Complex Networks in Climate Science: Progress, Opportunities and Challenges

Karsten Steinhaeuser

Nitesh V. Chawla
Auroop R. Ganguly

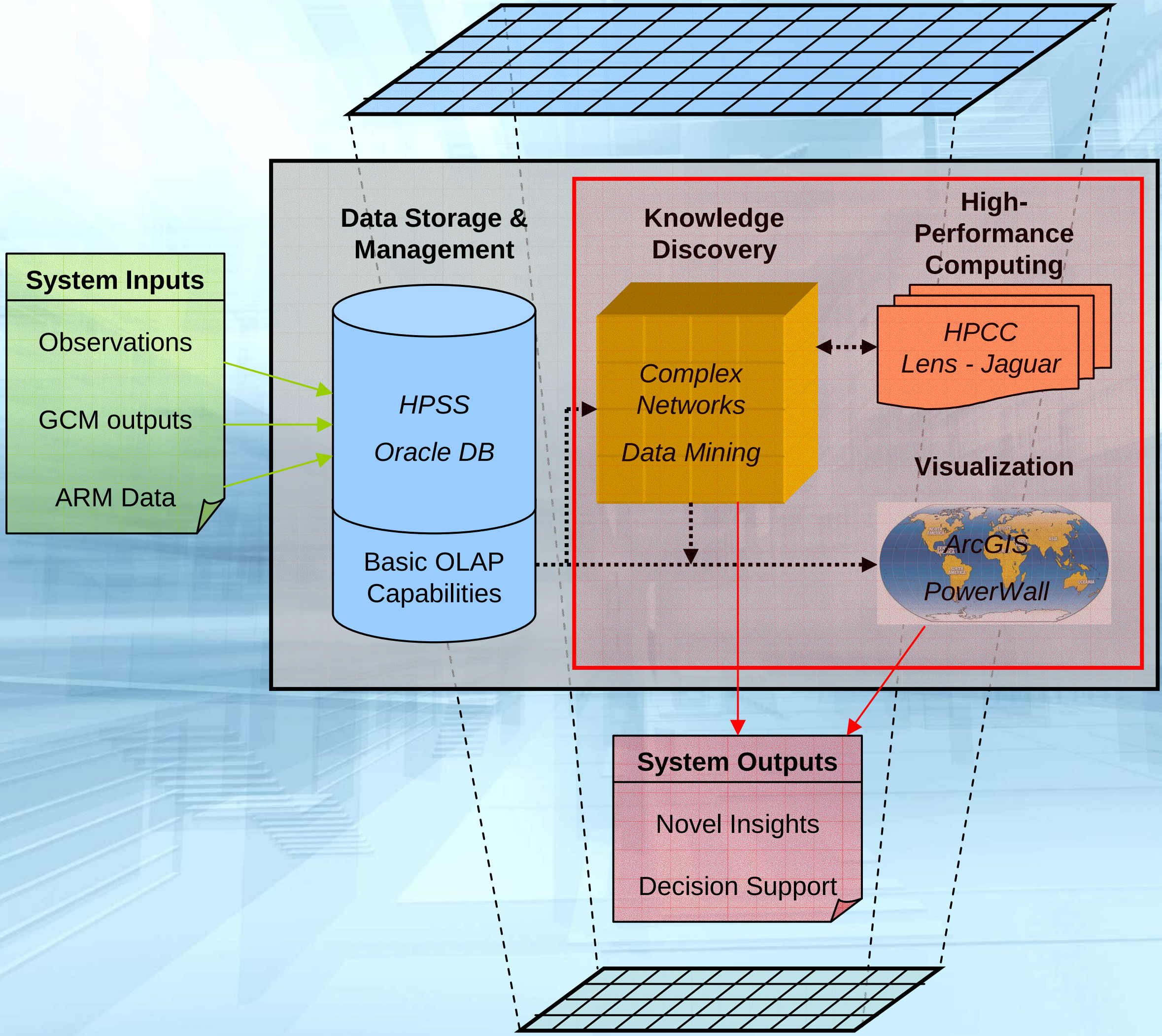
Data, Inference, Analysis and Learning Lab @ ND
University of Notre Dame
Oak Ridge National Laboratory
iCeNSA Interdisciplinary Center for Network Science & Applications

