Descriptive and Predictive Analysis of Climate Data

Karsten Steinhaeuser, Ph.D. Candidate

Department of Computer Science & Engineering
Data, Inference, Analysis & Learning Laboratory
Also affiliated with the Interdisciplinary Center for Network Science and Applications (ICeNSA)
University of Notre Dame

Advisor: Nitesh V. Chawla

Department of Computer Science & Engineering
Data, Inference, Analysis & Learning Laboratory
Associate Director of the Interdisciplinary Center for Network Science and Applications (ICeNSA)
University of Notre Dame

Co-Advisor: Auroop R. Ganguly

Geographic Information Science and Technology Group (GIST)
Computational Sciences& Engineering Division
Oak Ridge National Laboratory

Motivation & Context

Why Study Applications in Climate?

- Exciting new challenges for computational thinking
- Bridges interdisciplinary gap with societal impact

Grand Challenges – DOE Strategic Plan

1. Characterize the Earth's current climate and its evolution over the last century to its present state *(descriptive analysis)*
2. Predict regional climate change for the next several decades *(predictive analysis)*
3. Simulate Earth System changes and their consequences over centuries *(not addressed)*

Climate Data

Historical Observations

- Paleoclimate indicators (tree rings, ice cores)
- Measurements (in-situ / remote sensors)
- Reanalysis Products (post-processed data)

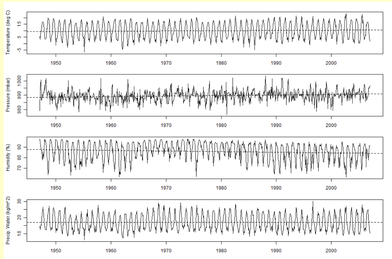
Climate Models

- Physically-based simulations of earth system
- Initialized with known conditions
- Run in hindcast (tuning, verification) or forecast (projection) modes

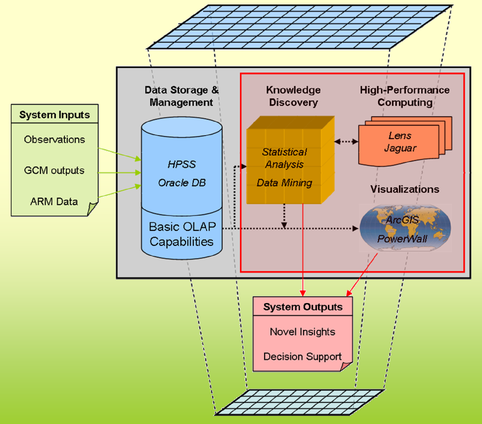
NCEP/NCAR Reanalysis

Best Available Proxy for a Global Record

Monthly measurements over a 60-year period (1948-2007) fitted to 2.5°x2.5° spatial grid.



Climate KD Framework



Descriptive Analysis

Objective: Discover Patterns in Historical Climate Data

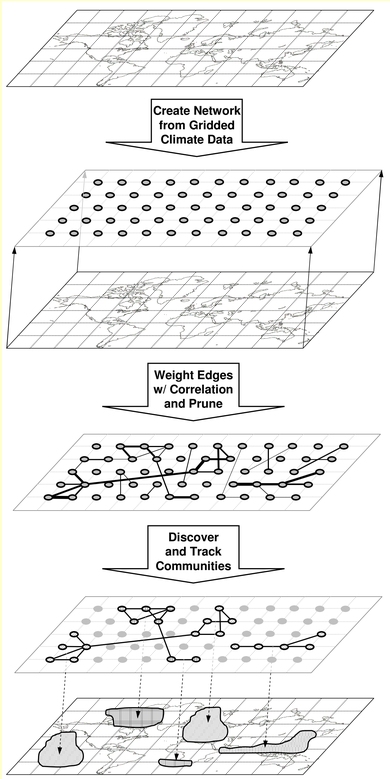
Traditionally hypothesis-driven, but better historical records are becoming available for analysis.

We propose to employ a ***data-driven approach***.

Challenges

Lots of different potential data are available for mining.

Moreover, many different types of relations are known to exist, including spatially disjoint regions (teleconnections), non-linear relationships, and dependence between extreme events.



Methodology

Complex Networks

Loosely defined as a set of related entities represented by nodes and connected by (weighted) edges.

Networks have measurable properties that provide rich information about the data and enable the analysis of complex relationships.

Climate Networks

A Versatile and Flexible Data Representation

Nodes represent geographic locations and weighted edges are placed between them based on the climatic relationship (no spatial constraints).

This network is analyzed in terms of its fundamental properties, clustering, and changing characteristics over time.

Network Construction

Capture Interaction Between Multiple Climate Variables

Edge weighting is critical to exploit the analytic power of networks, but is often ignored or done in an *ad hoc* manner.

We compute the pairwise cross correlation between all climate variables at each location (in this case ${}_4C_6 = 6$ pairs) and compute similarity between locations as Euclidean distance in this new R^6 feature space.

