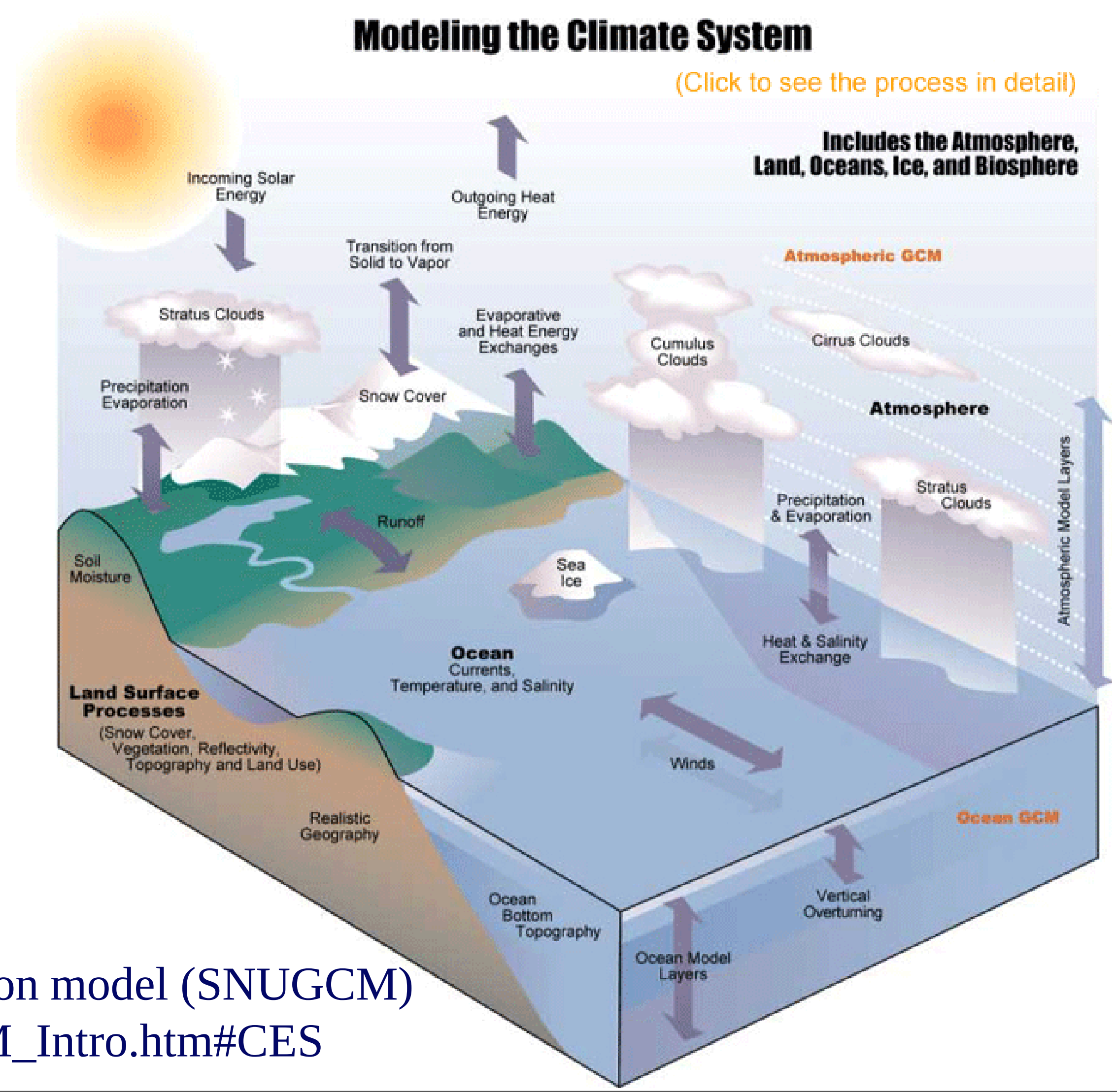
General Circulation Models

GCMs

Models of the atmosphere built from fundamental conservation laws governing the physical behavior of the atmosphere, and use numerical methods to obtain the solution to the system of coupled governing equations.

Physics in part of GCM calculate the forcing terms in governing equations (Partial Differential Equations).



Seoul National University general circulation model (SNUGCM)
http://climate.snu.ac.kr/gcmdocu/GCM_Intro.htm#CES

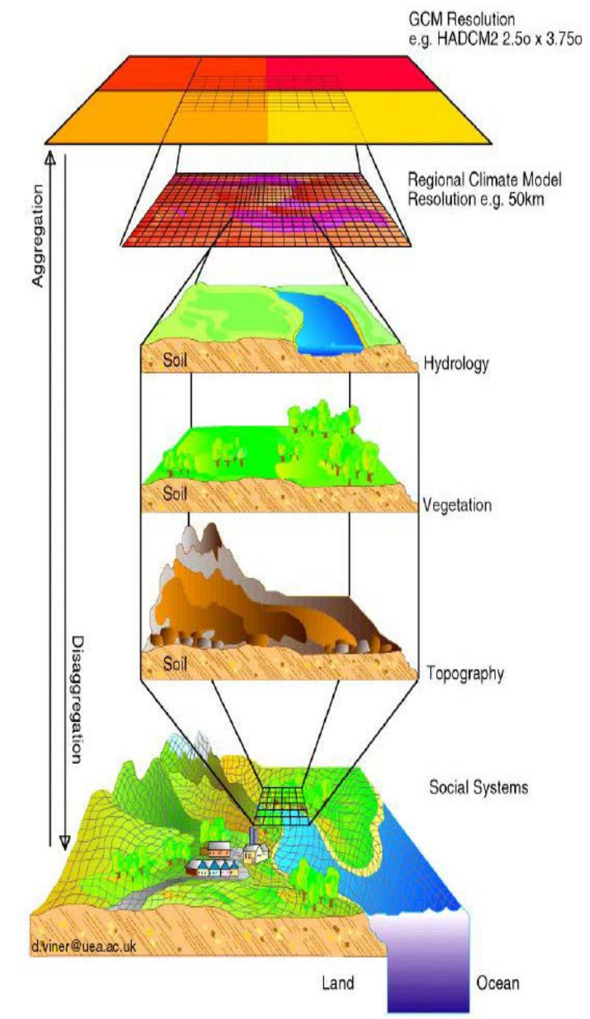
What are the difficulties with GCMs?

GCM for Climate Change Studies

Difficulties with Downscaling Methods

- Not able to address the hydrological **variability and extremes well**
- Dynamical downscaling methods can only run in **a short period of time** and are limited to address **single or a few GCM outputs**

Provide initial information for public decision makers and cannot be applicable for risk /reliability analysis



http://www.cccsn.ec.gc.ca/images/downscaling01.jpg

University of Central Florida

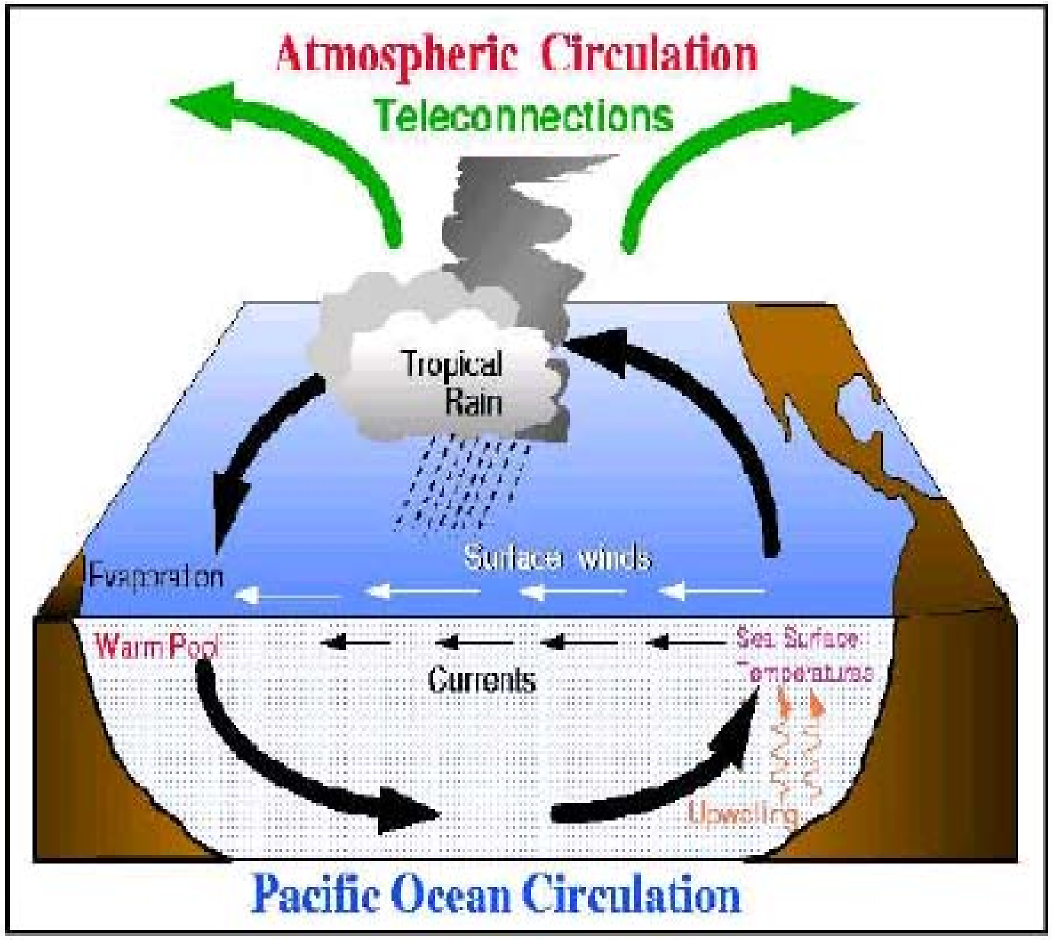
Definition of Teleconnection

Hydroclimatic Teleconnection

Determining the **statistical relationship** between **hydrological variables** and the **Atmospheric / Oceanic variables** separated by **large distances**

- Teleconnection across a wide region at sub-continental scale can be hardly analyzed by using linear analysis directly.

- Existence of non-stationary signals makes the identification of teleconnection complicated at a local scale.



http://www.sarcs.org/new/issp/Isspoe19.jpg

University of Central Florida

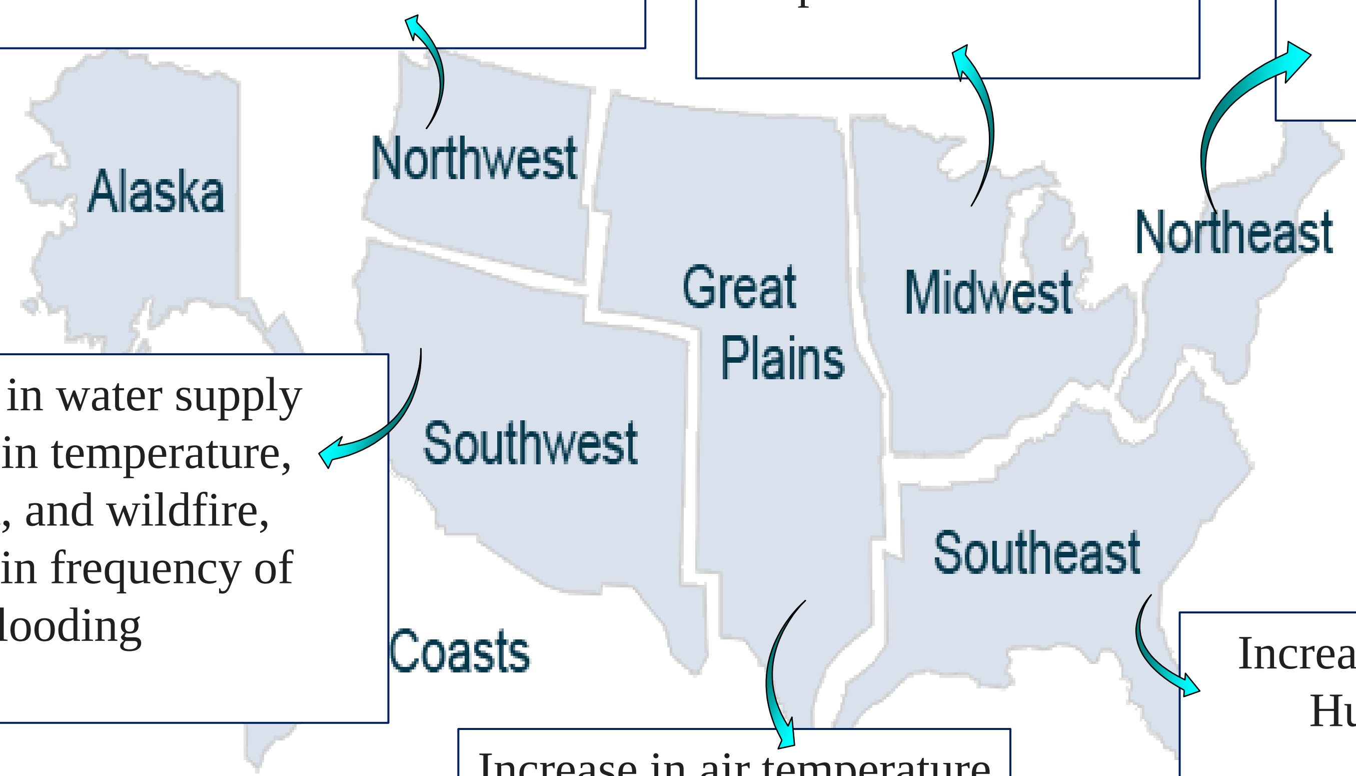
Current Prediction of Climate Changes

2009 Key Climate Issues - United States Global Change Research Program

Decline in springtime snowpack streamflow
Increase in sea-level rise

Increase in precipitation in winter / spring
Evaporation in summer

Severe flooding
Heavy downpours
Reduction in snow cover

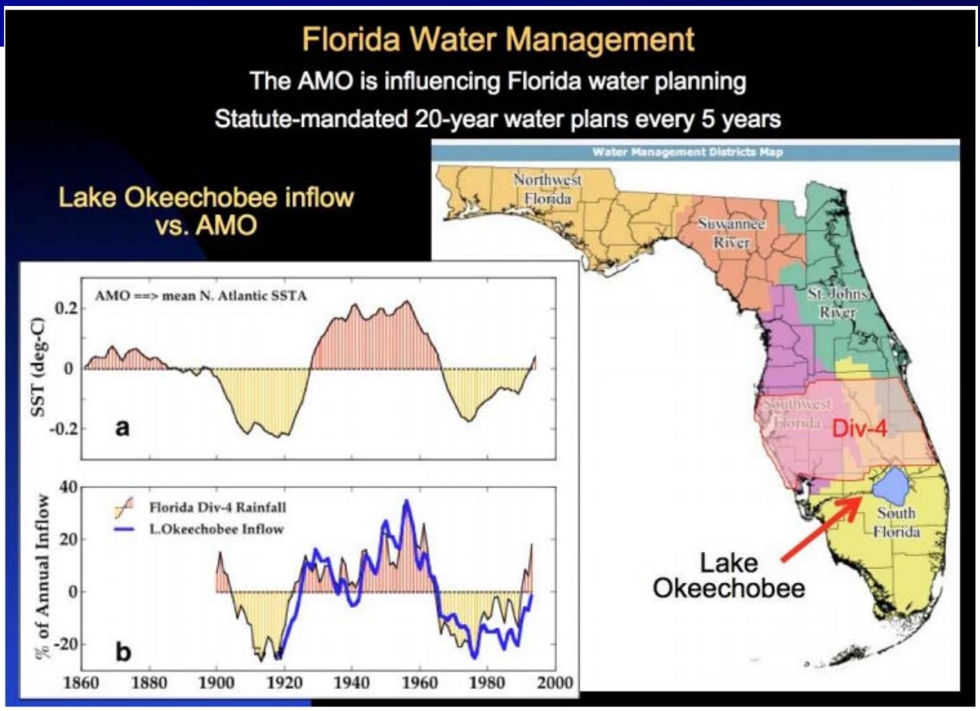


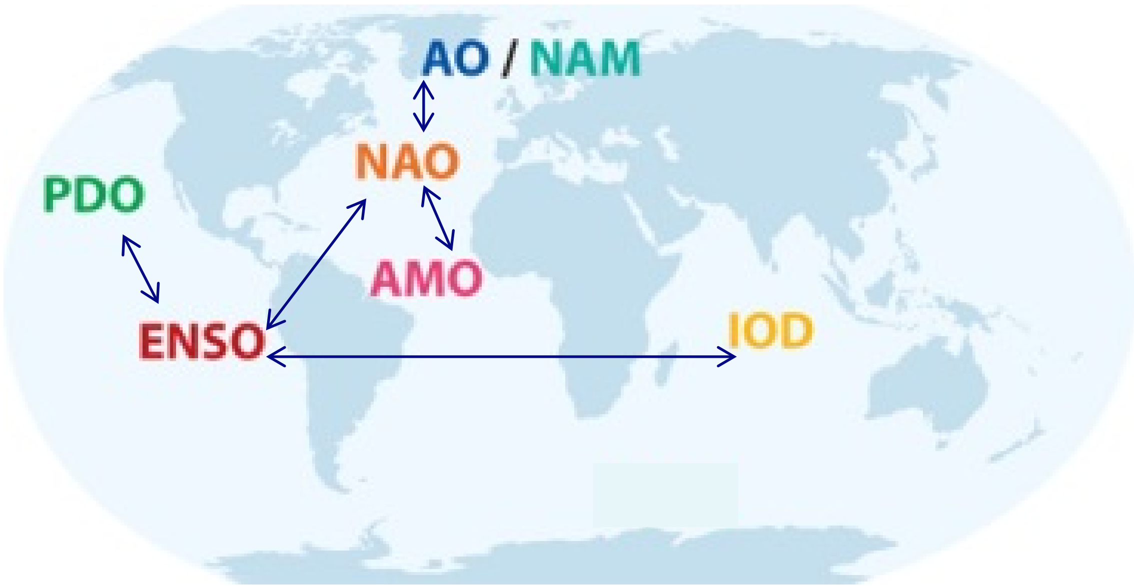
Decrease in water supply
Increase in temperature, drought, and wildfire,
Increase in frequency of flooding

Increase in air temperature
Evaporation and drought severity

Increase in air temperature
Hurricane intensity
Storm surge
Decrease in water availability

Known or Leading Teleconnection Patterns





Teleconnection Pattern	Abbrev.	Time scale
Arctic Oscillation	AO	decadal time scale
North Atlantic Oscillation	NAO	NAO can occur on a yearly basis, or the fluctuations can take place decades apart.
Atlantic Multi-decadal Oscillation	AMO	decadal time scale
Pacific Decadal Oscillation	PDO	decadal time scale
El Nino – Southern Oscillation	ENSO	El Nino and La Nina episodes typically occur every 3-5 years. However, in the historical record this interval has varied from 2 to 7 years.
Indian Ocean dipole	IOD	Every 30-year period

The Most Well-known Teleconnection Pattern

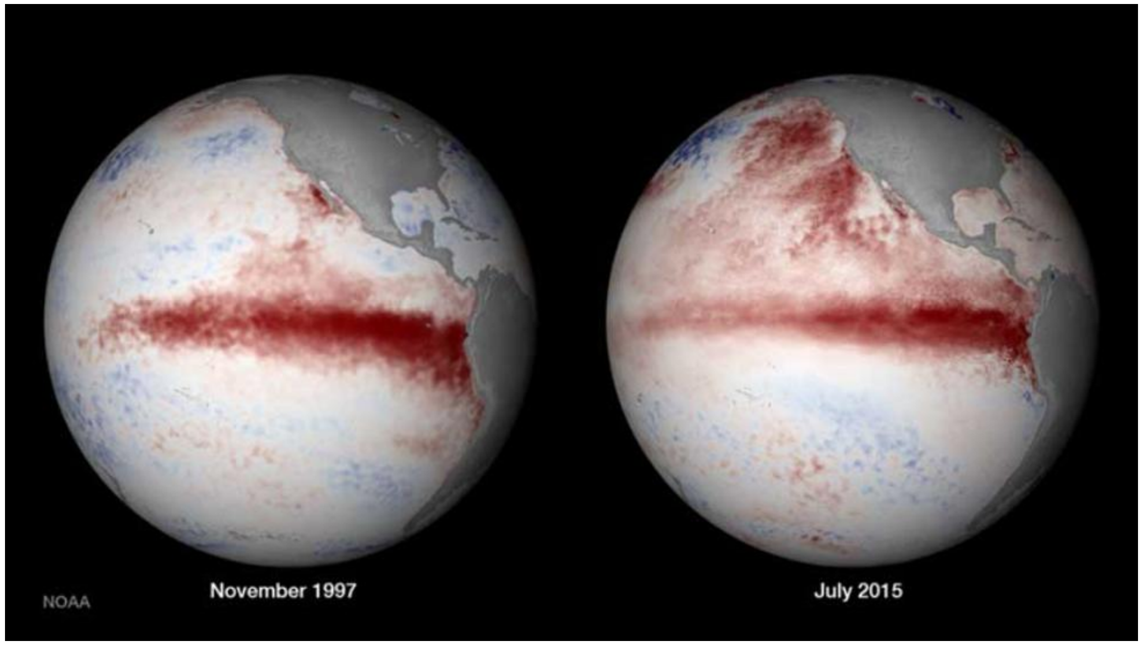
EL- NINO

- Strong years:

Winter 1982/1983

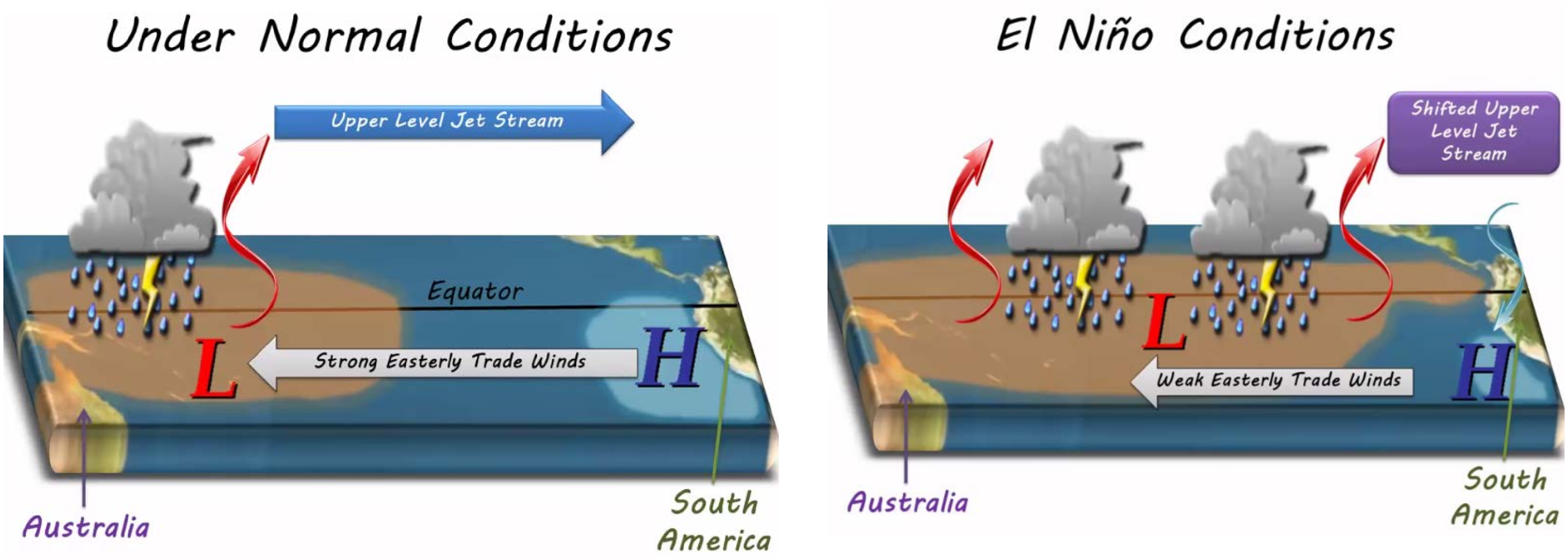
Winter 1997/1998

- Widescale climate effects
- Increased tropical cyclone activity in Pacific Ocean
- Affects tropical cyclone formation in the Atlantic
- Increasing cooling and precipitation during winter months in southern U.S.
- Affects Florida during the winter months



