Construction and Analysis of Climate Networks

Karsten Steinhaeuser
University of Minnesota

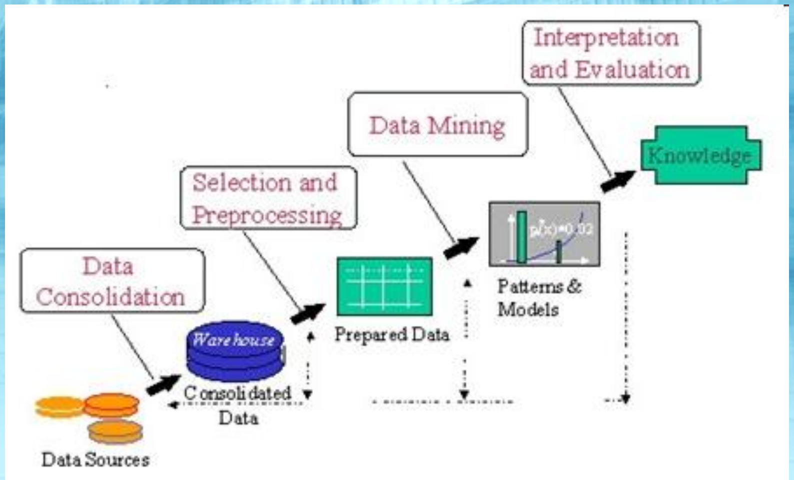
Workshop on Understanding Climate Change from Data
Minneapolis, MN

August 15, 2011

Working Definitions

- *Knowledge Discovery* is the process of identifying valid, novel, potentially useful, and ultimately understandable structure in data (source: ACM SIGKDD)

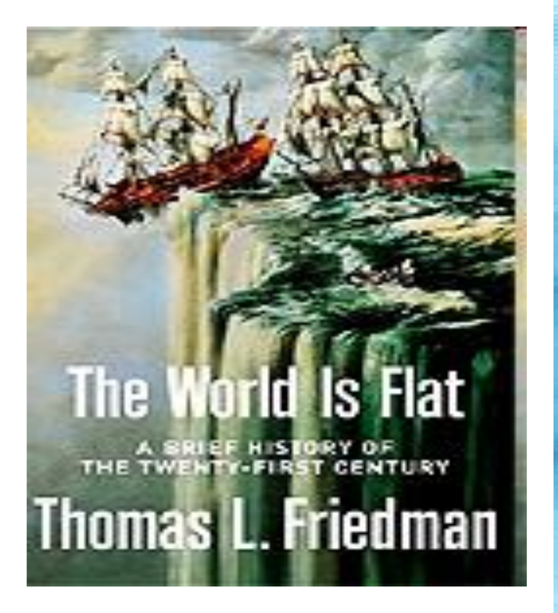
- *Data Mining* is the step in the knowledge discovery process concerned with identifying patterns in data and building models to represent those patterns



Mining Complex Data

- Complex spatio-temporal data pose unique challenges

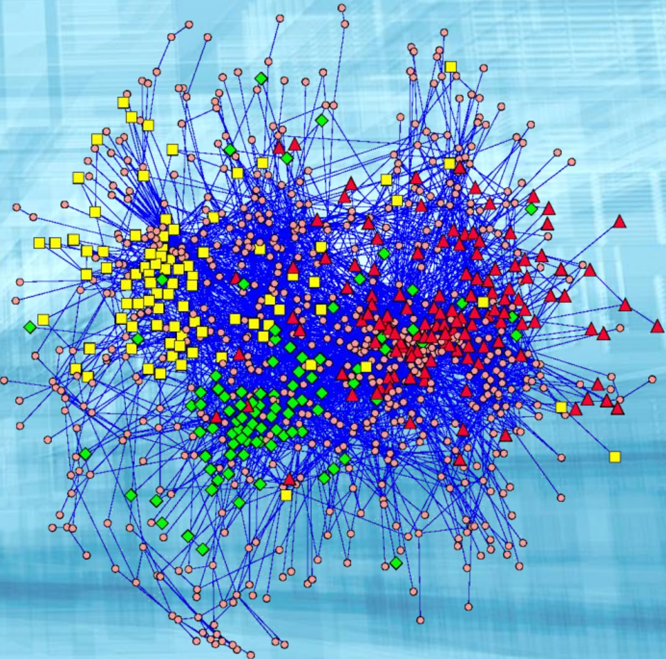
- Tobler's First Law of Geography: "Everything is related, but near things more than distant."
  - But are all near things equally related?
  - Are there phenomena explained by interactions among distant things? (teleconnections)



Networks Primer

What is a Network?

