Sensors**
  - Power Laws and Nonlinear Correlations
  - Patterns of Normal Behavior and Anomalies

Transportation Security and Safety

Predictive Insights on Natural Disasters for Strategic Policy-Making

Case Study with Rainfall Extremes in South America

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

UT-BATTELLE

Excellent Team Work

Names listed in alphabetical order per category

ORNL Staff

- **Marcia Branstetter, CSMD**
- **David Erickson, CSMD (Co-PI)**
- **Auroop Ganguly, CSED (PI)**
- **George Ostrouchov, CSMD (Co-PI)**

Consultants

- **Tailen Hsing, Ohio State**
- **Rick Katz, NCAR**

ORAU Students

- **Chris Fuller[1]**
- **Shiraj Khan[2]**
- **Gabriel Kuhn[2]**
- **Aarthy Sabesan[2]**

1: RAMS Intern

2: CSED Post-Master

Funding provided by SEED money funds of the Laboratory Directed Research and Development (LDRD) program at ORNL

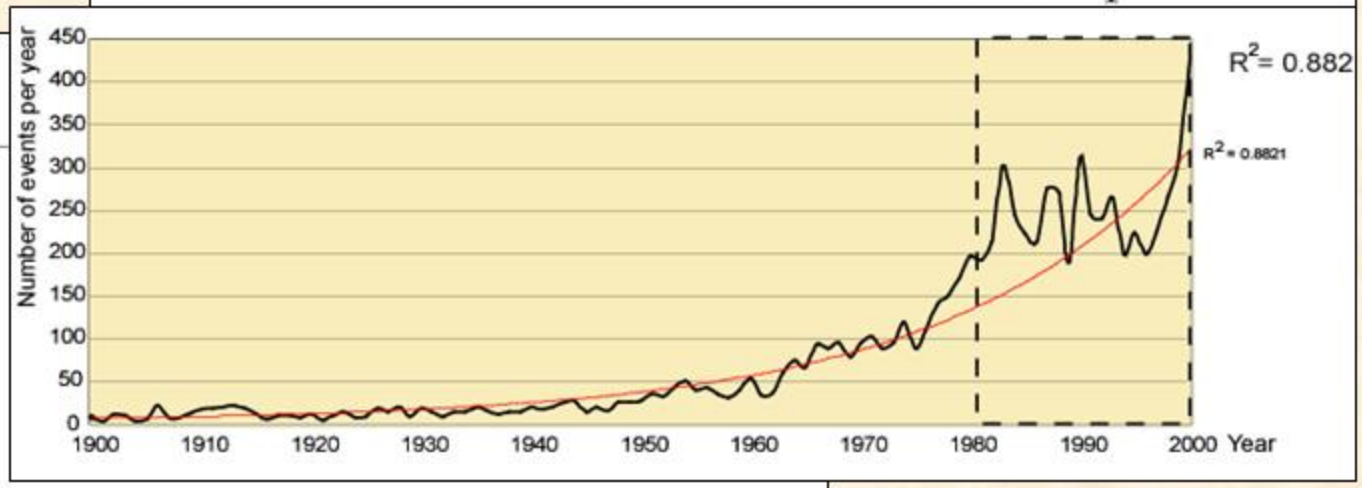
United Nations Development Programme

UNDP Report: "Reducing Disaster Risk: a challenge for development"

National and Global Problem

1.3 Disaster Losses are Increasing

Over the last quarter century, the number of reported natural disasters and their impact on human and economic development worldwide has been increasing yearly. Existing records, while less reliable before 1980, can be traced back to 1900. This longer time period also shows a relentless upward movement in the number of disasters and their human and economic impacts.[2]



The New York Times
nytimes.com

January 22, 2007

OP-ED CONTRIBUTOR

Before the Flood

By JACQUES LESLIE

San Francisco

WHEN the 40-foot-high Kaloko dam collapsed on the Hawaiian island of Kauai last March, its reservoir released a 70-foot-high, 200-foot-wide, 1.6-million-ton wave that carried away 16 cars, hundreds of trees and a cluster of houses, drowning all seven occupants. At least two bodies were swept three-quarters of a mile to the ocean; four were never recovered.

It is tempting to dismiss Kaloko's collapse as an isolated event, but given the perilous state of the nation's dams, it is more likely a harbinger. In 2005, the American Society of Civil Engineers gave United States dams a D, a grade that is still justified two years later.

For starters, the nation's dam stock is rapidly aging. Most dams need major repairs 25 to 50 years after they're built, and most United States dams are at least 25 years old; some, like the 116-year-old Kaloko, were built more than a century ago.

The New York Times

Op-Ed (Editorial) Article on Floods and Dams

Decaying Civil Infrastructures

Science Can Provide Solutions:
A. Risk Map / Database

Don't blame God. Better planning could make natural disasters much less disastrous, experts say.

SCIENCE VOL 310
23 December 2005

Science
AAAS

"Envisions a Public Database Like Google Earth"

"Risk Landscape down to Zip Code Level"

"Because the causes and impacts of disasters are so broad … we need teams of geophysicists who can talk fluently with epidemiologists, and engineers with psychologists"

"One thing is all but certain: Even worse years lie ahead"

"Looking back over 2005 … these disasters should be taken as "opportunities to learn""

Science Can Provide Solutions:
B. Preventive Measures

VIEWPOINT

Refocusing Disaster Aid

Joanne Linnerooth-Bayer, Reinhard Mechler, Georg Pflug

Science
AAAS
SCIENCE VOL 309
12 August 2005

With new modeling techniques for estimating and pricing the risks of natural disasters, the donor community is now in a position to help the poor cope with the economic repercussions of disasters by assisting before they happen. Such assistance is possible ransferring catastrophe risks to the ansfer programs not only would uld free recipient countries from . Both donors and recipients stand e closely coupled with preventive

"Post disaster assistance … has failed to meet the needs of developing countries"

"Free from … vagaries of post disaster assistance"

"Transferring catastrophe risks to global financial markets "

"Need … has become more pressing in light of climate-change negotiations"

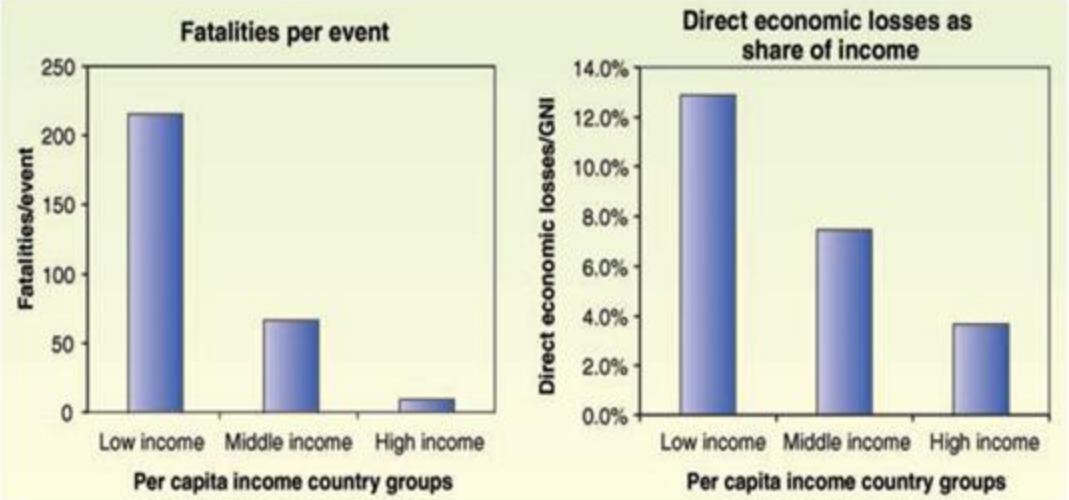


Fig. 1. Fatalities per event and direct economic losses as a share of national per capita income over the period 1980 to 2004. Country income groups according to World Bank classification using GNI per capita. Lower-middle and upper-middle income groups were combined.

Progress Achieved

Publications

Fuller, C.T., Sabesan, A., Khan, S., Kuhn. G., Ganguly, A.R., Erickson, D., and G. Ostrouchov (2006): *Quantification and visualization of the human impacts of anticipated precipitation extremes in South America*. **American Geophysical Union**, Fall Meeting, San Francisco, CA.

Khan, S., Kuhn, G., Ganguly, A.R., Erickson, D.J., and G. Ostrouchov (2006): *Spatio-temporal variability of daily and weekly precipitation extremes in South America*, **Water Resources Research** (*In Revision*).

... and many more

Publicity

SUPERCOMPUTINGONLINE.COM

EARTHtimes.org

upi United Press International® News. Analysis. Insight.™

AGU

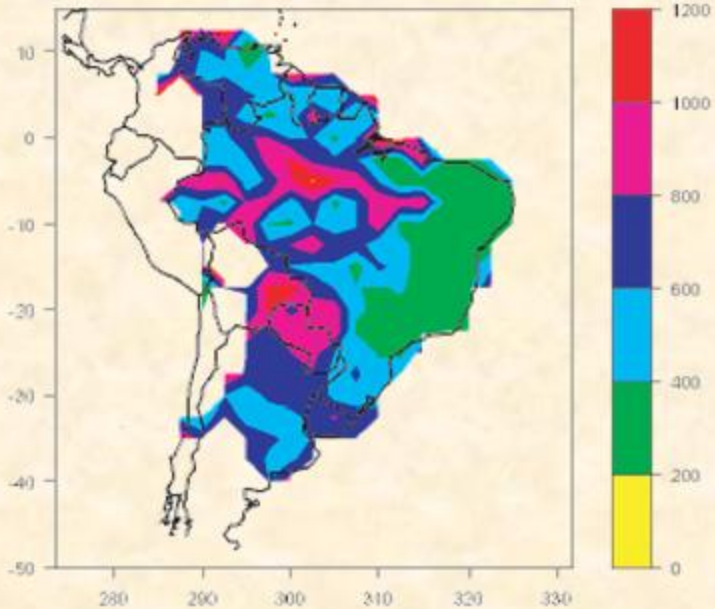
OAKRIDGER.com

... and many more

Case Study Region: South America

Population

GDP



Precipitation

Why South America?

1. Recently available precipitation dataset from NOAA…

Why South America?

