TELECONNECTION SIGNALS EFFECT ON TERRESTRIAL PRECIPITATION: BIG DATA ANALYTICS VS. WAVELET ANALYSIS

Yahui Di[1], Wei Ding[1], Sanaz Imen[2], Ni-Bin Chang[2]

1 Department of Computer Science, University of Massachusetts Boston, Boston MA, USA

2 Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, USA

Introduction

Purpose:

- Determine the associations between hydro-climatic variables and the atmospheric / oceanic variables separated by large distances, which are known as the phenomenon of **hydro-climatic teleconnection**.
- Discover physically meaningful patterns from big climate databases.

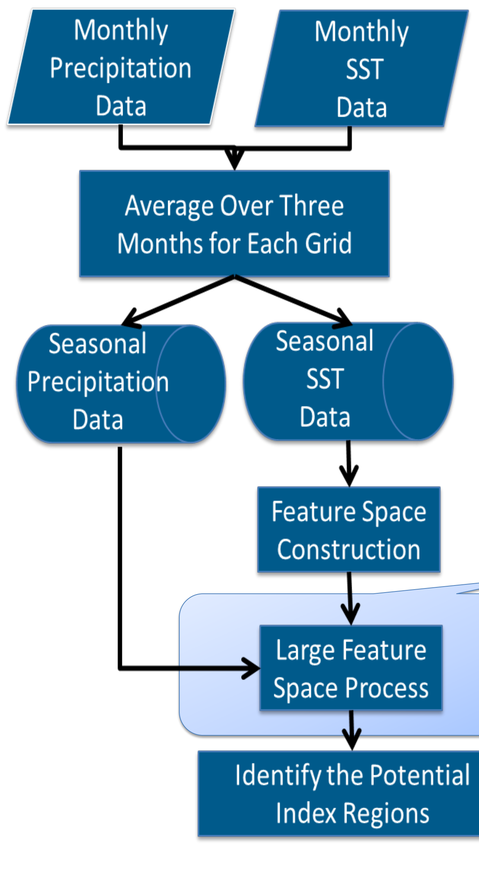
Methodology: develop efficient data-driven approaches with the aid of machine learning, signal processing, and domain knowledge for constrained search.

- **Big Data Analytics:** extract hydro-climatic variables from large temporal and spatial feature space and formulate the global search for teleconnection signals effect on terrestrial precipitation as feature selection in machine learning aspect.
- **Wavelet Analysis:** retrieve the scale-averaged wavelet power to signify the teleconnection signals via a pixel-wise linear lagged correlation analysis.

Methodology

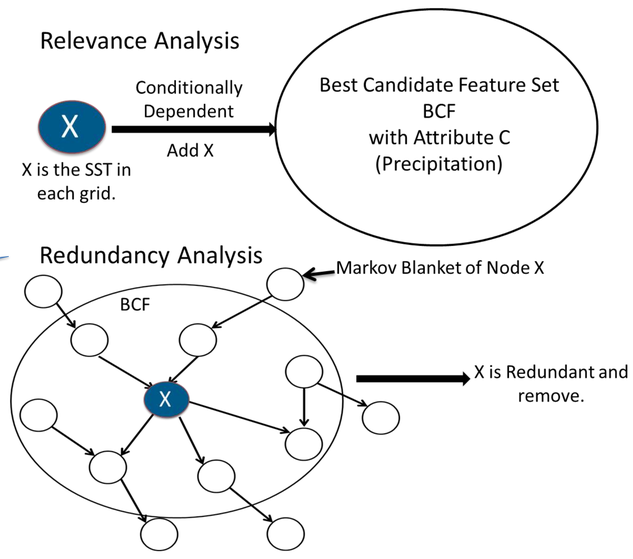
Big Data Analytics

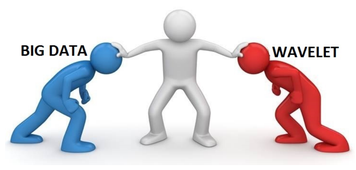
Use an efficient streaming feature selection method [3] to identify strongly relevant non-redundant features from extremely high feature space.



Build a local Bayesian network and efficiently selects strongly relevant features using Markov blankets.

- Process the large volume of features sequentially evaluated one feature at a time—a feature is evaluated upon its arrival and is decided whether it can be added into the current pool of selected features.
- Discover Markov blankets for each feature seen so far using the direct causes and direct effects that produces the skeleton of a Bayesian network to avoid the expensive calculation of a full Bayesian network.