October 5, 2010

Outline

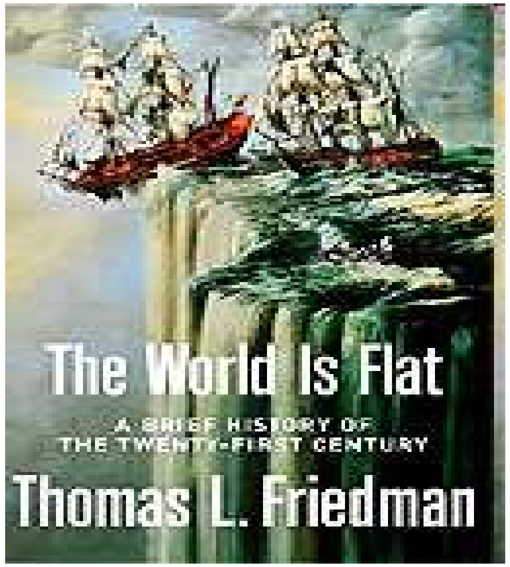
- Motivation
- Data & Methodology
- Network Structure
- Predictive Modeling
- Computational Issues
- Open Questions

Knowledge Discovery for Climate



Mining Complex Data

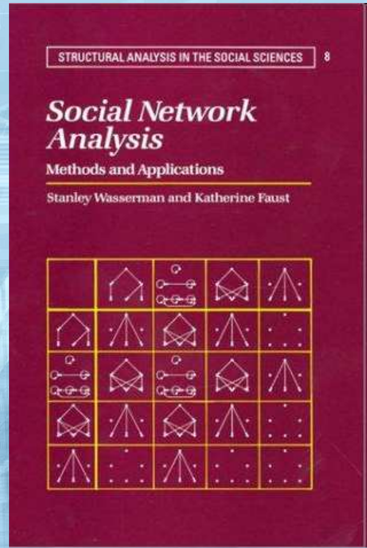
- Complex spatio-temporal data pose unique challenges
- Tobler's First Law of Geography: "Everything is related, but near things more than distant."
  - But are all near things equally related?
  - Are there phenomena explained by interactions among distant things? (teleconnections)



"Networked Thinking"