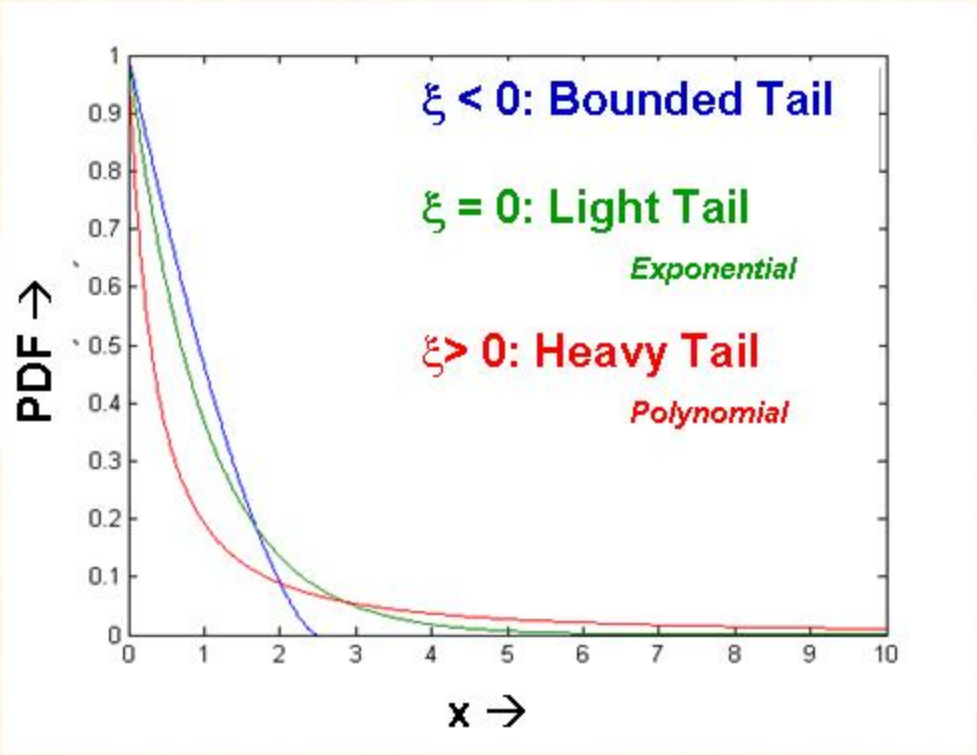
2. Countries more easily comparable (financial, geo-political, cultural…)

Case study can be repeated for other regions

Extreme Value Theory

The Generalized Pareto Distribution

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



Plots: Courtesy "The Mathworks"

EXCEEDENCES OVER THRESHOLD

Prob. (X – u | X > u)

PARAMETERS ESTIMATED BY MAXIMUM LIKELIHOOD

Scale Parameter: σ

Shape Parameter: ξ

Extreme Value Theory

Return Level

- **T-year Return Level, RL(T)**
  - Level that will be exceeded once in every T years
  - Probability of exceeding RL(T) in any year: 1/T

$$RL_N = \begin{cases} u + \frac{\sigma}{\xi}[(Nn_y\zeta_u)^\xi - 1], & \xi \neq 0 \\ u + \sigma \log(Nn_y\zeta_u), & \xi = 0 \end{cases}$$

N-year Return Level
u: Threshold exceeded in a year
n_y: Observations in a year
ζ_u: Probability of an individual observation exceeding u

A Measure of Surprise

"Surprise" caused by "Truly Unusual" Extremes

- **"Design Extremes": *t*-year Return Level, RL(*t*)**
  - *t*: Design return period
  - Example: Dams designed for "50-year floods", or 1/50 chance
- **"Truly Unusual Extremes": *T*-year RL, RL(*T*)**
  - *T*: Higher bound on anticipated return level
  - Example: A 500-year flood is rare but has 1/500 chance
- **Measure of "surprise": RL(*T*) / RL(*t*)**
  - How surprised will one be by a 500-year event when one is prepared for a 50-year event?
  - If the "truly unusual" extremes are close to the "design extreme", anticipated surprise is low, and vice versa

Extreme Volatility Ratio (EVR)

A "new" measure of surprise

- **"EXTREMES VOLATILITY RATIO" (EVR): RL(*T*) / RL(*t*)**
  - *EVR = 1 ➔ No possibility of surprise*
  - *Higher values of EVR ➔ Greater degree of surprise*

$$\frac{RL_T}{RL_t} \sim \begin{cases} (T/t)^{\xi}, & \xi > 0 \\ \log(T)/\log(t), & \xi = 0 \\ 1, & \xi < 0 \end{cases}$$

- **Not so "new":** *EVR relates directly to the shape parameter, ξ*
- **Much more interpretable:** *Easier to understand and visualize*

Extremes Volatility Index (EVI)

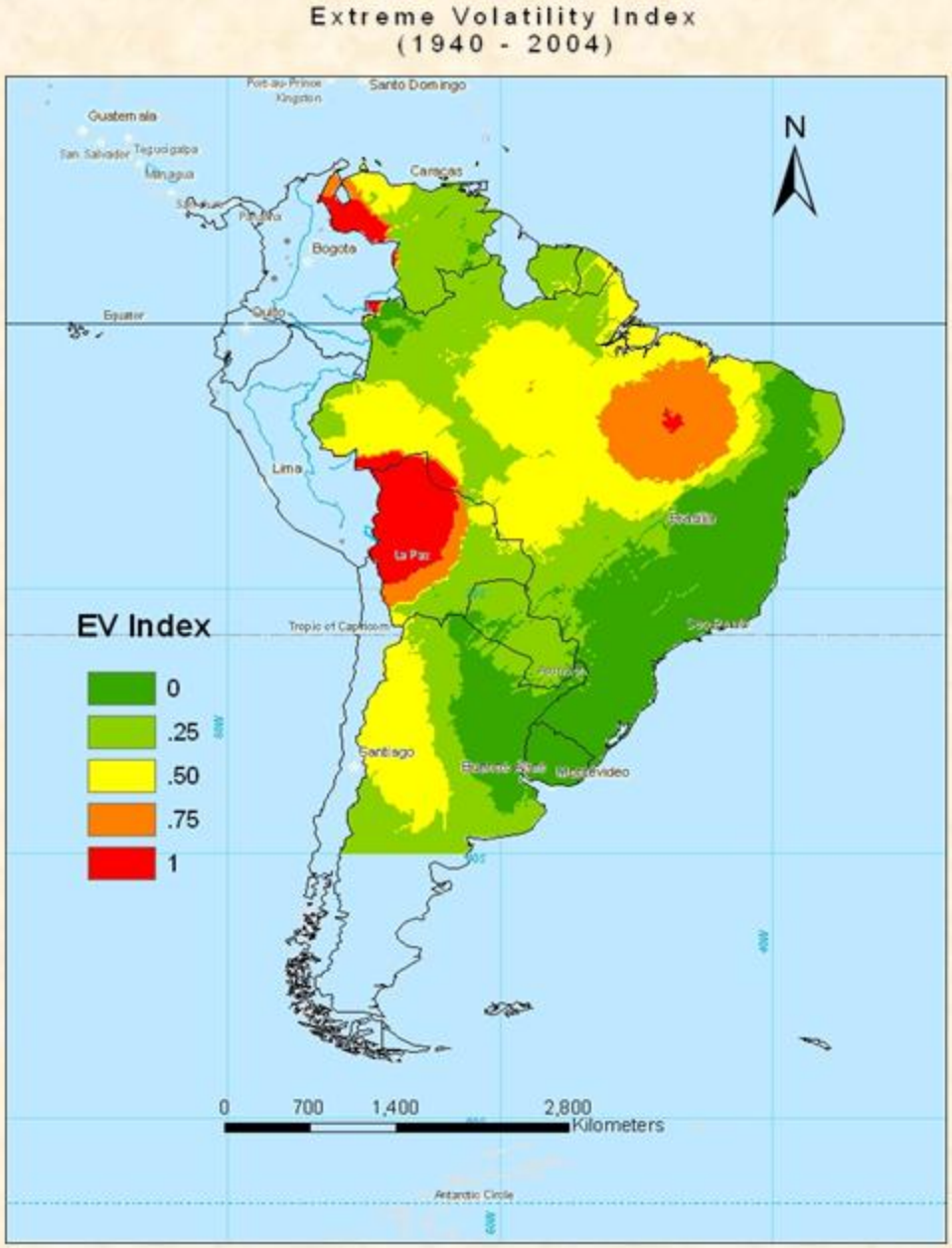
Normalized vulnerability measure

- **EVI = 1 – (1/EVR)**

EVI = {1 – RL(t) / RL(T)}

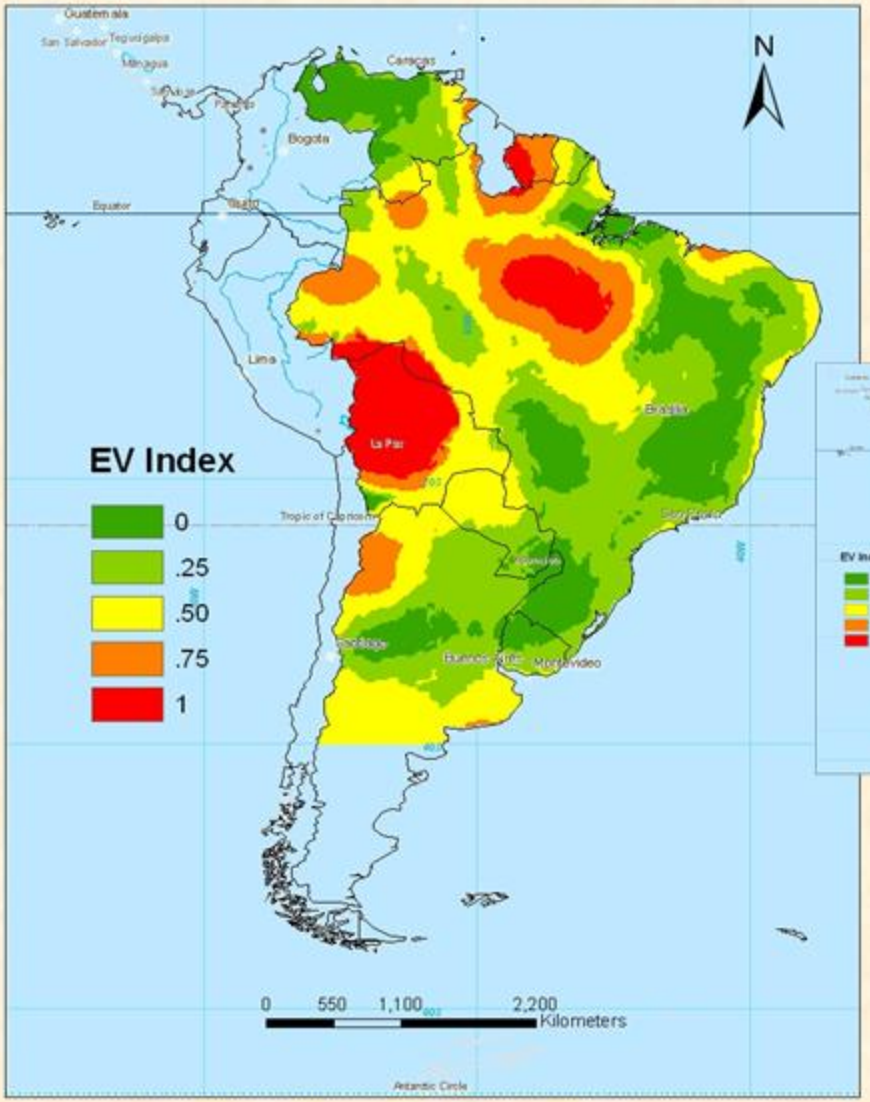
EVI ➔ (0, 1)

  - Normalized measure of surprise
  - Non-rigorous interpretation: "Probability measure"