- Oxford English Dictionary:
  network, n.: Any netlike or complex system or collection of interrelated things, as topographical features, lines of transportation, or telecommunications routes (esp. telephone lines).



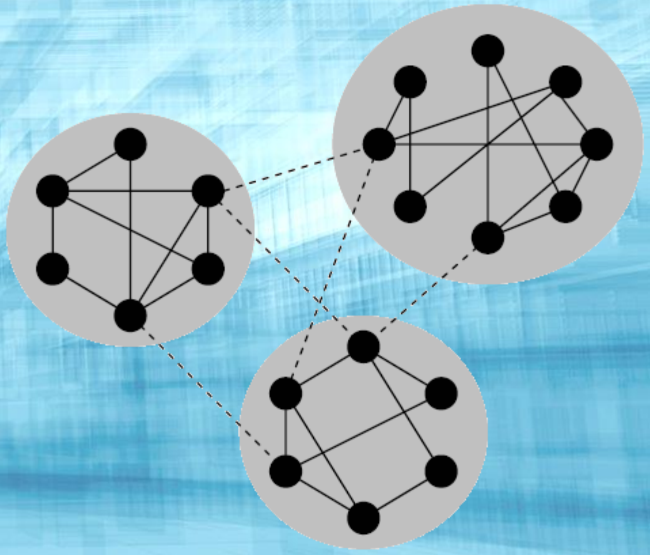
- My working definition:
  Any set of items that are connected or related to each other.
  (“items” and “connections” can be concrete or abstract)

Networks Primer

Community Detection in Networks

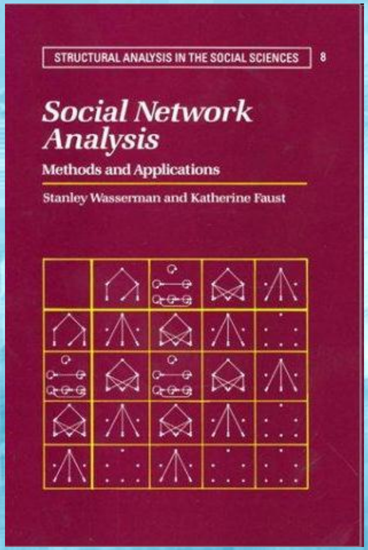
- Identify groups of nodes that are relatively more tightly connected to each other than to other nodes in the network

- Computationally challenging problem for real-world networks



Climate + Networks?