- Networks are pervasive in social science, technology, and nature

- Many datasets explicitly define network structure

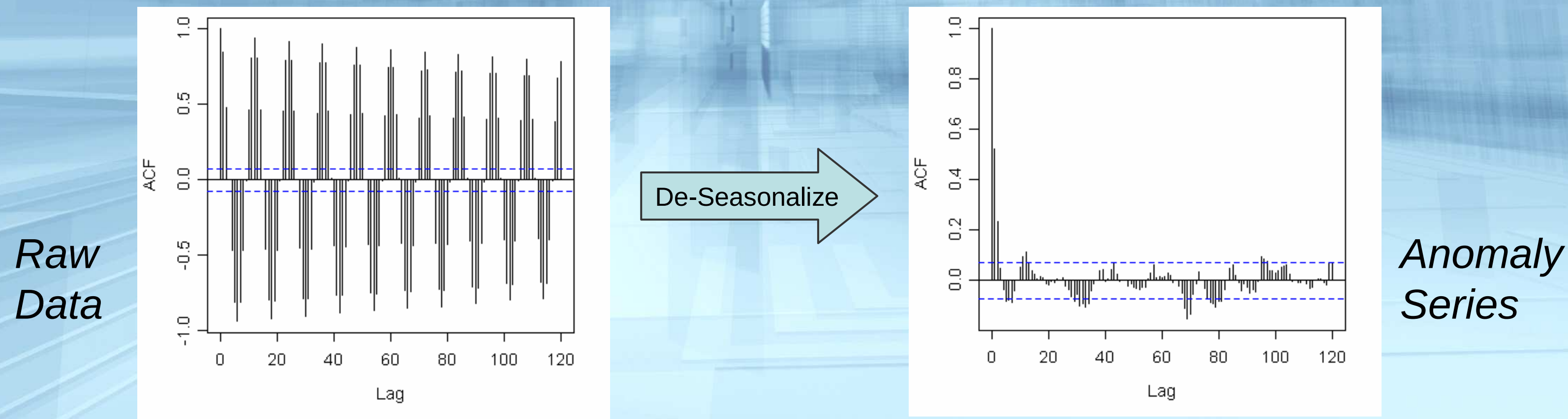


- But networks can also represent other types of data, framework for identifying relationships, patterns, etc.

Historical Climate Data

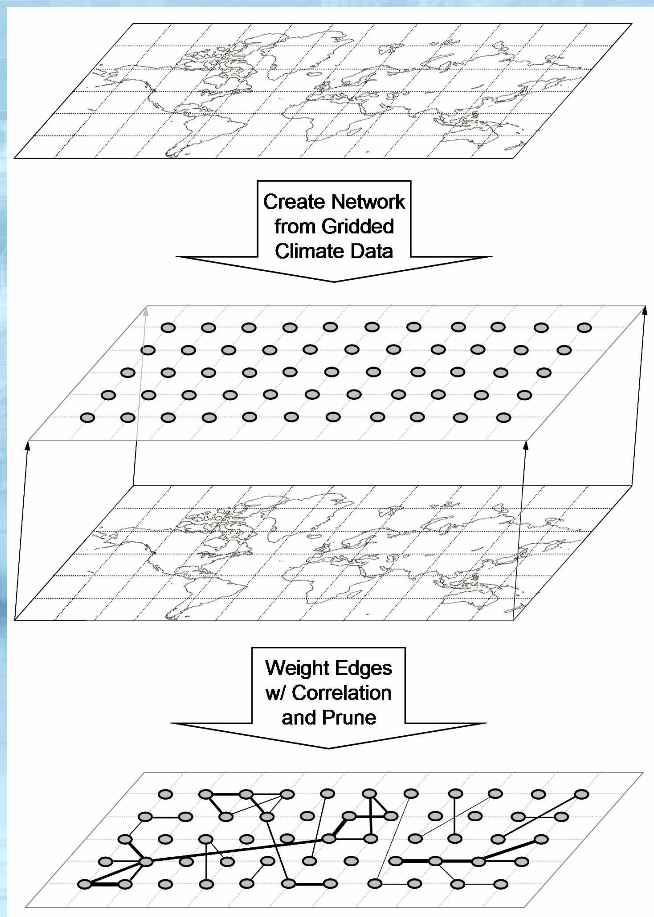
- NCEP/NCAR Reanalysis – proxy for observation
- Monthly for 60 years (1948-2007) on 5ºx5º grid

Sea surface temperature (SST), Sea level pressure (SLP)
Geopotential height (Z), Precipitable water (PW), Relative Humidity (RH), Horizontal (WSPD) / Vertical (w) wind speed



Network Construction

- View the global climate system as a collection of interacting oscillators
  - Nodes represent physical locations in space
  - Edges denote correlation in climate variability
- Link strength estimated by correlation, low-weight edges are pruned from the network



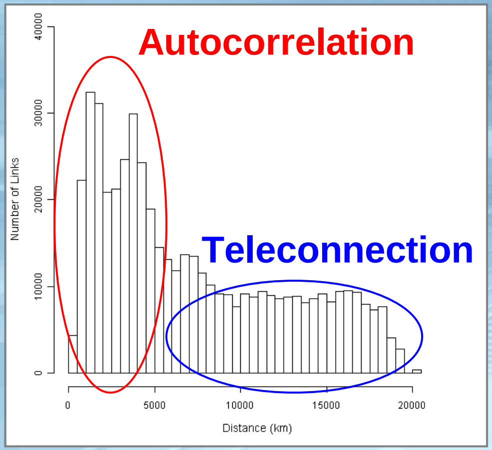
Network Topology

- Small-World (high clustering, short paths)
- Not scale-free (power law exponent α ~ 1)