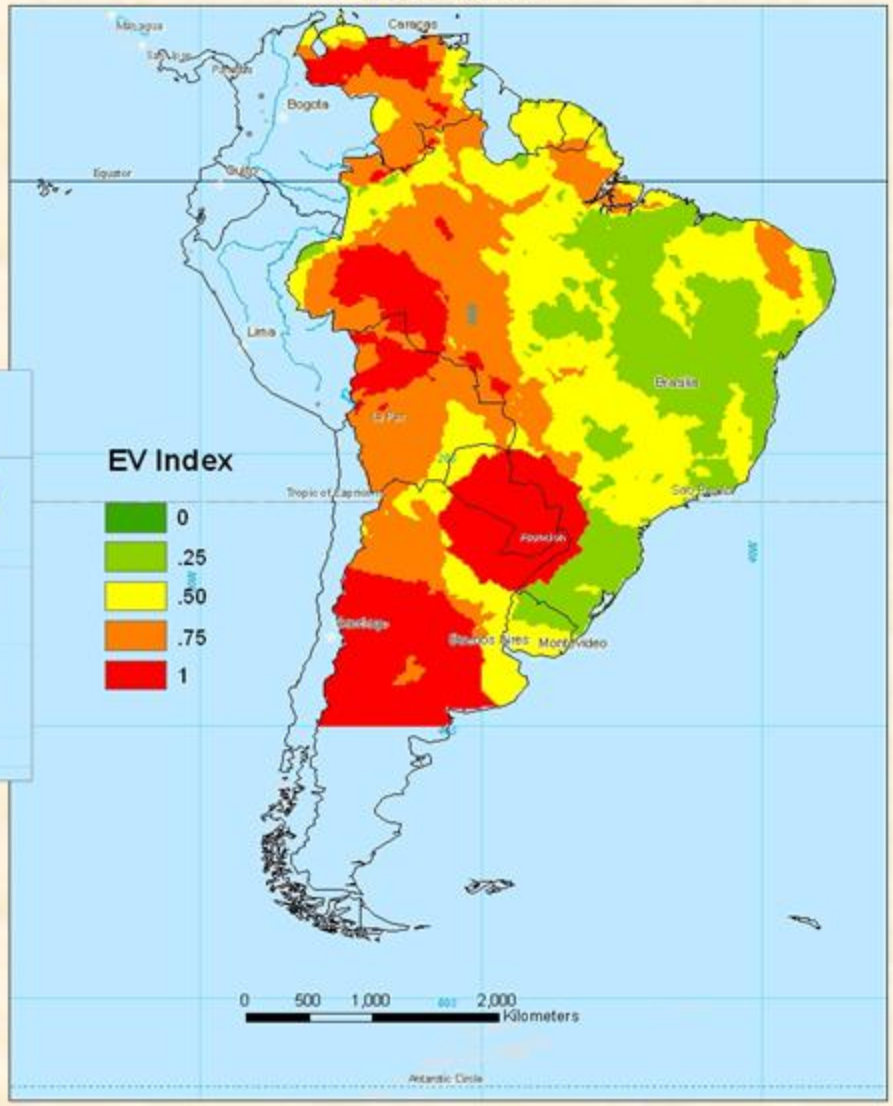


Precipitation EVI Trends

Extremes Volatility Ratio
(1970 - 1989)



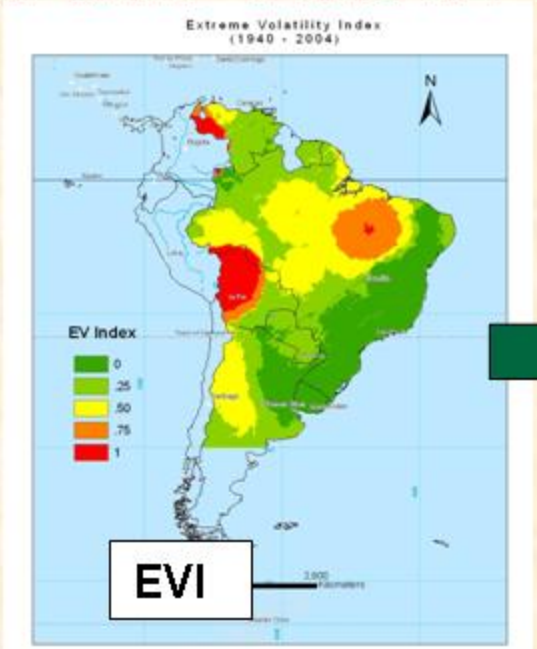
Extremes Volatility Ratio
(1985 - 2004)

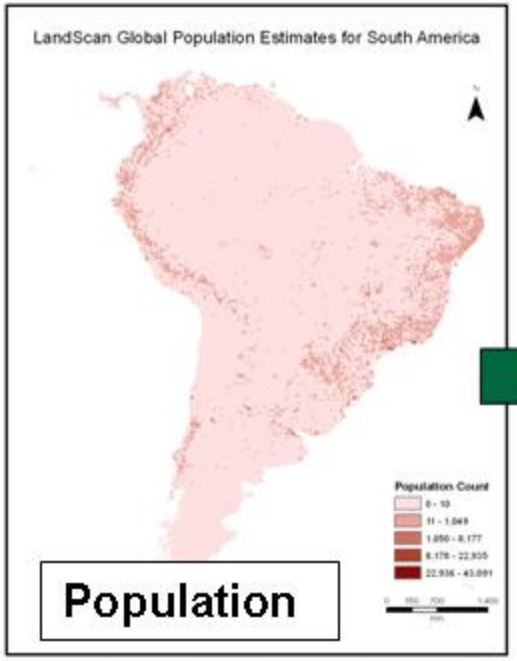


Indices for Risks and Impacts of Precipitation Extremes

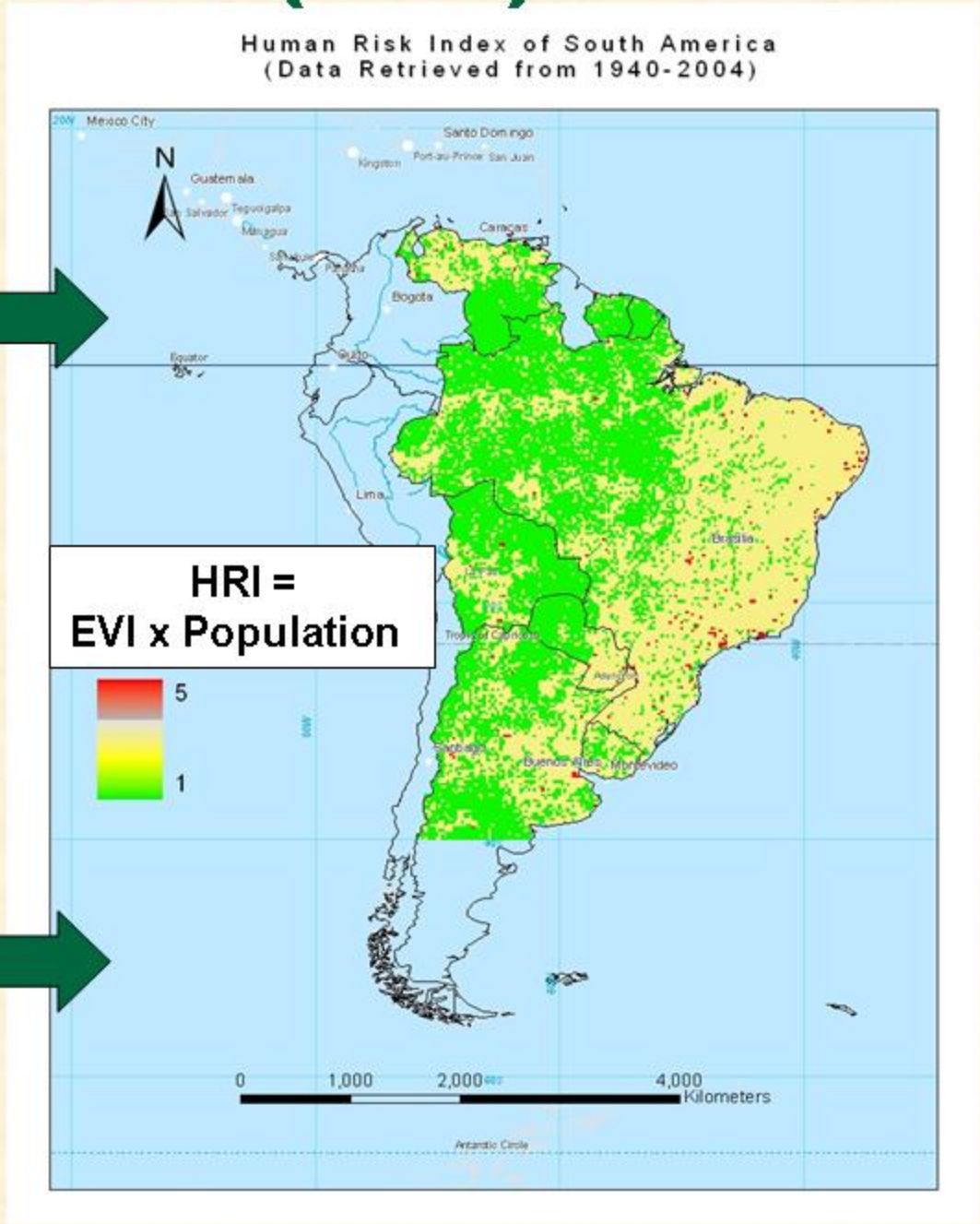
- **“Probability” of Surprise: Precipitation EVI**
  - Probability of Truly Unusual Extremes
- **Potential Risk: Probability X Potential Cost**
  - Potential Cost ~ Population Density
- **Disaster Impact: Potential Risk / Ability to Respond**
  - Ability to Respond ~ GDP

Human Risk Index (HRI)

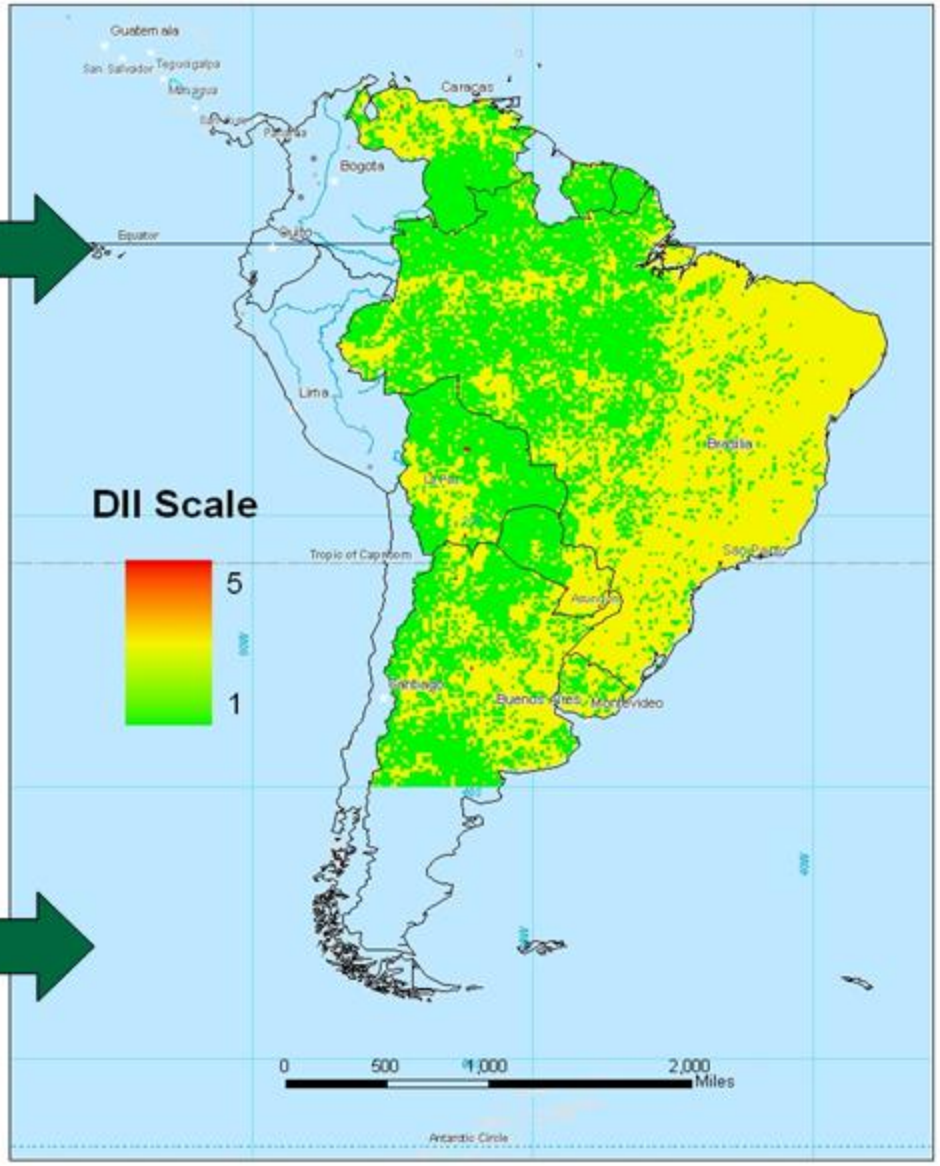




LandScan
(ORNL)