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

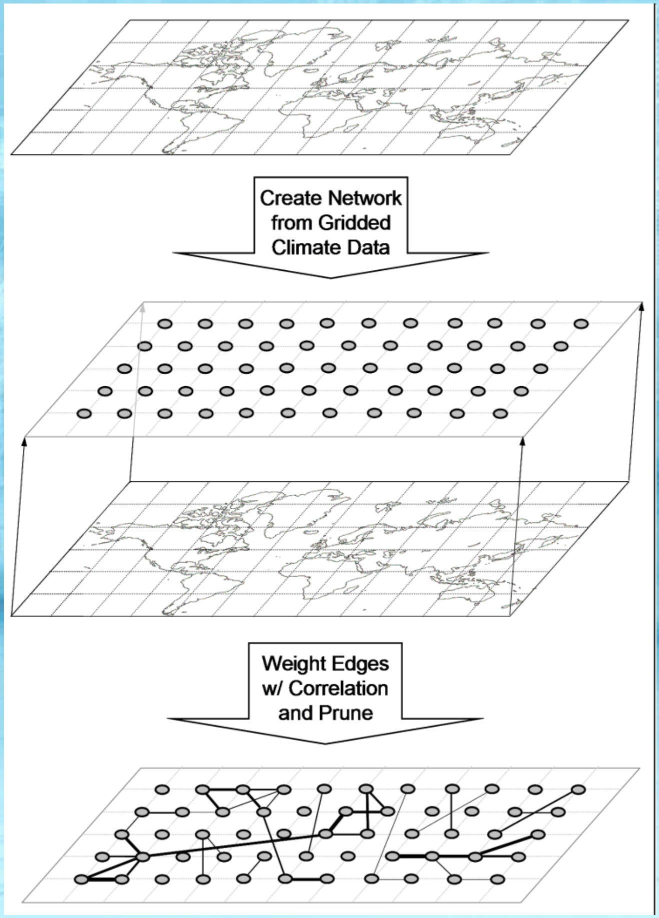


- Many datasets explicitly define network structure

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

Network Construction

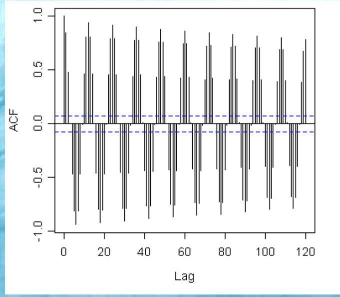
- View the global climate system as a collection of interacting oscillators [Tsonis & Roebber, 2004]
  - Vertices represent physical locations in space
  - Edges denote correlation in climate variability
- Link strength estimated by correlation, low-weight edges are pruned from the network



Example: Historical Data

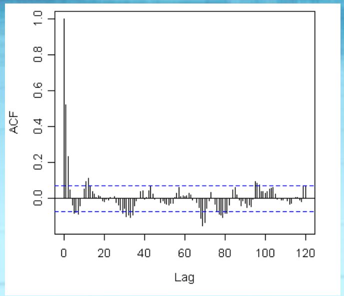
- NCEP/NCAR Reanalysis (proxy for observation)
- Monthly for 60 years (1948-2007) on 5ºx5º grid
- Seven variables:

  Sea surface temperature (SST)
  Sea level pressure (SLP)
  Geopotential height (GH)
  Precipitable water (PW)
  Relative Humidity (RH)
  Horizontal wind speed (HWS)
  Vertical wind speed (VWS)