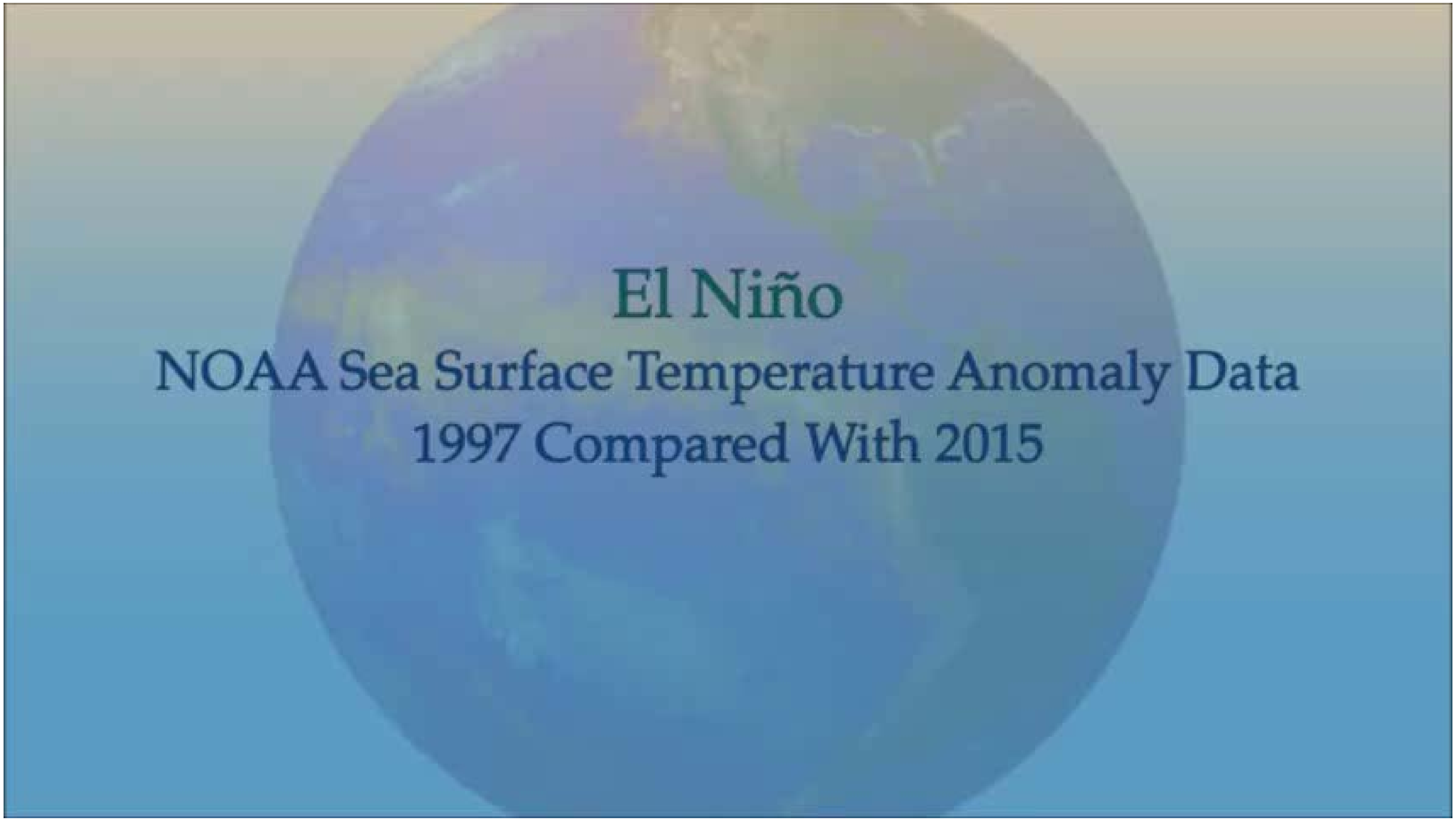
Comparison of sea surface temperature during El Nino 1997 (left) and current El Nino (right) images from NOAA satellites

The El- Nino

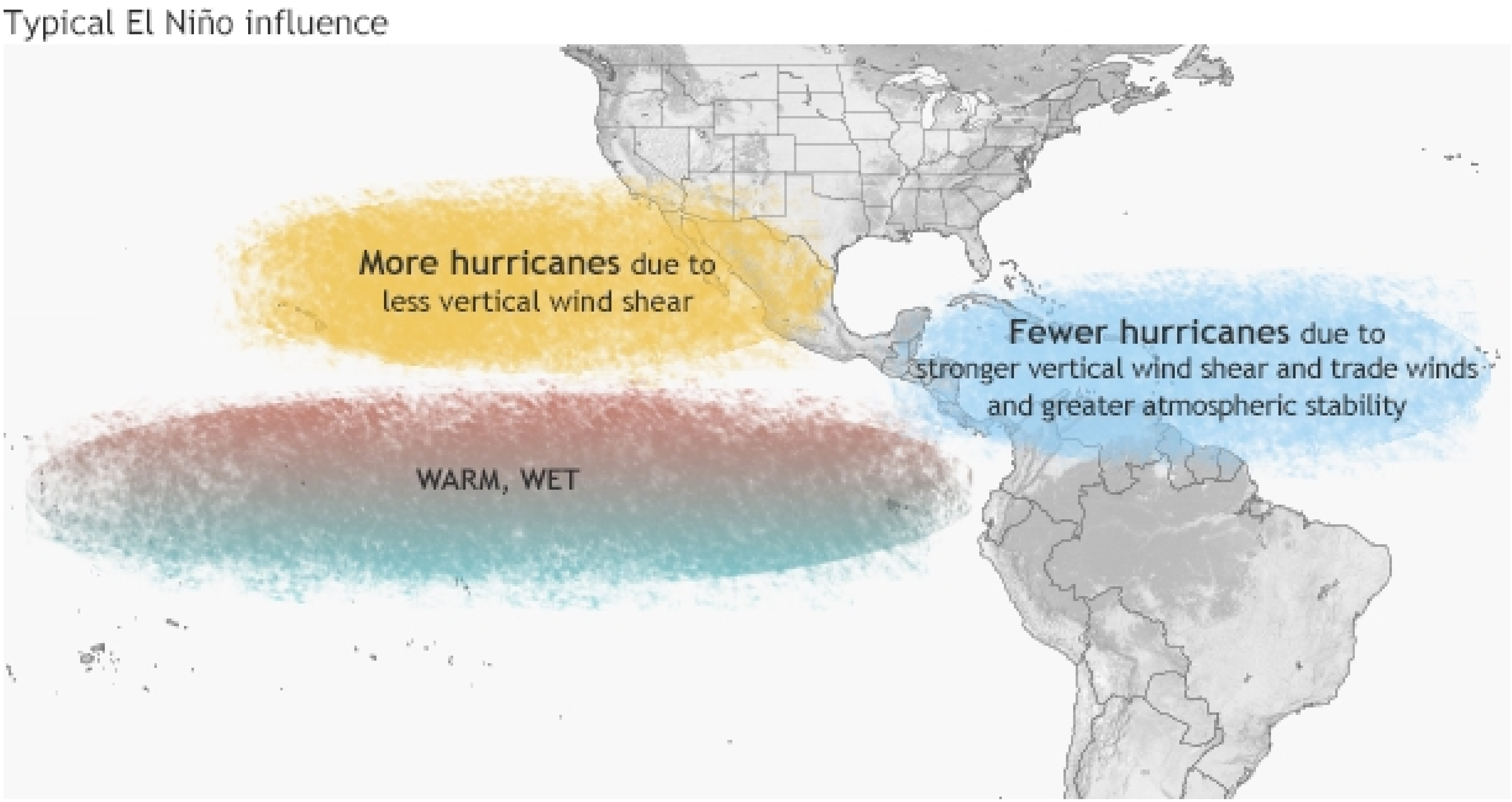


Source: National Weather Service Bismarck

The El- Nino


El Niño
NOAA Sea Surface Temperature Anomaly Data
1997 Compared With 2015

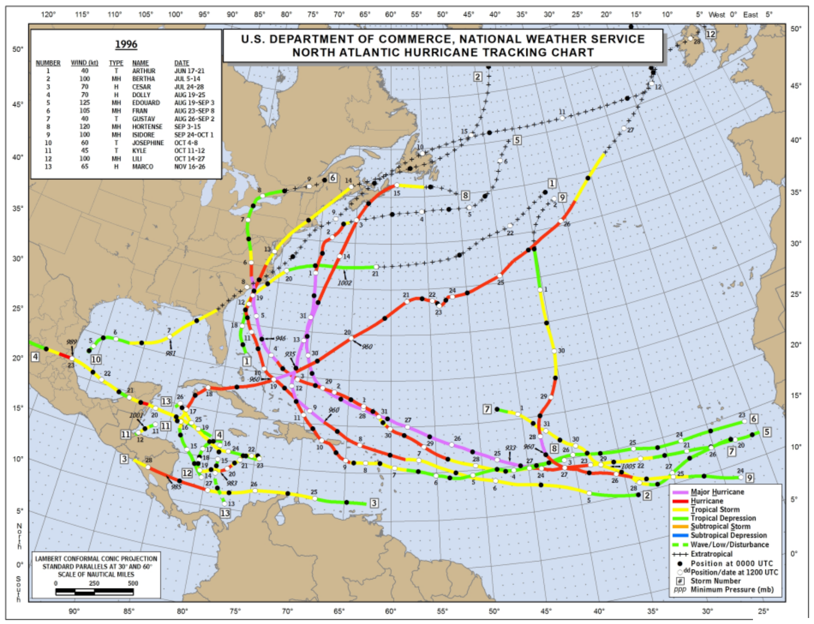
El Ninos & Atlantic Tropical Cyclones



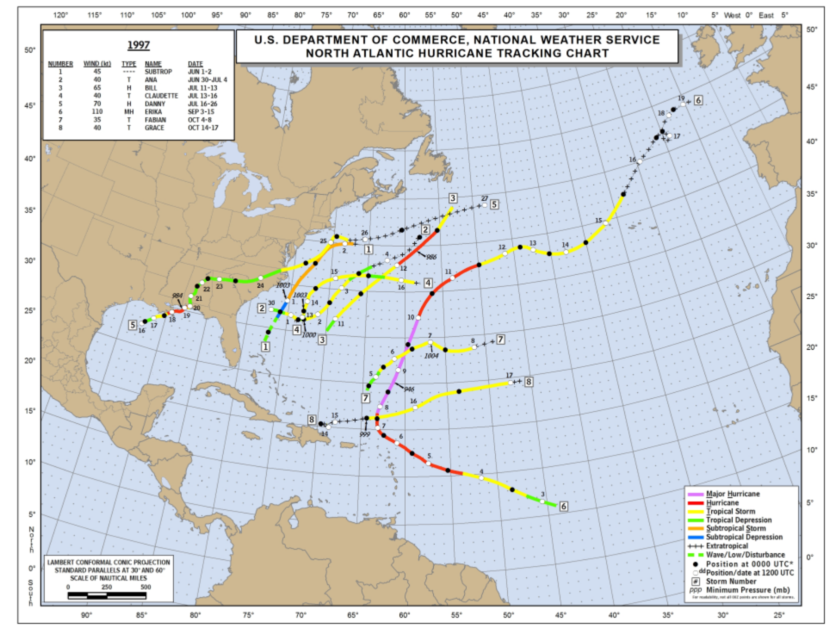
Source: Climate.gov (NOAA)

El Ninos & Atlantic Tropical Cyclones

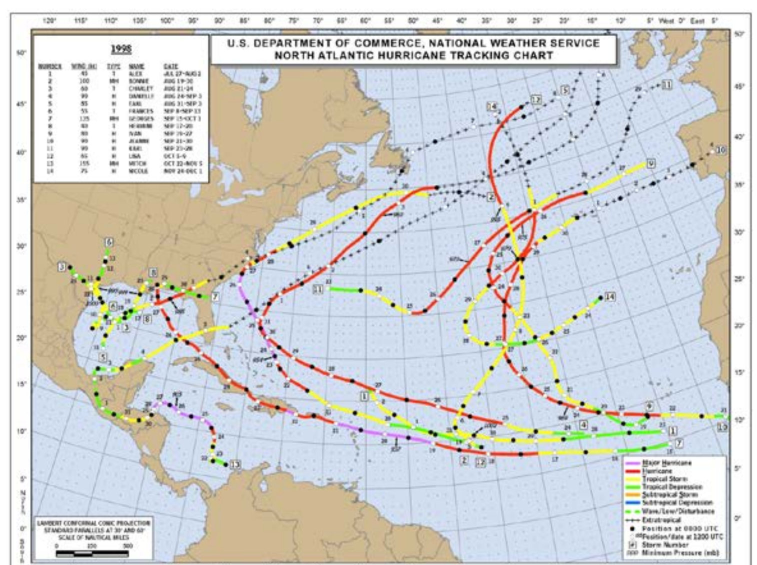
El Nino & Atlantic Tropical Cyclones



1996-Non El Nino year



1997-El Nino year



1998-Non El Nino year(post-1997 El Nino)

General Circulation Models

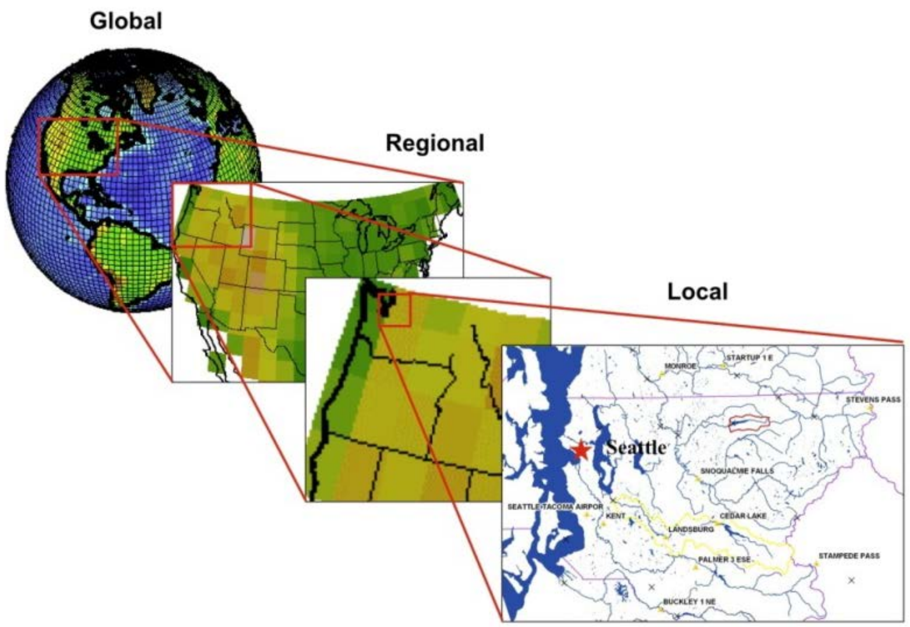
Difficulties with GCMs

GCMS highly successful at understanding large-scale climate processes, but are limited by:

(1) The inability to find a direct relationship between **local terrestrial responses and global atmospheric circulation**

(2) Spatial and temporal resolutions of GCMs are **too coarse to be applied in a regional scale**

Downscaling Methods



https://depts.washington.edu/nwst/publish/stories/2008-fall/2008-fall-env-2-downscaling.jpg

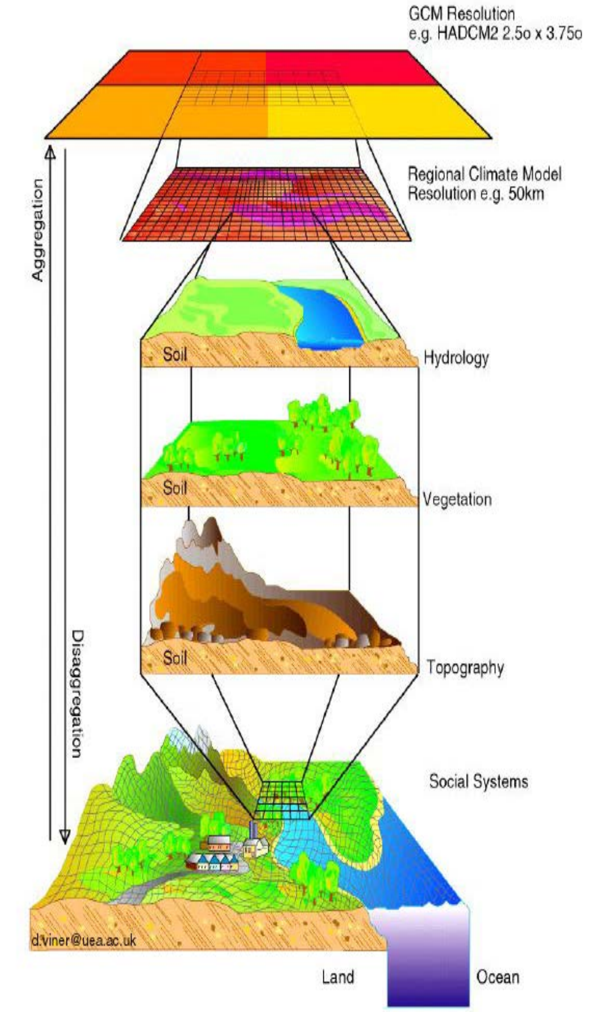
University of Central Florida

General Circulation Models

Difficulties with Downscaling Methods

- Often not able to address the hydrological **variability and extremes well**
- Dynamical downscaling methods can only run in **a short period of time** and are limited to address **single or a few GCM outputs**

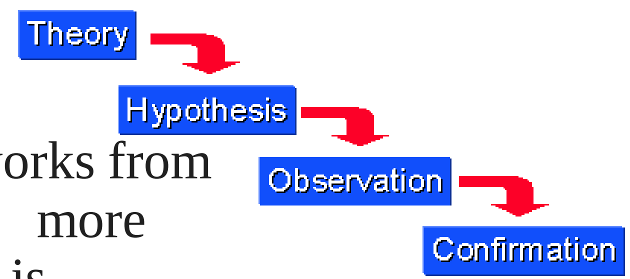
Provide initial information for public decision makers and cannot be applicable for risk /reliability analysis



http://www.cccsn.ec.gc.ca/images/downscaling01.jpg

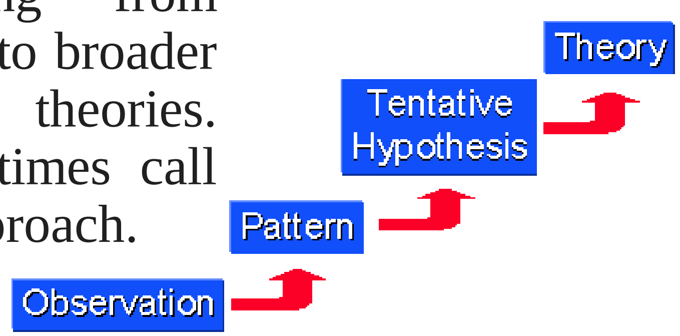
University of Central Florida

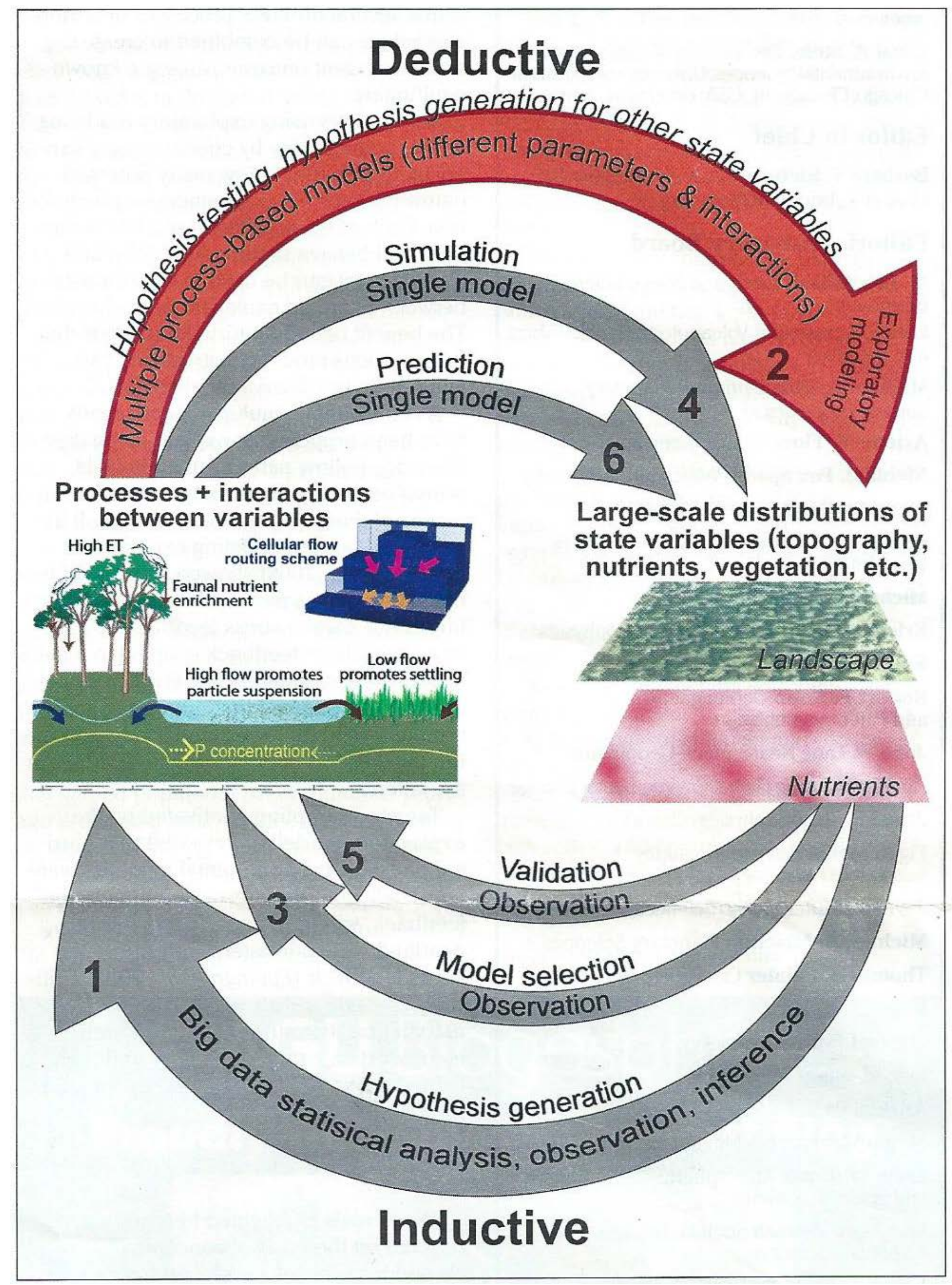
Extract Causality from Complexity



Deductive reasoning works from the more general to the more specific. Sometimes this is informally called a "top-down" approach.