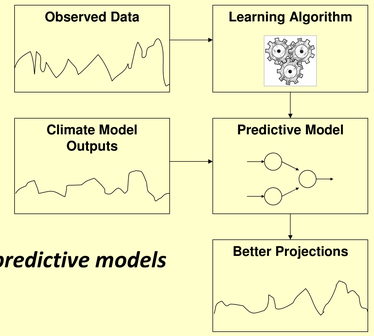
$$CCF(A,B,d) = \frac{\sum_{i=i}^{t}\left[(a_i-\bar{a})(b_{i-d}-\bar{b})\right]}{\sqrt{\sum_{i=1}^{t}(a_i-\bar{a})^2\sum_{i=1}^{t}(b_{i-d}-\bar{b})^2}}$$

Predictive Analysis

Objective: Predict Future Climate

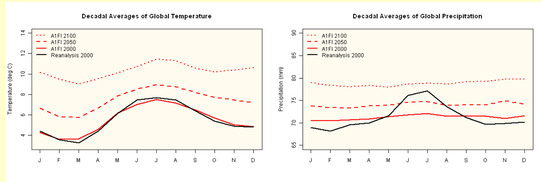
Traditionally uses physically-based simulations of the earth systems, but limited by our understanding and ability to model the physical complex physical processes.

We propose to use ***network-based predictive models*** to refine climate model outputs.

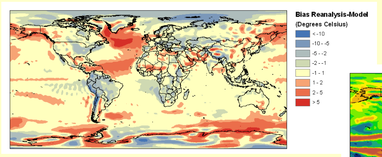


Challenges

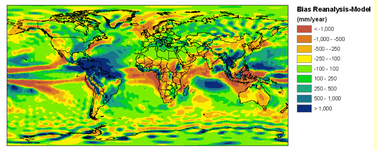
Limited understanding of the physical processes; lack of validation data; changing conditions in the future – scenarios.



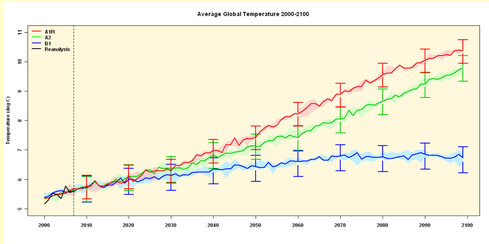
Prognostic vs. Diagnostic Variables



Geographic Variability



Emissions scenarios explore space of hypothetical future states of the world.



Intellectual Merits

Descriptive Analysis

Representing climate data as complex network without spatial constraints; applying non-linear correlation and extremes dependence in this context; using graph-based algorithms to discover space-time patterns from data.

Predictive Analysis

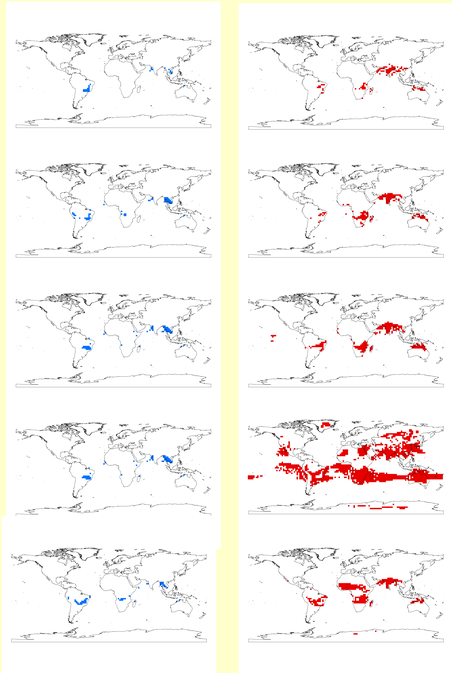
Incorporating observed data with models to improve predictive insights; reduced uncertainty in variables poorly captured by models; leveraging high-performance computing resources for spatio-temporal prediction.

Acknowledgement This work was supported in part by National Science Foundation Grant CNS-07-20813. This research was also funded in part by the project entitled "Uncertainty Assessment and Reduction for Climate Extremes and Climate Change Impacts" under the initiative "Understanding Climate Change Impact: Energy, Carbon, and Water Initiative" within the Laboratory Directed Research and Development (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.

Preliminary Results

Clustering and Network Dynamics

- Divide data into 5-year windows
- Construct network and prune edges
- Identify clusters and track over time
- Visualize using GIS capabilities



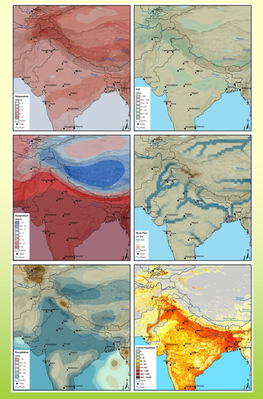
Challenges

Algorithmic Issues

Variable selection; principled methods for network construction and edge weighting; alternate measures of similarity; dynamics.

Computational Issues

Computing pairwise similarity between all locations (10K grid points ~ 55M pairs!) may need high-performance computing solution.

Broader Significance

- Improved inputs for impacts assessment and policy discussions
- Guide the development of climate models
- Undergraduate teaching and mentorship