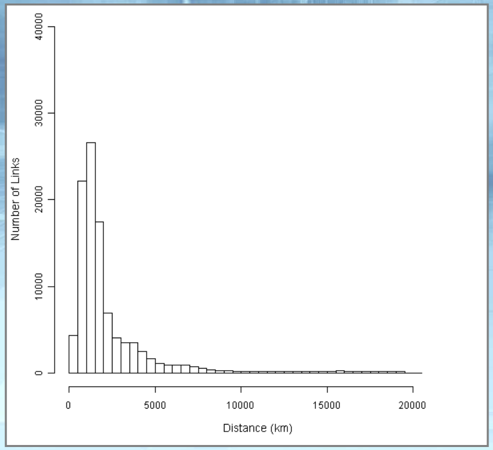
Var.	Nodes	Edges	C	L	α
SST	1,701	132,469	0.541	2.437	1.089
SLP	1,701	175,786	0.629	2.547	1.028
Z	1,701	249,322	0.673	2.436	1.286
PW	1,701	50,835	0.582	4.281	1.152
RH	1,700	25,375	0.559	4.063	1.150
WSPD	1,699	31,615	0.554	4.826	1.127
W	1,701	71,458	0.342	2.306	1.033

Geographic Properties

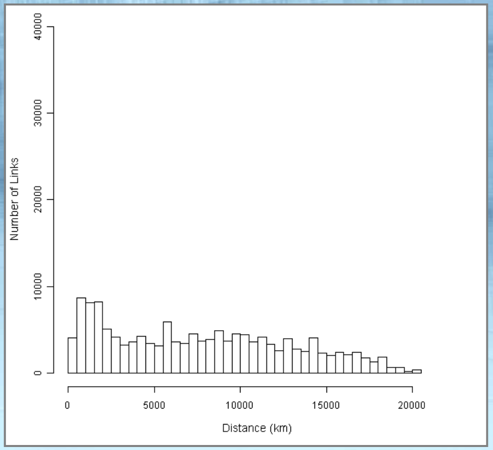
- Examine network structure in spatial context
  - Link lengths computed as great-circle distance
  - Compare autocorrelation / de-correlation lengths for different variables, interpret within the domain