Inductive reasoning works the other way, moving from specific observations to broader generalizations and theories. Informally, we sometimes call this a "bottom up" approach.





Courtesy of AGU

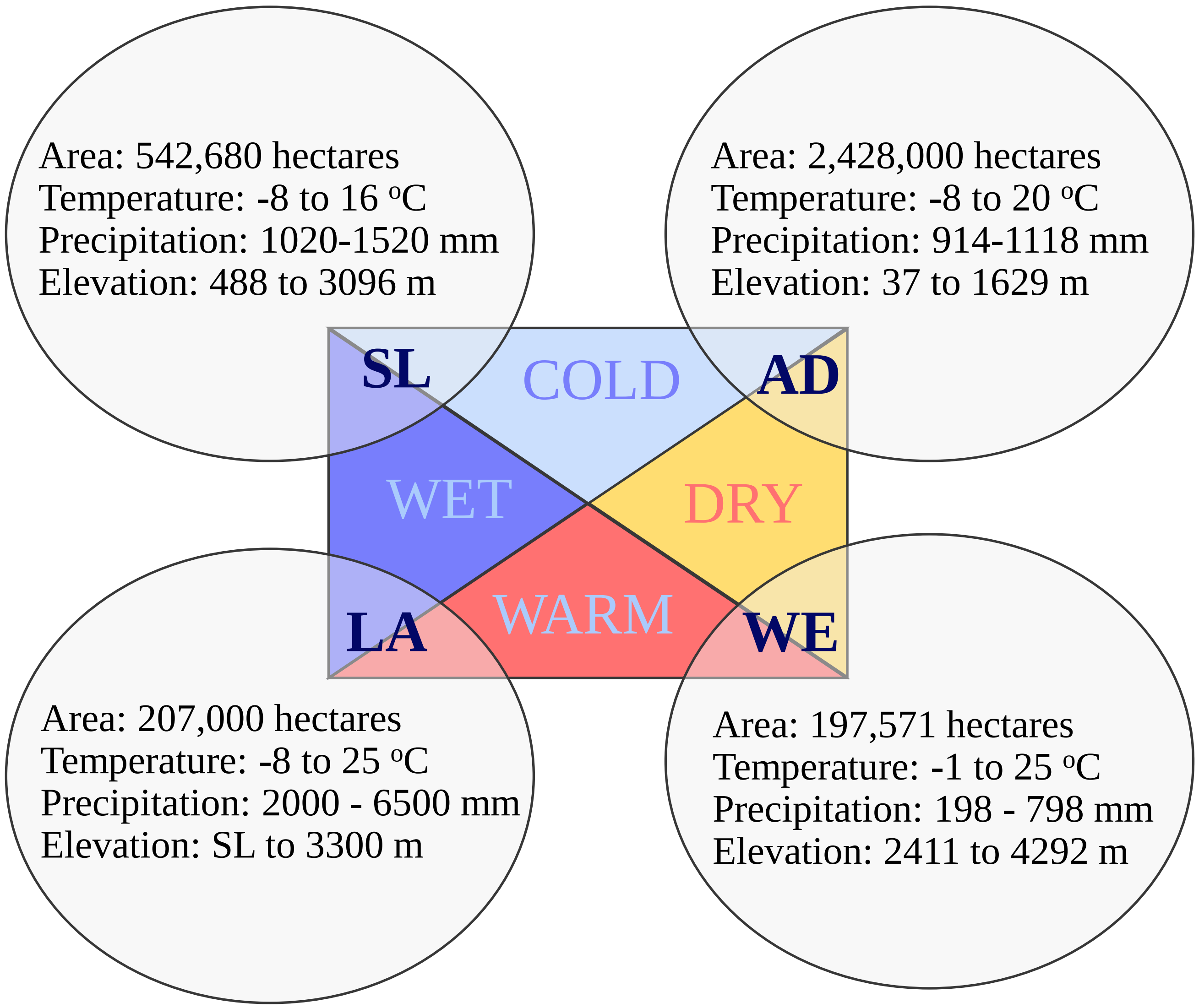
Science Questions

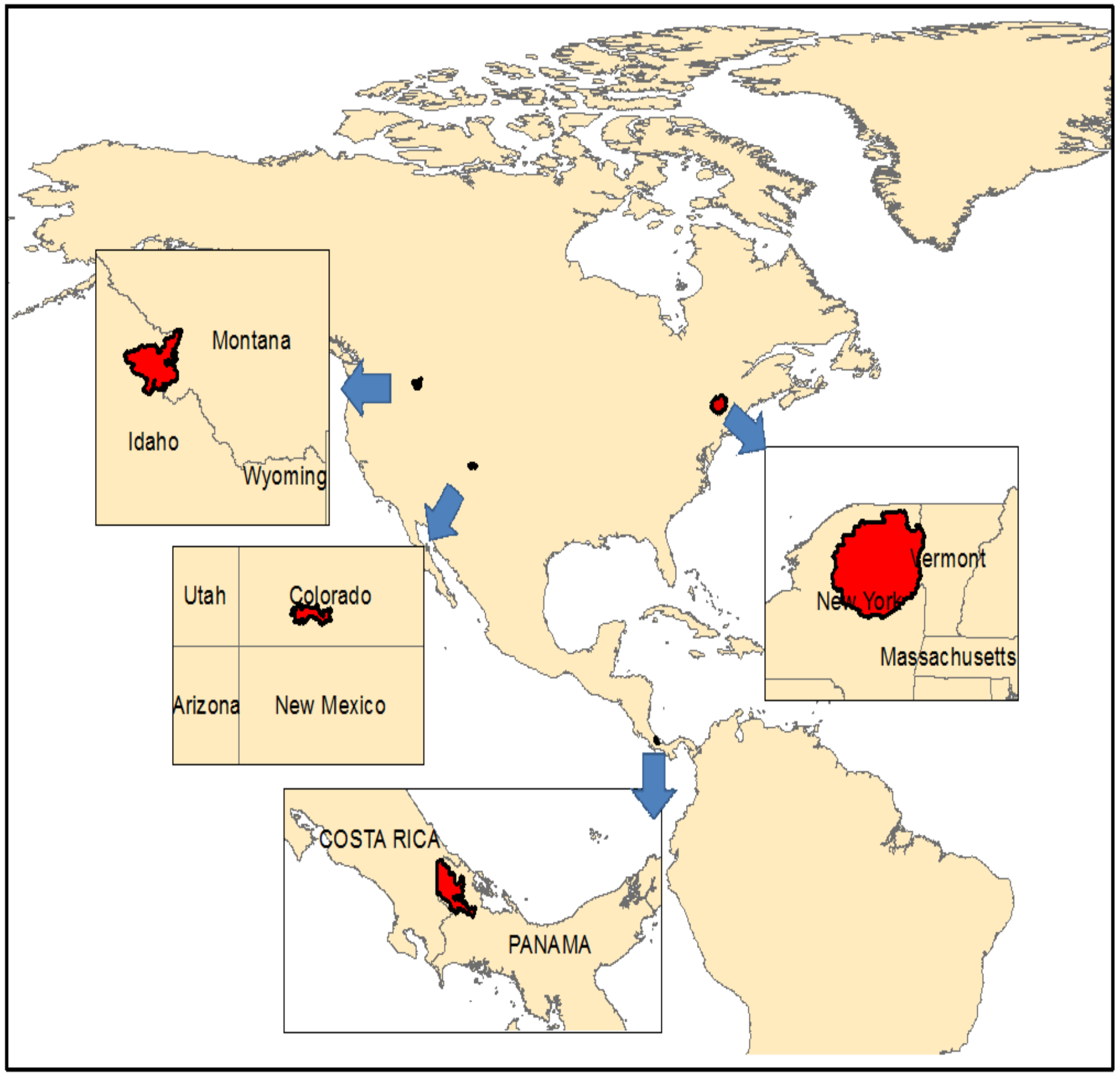
- **Objective**: Develop a hybrid inductive approach to supplement the GCM modeling framework with the aid of empirical mode decomposition, wavelet analysis, and extreme learning machine to quantify the possible impact from the leading and non-leading SST teleconnection signals on terrestrial precipitation.
- **Science Question 1:** Is there any non-leading teleconnection patterns affect terrestrial precipitation more than the known leading teleconnection patterns? if so, to what extent?
- **Science Question 2**: Is there any commonality among the selected four study sites, with different geographical context, with respect to their precipitation trends affected by global SST?

Study Sites

Sites		Adirondack State Park	Selway-Biterroot wilderness	La Amistad International Park	Weminuche Wilderness
Features	Elevation (m)	37-1,629	488-3,096	3,300	2,400 - 4,000
	Area (ha)	2,428,000	542,680	207,000	197,600
	Mean Annual Precipitation (mm)	914-1,118	1,020-1,520	2,000-6,500	198 - 798
	Temperature(ºc)	-8 to 20	-8 to 16	-8 to 25	-10.5 to 15
	Location	Northern New York	Border of Idaho and Montana State	Panama	Colorado
	Establishment date	1892	1964	1982	1975

Study Sites





*AD stands for Adirondack international Park; WE stands for Weminuche Wilderness; SL stands for Selway Bitterroot; LA stands for La Amistad International Park;

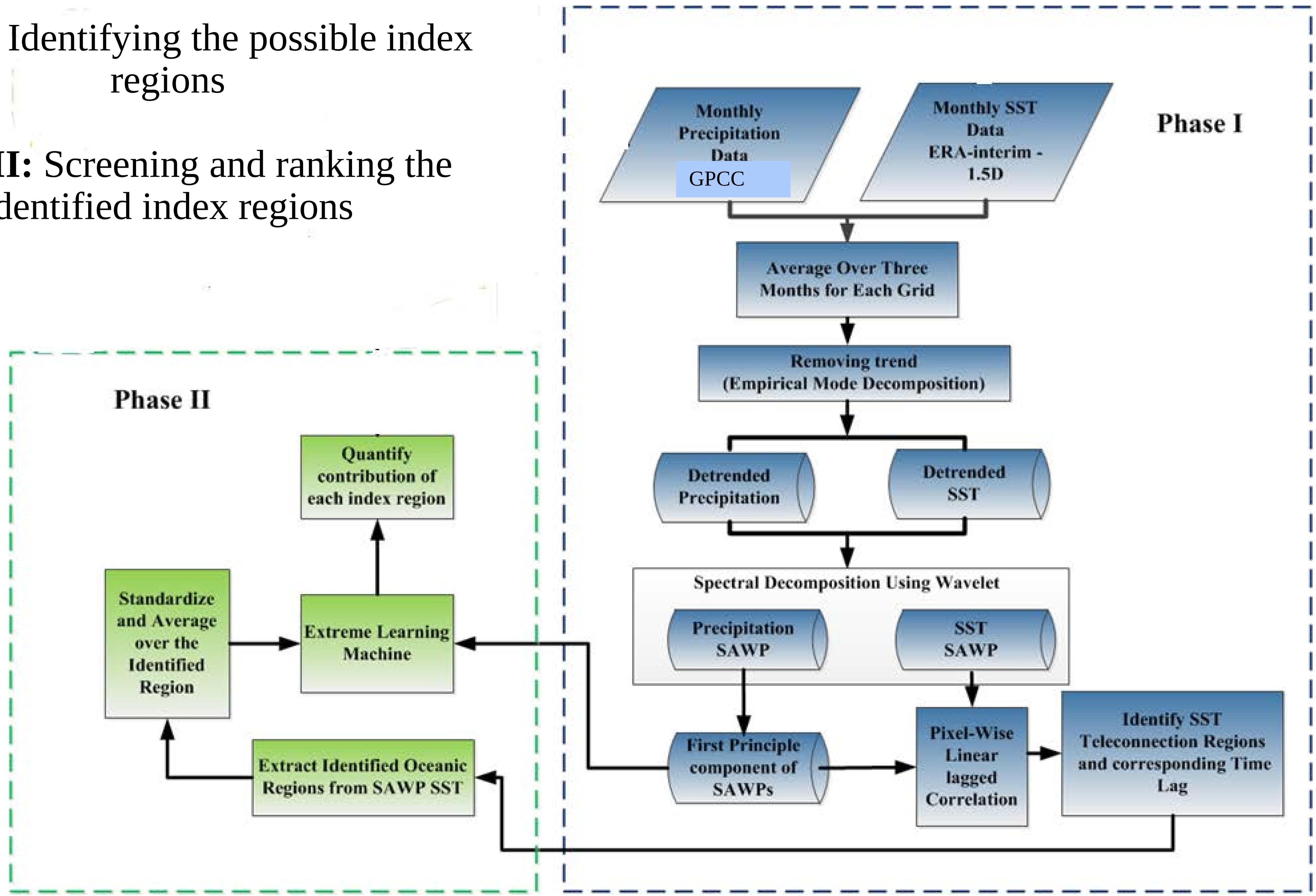
NE: Adirondack Park
NW: Selway-Bitterroot Wilderness
SW: Weminuche Wilderness
Central America: La Amistad International Park

University of Central Florida

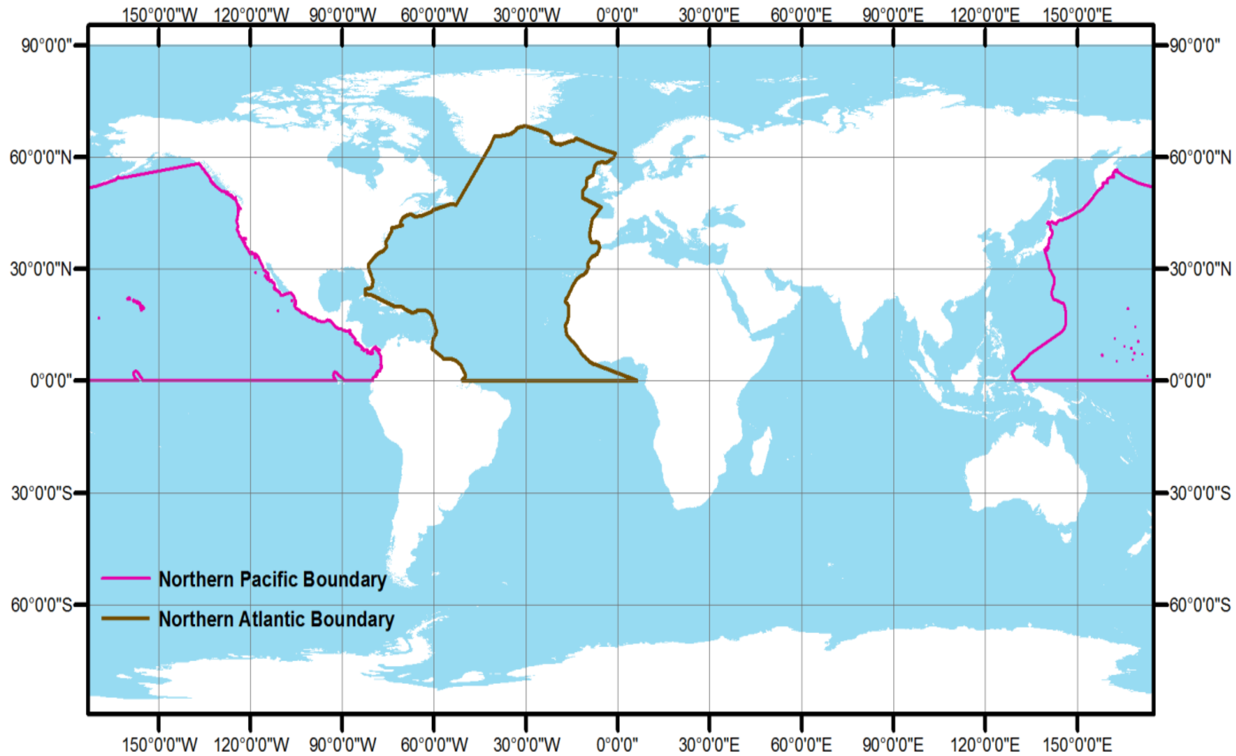
Research Framework

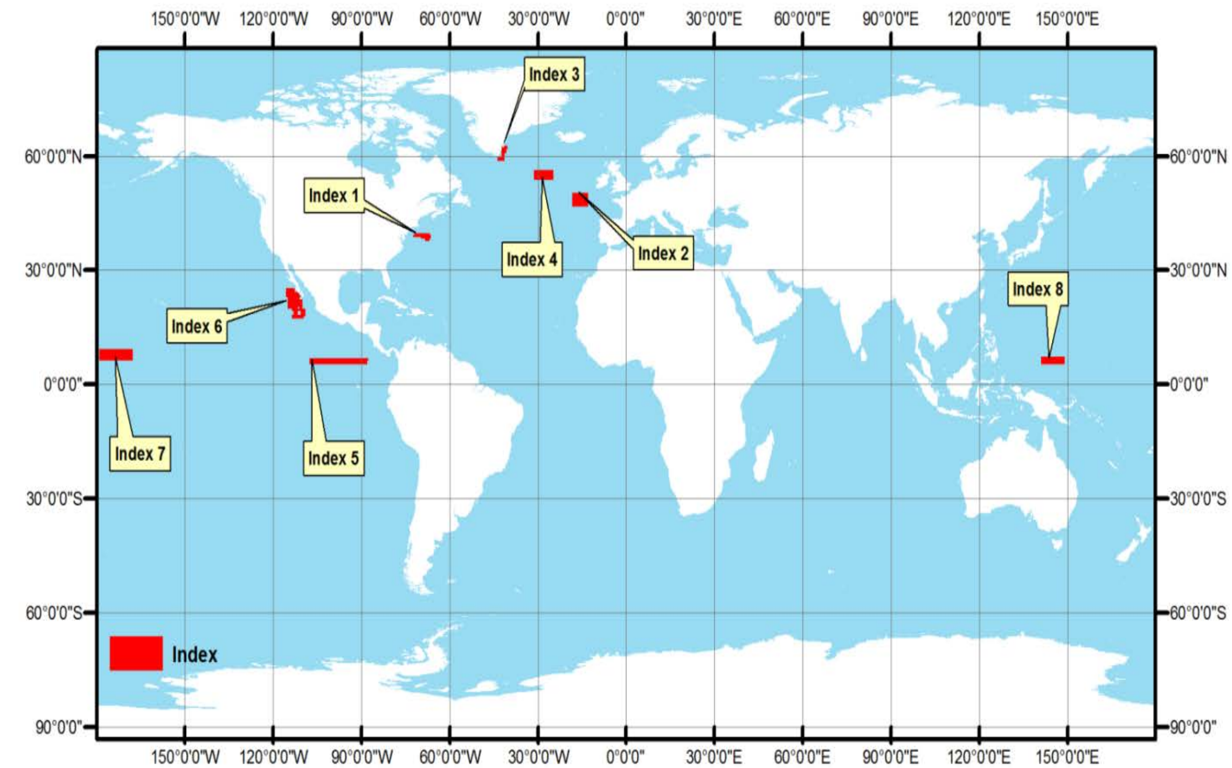
Phase I: Identifying the possible index regions

Phase II: Screening and ranking the identified index regions



Examples of Significant Oceanic Indices





Data Sources

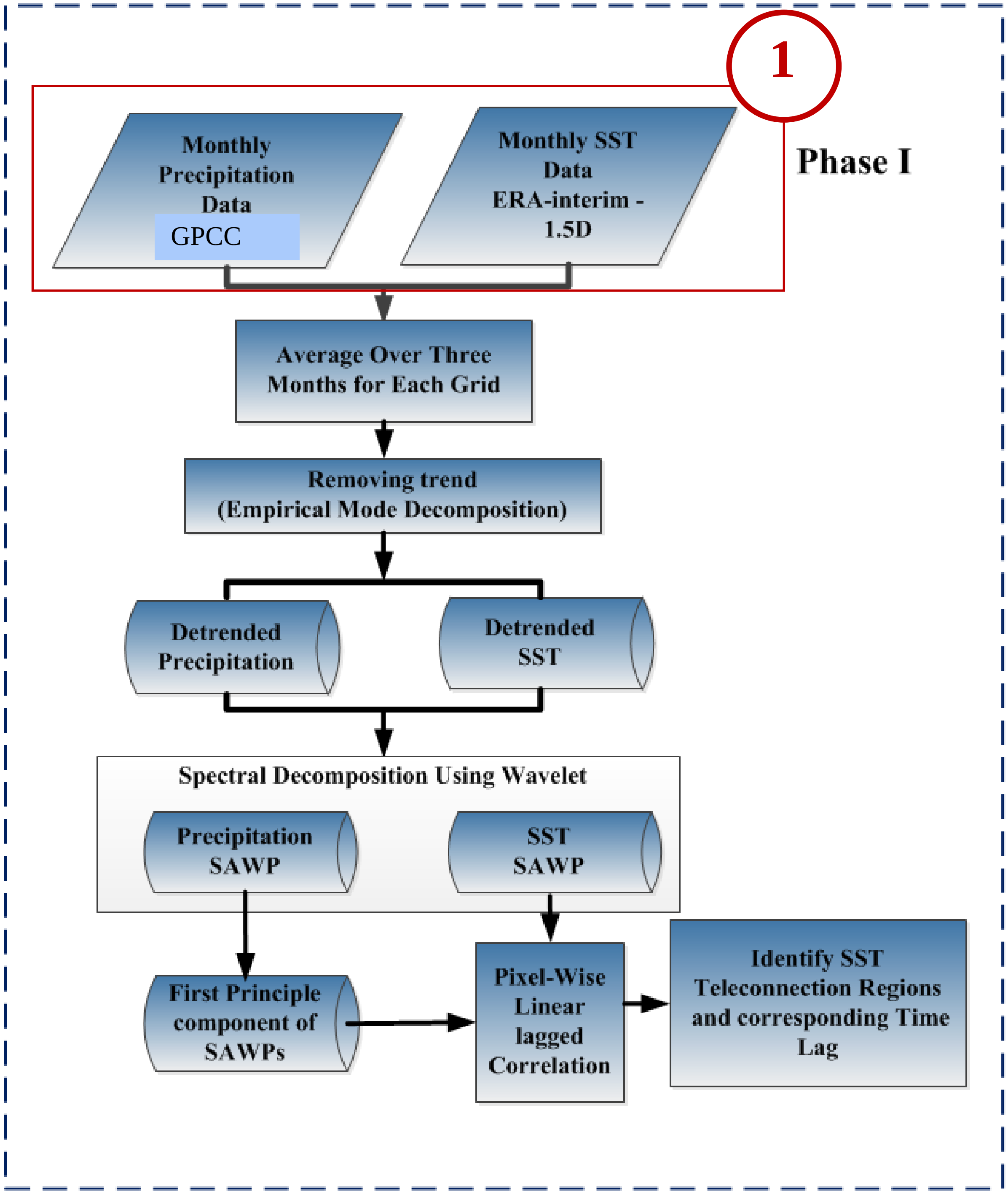
Data Sources

Precipitation Data

- ❑ Full data product of Global Precipitation Climatology Center (GPCC-V6)
- ❑ Gridded gauge-analysis products based on quality controlled data from 67,200 stations world-wide for the period of 1901-2010.
- ❑ Spatial resolution: 0.5º ×0.5º