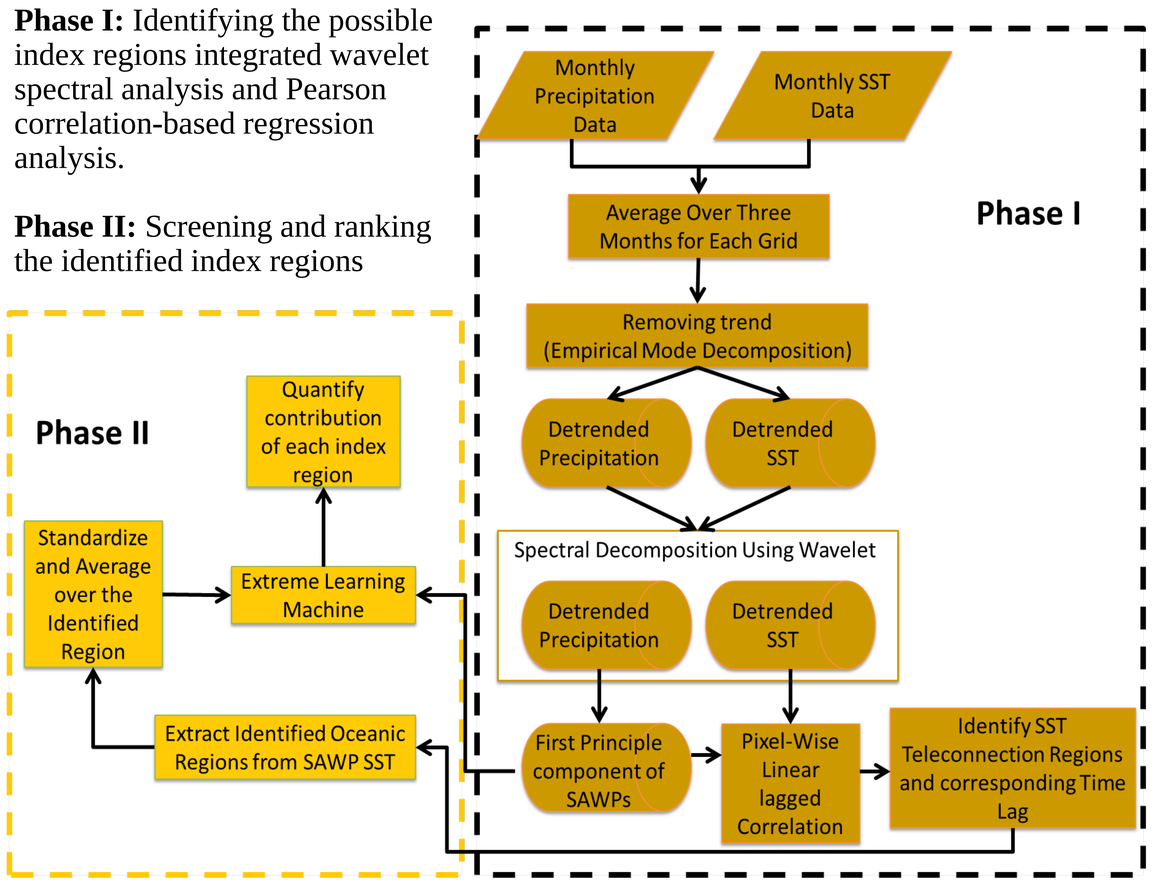


Wavelet Analysis

Use integrated wavelet spectral analysis and Pearson correlation-based regression analysis to search for nonlinear and non-stationary signals of climate teleconnections associated with terrestrial sites [2].

Phase I: Identifying the possible index regions integrated wavelet spectral analysis and Pearson correlation-based regression analysis.

Phase II: Screening and ranking the identified index regions



Preliminary Results

Data Description

Adirondack, NY was chosen as the study area.

Precipitation Data

- Full data product of Global Precipitation Climatology Center (GPCC-V6).
- 1980-2010
- Spatial resolution: 0.5º ×0.5º

Sea Surface Temperature (SST) Data

- ERA-Interim reanalysis product
- 1979-2010
- Spatial resolution: 1.5º ×1.5º

SSTs time series with time lags from 0 to 12 months.
E.g. March-April-May (Spring), the associated SST time series starting with March-April-May (0 month lag), February-March-April (1 month lag), January-February-March (2 months lag), ..., March-April-May (12 months lag).