Disaster Impact Index (DII)



Related Research on Natural Processes and Hazards Mitigation

Hydro-Climate Processes and Impact Analyses

Our Related Research

- **Geospatial-Temporal Dependence among Hydrologic Extremes**
- **Nonlinear "Teleconnection" in Space and Time among Climate and Hydrological Variables**
- **Detection of hidden nonlinear dynamical behavior from short and noisy hydrologic time series**
- **Online Change Detection in Space and Time from Remotely Sensed Observations**
- **Weather Prediction based on Remotely Sensed Observations and Numerical Weather Models**
- **Evaluation Metrics and Uncertainty in High-Resolution Population Estimates**

Geospatial-Temporal Dependence among Climate/Weather Extremes

A new measure to describe the relationship among extremes in space- time

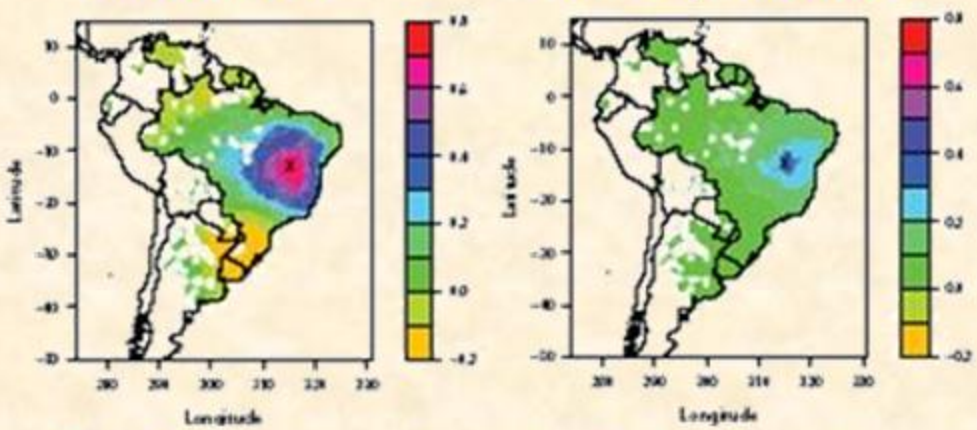
Simultaneous occurrence of 100-year return levels

- No dependence ➔ 10000-yr event
- Perfect dependence ➔ 100-yr event

Kuhn, G., Khan, S., Ganguly, A.R., and M. Branstetter (2006): *Geospatial temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America*, **Advances in Water Resources** (*In Review*).

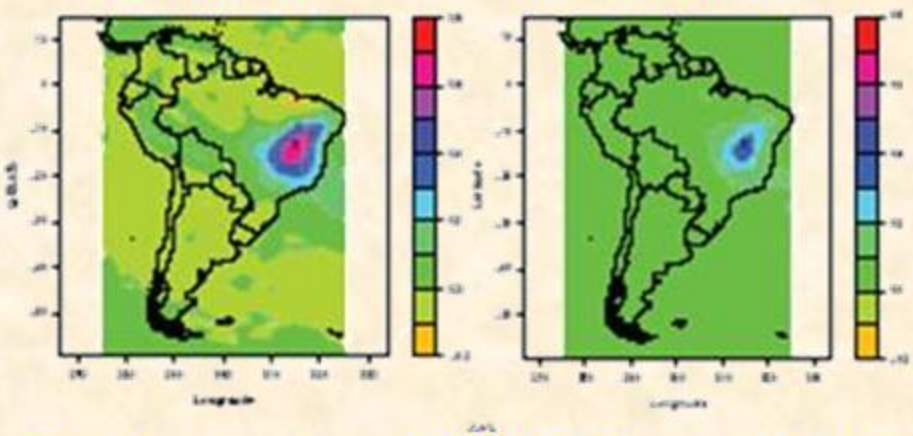
Funding: LDRD/SEED (ORNL)

Correlation **Tail Dependence**



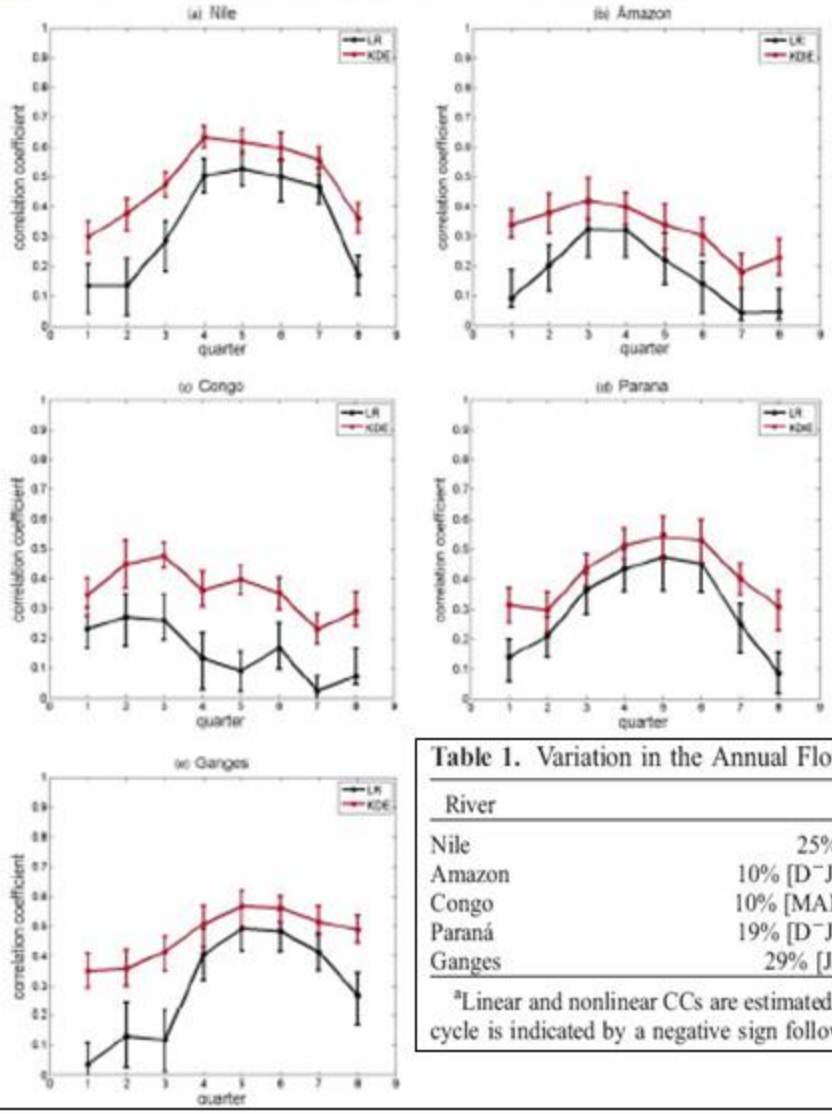
Rainfall Observations in South America

Correlation **Tail Dependence**



Simulations from CCSM3 Climate Model

Nonlinear Climate-Hydrology Teleconnection



Stronger correlation between the El Nino climate index and the tropical hydrological cycle obtained through new adaptations of nonlinear statistical tools

Funding: LDRD/SEED (ORNL)

Table 1. Variation in the Annual Flow of Rivers Associated With ENSO[a]

River	Previous Studies	Linear CC	Nonlinear CC
Nile	25% [SON] [*Eltahir*, 1996]	28% [ASO]	40% [MJJ]
Amazon	10% [D⁻JF] [*Amarasekera et al.*, 1997]	11% [A⁻S⁻O⁻]	18% [A⁻S⁻O⁻]
Congo	10% [MAM] [*Amarasekera et al.*, 1997]	7% [F⁻M⁻A⁻]	23% [M⁻J⁻J⁻]
Paraná	19% [D⁻JF] [*Amarasekera et al.*, 1997]	23% [DJF]	29% [DJF]
Ganges	29% [JJA] [*Whitaker et al.*, 2001]	24% [JAS]	32% [JAS]

[a]Linear and nonlinear CCs are estimated using LR and KDE, respectively. Months in a quarter are given in brackets. The month preceding the seasonal cycle is indicated by a negative sign following a month.

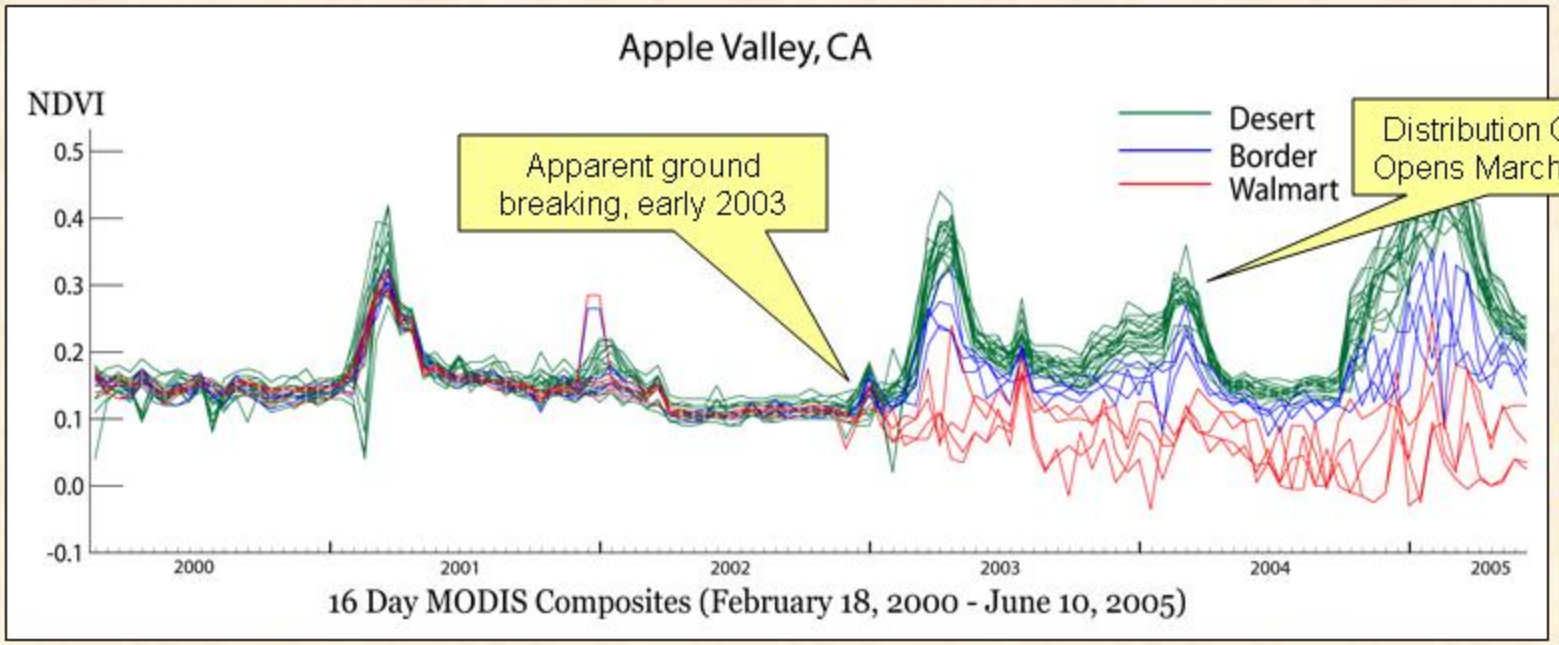
Khan, S., A. R. Ganguly, S. Bandyopadhyay, S. Saigal, D. J. Erickson, III, V. Protopopescu, and G. Ostrouchov (2006): *Nonlinear statistics reveals stronger ties between ENSO and the tropical hydrological cycle*, **Geophysical Research Letters**, 33, L24402, doi:10.1029/2006GL027941.