SST Data

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



Data Preprocessing

Seasonal Data

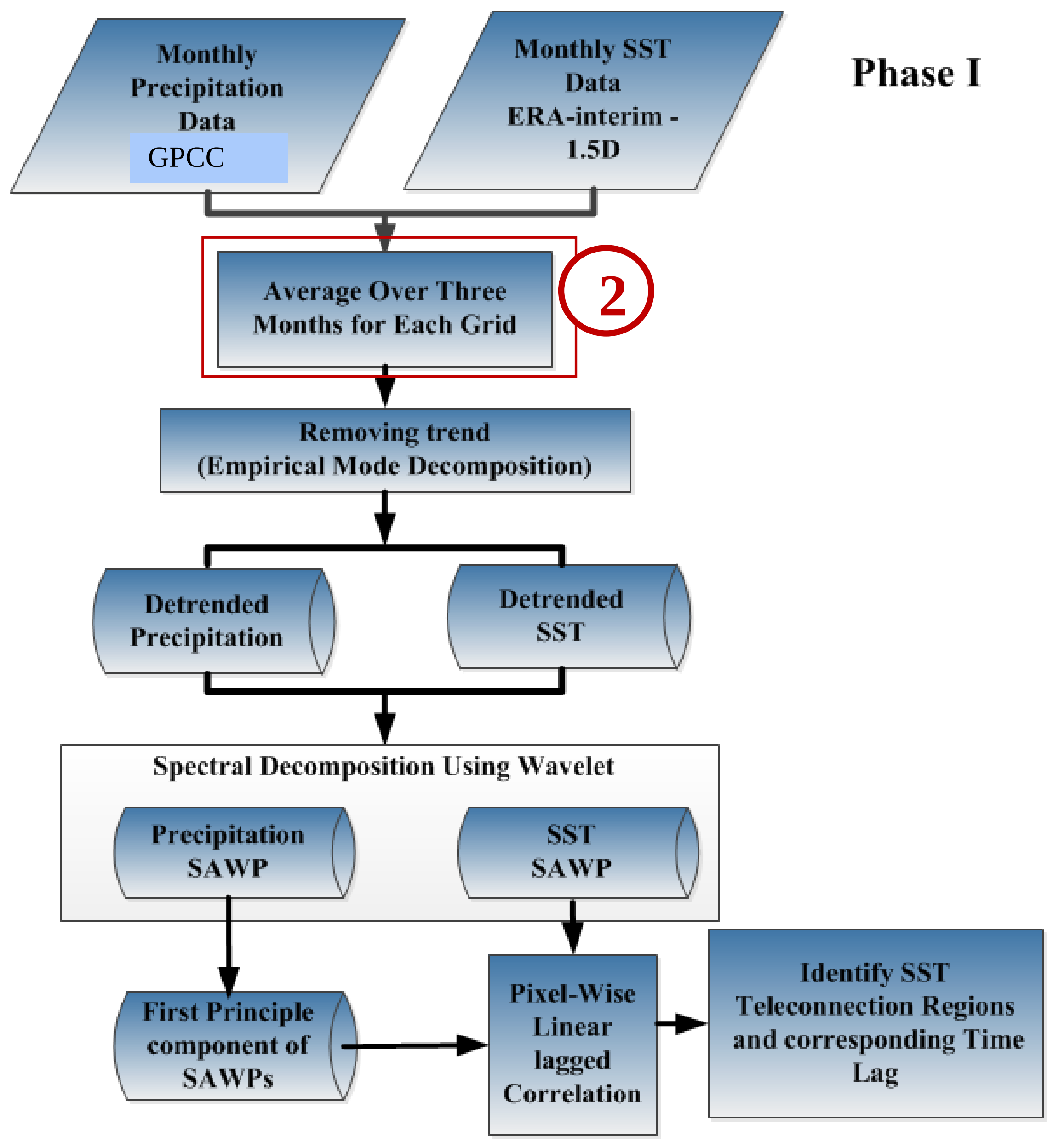
As a result of precipitation regimes, and the significant seasonality revealed between terrestrial precipitation variability and SST forcing, seasonal scale was selected in this study.

Precipitation

- ❑ computing every 3 months averages, namely MAM, JJA, SON, DJF.

SST

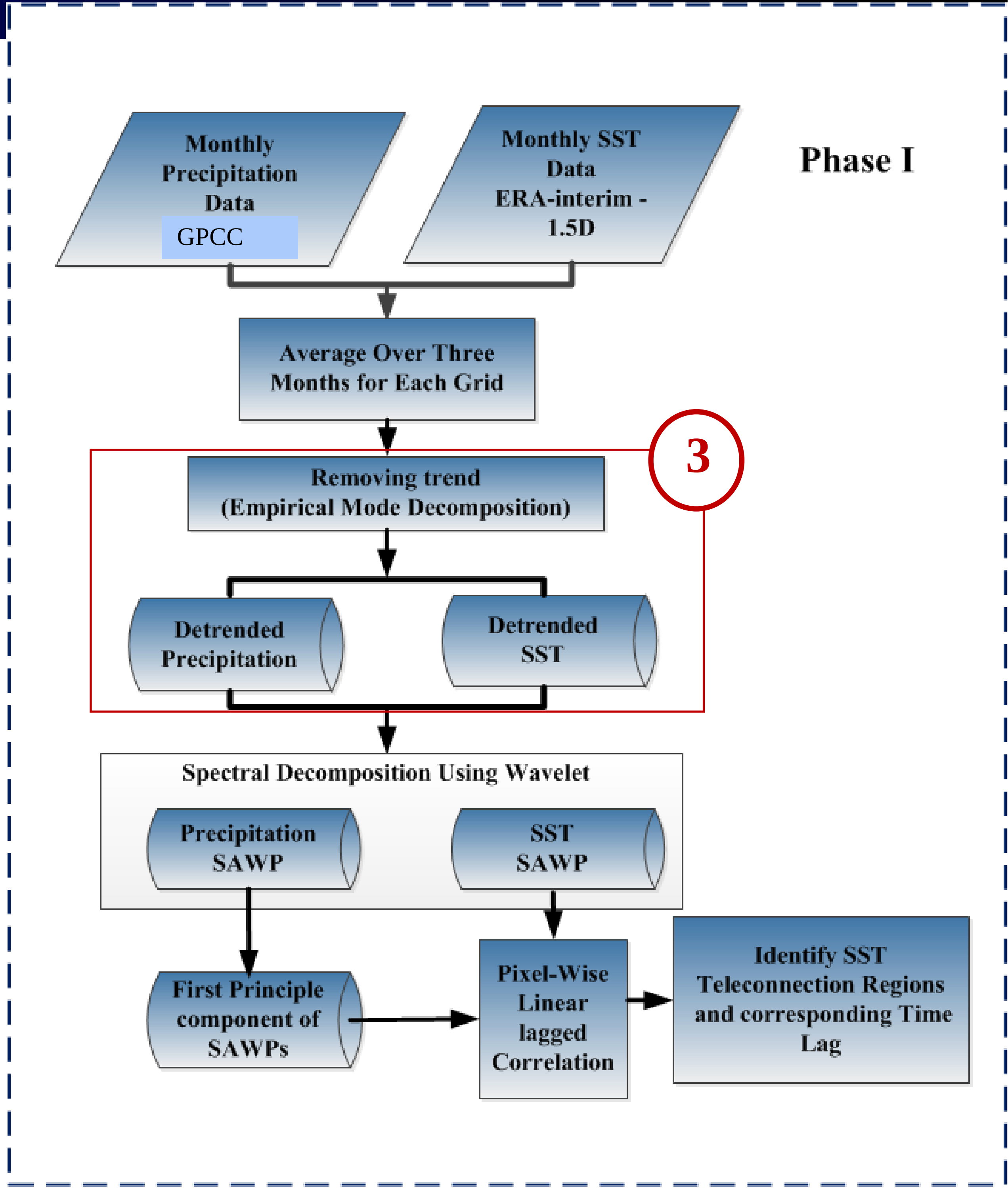
- ❑ 13 different SST time series are computed with time lags from 0 to 12 months for each season.



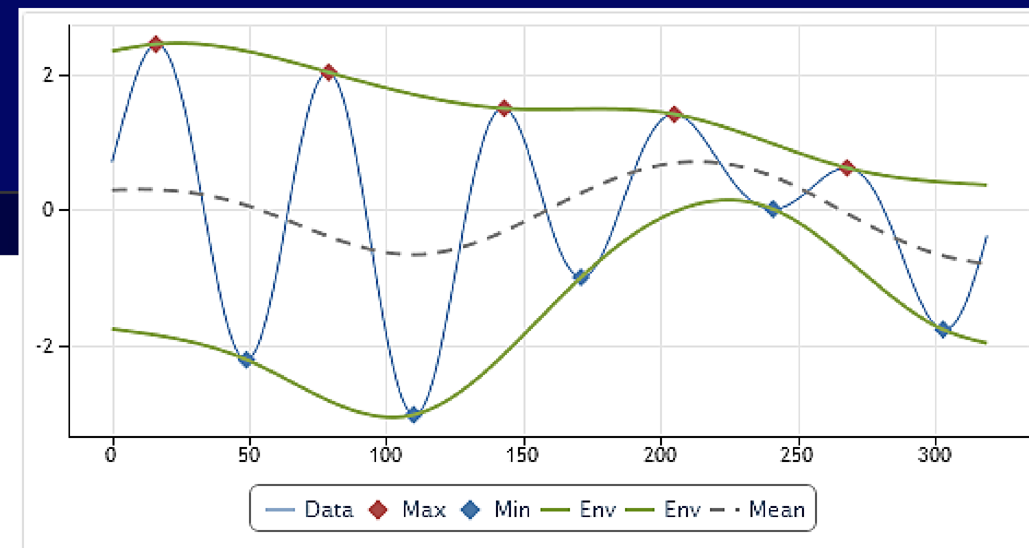
Data Preprocessing

Empirical Mode Decomposition

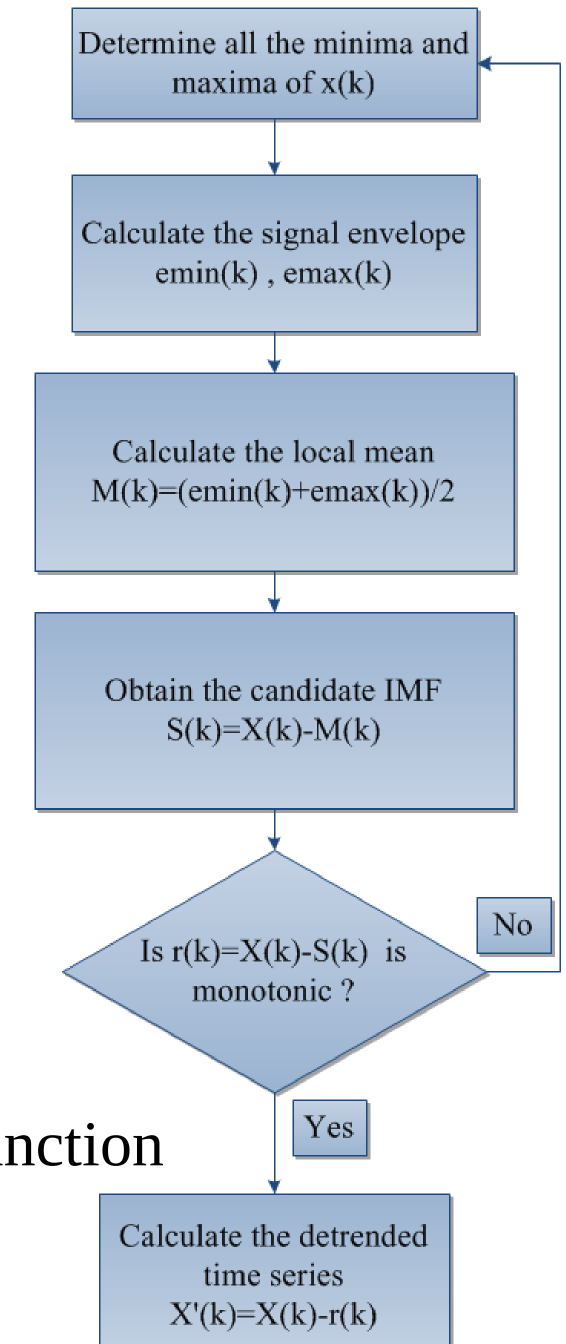
- Long-term trend causes large uncertainty in linear correlation analyses
- EMD approach was first suggested by Huang et al. (1998), and it has the ability of extracting the intrinsic and adaptive trends from non-linear and non-stationary time series.

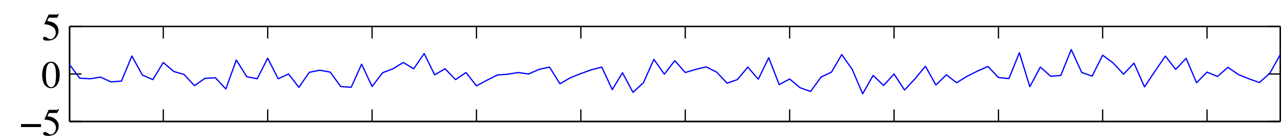


Data Preprocessing

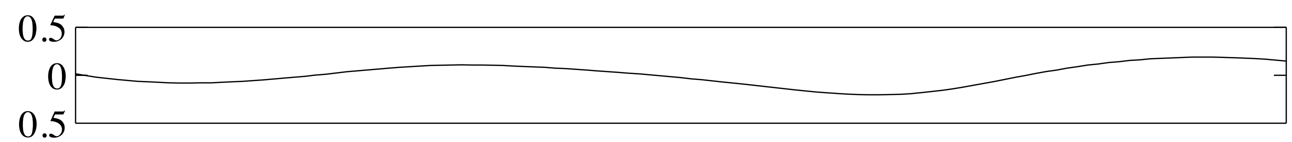


Detrending Algorithm





IMF:
Intrinsic Mode Function



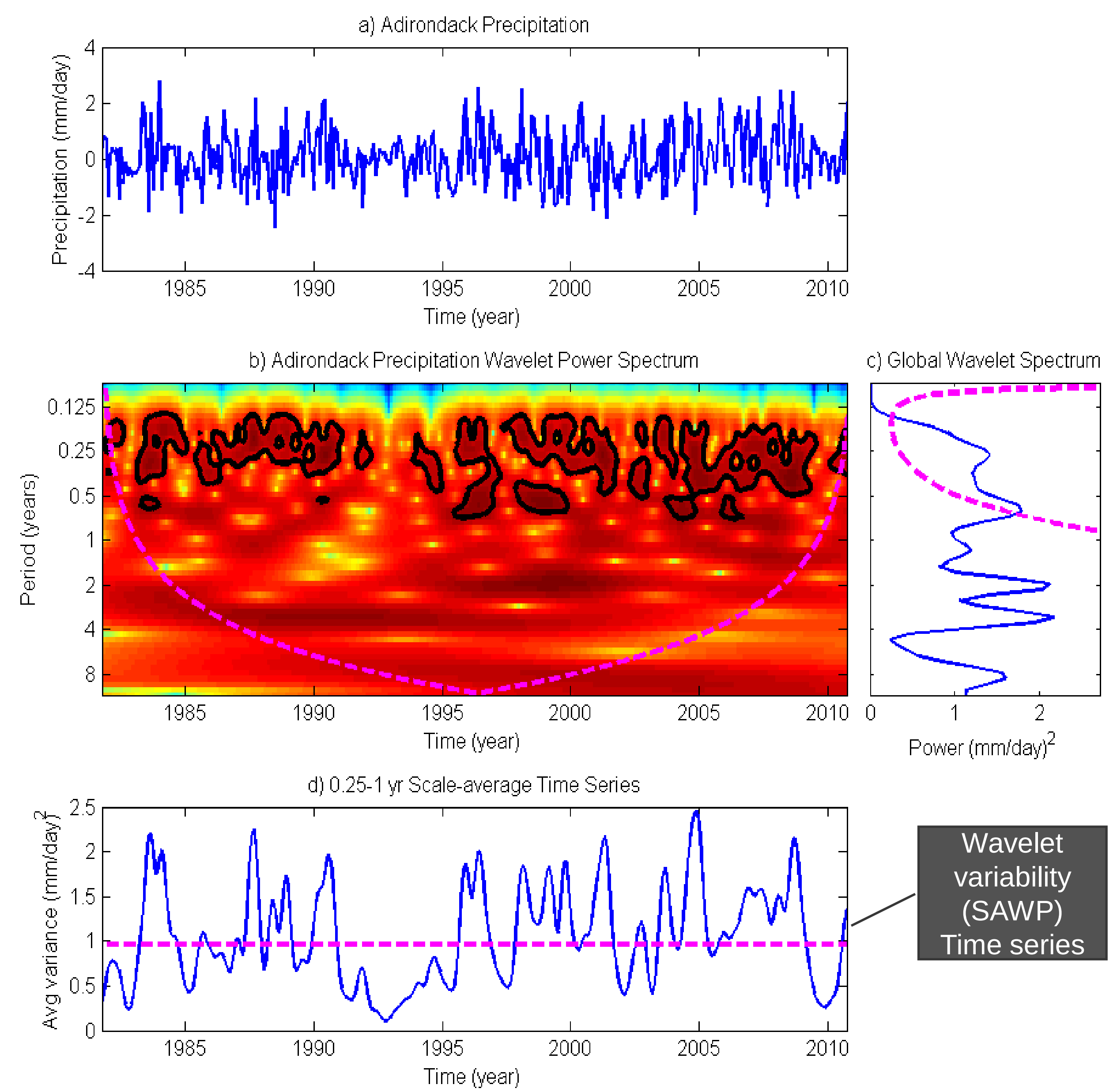
Wavelet Analysis

Continuous Wavelet Transform

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^*[(n'-n)\delta t/s]$$

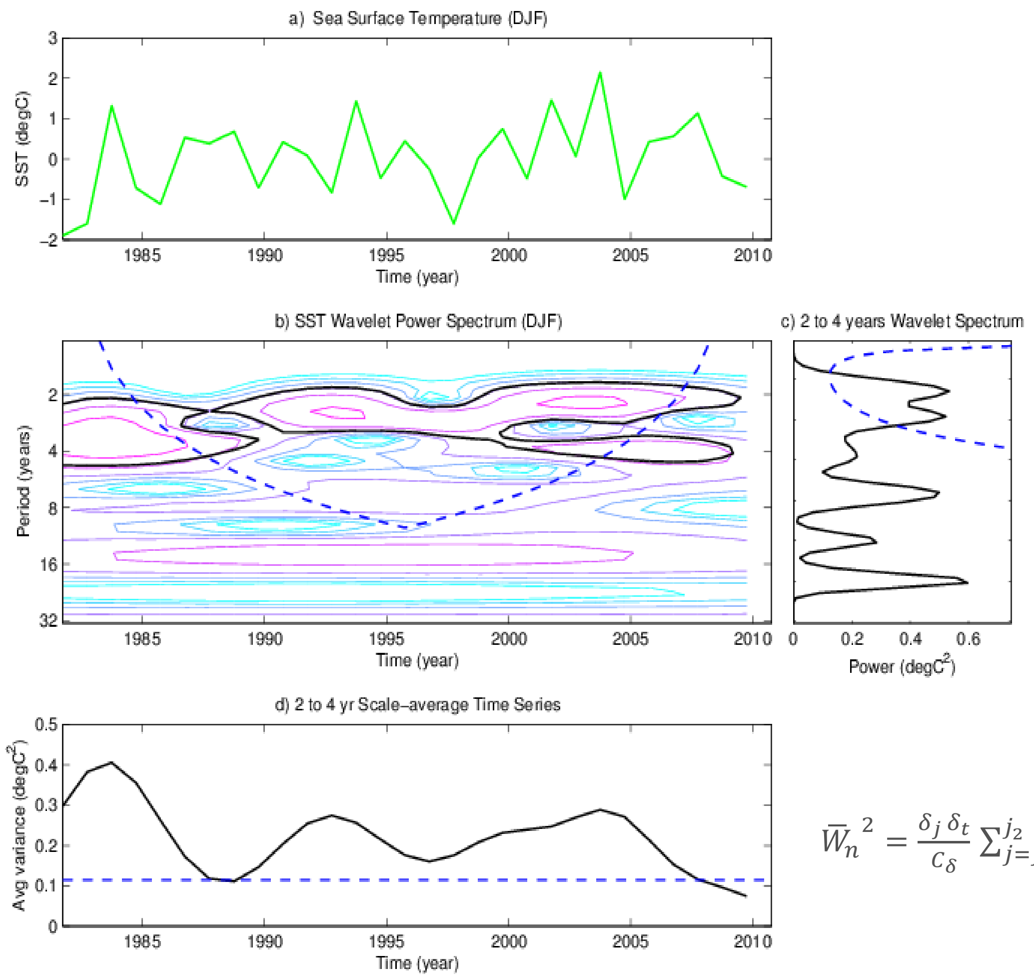
Where:
s = wavelet scale
x_n = discrete time series sequence
n = localized time index
n' = translated time index
ψ = normalized wavelet
* = complex conjugate

1. Anomalous Time Series Graph
2. Wavelet Power Spectrum Image
3. Global Significance Wavelet Spectrum
4. Scale Average Wavelet Power
   a) Weighted Sum of Wavelet Power Over Defined Scales
   b) Converts the original time series to a variance plot of the 0.25-1 year Frequency Band.
   c) Used to capture nonlinear, spectral information.