Results

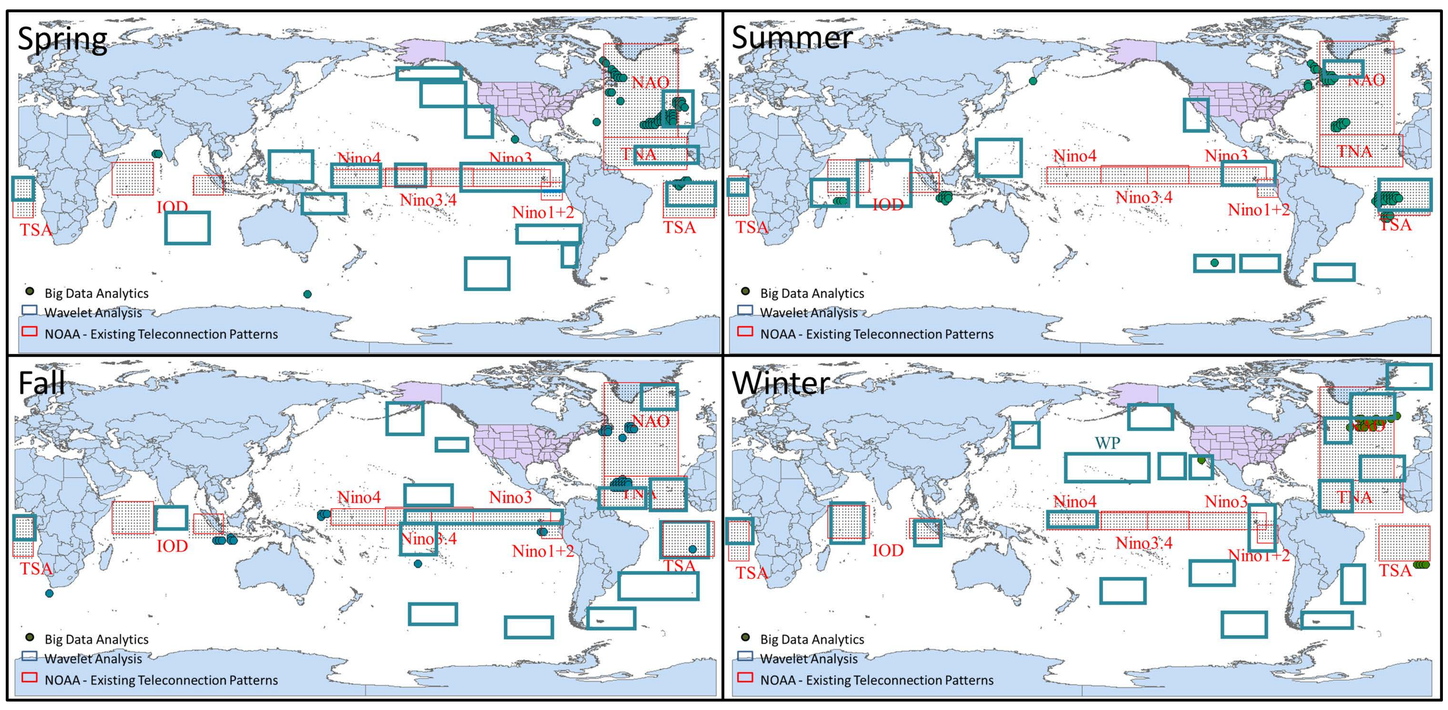


Figure 1 The comparison among NOAA-Existing Teleconnection Patterns and two different methods' findings.

The found index regions:

- Somewhat overlapped.
- Located in or near the area of NOAA ocean Patterns.

Next Step:

- Study how the different regions founded by two methods.
- Identify which regions contribute higher to the precipitation variability at Adirondack site.
- Integrate streaming feature selection and wavelet analysis to be more efficiently identify physically meaningful teleconnection patterns from big climate data.

Reference

[1] N. B. Chang, M. V. Vasquez, C.F. Chen, S. Imen, and L. Mullon. "Global nonlinear and nonstationary climate change effects on regional precipitation and forest phenology in Panama, Central America." Hydrological Processes 29, no. 3 (2015): 339-355.

[2] L. Mullon, N. B. Chang, Y. J. Yang, and J. Weiss. "Integrated remote sensing and wavelet analyses for screening short-term teleconnection patterns in northeast America." Journal of Hydrology 499 (2013): 247-264.

[3] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu. "Online feature selection with streaming features." Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, no. 5 (2013): 1178-1192

[4] Y. Di, W.Ding, Y.Mu, D.L. Small, S. Islam, and N. B. Chang. "Developing machine learning tools for long-lead heavy precipitation prediction with multi-sensor data." In Networking, Sensing and Control (ICNSC), 2015 IEEE 12th International Conference on, pp. 63-68. IEEE,2015.