Sea Level Pressure



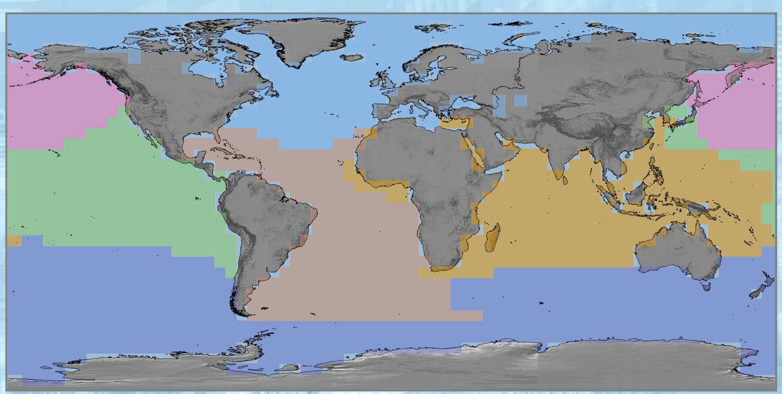
Precipitable Water



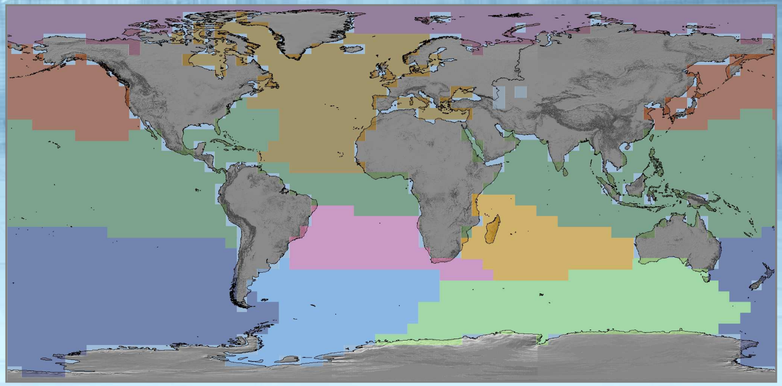
Vertical Wind Speed

Clustering Climate Networks

- Apply community detection to partition networks
- Visualize spatial pattern using GIS tools
- Cluster structure suggests relationships within the climate system



Sea Level Pressure



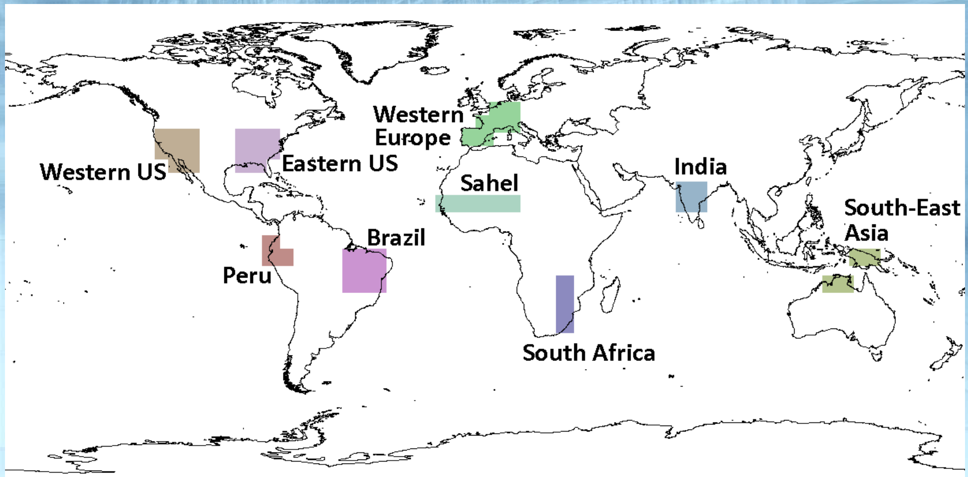
Precipitable Water

Descriptive → Predictive

- Network representation is able to capture interactions, reveal patterns in climate
  - Validate existing assumptions / knowledge
  - Suggest potentially new insights or hypotheses for climate science
- Want to extract the relationships between atmospheric dynamics over ocean and land
  - i.e., “Learn” physical phenomena from the data