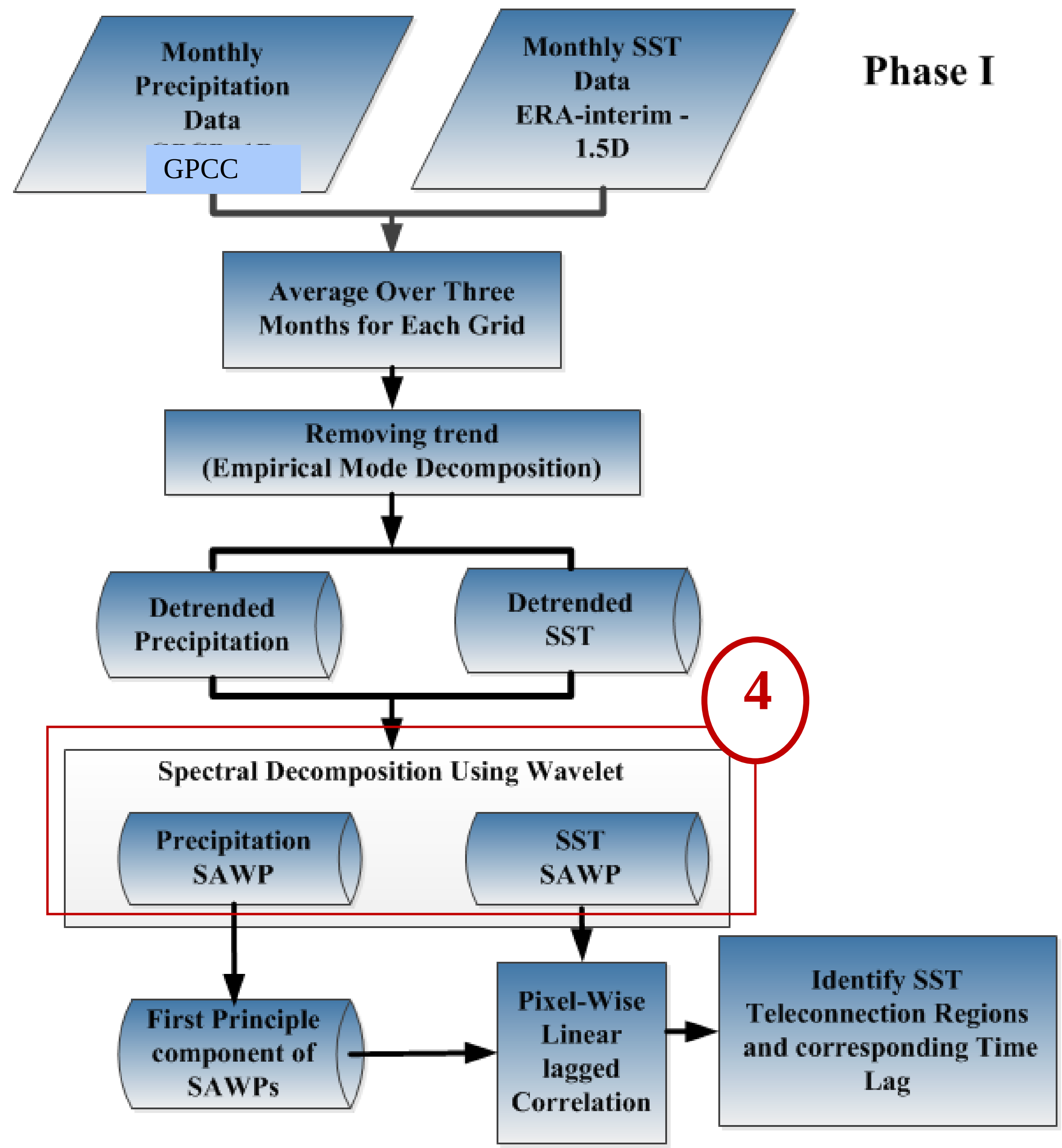
Wavelet Analysis

Dominant oscillation of SST and precipitation time series was detected in a certain band (2-4 years).



$$\bar{W}_n^{\,2} = \frac{\delta_j\,\delta_t}{C_\delta}\sum_{j=j_s}^{j_2}\frac{|W_n(s_j)|^2}{s_j}$$

The wavelet power spectrum is defined as $|W_n(S)|^2$ and the amplitude at each point $|W_n(S)|$. S is the scale. Cδ is the reconstruction factor that takes on values depending on the mother wavelet used, δj is a factor for scale averaging, j1 and j2 are scales over which the averaging takes place, and δt is the sampling period

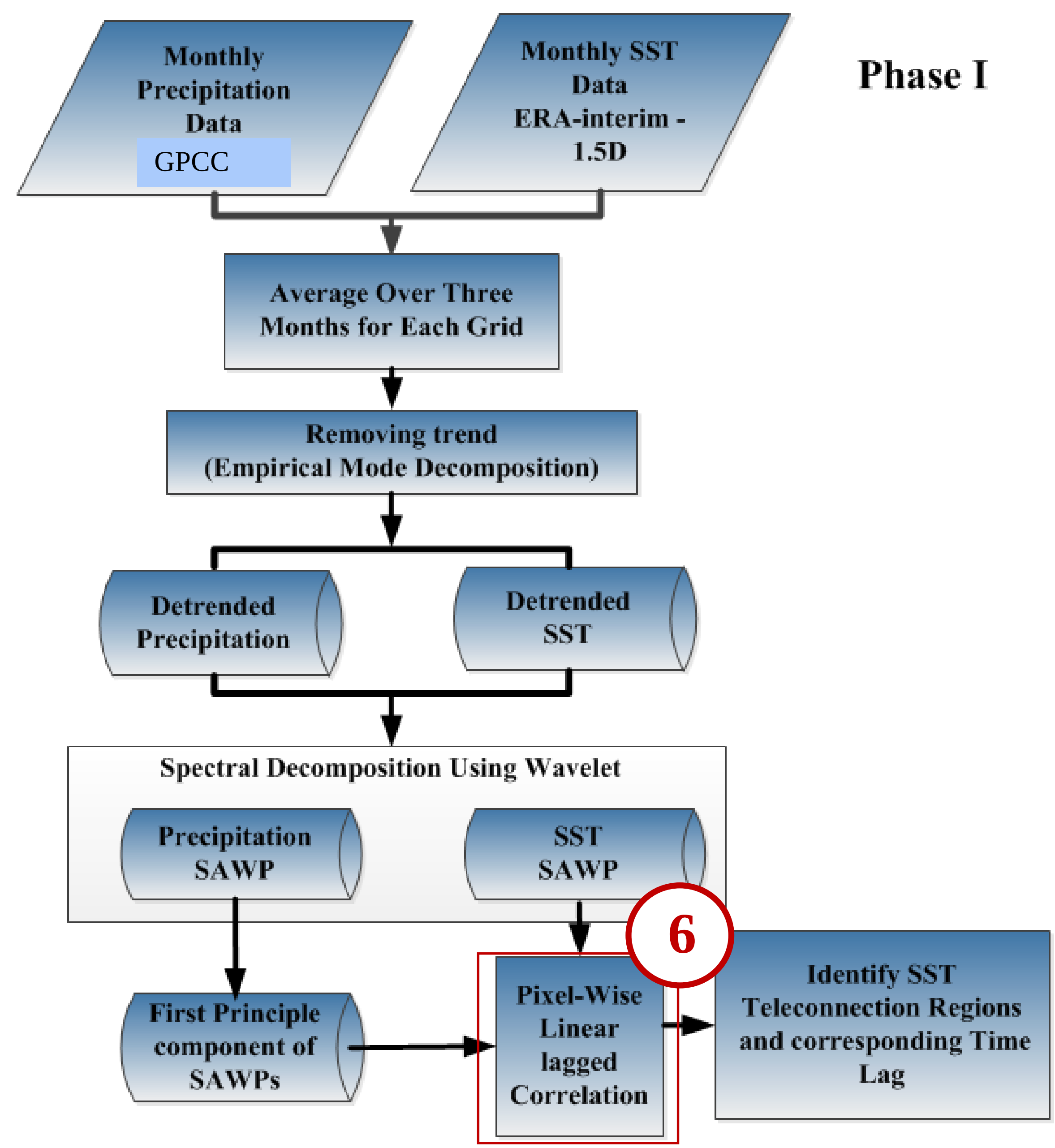


Identification of Index Regions

Pixel-wise Linear Lagged Correlation

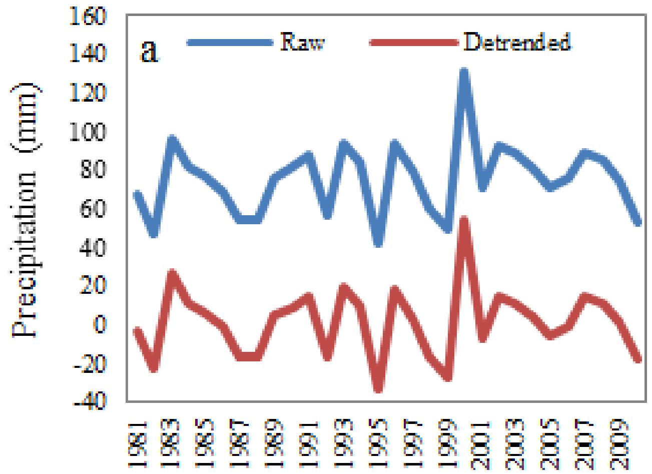
$$r = \frac{\sum_i^n[(x_i - \bar{x}) \times (y_{i+d} - \bar{y})]}{\sqrt{\sum_i^n[(x_i - \bar{x})]^2} \times \sqrt{\sum_i^n[(y_{i+d} - \bar{y})]^2}}$$

r :correlation coefficient,
n: number of data,
x : represents the precipitation dataset,
y: represents the SST dataset;
$\bar{x}$ and $\bar{y}$ are the mean;
d: time lag.

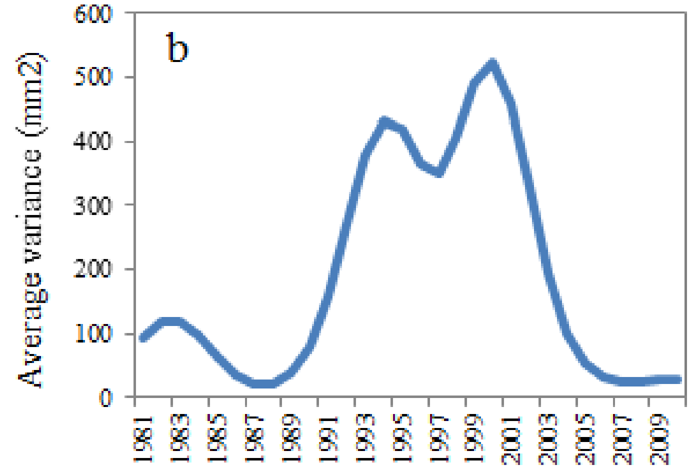


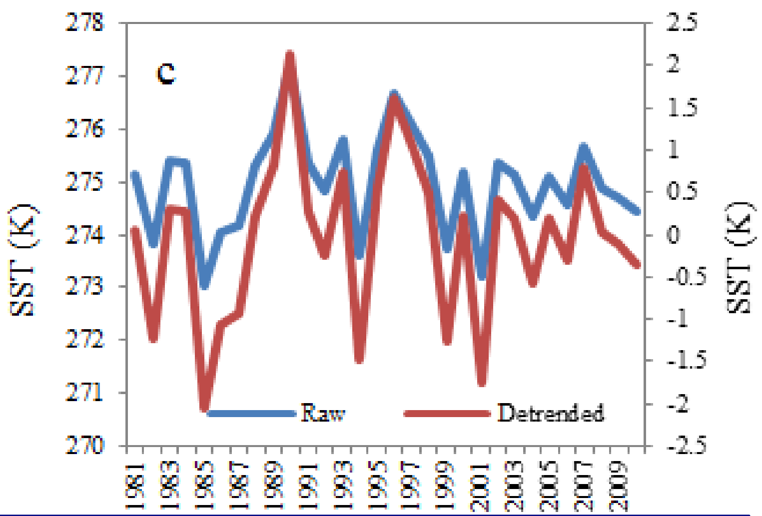
Identification of Index Regions

Raw and detrended Precipitation

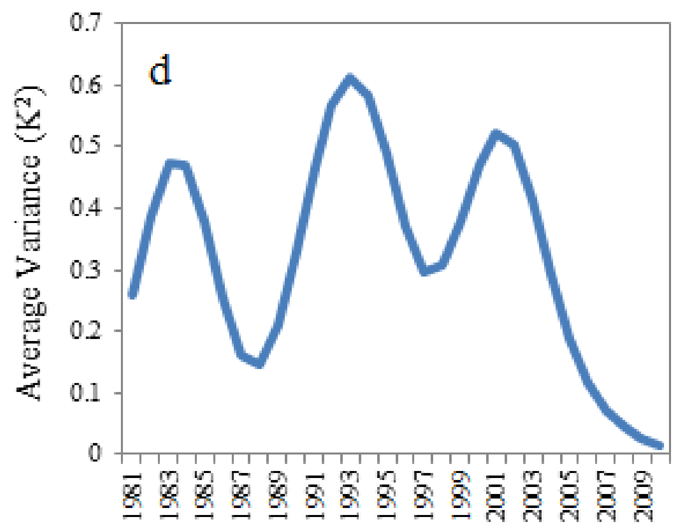


Precipitation SAWP





Raw and detrended SST



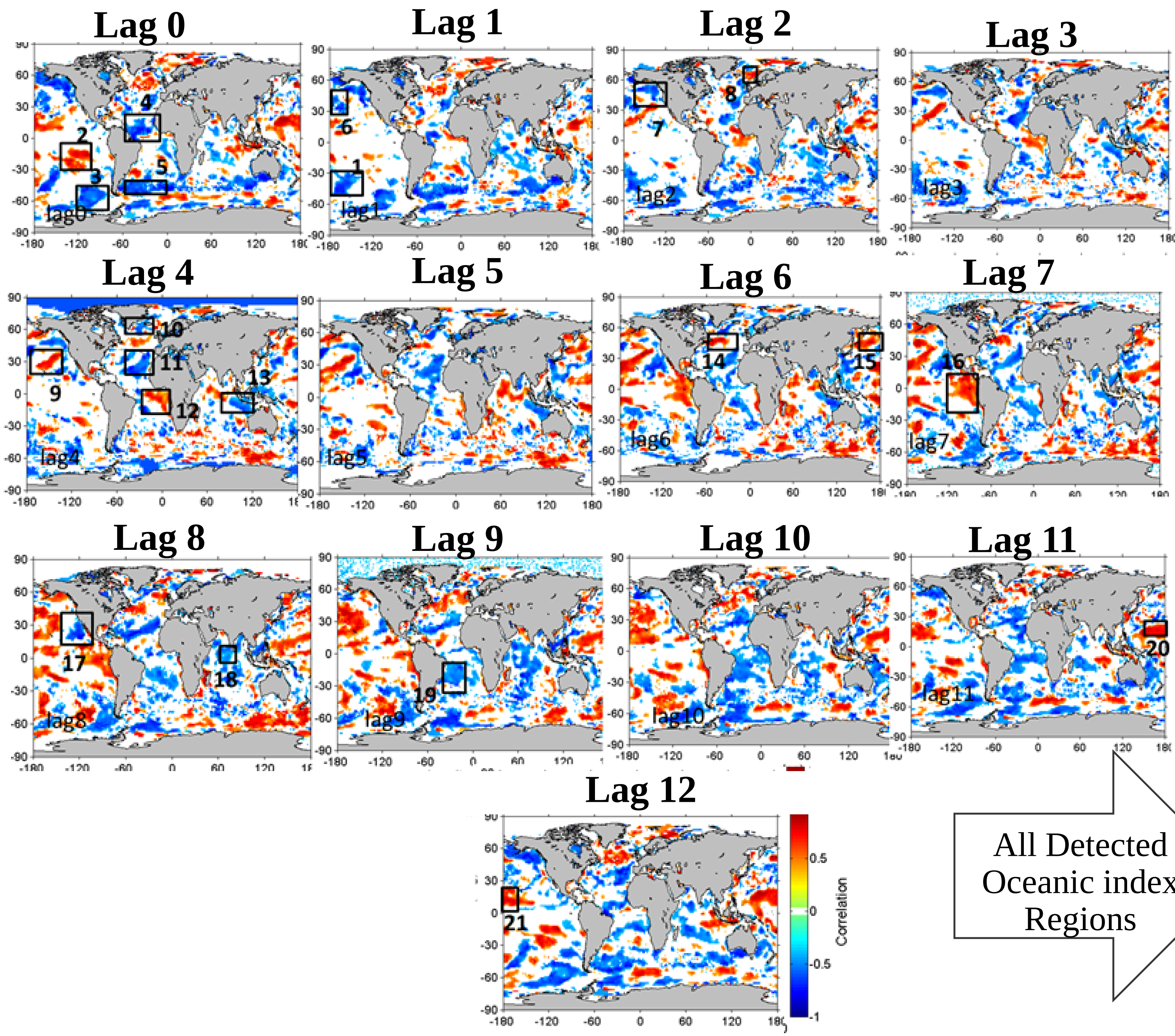
SST SAWP

Detrending process has well removed the long-term trends while reserved the oscillation characteristics of original signals.

SAWPs are totally different from the original and detrended time series, as they are reconstructed from the significant wavelet power at selected frequency band.

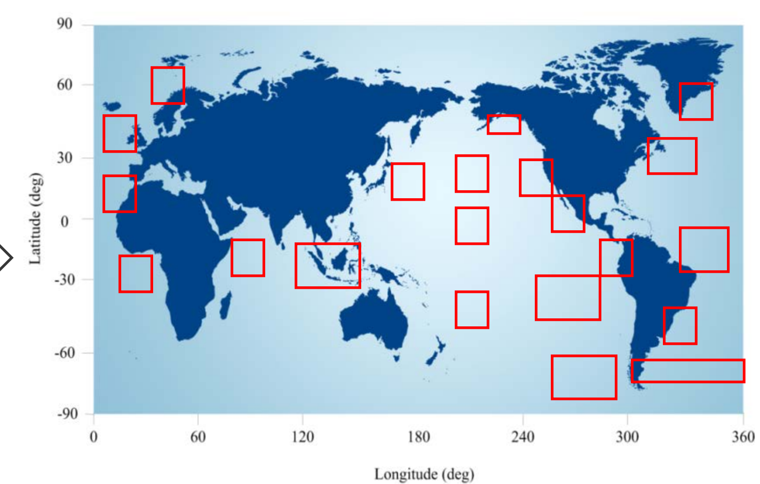
Method of Identifying the Dominant SST Teleconnection Regions

Adirondack – Winter Season Precipitation

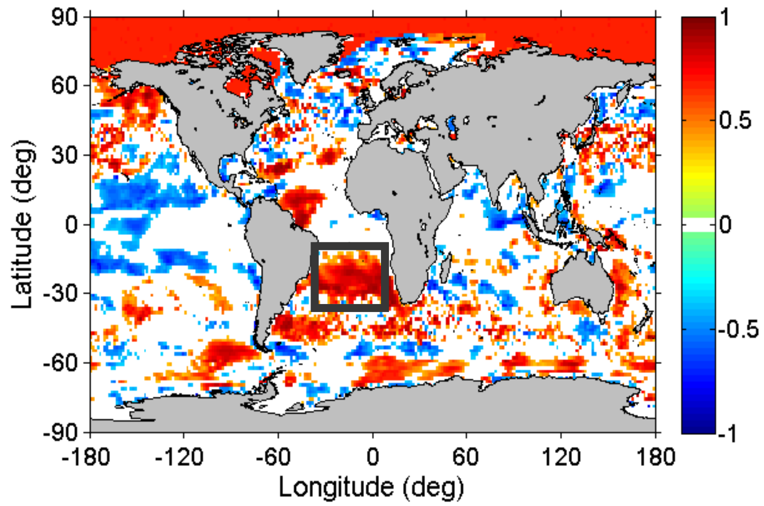


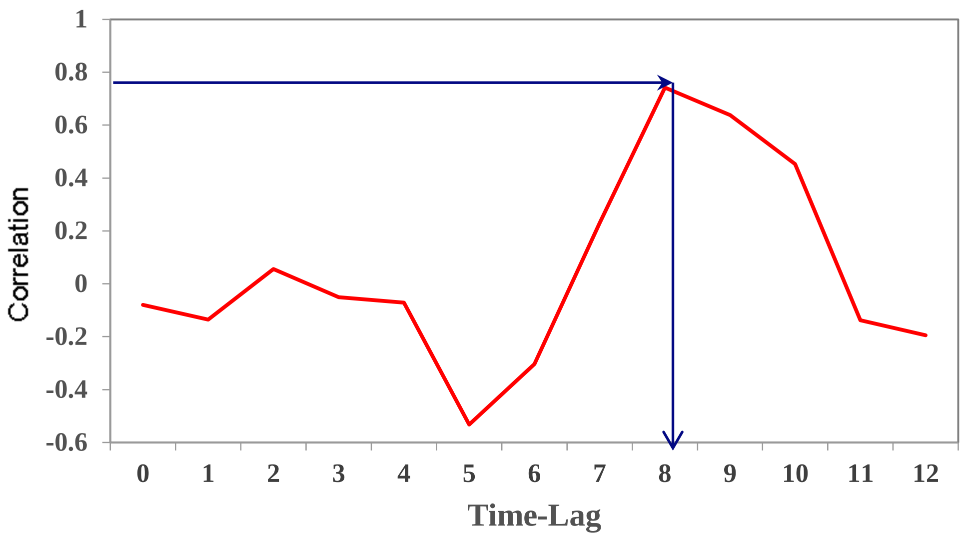
- Shaded colors show statistically significant correlation at the 95% confidence interval.
- Areas with a consistent significant correlation (lasting for more than 3 months) were extracted as possible forcing regions

All Detected Oceanic index Regions



Determining the corresponding times of Maximum and Minimum Correlation





- ❑ Associated time lags for each oceanic index regions were also identified by selecting the corresponding time lags with the maximum correlation coefficient between the oceanic index regions and precipitation.