Online Change Detection from Remotely Sensed Data

A new method for real-time detection of anomalies, outliers, change and change points in space and time

	CA	ME	NC
Alarm	75	52	35
Change Point	70	43	30
Store Opening	93	76	55
Groundbreaking	68	Missing	Missing

Yi Fang, Auroop R. Ganguly, Nagendra Singh, Veeraraghavan Vijayaraj, Neal Feierabend, David T. Potere, "*Online change detection: Monitoring land cover from remotely sensed data*", icdmw, pp. 626-631, **Sixth IEEE International Conference on Data Mining – Workshops** (ICDMW'06), 2006.

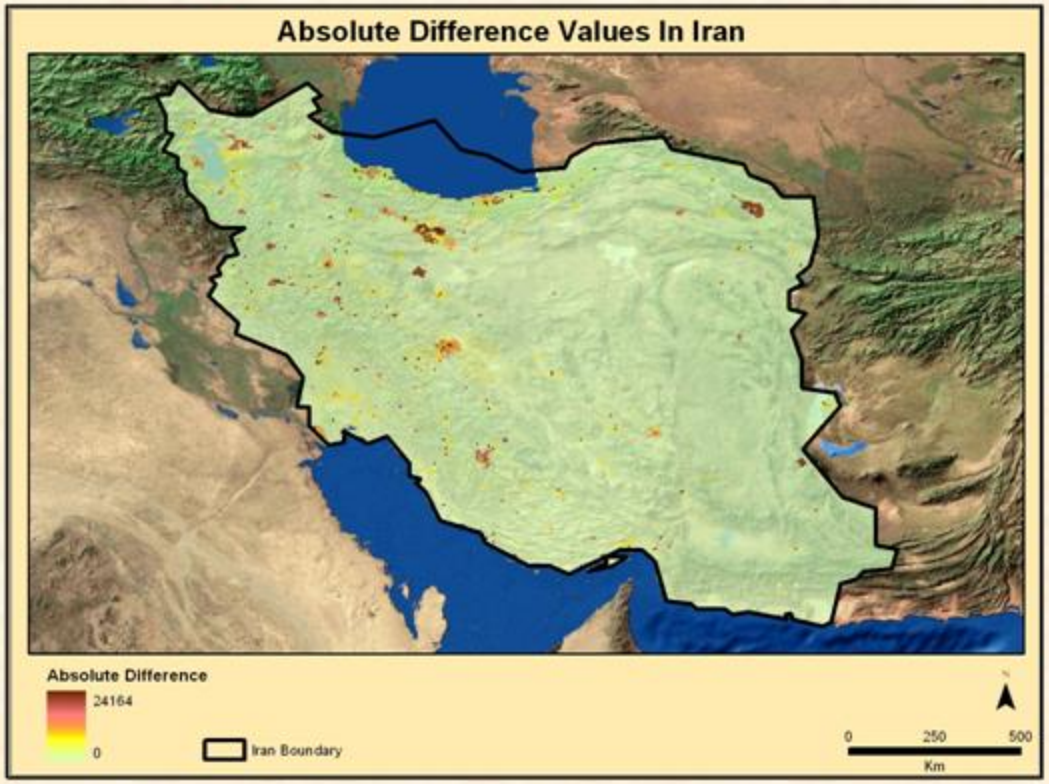


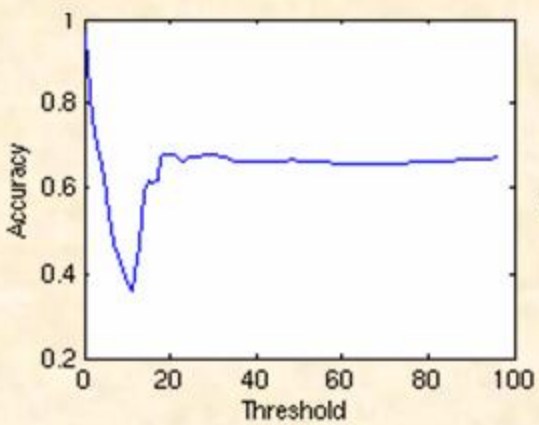
Funding: LDRD (ORNL)

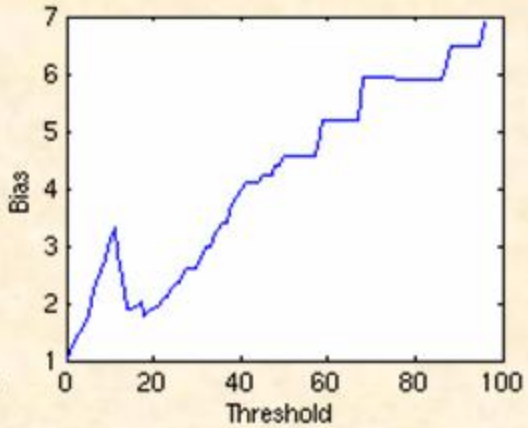
Land Cover Change Detection

MODIS NDVI data and spread of Wal-Mart stores used to validate

Evaluation and Uncertainty Estimation of High-Resolution Global Population Maps







Metrics for the evaluation of high-resolution, global population data

Process-based uncertainty estimation in global population database

Funding: LandScan (ORNL/ GIST Sponsors)

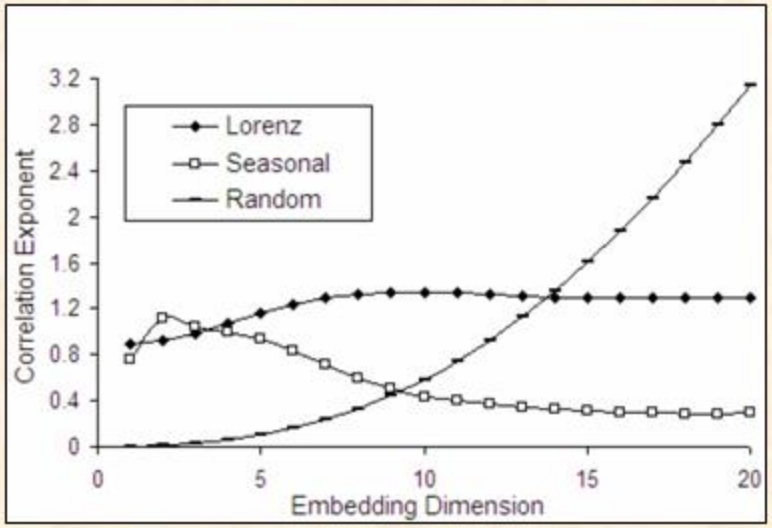
Sabesan, A., Abercrombie, K., Ganguly, A.R., Bhaduri, B.L., Bright, E.A., and P. Coleman (2006): *Metrics for the comparative analysis of geospatial datasets with applications to high-resolution grid-based population data*, **GeoJournal** (*Invited: In Review*).

Detection of hidden nonlinear dynamics from short/noisy hydrologic data

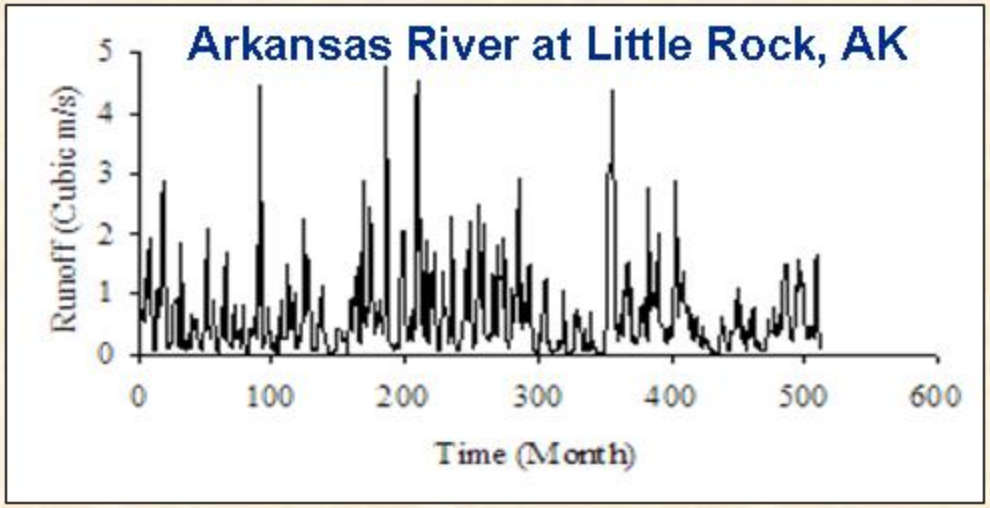
A new method to isolate and identify chaos and nonlinear dynamics from short and finite time series with noise & seasonality

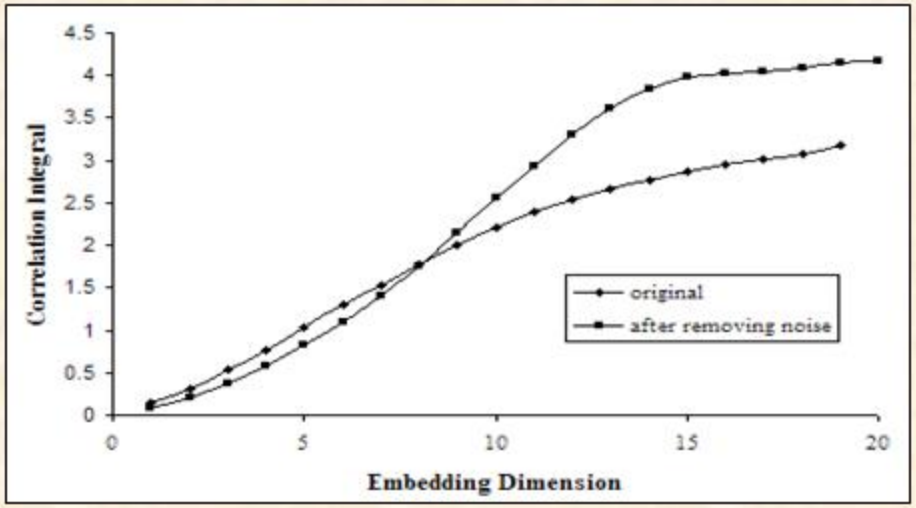
Simulated data suggests thresholds beyond which currently available tools cannot detect chaos

Frequency domain analysis identifies seasonality and short-term nonlinear prediction tools distinguish noise from the dynamical signal, if any