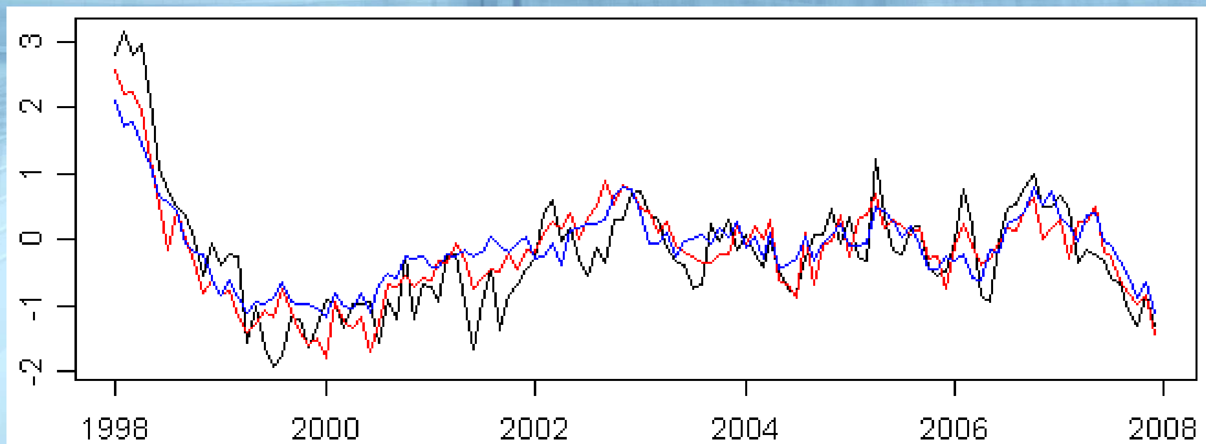
Case Study: Teleconnections

- Predictive models for ocean-based indicators
- Use network clusters as candidate predictors
- Create response variables for target regions (illustrated below)
- Build regression model relating ocean clusters to land climate



Illustrative Example

- Predictive model for air temperature in Peru
  - Long-term variability highly predictable due to well-documented relation to El Nino
- Small number of clusters have majority of skill
  - Feature selection (blue line) improves predictions



Results on Train/Test

	Region	Network Clusters	K-Means		
			$k = 5$	$k = 10$	$k = k_n$
Air Temperature	SE Asia	**0.541**	0.629	0.694	*0.886*
	Brazil	0.534	*0.536*	0.532	**0.528**
	India	**0.649**	0.784	*1.052*	0.791
	Peru	**0.468**	0.564	*0.623*	0.615
	Sahel	**0.685**	0.752	0.750	*0.793*
	S Africa	0.726	**0.711**	*0.968*	0.734
	East US	0.815	0.824	*0.844*	**0.811**
	West US	**0.767**	0.805	0.782	*0.926*
	W Europe	0.936	*1.033*	**0.891**	0.915
	Mean	**0.680**	0.737	*0.793*	0.778
	StdDev	±0.150	±0.152	±*0.165*	±**0.135**
Precipitation	SE Asia	**0.665**	0.691	*0.700*	0.684
	Brazil	**0.509**	0.778	*0.842*	0.522
	India	**0.672**	0.813	0.823	*0.998*
	Peru	**0.864**	*1.199*	1.095	1.130
	Sahel	**0.533**	*0.869*	0.856	0.593
	S Africa	**0.697**	*0.706*	0.705	0.703
	East US	0.686	0.750	*0.808*	**0.685**
	West US	**0.605**	0.611	*0.648*	0.632
	W Europe	**0.450**	*0.584*	0.549	0.542
	Mean	**0.631**	0.778	*0.781*	0.721
	StdDev	±**0.124**	±0.182	±0.156	±*0.207*
Friedman Test ($\alpha = 0.05$)			✓	✓	✓

Variations / Extensions

- Compare network approach to traditional clustering methods
  - k-means, k-medoids, spectral, EM, etc.
- Compare different types of predictive models
  - (linear) regression, regression trees, neural nets, support vector regression
- Evaluate both model fit and performance on split train/test sets

Computational Issues

- Network Construction
  - 5°x5° network has O(10^6) pairs of nodes, calculating correlations takes thousands of CPU-hours
  - High-resolution data available (both space and time) increases computational complexity
- Predictive Modeling
  - Currently considering only 18 sample target regions, want to predict everywhere (thousands of locations)

Open Questions

- Nonlinearity – known to exist in climate, but relevance in network context not fully explored
- Multivariate Relationships – must be integrated within networks for realistic representation
- Spatio-Temporal Dynamics – capture relative stability and/or changes in structure over time
- Predictive Modeling – work with domain experts to define relevant predictive tasks

Summary

- Complex networks provide a flexible data representation and powerful analysis tool
- Data mining methods able to extract complex relationships from observed data and potentially augment or suggest new domain knowledge
- Interdisciplinary approach required from problem definition to analysis and interpretation of results

Thanks & Questions

Acknowledgements
National Science Foundation Grant No. 0826958; LDRD Program of the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

DIAL Research Group
http://www.nd.edu/~dial

Interdisciplinary Center for Network Science & Applications
http://icensa.nd.edu