Raw Data

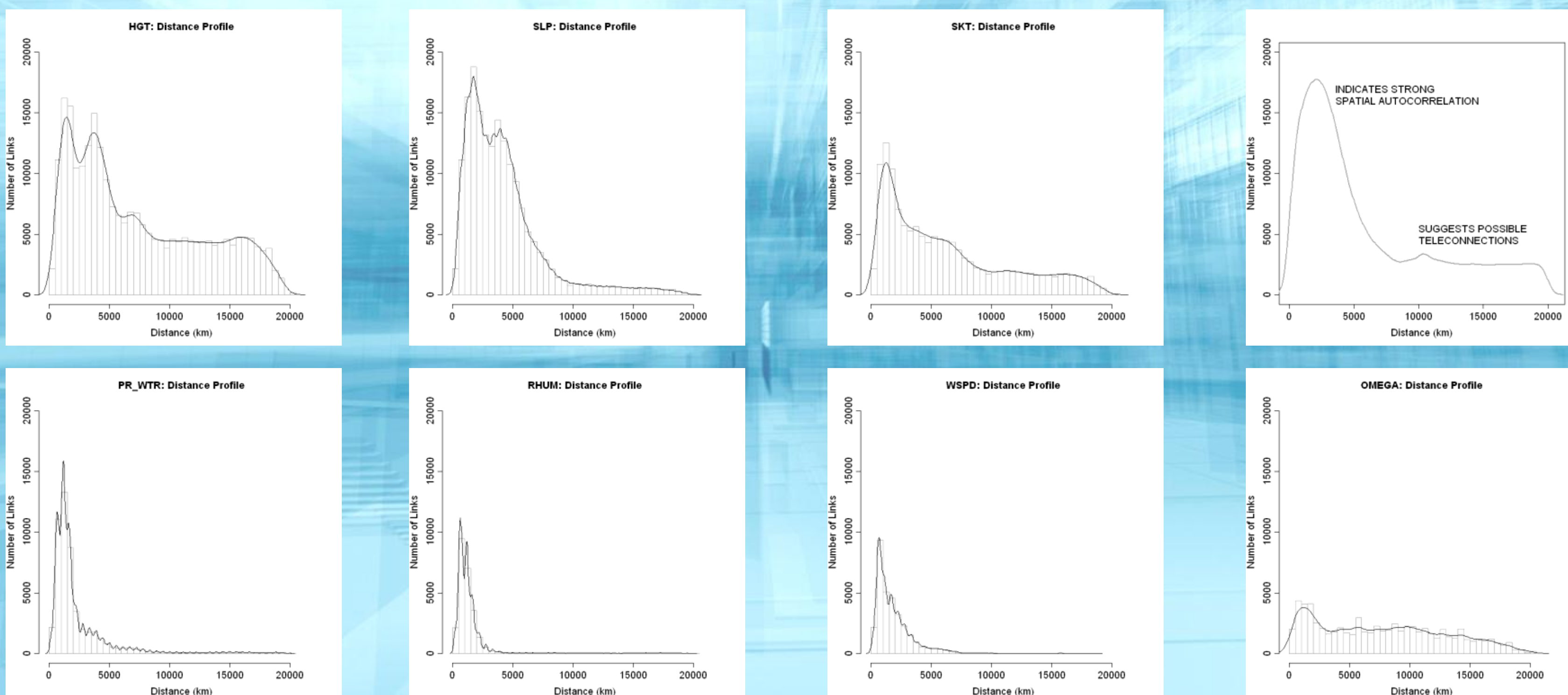
De-Seasonalize



Anomaly Series

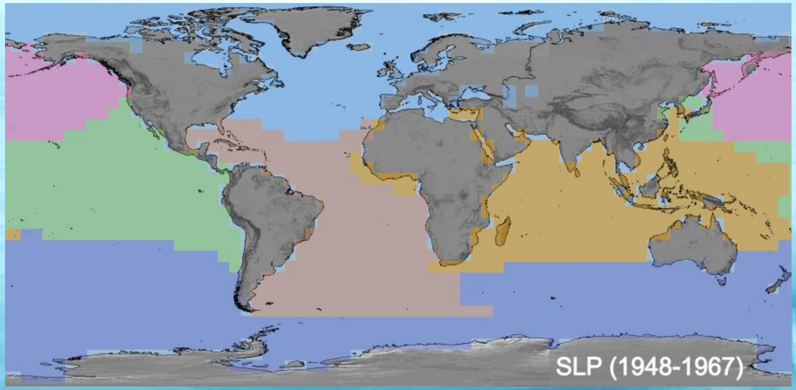
Geographic Properties

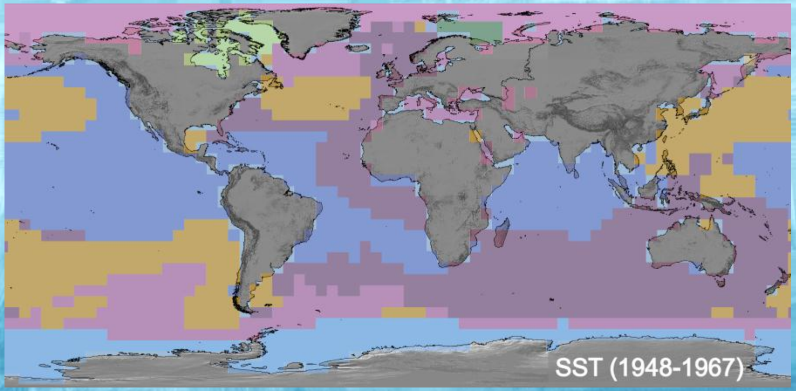
- Examine network structure in spatial context (link lengths computed as great-circle distance)