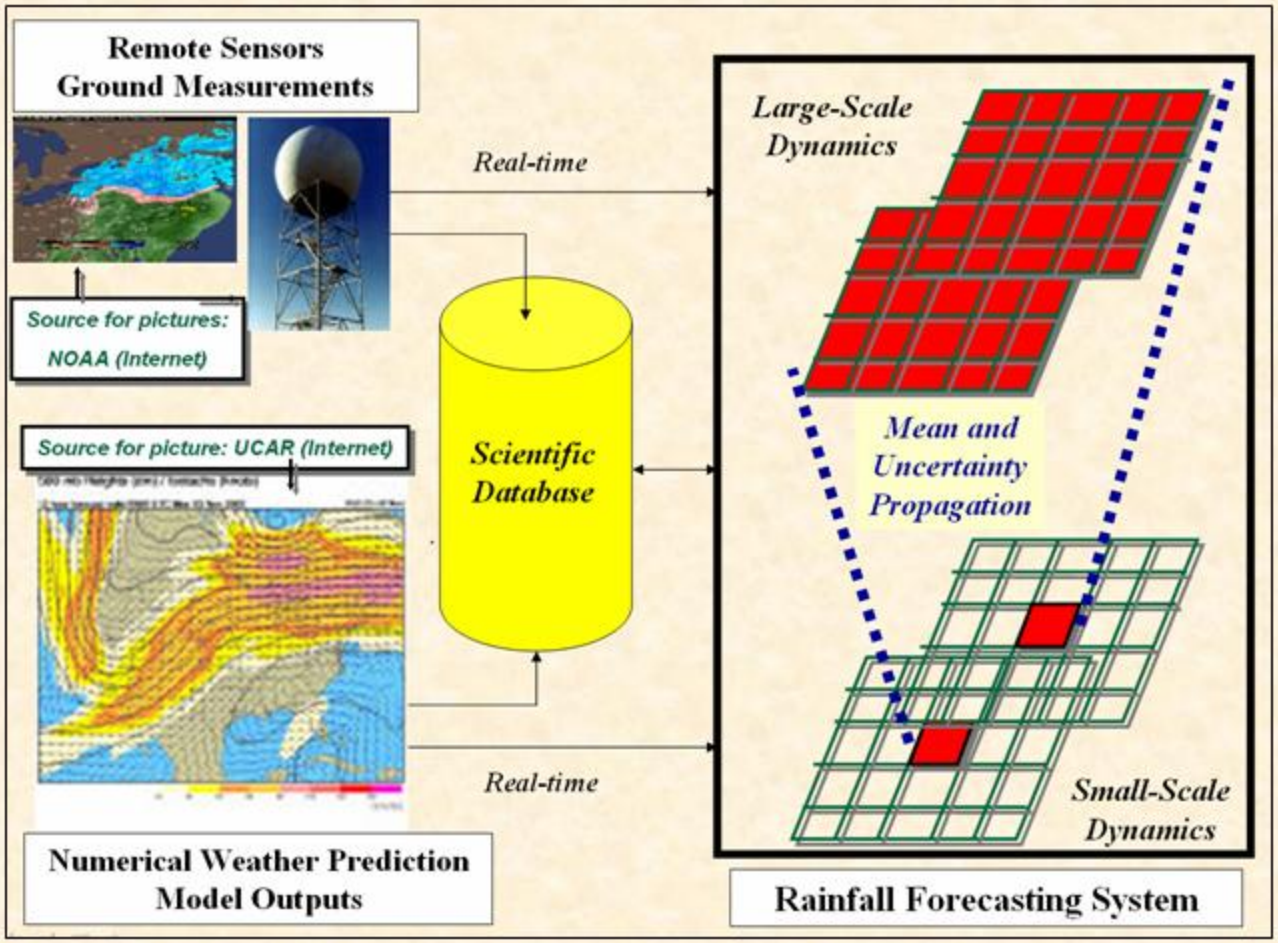
Funding: USF





Khan, S., Ganguly, A. R., and Saigal, S. (2005): *Detection and predictive modeling of chaos in finite hydrological time series*, **Nonlinear Processes in Geophysics**, 12: 41-53..

Weather Prediction from Remotely Sensed Data and Numerical Models



Ganguly AR, Bras RL (2003) *Distributed Quantitative Precipitation Forecasting Using Information from Radar and Numerical Weather Prediction Models.* **Journal of Hydrometeorology** 4(6): 1168

Improved prediction compared to state of the art approaches

Funding: MIT

The rainfall forecasting problem was partitioned into component processes in a way such that weather physics and multiple data dictated tools could be utilized where each fit best

Knowledge discovery from wide area sensor data for security applications

Case Study with Truck Weigh Station Sensors and Transportation Security

Multidisciplinary Team

Names listed in alphabetical order per category

ORNL Staff

- **Auroop Ganguly, CSED (PI)**
- **George Ostrouchov, CSMD (Co-PI)**
- **Vladimir Protopopescu, CSED (Co-PI)**

External Collaborators

- **Arindam Banerjee, UMN**
- **Amrudin Agovic[3], UMN**

1: CSED Post-Master

2: CSED Post-Doc

3: UMN PhD Student

ORAU Students

- **Yi Fang[1]**
- **David Gerdes[1]**
- **Olufemi Omitaomu[2]**

ORNL SensorNet®

- **Frank DeNap**
- **Ian Gross**
- **Bruce Patton**
- **Steven Saavedra**
- **Randy Walker**

Funding provided by the Director's R&D funds of ORNL's LDRD program

Radiation Detection Systems

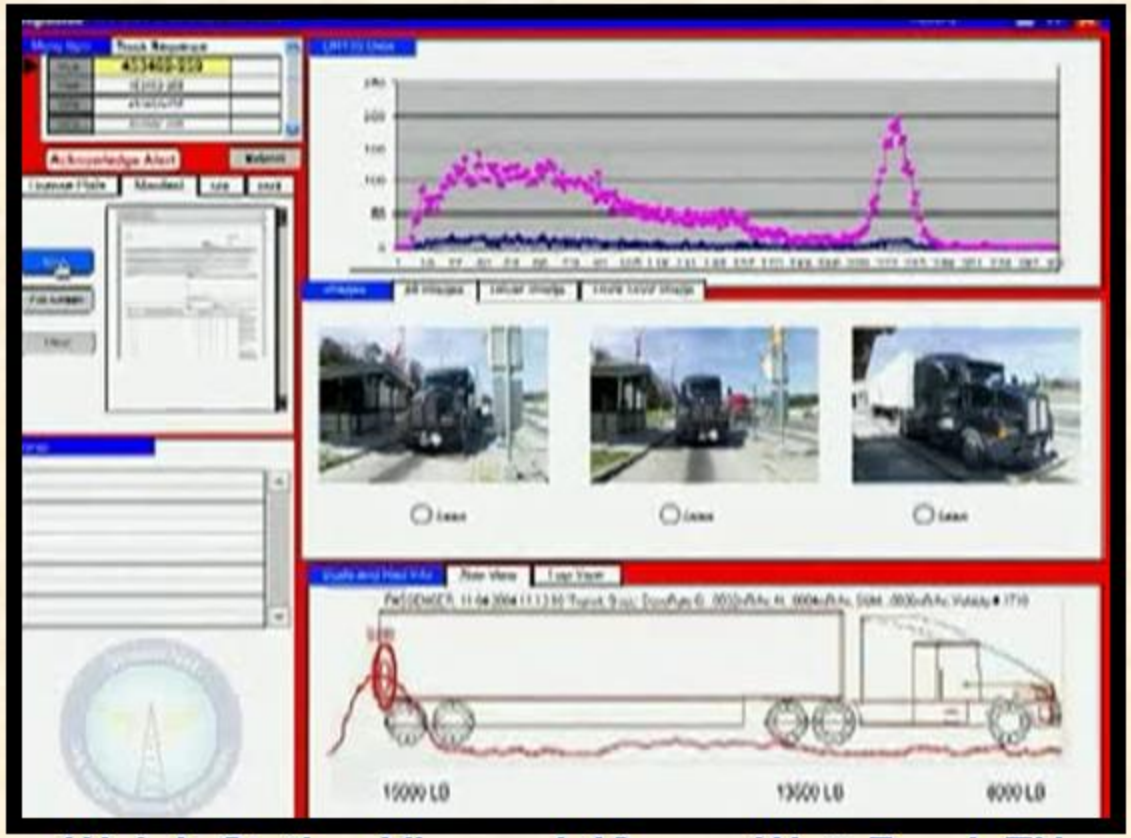
www.sensornet.gov

Inspect

?

Pass

Static Scale

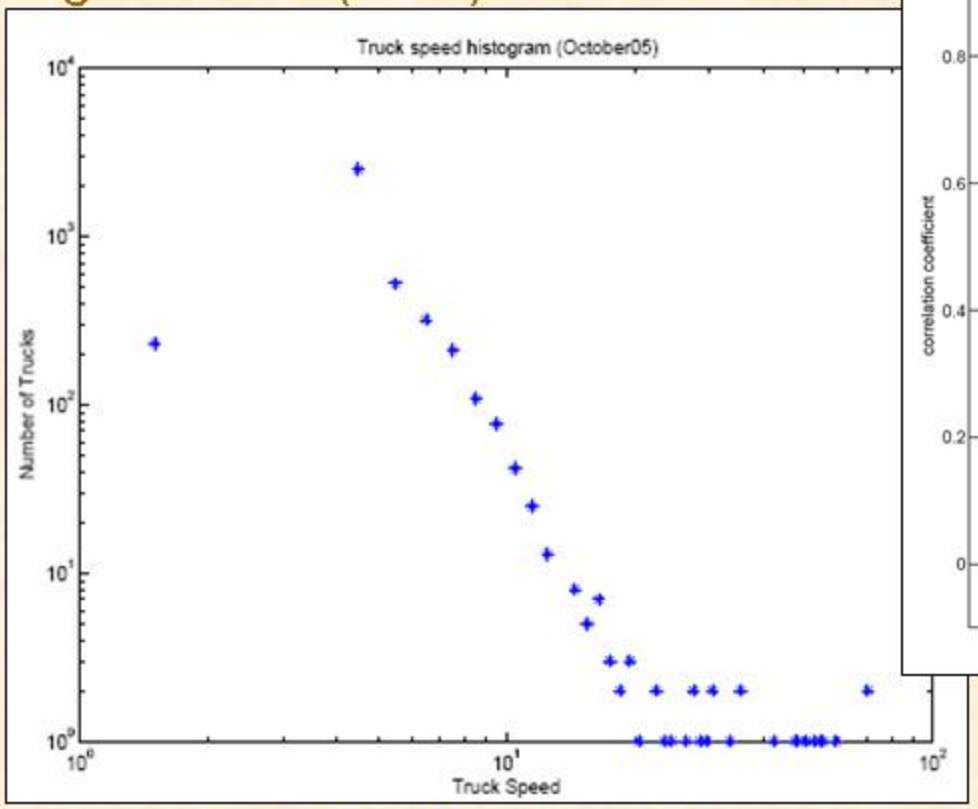


Weigh Station Viewer: I-40 near Watt Road, TN

Descriptive Analysis of Static Scale Data
From Data to Questions

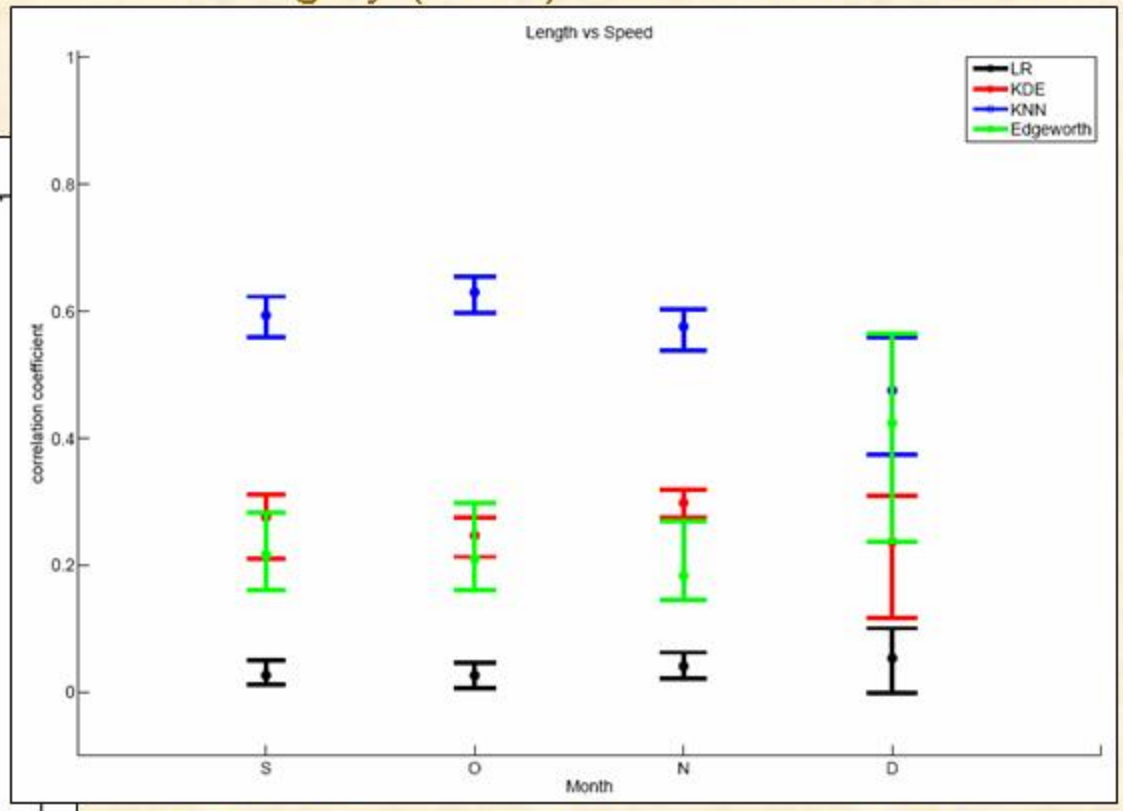
Advanced descriptive analysis of static scale data raises intriguing questions