Adirondack

Spring

Climate Variable	Lag Time												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Nino3.4	+												
SP *	+												
AMO		+											
AMO			+										
Nino 3				+									
PDO				+									
IOD						+							
Nino 4									+				
SP*		+											
PDO											+		
SP*											+		
IOD												+	
NAO							+						
WP						+							
SP*				+									
Nino4			+										
PDO				+									

Summer

Climate Variable	Lag Time												
	0	1	2	3	4	5	6	7	8	9	10	11	12
PDO												+	
IOD				+									
PDO				+									
SP*					+								
SP*						+							
AMO						+							
IOD					+								
Nino 1+2, Nino 3									+				
Nino4											+		
Nino 3.4											+		
SA*				+									
SP*		+											

SP: South Pacific
SA: South Atlantic
***: Non-Leading Teleconnection Pattern**

Adirondack

Winter

Climate Variable	Lag Time												
	0	1	2	3	4	5	6	7	8	9	10	11	12
SP *		+											
SP*	+												
SP *	+												
AMO	+												
SA *	+												
PDO										+			
PDO	+												
AO			+										
PDO						+							
AO				+									
NAO				+									
SA					+								
IOD					+								
NAO							+						
WP							+						
Nino1+2, Nino3								+					
PDO								+					
IOD						+							
SA											+		
Nino 4	+												
PDO												+	

Fall

Climate Variable	Lag Time												
	0	1	2	3	4	5	6	7	8	9	10	11	12
SP*	+												
AO	+												
NAO		+											
IOD												+	
Nino3.4, Nino3							+						
SA*							+						
AMO								+					
PDO		+											
SA*										+			
SP *										+			
SP *												+	
PDO						+							
PDO								+					
IOD			+										
PDO									+				

SP: South Pacific
SA: South Atlantic
***: Non-Leading Teleconnection Pattern**

Sensitivity Analysis

Extreme Learning Machine

For N arbitrary samples (xi , yi):

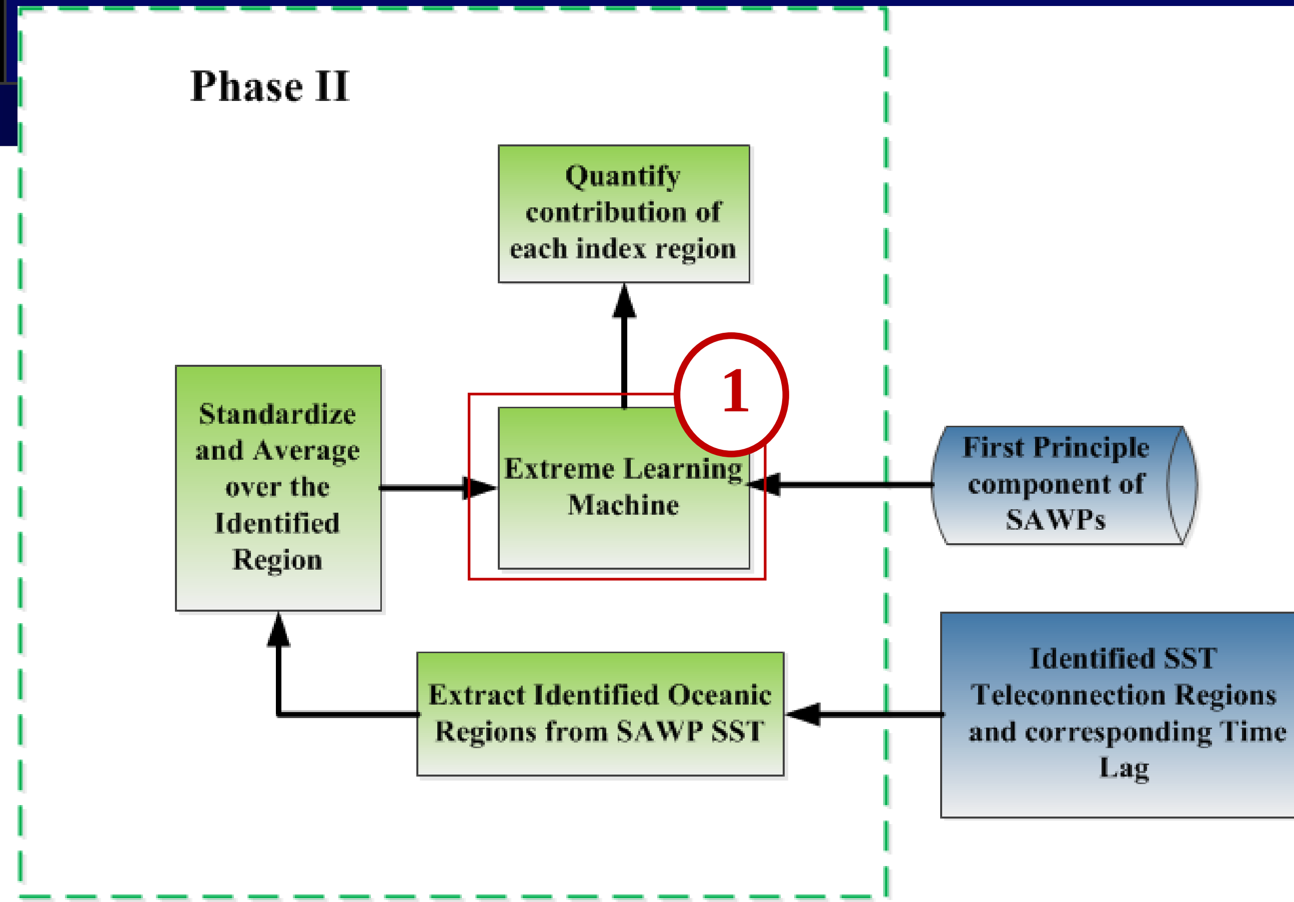
$$\sum_{i=1}^{N} v_i g(a_i \cdot x_j + b_i) = y_j \qquad j = 1, \dots, N$$

It can be written as

$$HV = Y$$

$$V = \begin{bmatrix} v_1^T \\ \vdots \\ v_L^T \end{bmatrix}_{L\times m} \qquad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N\times m}$$

$$H = \begin{bmatrix} g(a_1, b_1, x_1) & g(a_2, b_2, x_1) \dots & \dots g(a_L, b_L, x_L) \\ \vdots & \vdots & \vdots \\ g(a_1, b_1, x_N) & g(a_2, b_2, x_N) \dots & \dots g(a_L, b_L, x_N) \end{bmatrix}$$



g(.): Activation Function
a_i: the weight vector connecting the i^{th} hidden node and input nodes
v_i: the weight vector connecting the i^{th} hidden node and output nodes
b_i: is the biases of the i^{th} hidden node
y_j: is the expected value of the j^{th} output node.
H: Hidden layer output matrix
V: Output weight matric

Sensitivity Analysis

Extreme Learning Machine

(1) Given a training set, activation function g(.), and hidden node number (L),

(2) Randomly assign input weights a_i and bias b_i,

(3) Calculate the hidden layer output matrix H,

(4) Calculate the output weight V from

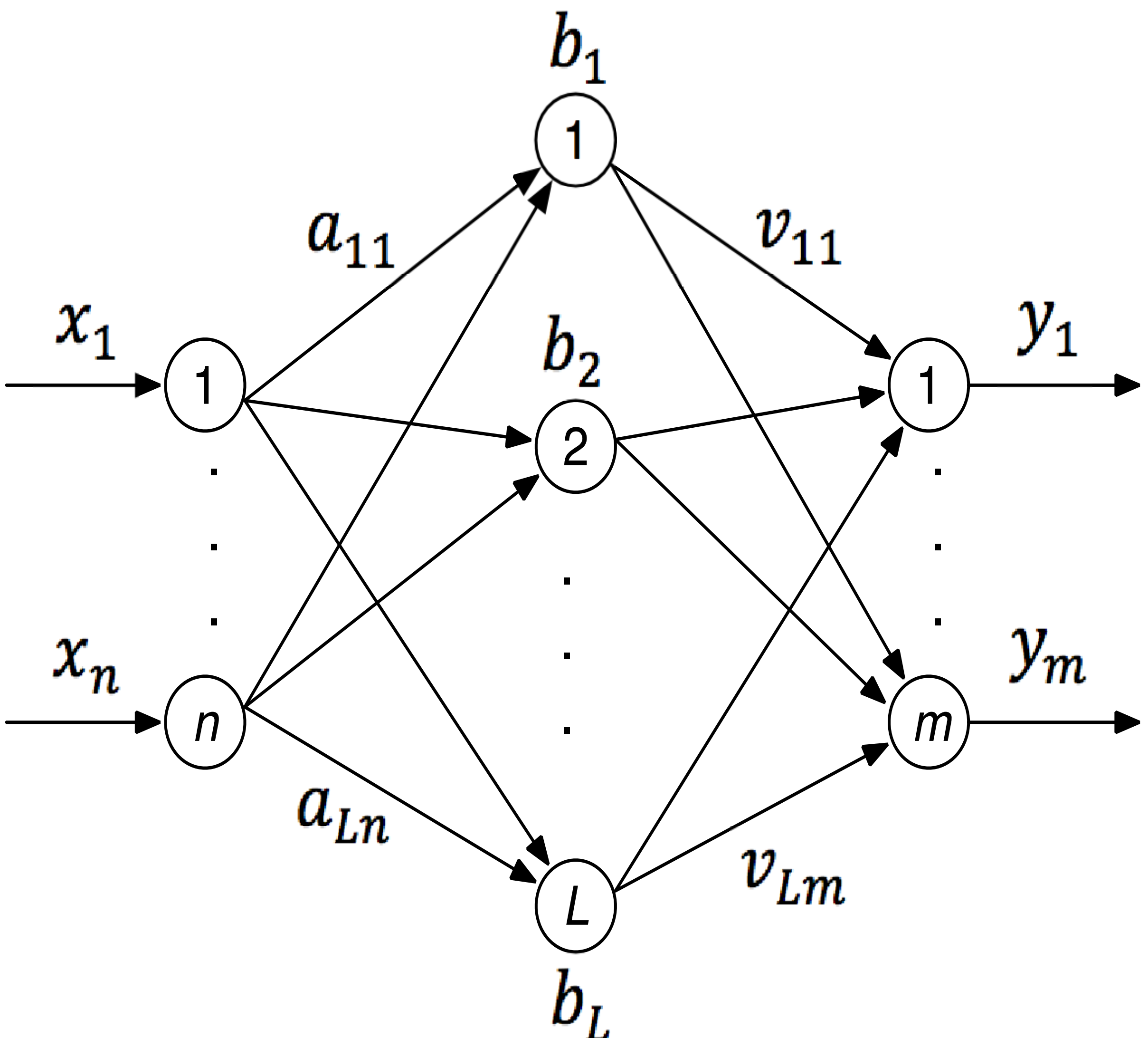
$V = YH^+$,

H^+: Moore – Penrose matrix inverse

Train : 70% of samples

Test: 30% of samples (unseen data)

Training process will be repeated until correlation coefficient reaches to 90%.



Typical scheme of single layer feedforward neural networks (Chang et al., 2010)

The ELM was performed using MATLAB, with code developed by Nanyang Technological University in Singapore.

Sensitivity Analysis

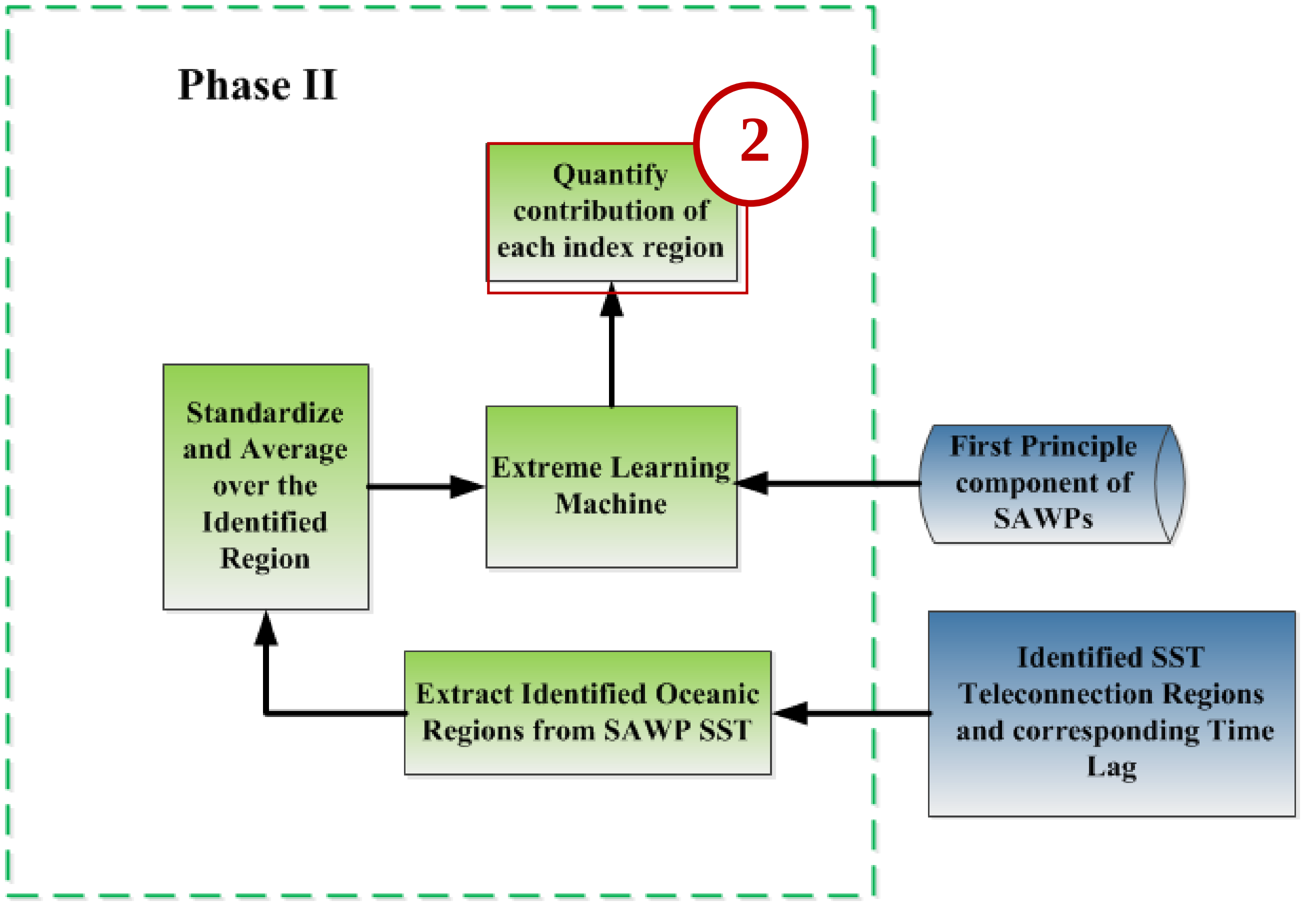
Sensitivity Analysis

(1) Including all the identified oceanic index regions,

(2) Excluding one of the identified oceanic index regions and including the rest indices,

(3) Residuals between the two simulated precipitation time series are defined as the precipitation responses to the excluded index.

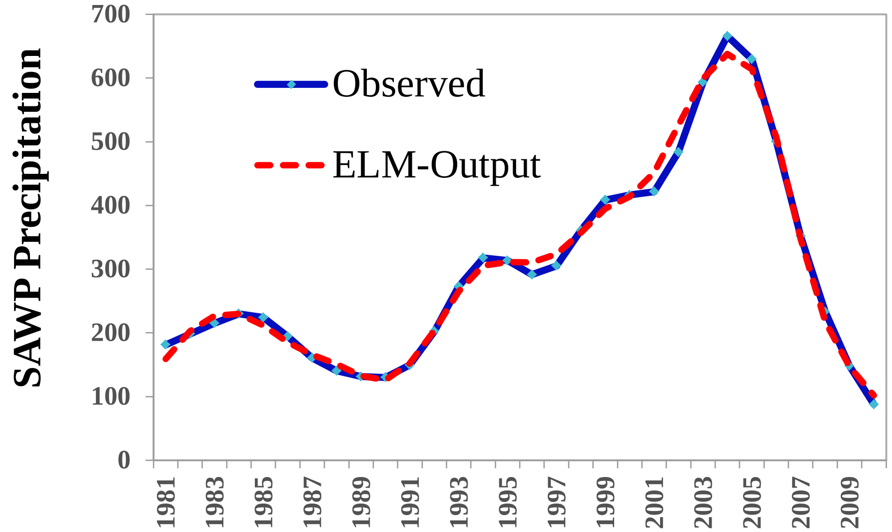
(4) To reduce the stochastic error, this procedure is repeated for 200 times for each index, and the average of these results is considered as the contribution of each index to precipitation at each site.



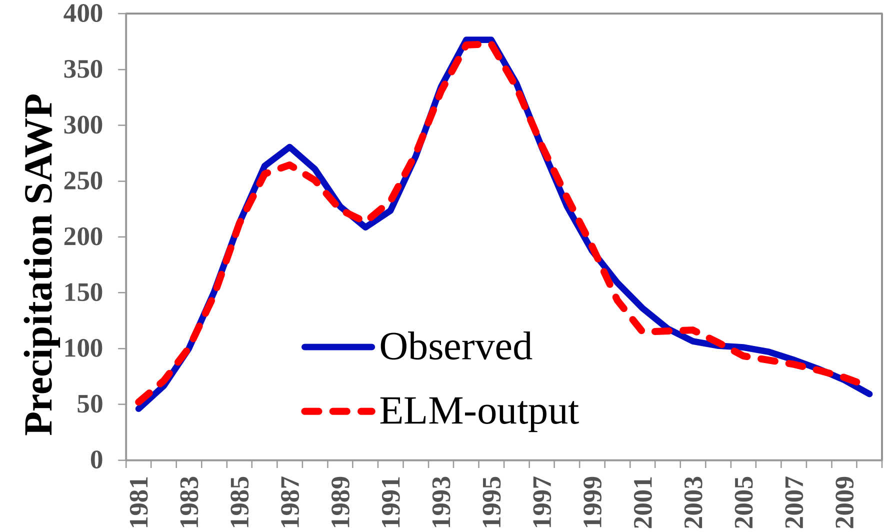
Sensitivity Analysis

Comparisons between observed and simulated precipitations by ELM in fall season

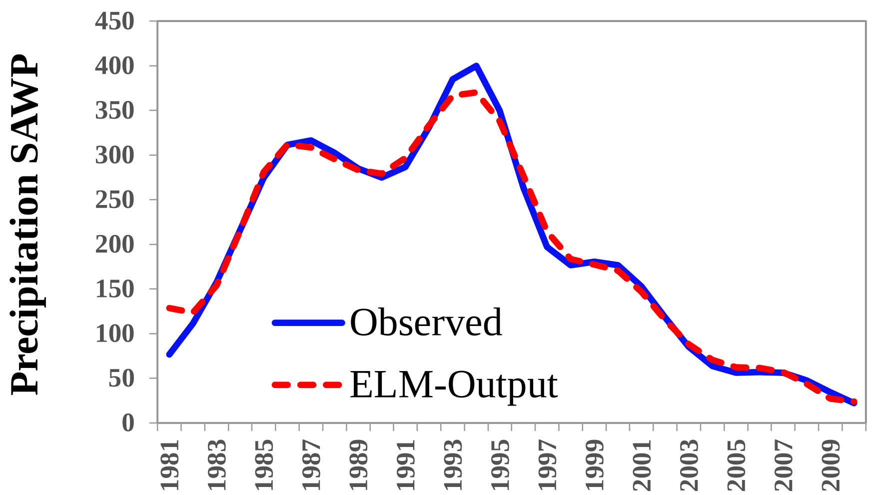
Adirondack



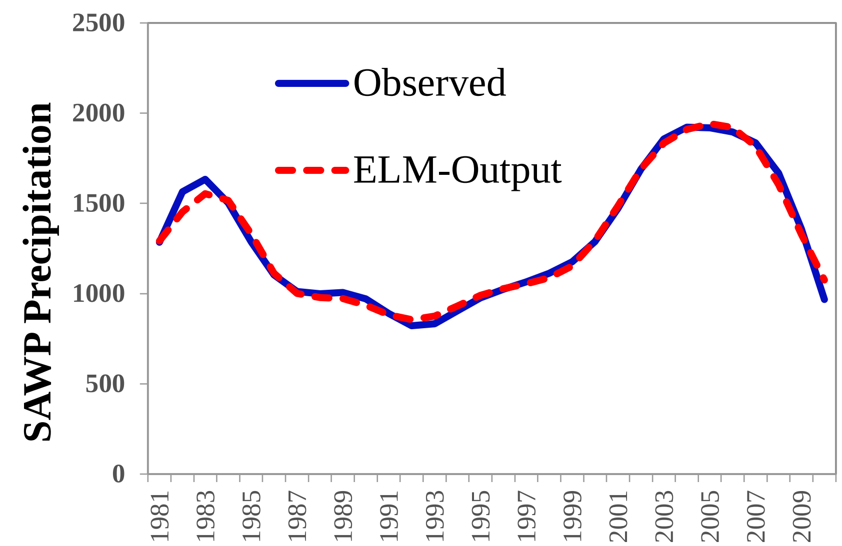
Selway-Bitterroot



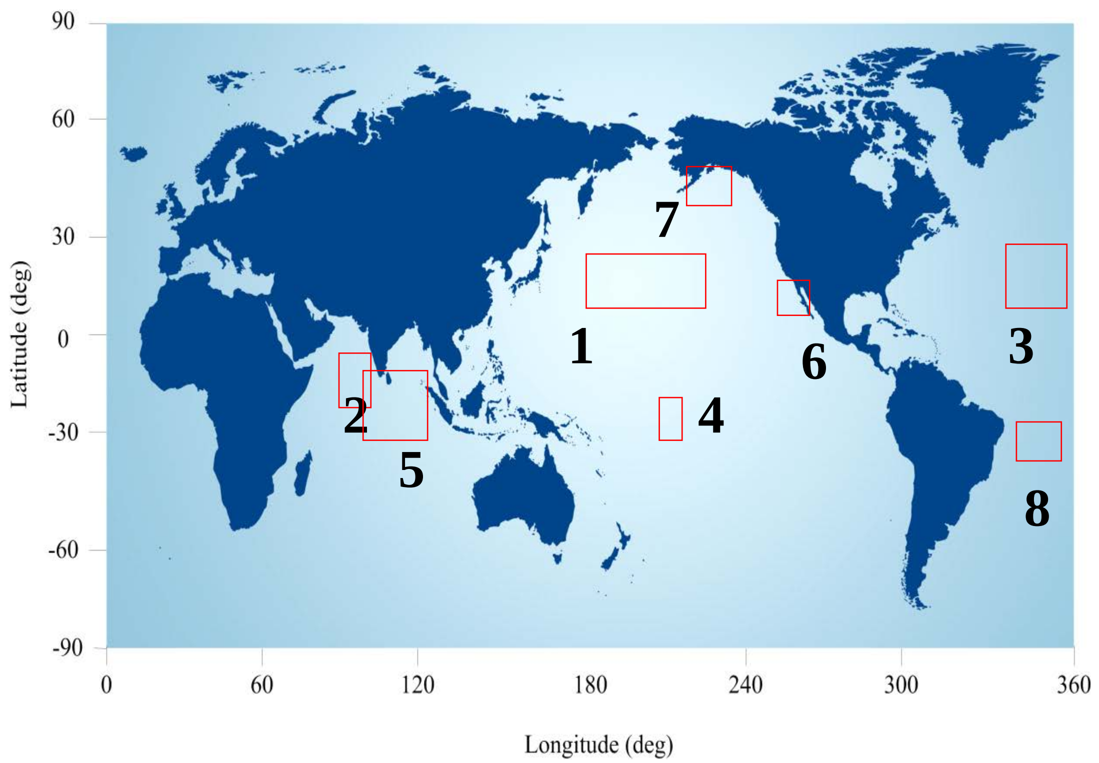
Weminuche Wilderness



La Amistad International Park



Adirondack



Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Adirondack State Park	48%	25%	55%	24%	75%	11%	63%	*16%

*** non-leading teleconnection patterns**

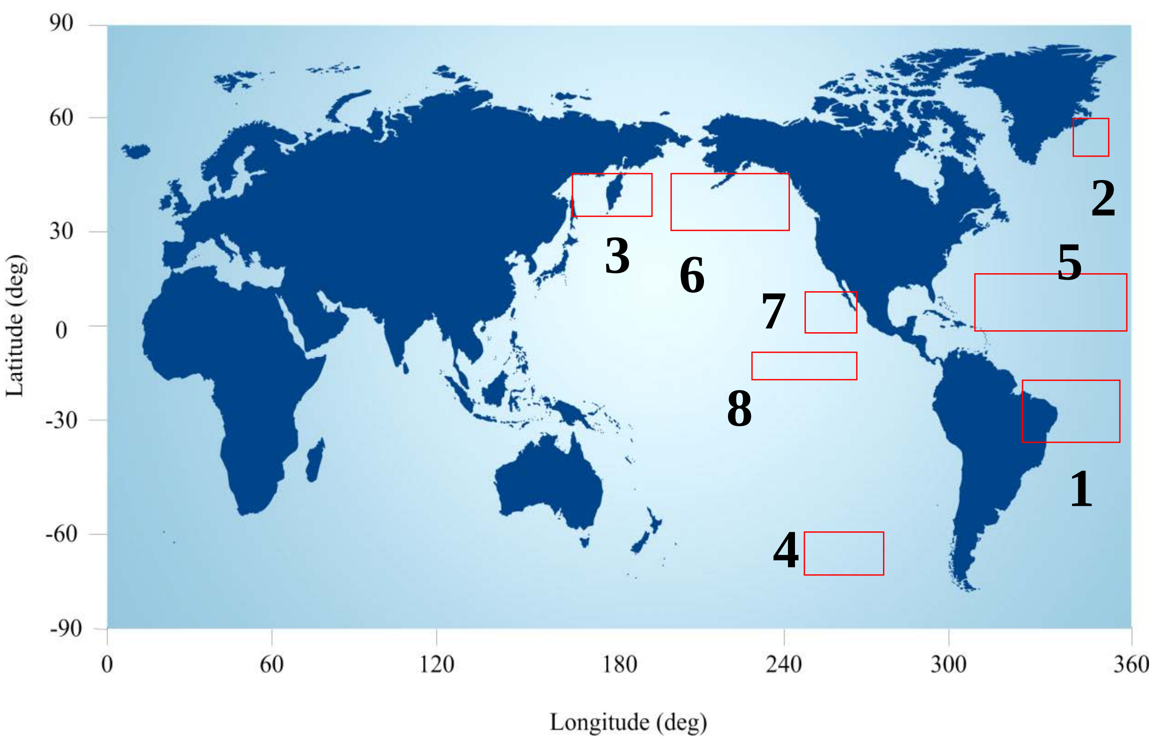
Contribution of each index on precipitation of four sites (percentage)

Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Adirondack State Park	WP	IOD	EA	Nino3.4	IOD	PDO	PDO	SA *

Indexed regions associated with the climate teleconnection patterns.