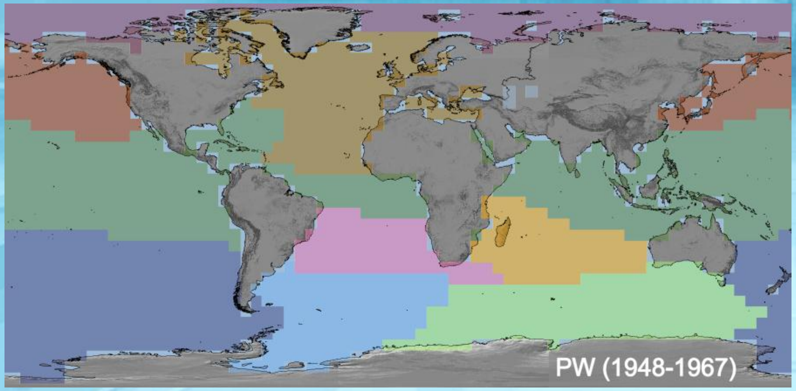


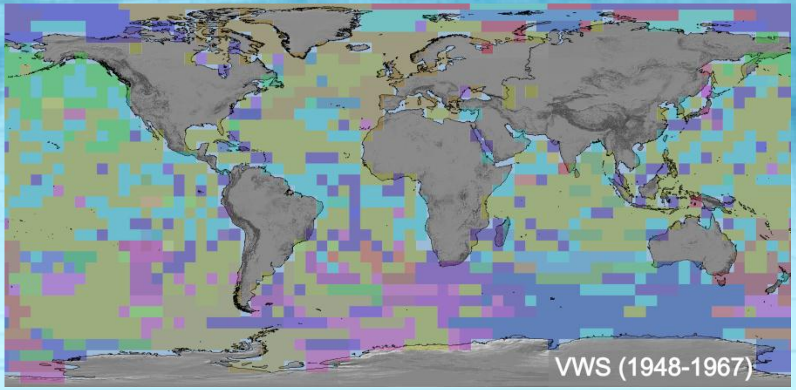
Clustering Climate Networks

- Apply community detection to partition networks into “clusters”