Agovic et al. (2006): Ref. Next Slide



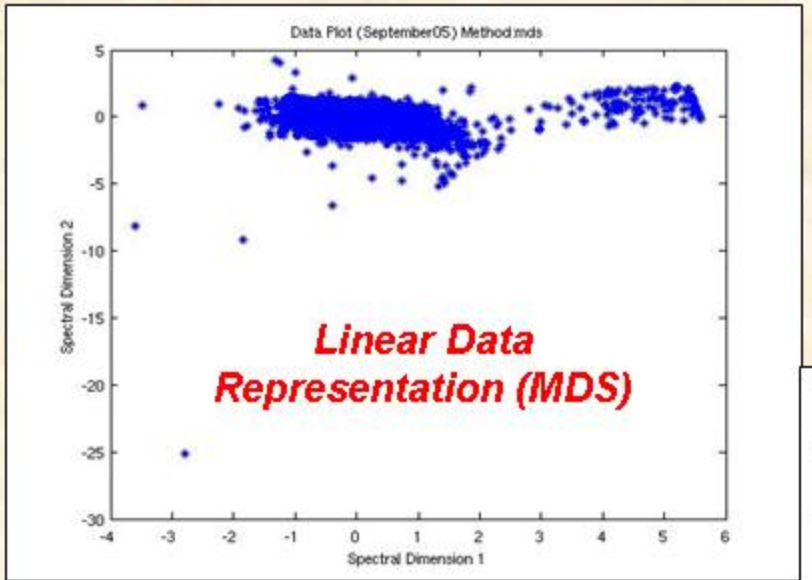
Power law behavior in truck speeds?

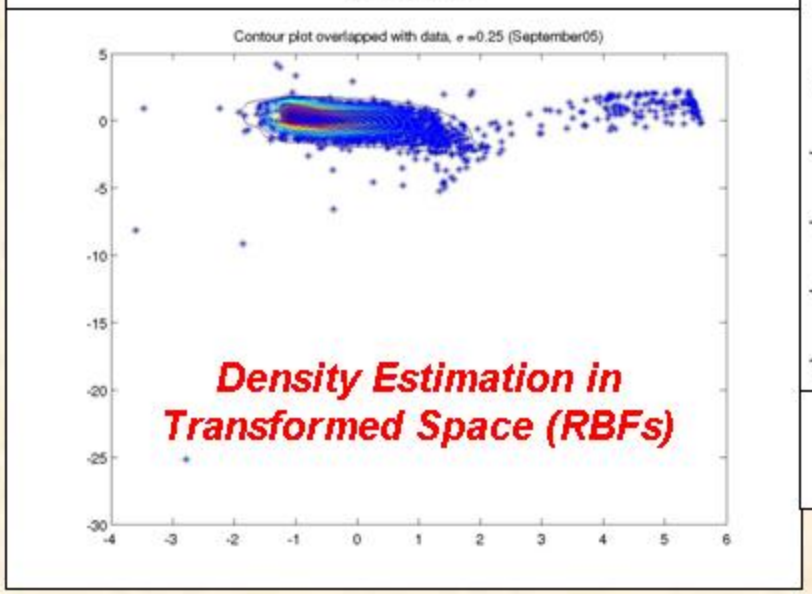
David Gerdes, Shiraj Khan, Auroop Ganguly (2006): ORNL Tech Manual



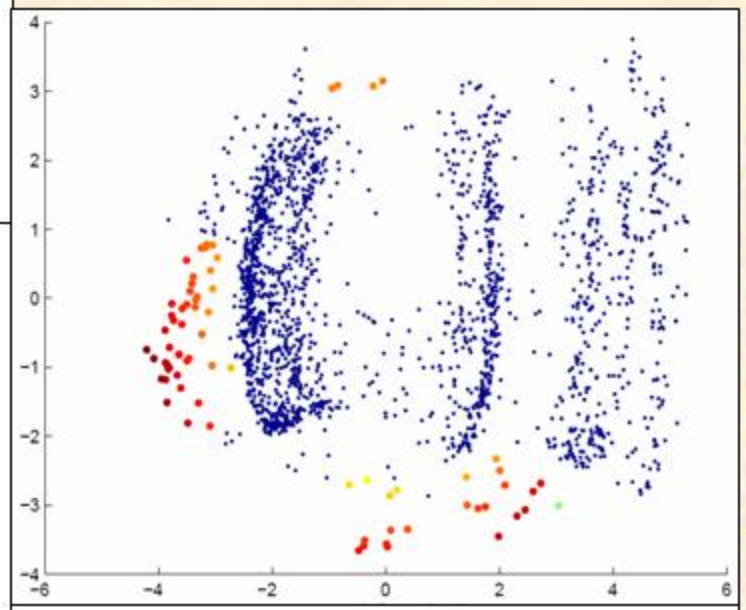
Significant nonlinear correlation between truck length and speeds even though linear correlation is absent?

Anomaly Analysis from Static Scale Data
Nonlinear Dimensionality Reduction





Nonlinear dimensionality reduction followed by anomaly analysis on the manifold

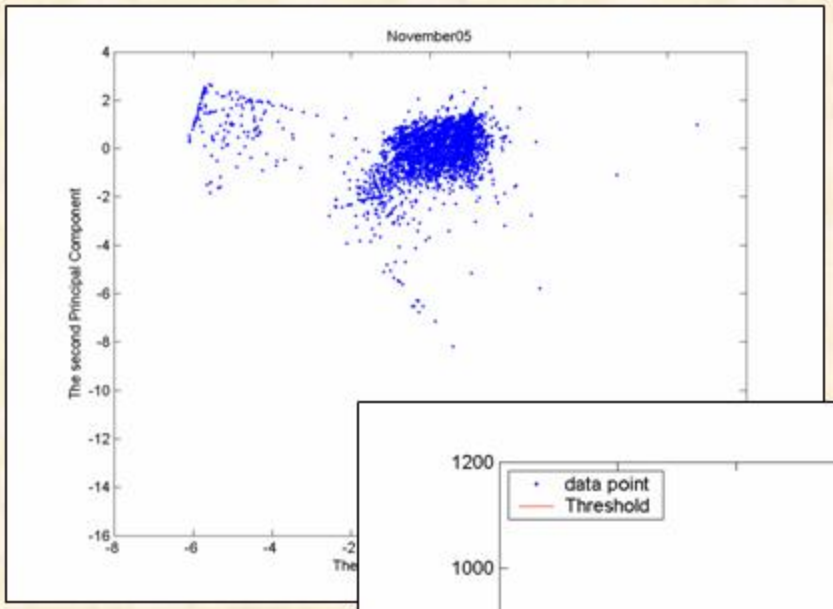


Nonlinear Data Representation (ISOMAP) with Outliers

Agovic, A., Banerjee, A., Ganguly, A., and V. Protopopescu (2006): *Anomaly Analysis with Manifold Embedding and Its Applications to Transportation Corridors*, ORNL: To be submitted.

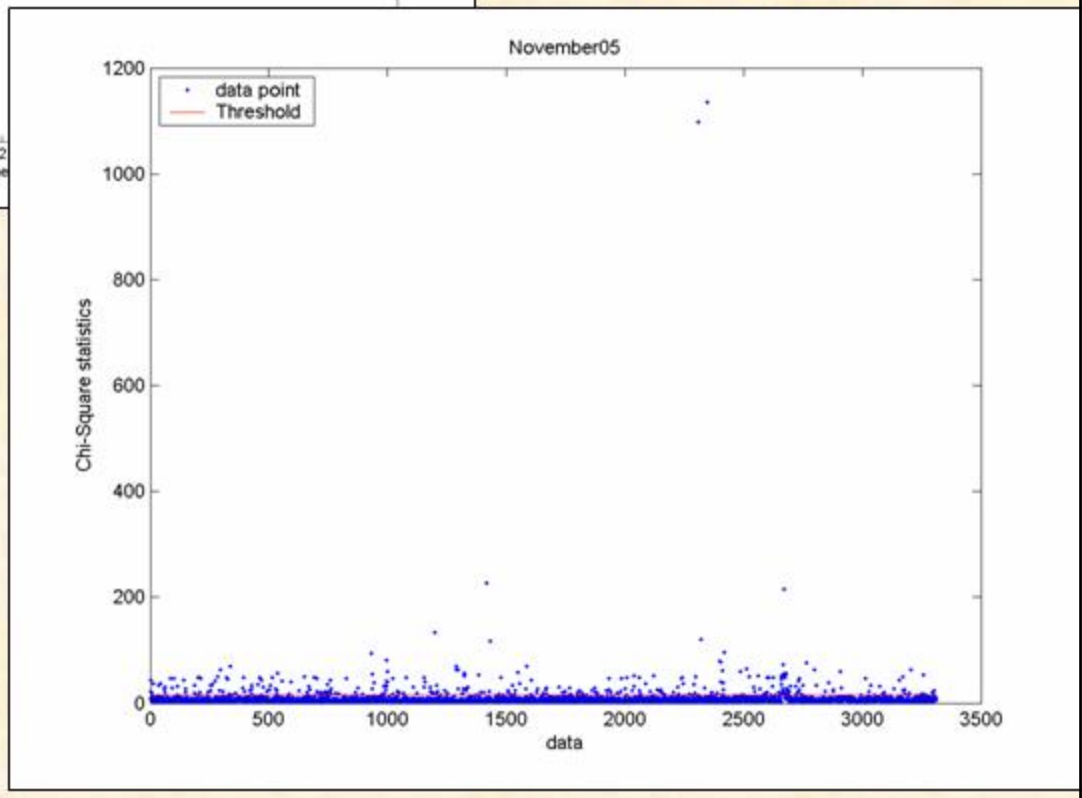
Online Anomalies from Static Scale Data
Probabilistic PCA and Chi-Squared

Online anomaly detection based on probabilistic dimensionality reduction and a goodness of fit measure



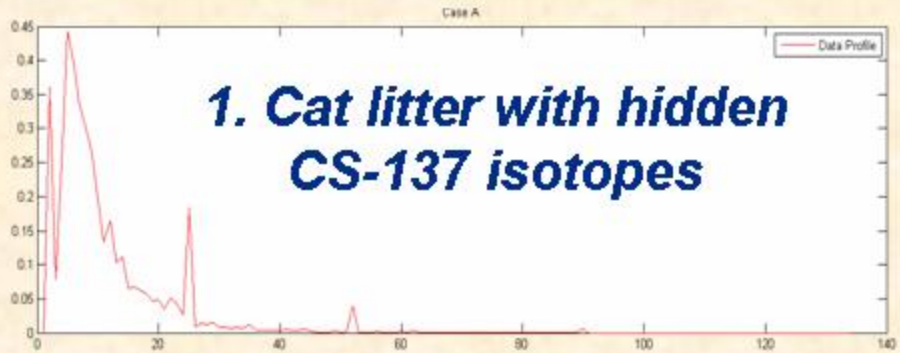
Probabilistic PCA
(Visually same as PCA of Agovic et al.)