EA: East Atlantic; IOD: Indian Ocean Dipole; PDO: Pacific Decadal Oscillation
SA: South Atlantic; WP: West Pacific; * non-leading teleconnection patterns

Selway-Bitterroot Wilderness



Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Selway Bitterroot Wilderness	75%	13%	64%	*16%	73%	13%	62%	15%

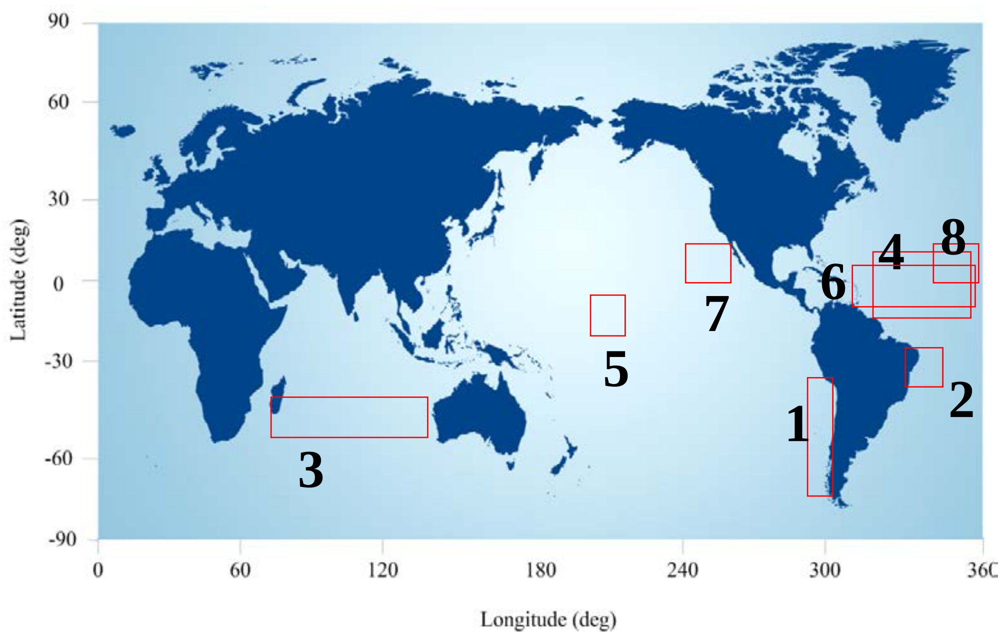
Contribution of each index on precipitation of four sites (percentage)

Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Selway Bitterroot Wilderness	TSA	AO	WP	SP *	TNA	PDO	PDO	Nino3

Indexed regions associated with the climate teleconnection patterns.

TSA: Tropical South Atlantic; SP: South Pacific; WP: West Pacific; TNA: Tropical North Atlantic; PDO: Pacific Decadal Oscillation
*** non-leading teleconnection patterns**

Weminuche Wilderness



Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Weminuche Wilderness	*65%	17%	70%	18%	57%	22%	69%	14%

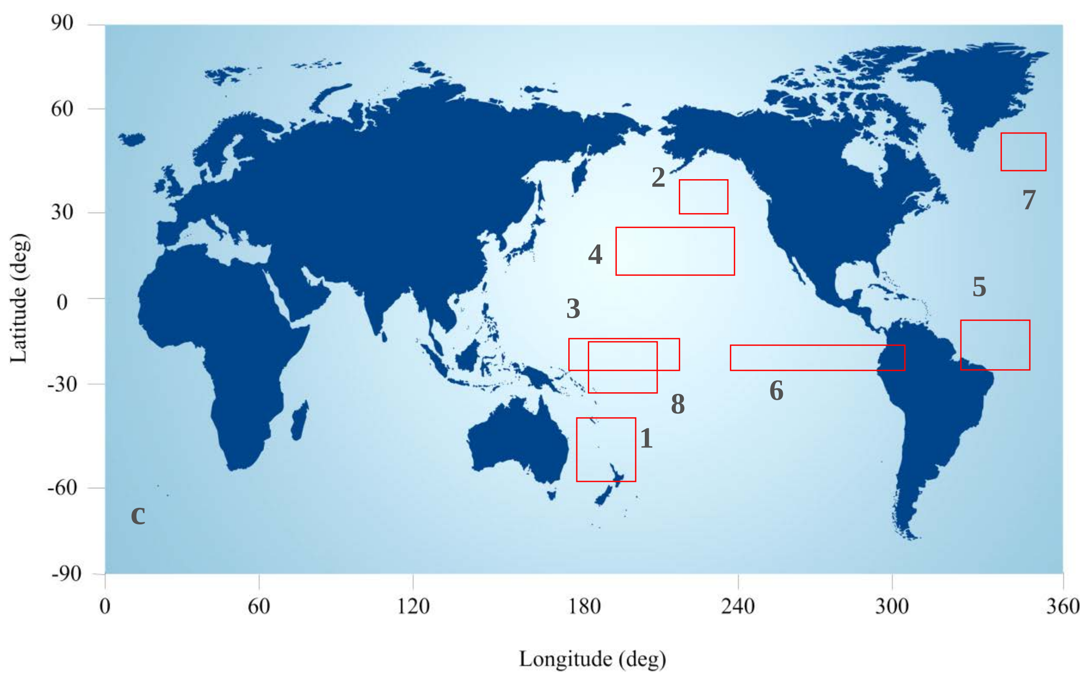
Contribution of each index on precipitation of four sites (percentage)

Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Weminuche Wilderness	SP *	TSA	IOD	TNA	Nino3.4	TNA	PDO	TNA

Indexed regions associated with the climate teleconnection patterns.

IOD: Indian Ocean Dipole; TSA: Tropical South Atlantic; SP: South Pacific; TNA: Tropical North Atlantic; PDO; Pacific Decadal Oscillation
*** non-leading teleconnection patterns**

La Amistad International Park



Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Weminuche Wilderness	*66%	16%	52%	26%	80%	9%	61%	18%

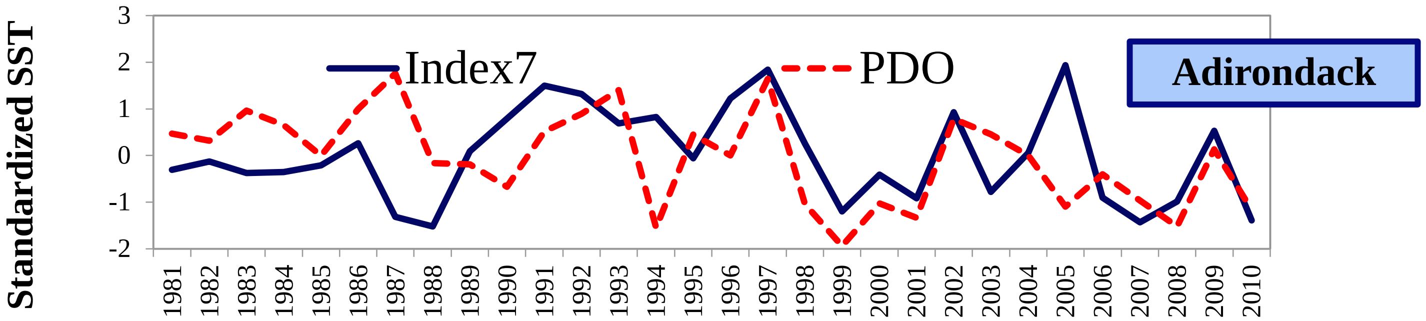
Contribution of each index on precipitation of four sites (percentage)

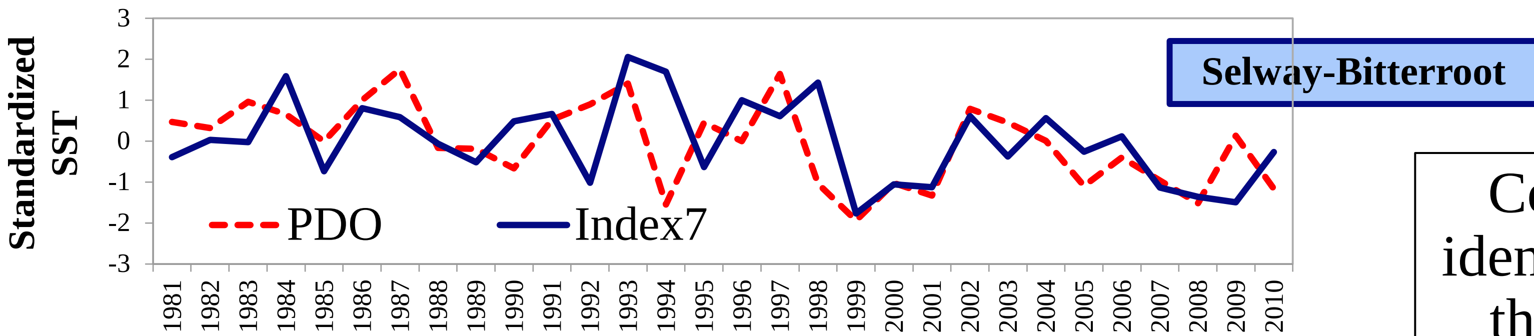
Study Area	Winter		Spring		Summer		Fall	
Index	1	2	3	4	5	6	7	8
Weminuche Wilderness	SP*	PDO	Nino3.4	PDO	TSA	Nino 1+2 Nino3	EA	Nino3.4

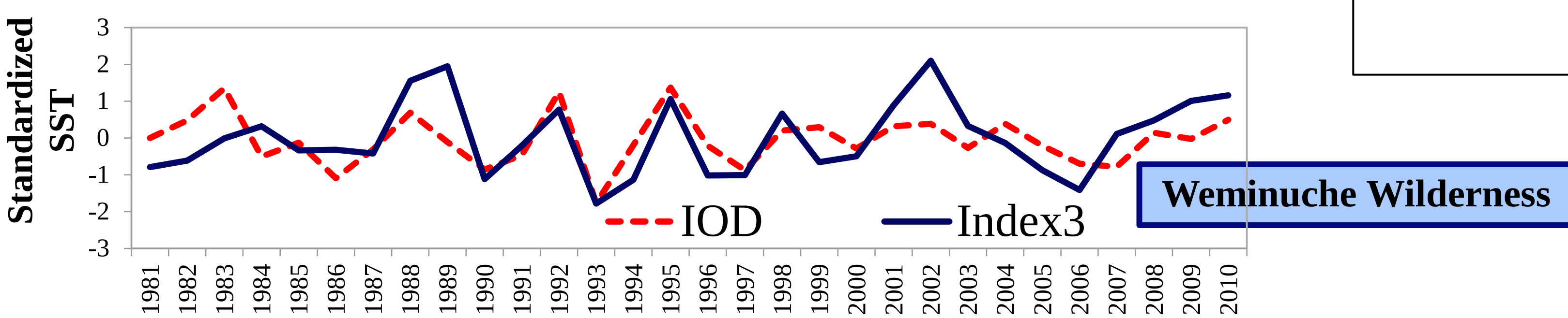
Indexed regions associated with the climate teleconnection patterns.

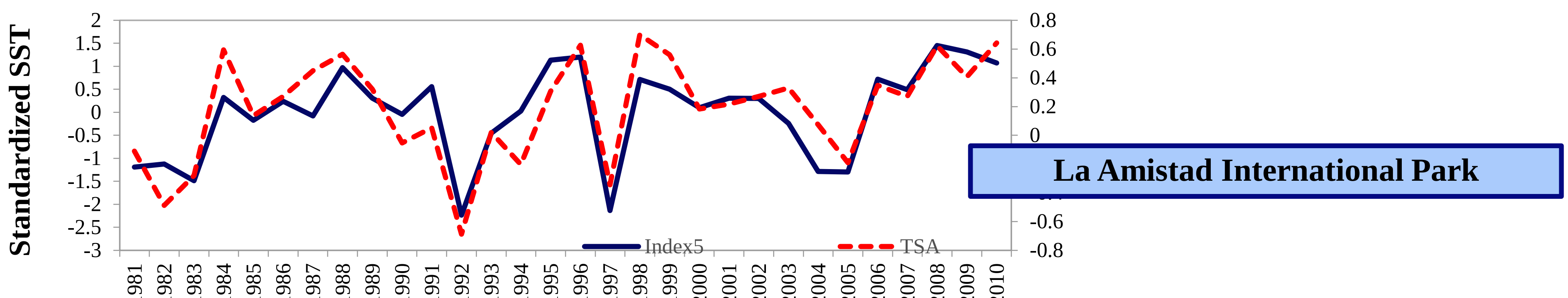
SP: South Pacific; TSA: Tropical South Atlantic; PDO; Pacific Decadal Oscillation; EA: East Atlantic
*** non-leading teleconnection patterns**

Model Validation







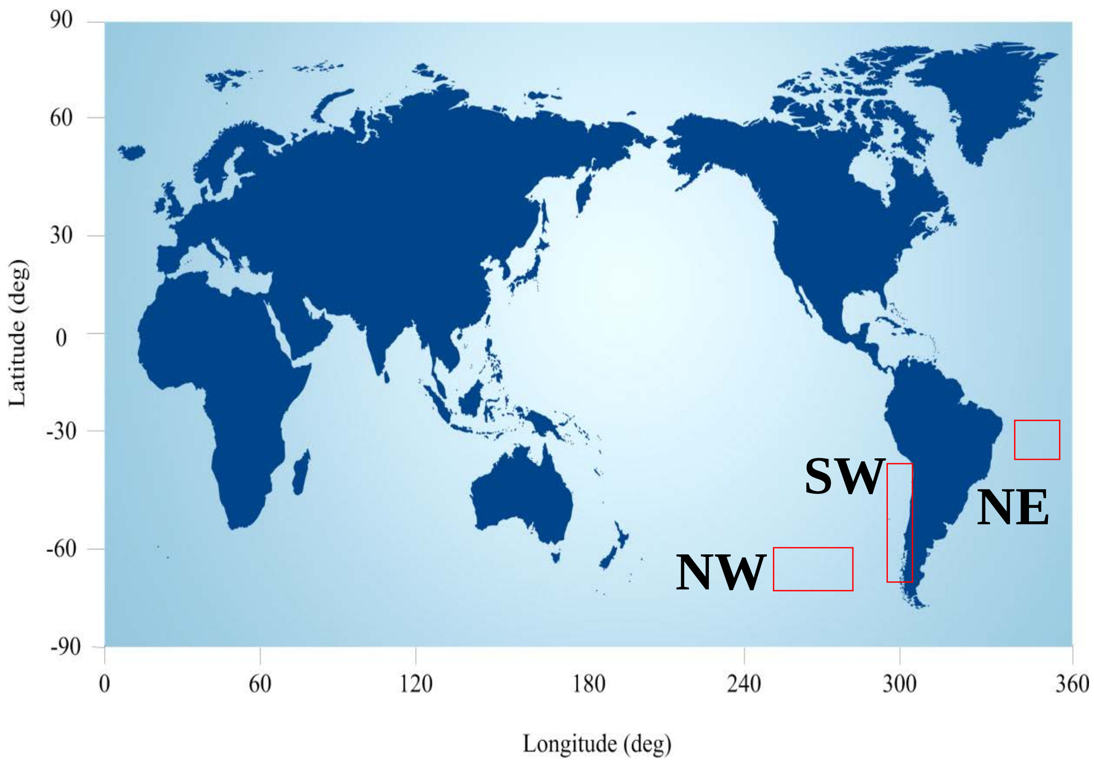


Comparisons between the identified oceanic indices and the known teleconnection patterns

Computational Time

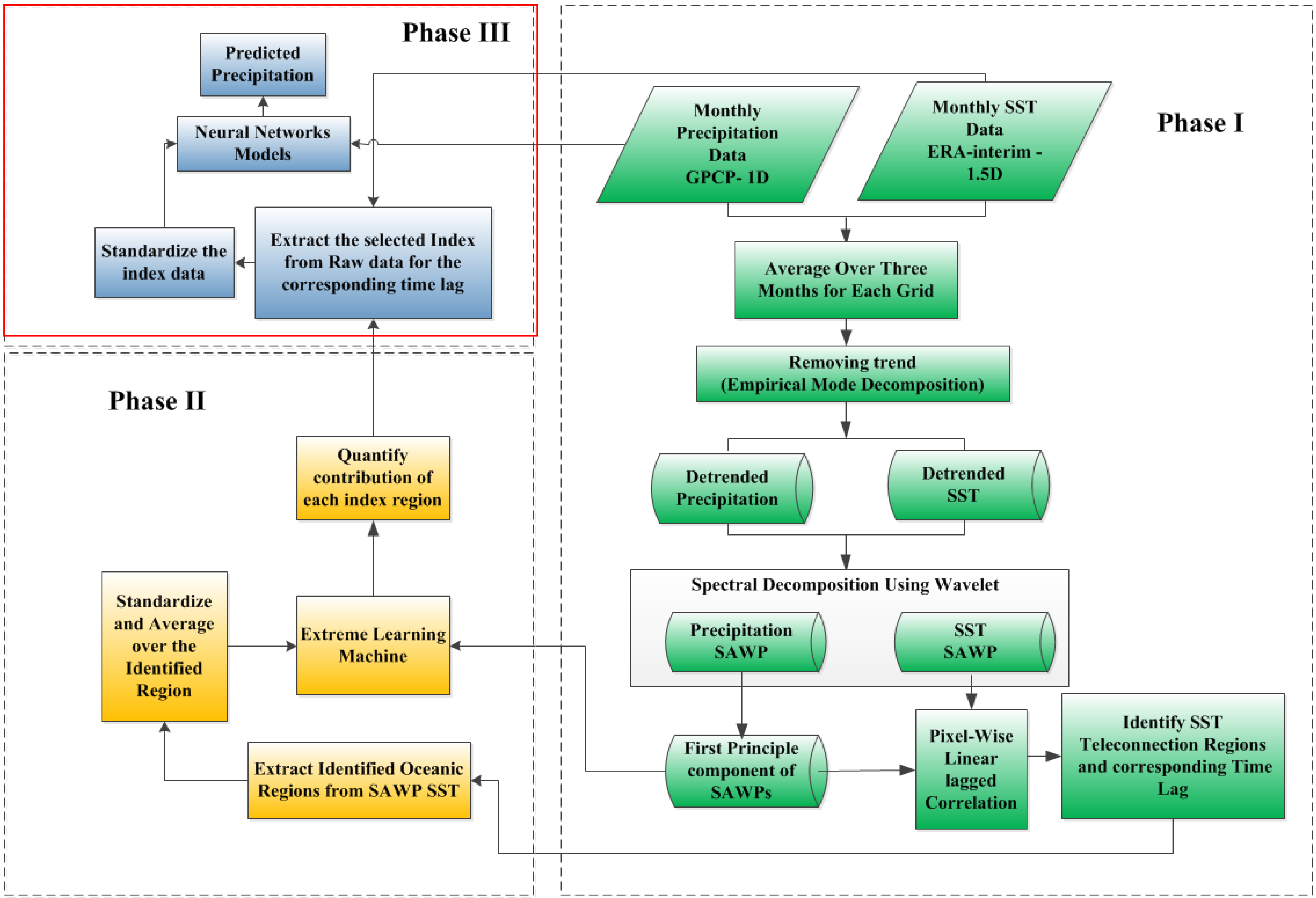
Phase	Task		CPU time per season (Second)	Total CPU time (Second)
Phase I	Detrending	SST	39,348	1,888,704
		Precipitation	AD(28),WE(7),SL(12),LA(16) *	252
	SAWP	SST	915	43,920
		Precipitation	AD(1),WE(1),SL(1),LA(1) *	16
	Linear Lagged Correlation		148	2,368
Phase II	ELM and Sensitivity Analysis		7	112

Scientific Findings



- ✓ Some of these non-leading oceanic regions had a much higher contribution to variability of precipitation compared to these known teleconnection patterns.
- ✓ It highlights the importance of considering the non-leading teleconnection signals as well as the known teleconnection patterns for precipitation forecasting.