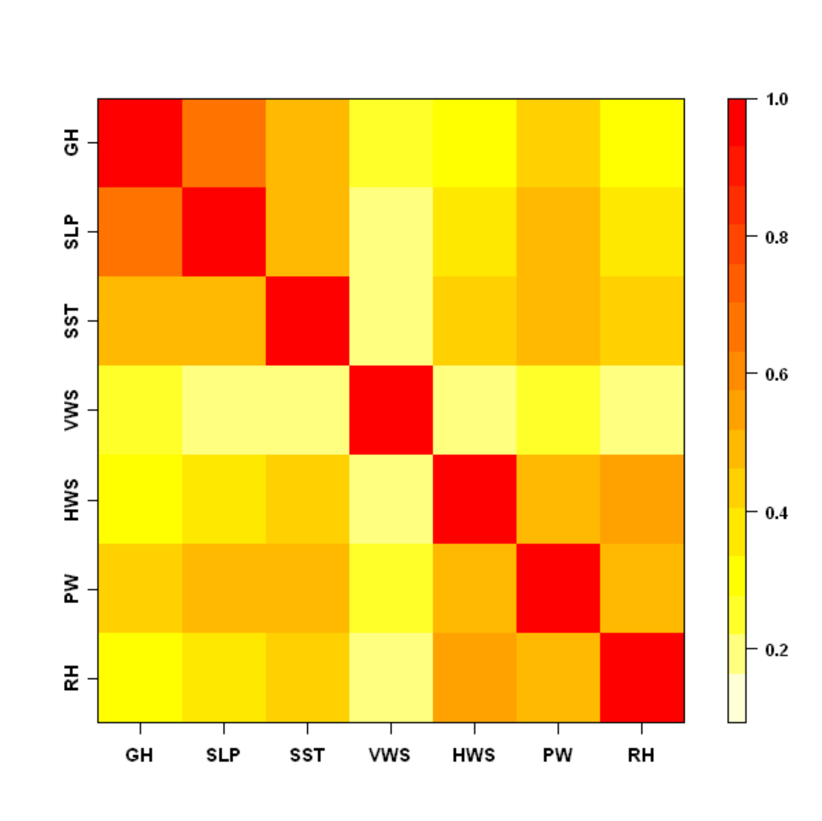






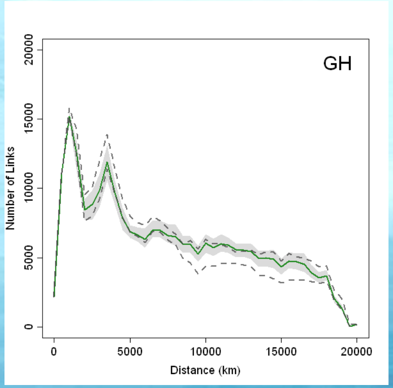
Cluster Similarity

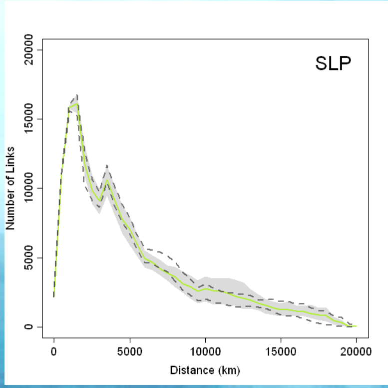
- Compare cluster structure between different variables
  - Computed Adjusted Rand Index (ARI) to quantify similarity
  - Groupings correspond to those identified from the profiles

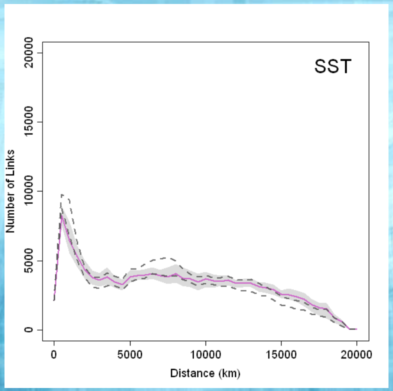


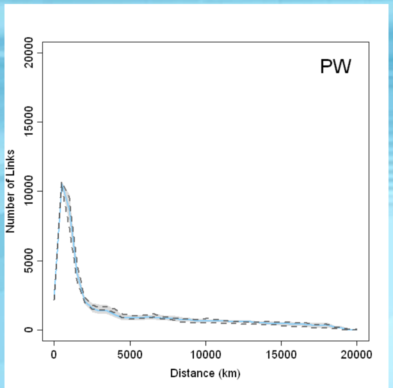
Stability over Time

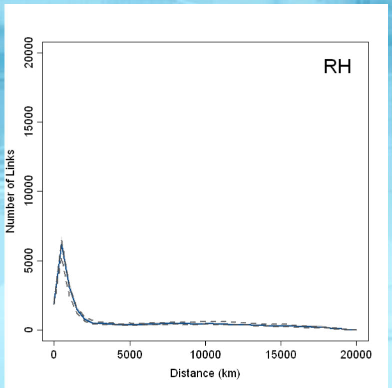
- Temporal evolution of profiles from sliding windows