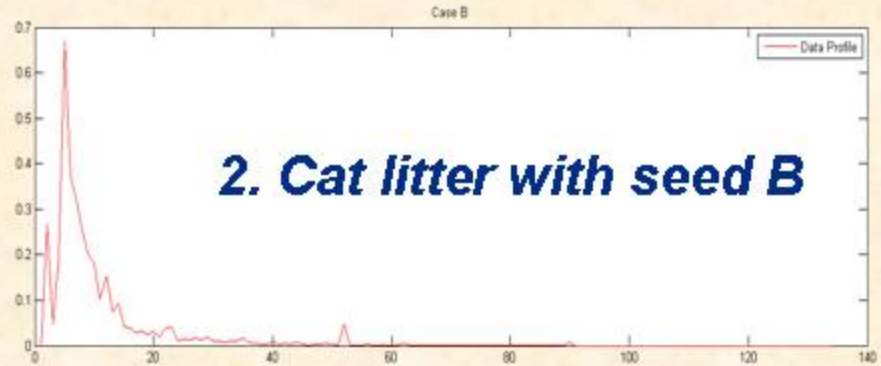
Chi-Squared

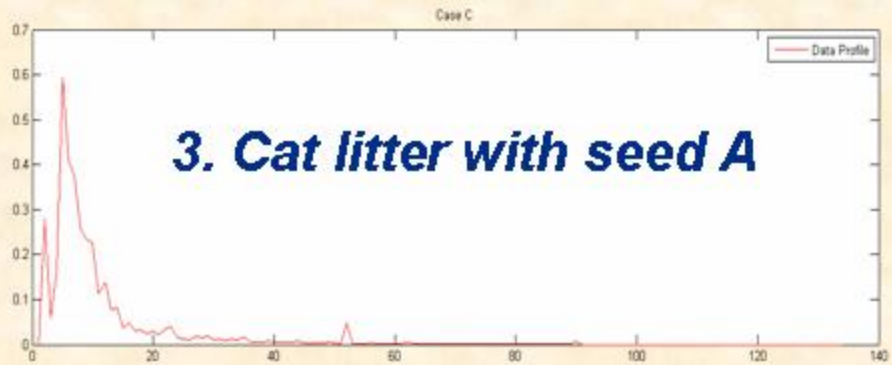


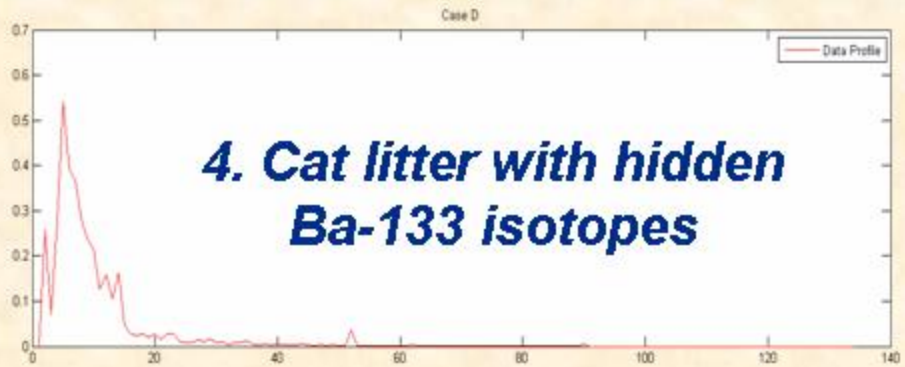
Yi Fang, Auroop Ganguly (2006): ORNL Tech Manual

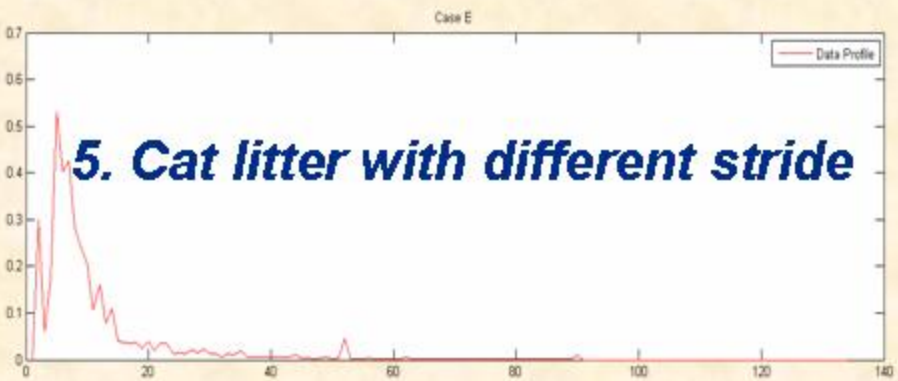
Spectroscopy data Simulated by MCNP







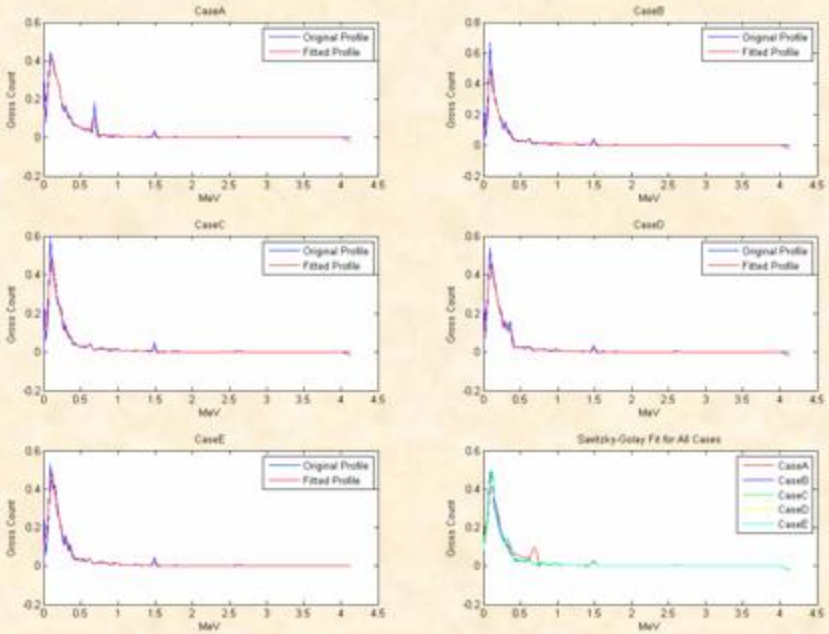




Simulated Data Generated by Bruce Patton

SAMPLE RESULTS

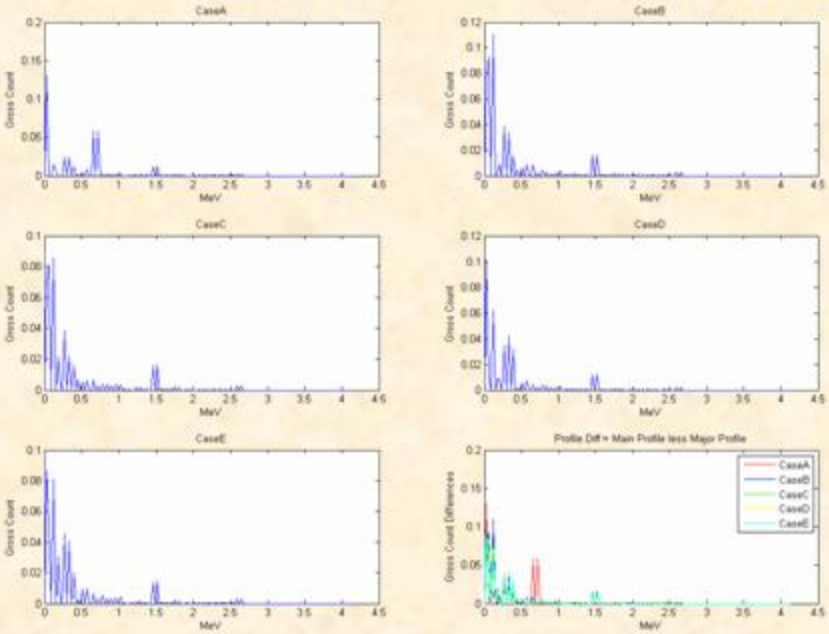
Olufemi Omitaomu, Auroop Ganguly, Vladimir Protopopescu

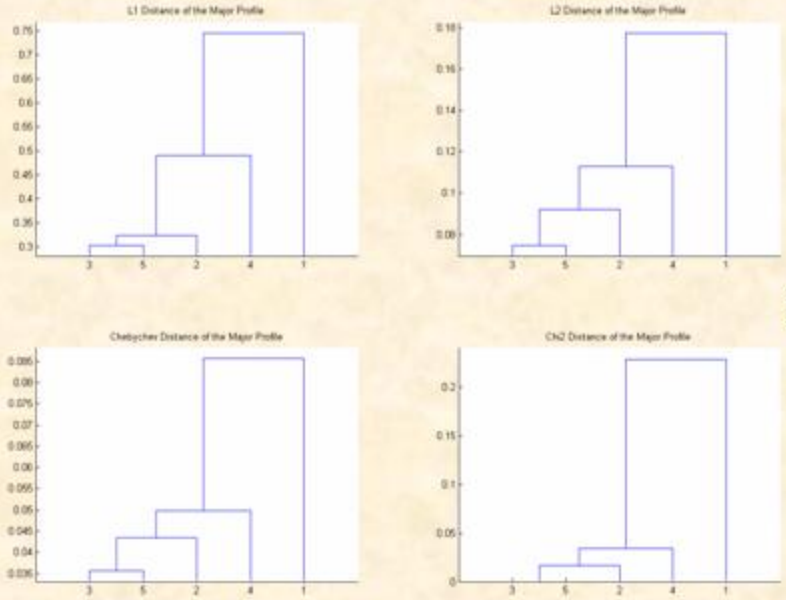


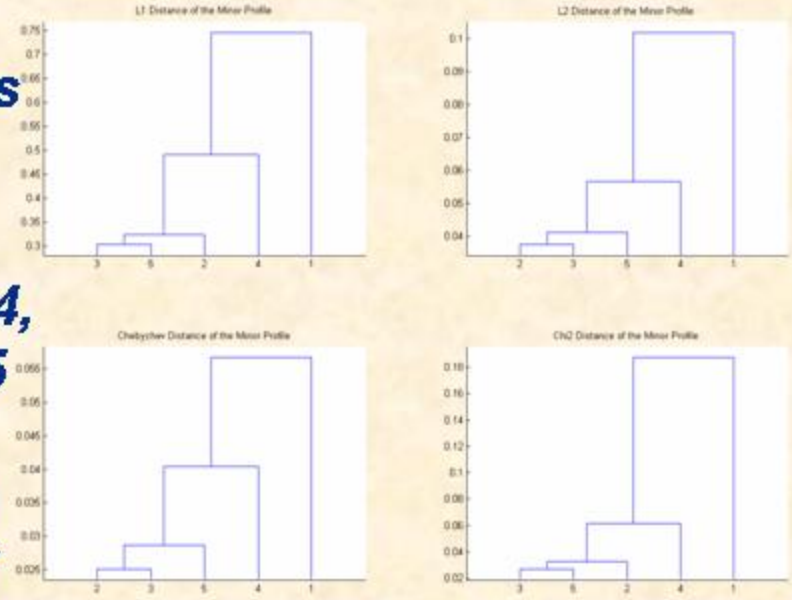
Savitzky-Golay

← Major Profile

Minor → Profile





Dendograms

1 is most distinct, followed by 4, while 2, 3, 5 appear closest to each other