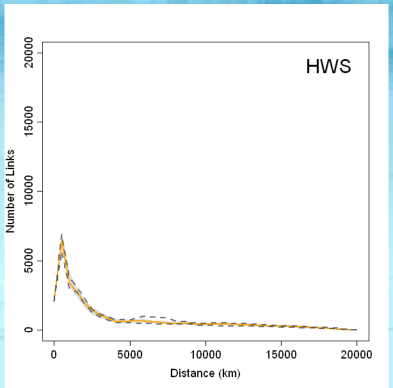


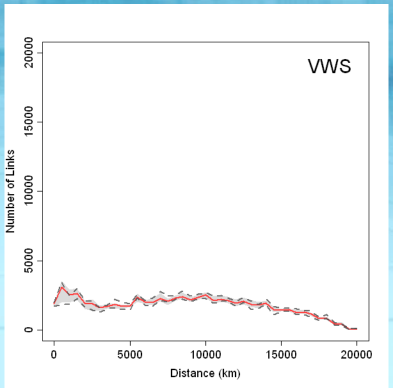










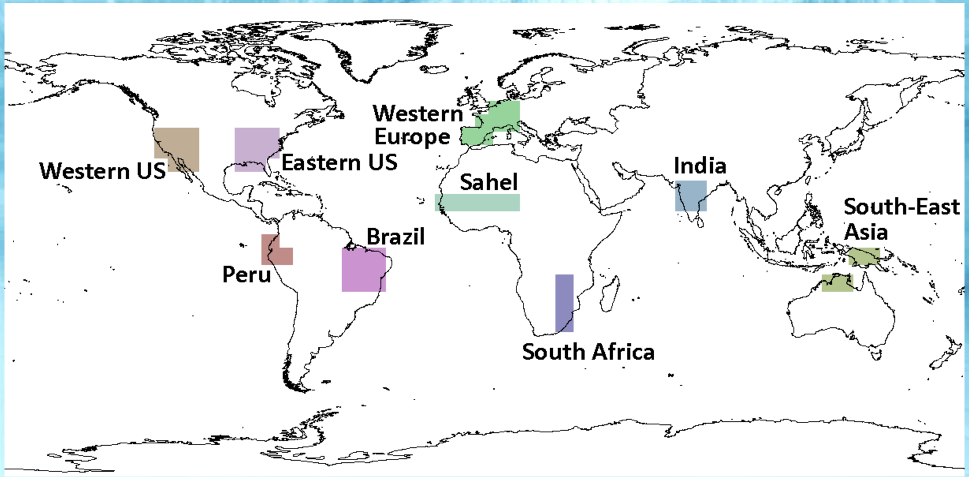


Predictive Modeling

- 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

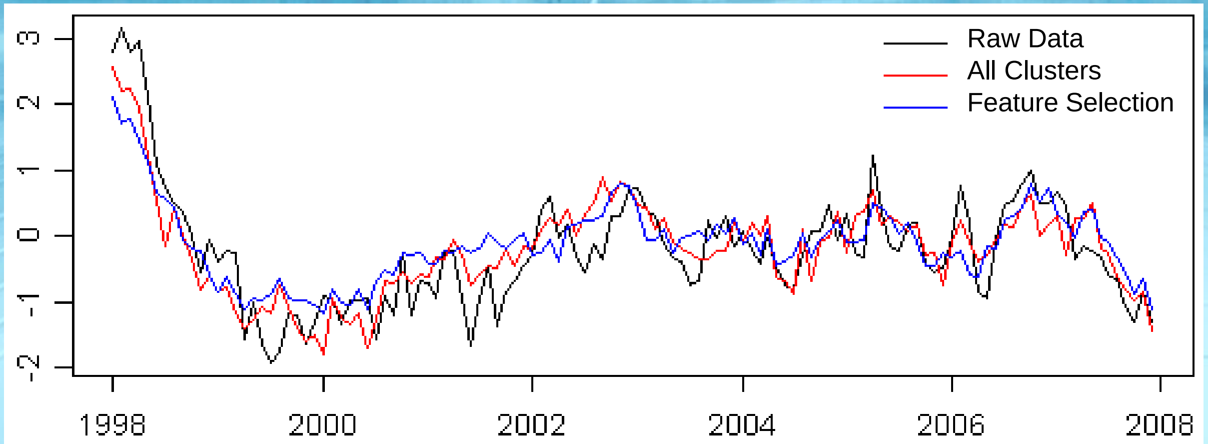
Predictive Modeling

- Use network clusters as candidate predictors
- Create response variables for target regions
- Build regression models 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-Means $k=10$	K-Means $k=k_n$
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 ($\alpha = 0.05$)			✓	✓	✓

Work-in-Progress

- More thoroughly incorporate multivariate relationships, nonlinearity, and temporal lags into network construction and predictive models
- Work with domain experts to integrate our (data-guided) predictive models with (physically-based computational) climate models
- Address computational issues arising from predictions at higher model resolution (both spatial and temporal), multiple variables, mean and extreme events, etc.

Upcoming Events

1. First International Workshop on Climate Informatics, New York, NY, August 26, 2011
   http://www.nyas.org/climateinformatics

2. NASA Conference on Intelligent Data Understanding (CIDU), Mountain View, CA, Oct 19-21, 2011
   https://c3.ndc.nasa.gov/dashlink/projects/43/

3. IEEE ICDM Workshop on Knowledge Discovery from Climate Data, Vancouver, Canada, December 10, 2011
   http://www.nd.edu/~dial/climkd11/

Thanks & Questions

Contact:

ksteinha@umn.edu

Personal Homepage

http://www.umn.edu/~ksteinha

NSF Expeditions Homepage

http://climatechange.cs.umn.edu/