Future Work

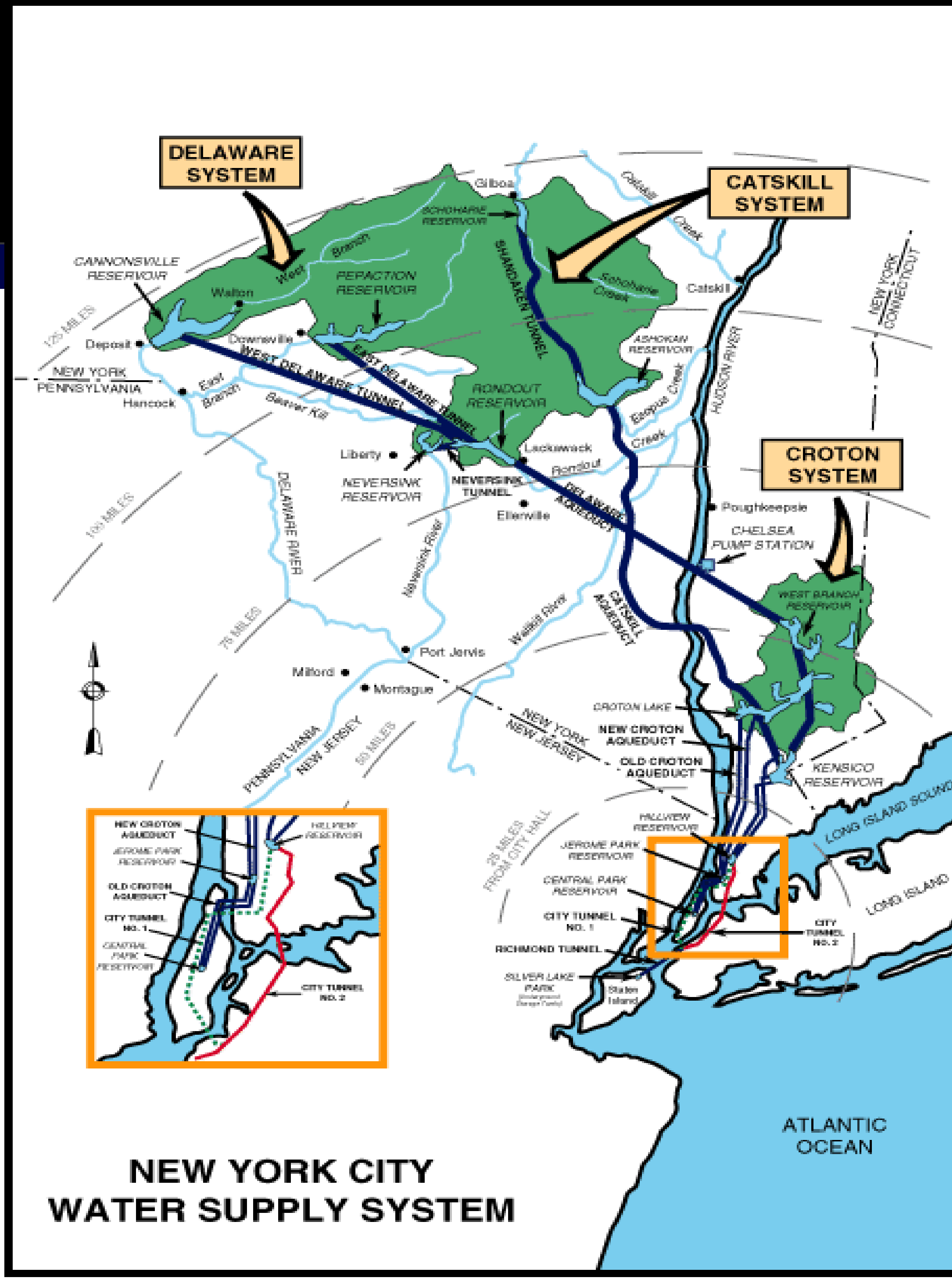

Phase III
Predicted Precipitation
Neural Networks Models
Standardize the index data
Extract the selected Index from Raw data for the corresponding time lag
Phase I
Monthly Precipitation Data GPCP- 1D
Monthly SST Data ERA-interim - 1.5D
Average Over Three Months for Each Grid
Removing trend (Empirical Mode Decomposition)
Detrended Precipitation
Detrended SST
Spectral Decomposition Using Wavelet
Precipitation SAWP
SST SAWP
First Principle component of SAWPs
Pixel-Wise Linear lagged Correlation
Identify SST Teleconnection Regions and corresponding Time Lag
Phase II
Quantify contribution of each index region
Standardize and Average over the Identified Region
Extreme Learning Machine
Extract Identified Oceanic Regions from SAWP SST

Future Work

- Source waters of New York City water supply system: 1) Catskill system, 2) Delaware System, 3) Croton System

- provide water for 8 million residents in New York City, as well as 1 million residents north of the city (Provide 1.3 billion gallons per day)

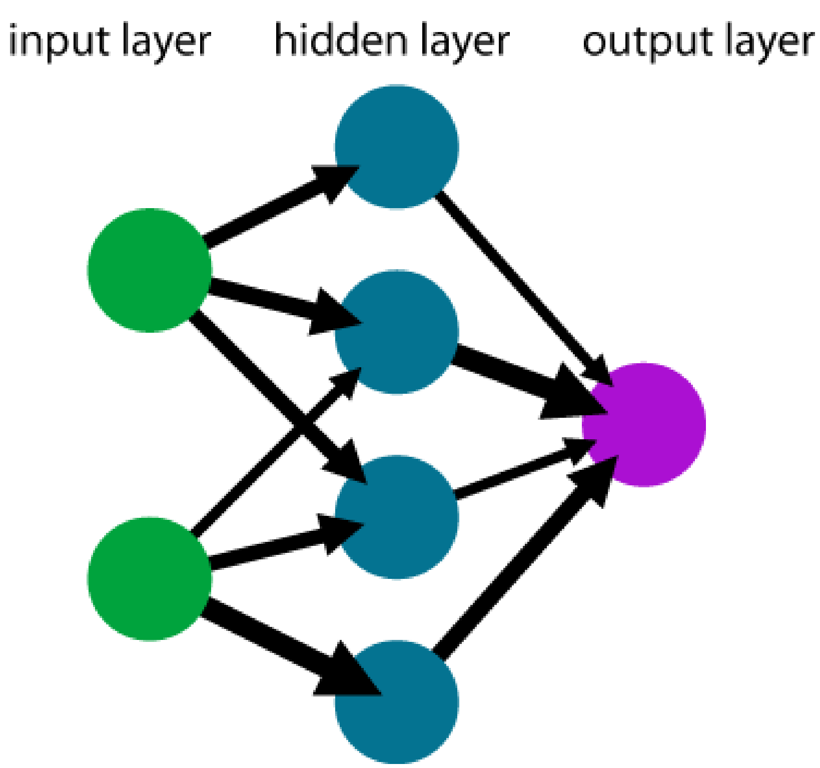
  - 40% is derived from the Catskill system, 50% from the Delaware System, and 10% from the Croton System



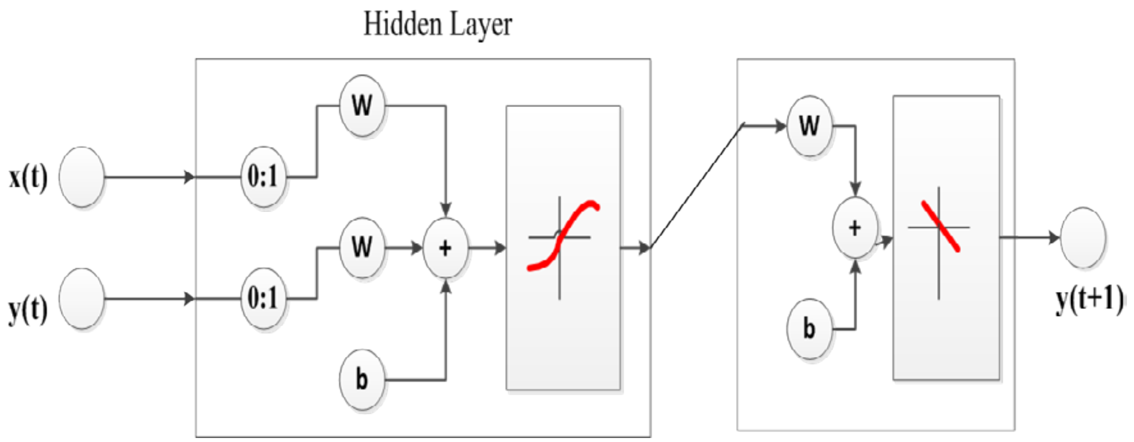
http://itsaboutthehudsonvalley.com/Reservoirs/reservoirs.html

Future Work

ANN
Artificial Neural Network Model



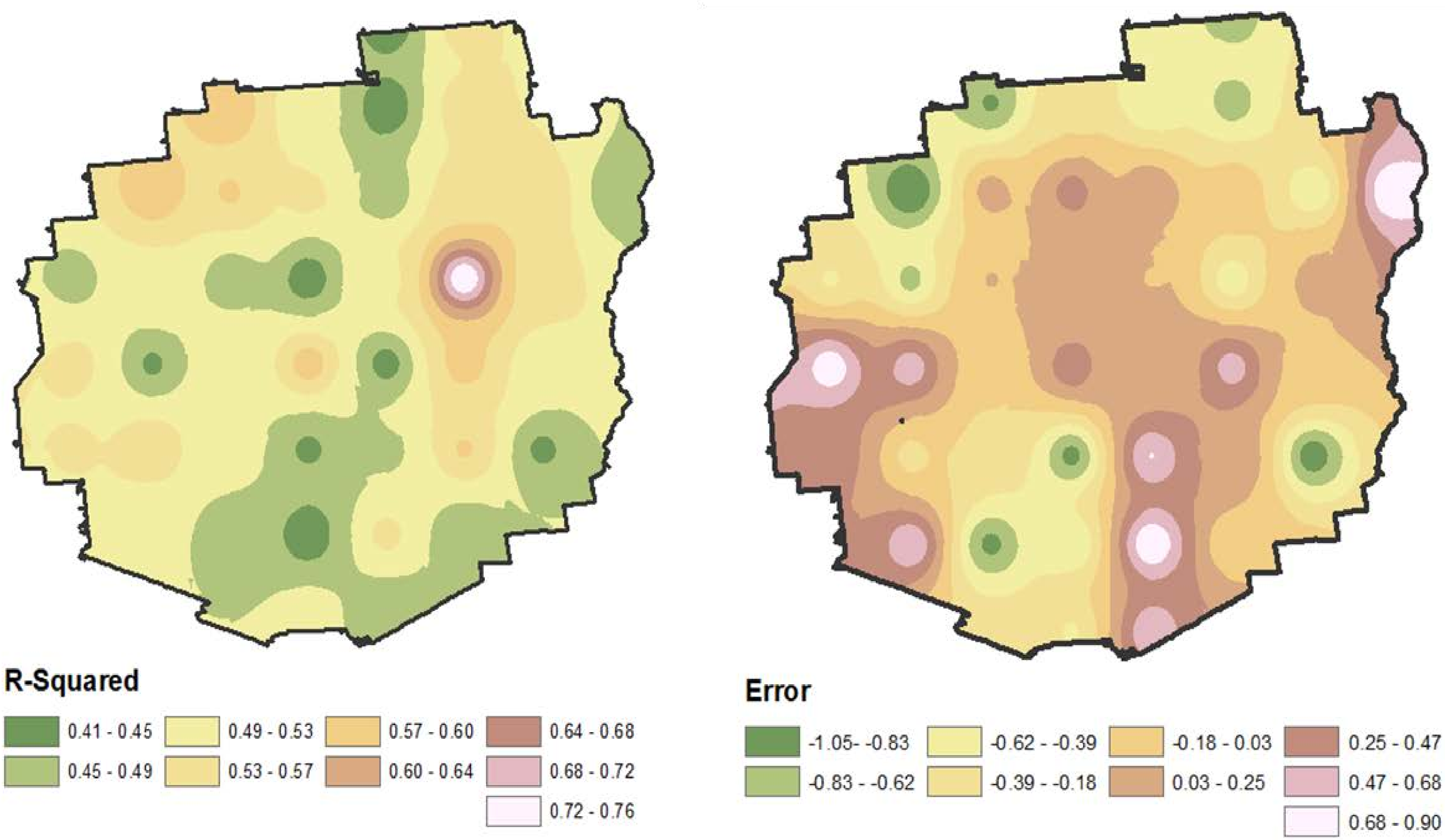
NARXNET
Nonlinear Autoregressive Neural Network with External Input



$$y(t) = f(y(t-1), \ldots \ldots, y(t-d), x(t-1), \ldots \ldots, x(t-d))$$

Where y(t) is the predicted time series, x(t) is the time series for each of the input variables (i.e. spectral reflectance values, meteorological parameters, and reservoir elevation), d is the input and feedback delay node.

Future Work



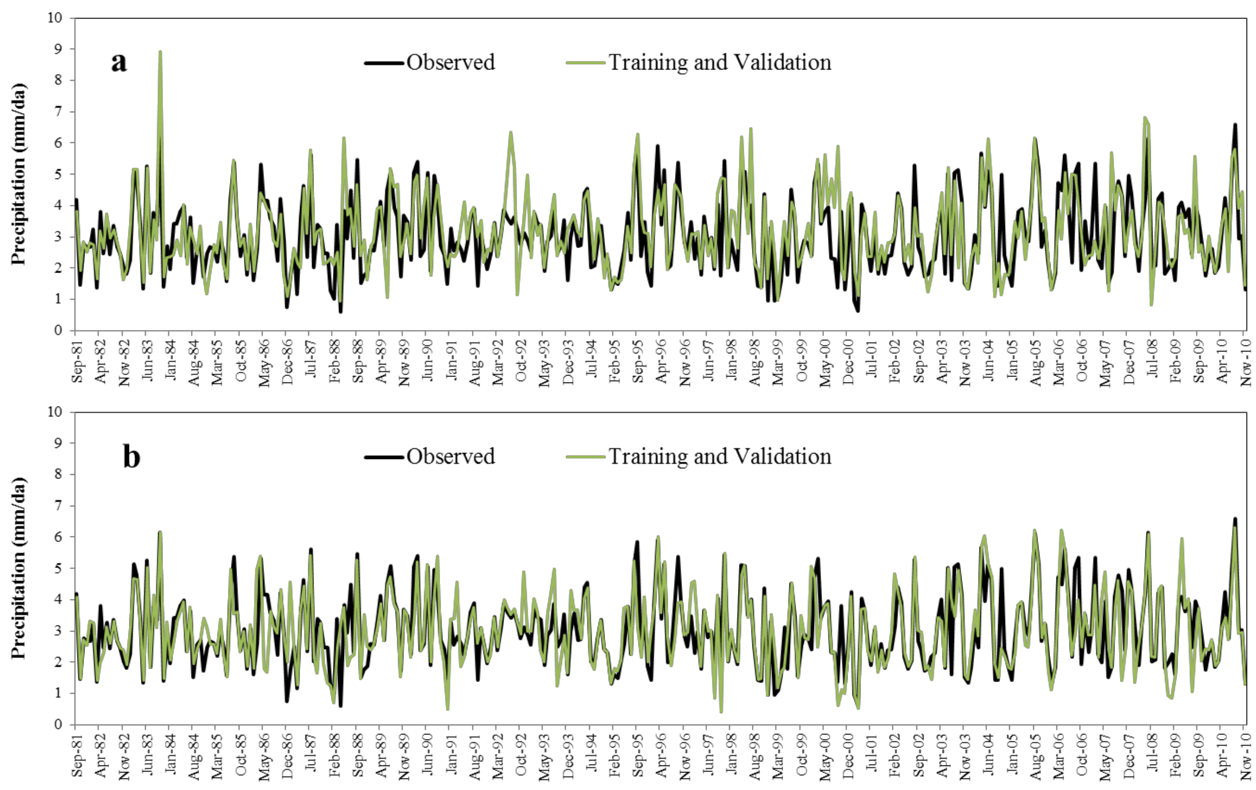
NARXNET precipitation forecasting model was applied to all 39 Adirondack precipitation grids in the year 2011. The image on the left displays the spatial correlation ('R-squared') contours. The image on the right displays the forecast error contours in mm/day.

Future Work

Scenario 1: It only includes the known teleconnection patterns,
Scenario 2: It contains both known and unknown teleconnection patterns.

R-squared and root-of-mean square error (RSME) of NARXNET and ANN models for the proposed scenarios

Models	Scenario	R-squared	RMSE Training	RMSE Validation
ANN	1	0.41	0.78	1.00
	2	0.50	0.49	0.97
NARXNET	1	0.60	0.38	0.95
	2	0.65	0.24	0.88



Comparison between the model output with observed precipitation for scenario 2 using (a) ANN model and (b) NARXNET model.

TELECONNECTION SIGNALS EFFECT ON TERRESTRIALPRECIPITATION: 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

DISCOVER PHYSICALLY MEANINGFUL TELECONNECTION PATTERNS WHICH EFFECT THE PRECIPITATION

INPUT
- 1980-2010 Monthly precipitation data
- 1979-2010 Monthly Sea Surface Temperature data

PROCESS 1

Big Data Analytics

Feature Construction with Lag Time(0-12 Months)

Streaming Feature Selection

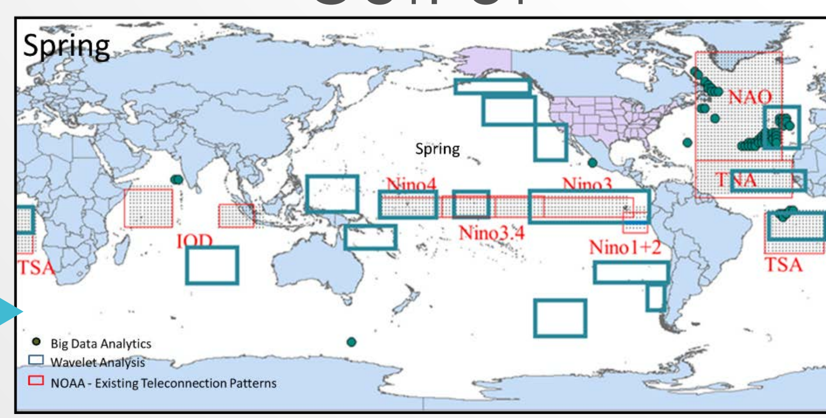
PROCESS 2

Wavelet Analysis

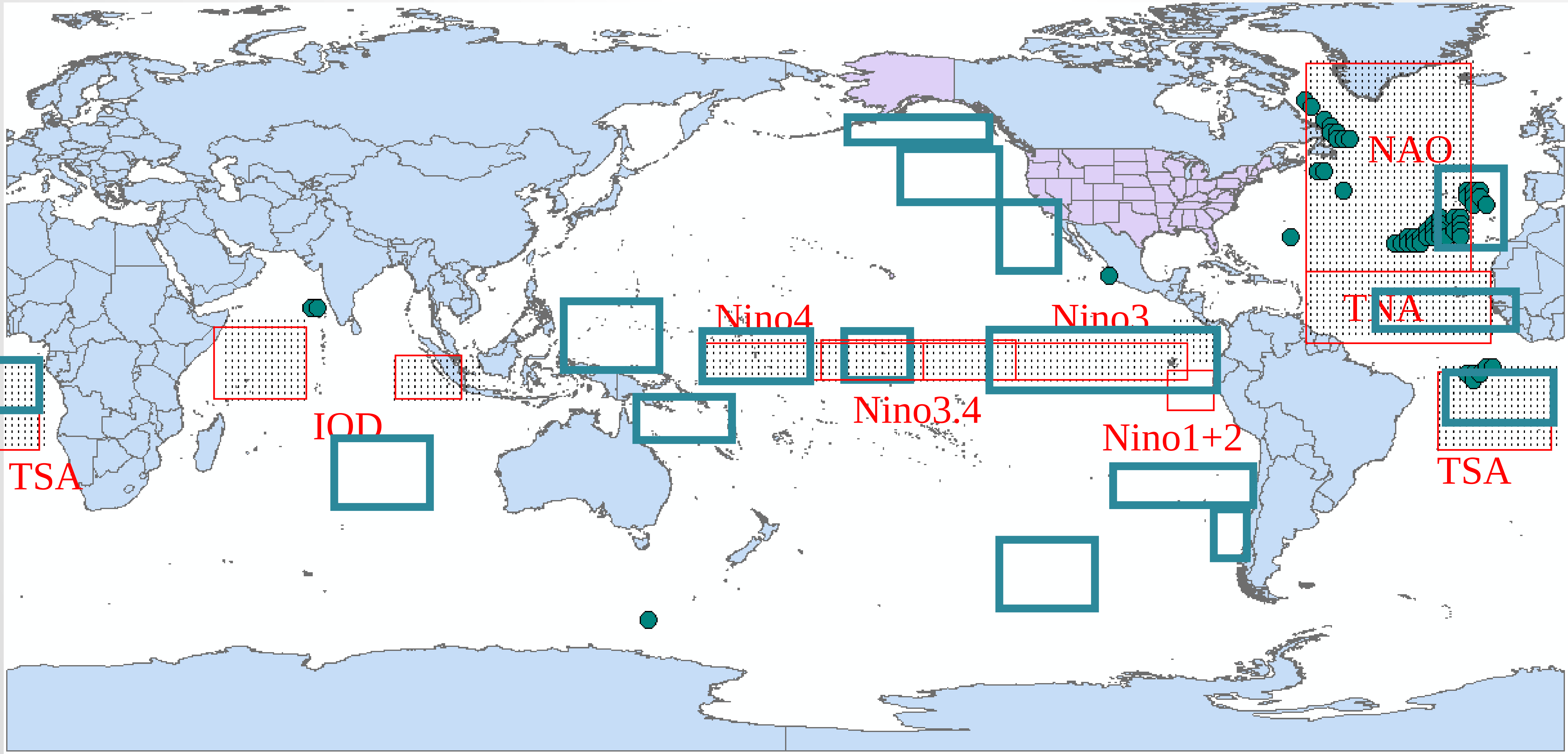
Empirical Mode Decomposition & Wavelet Spectral Analysis

Pearson Correlation-based Regression Analysis

OUTPUT



Spring 0.005

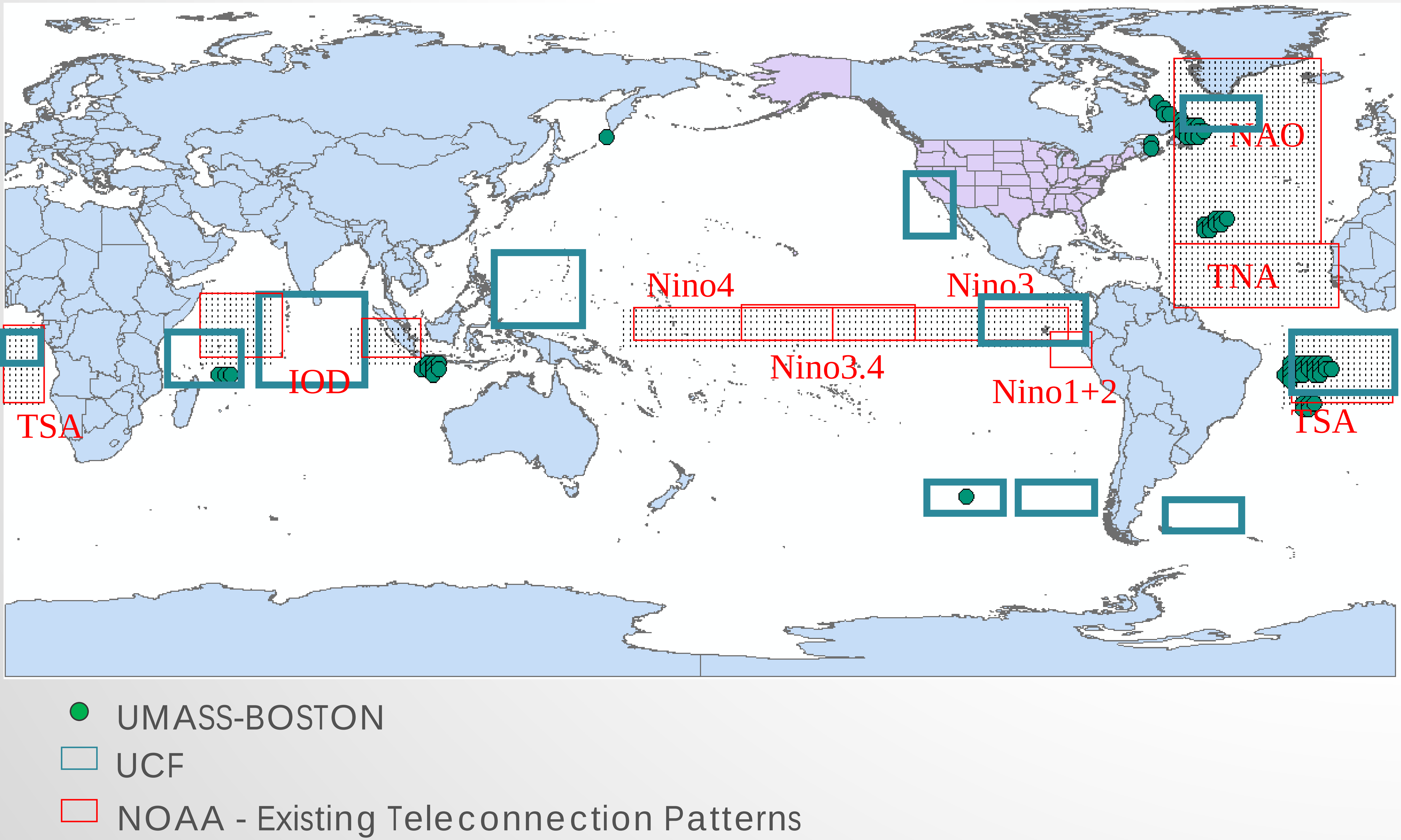


UMASS-BOSTON

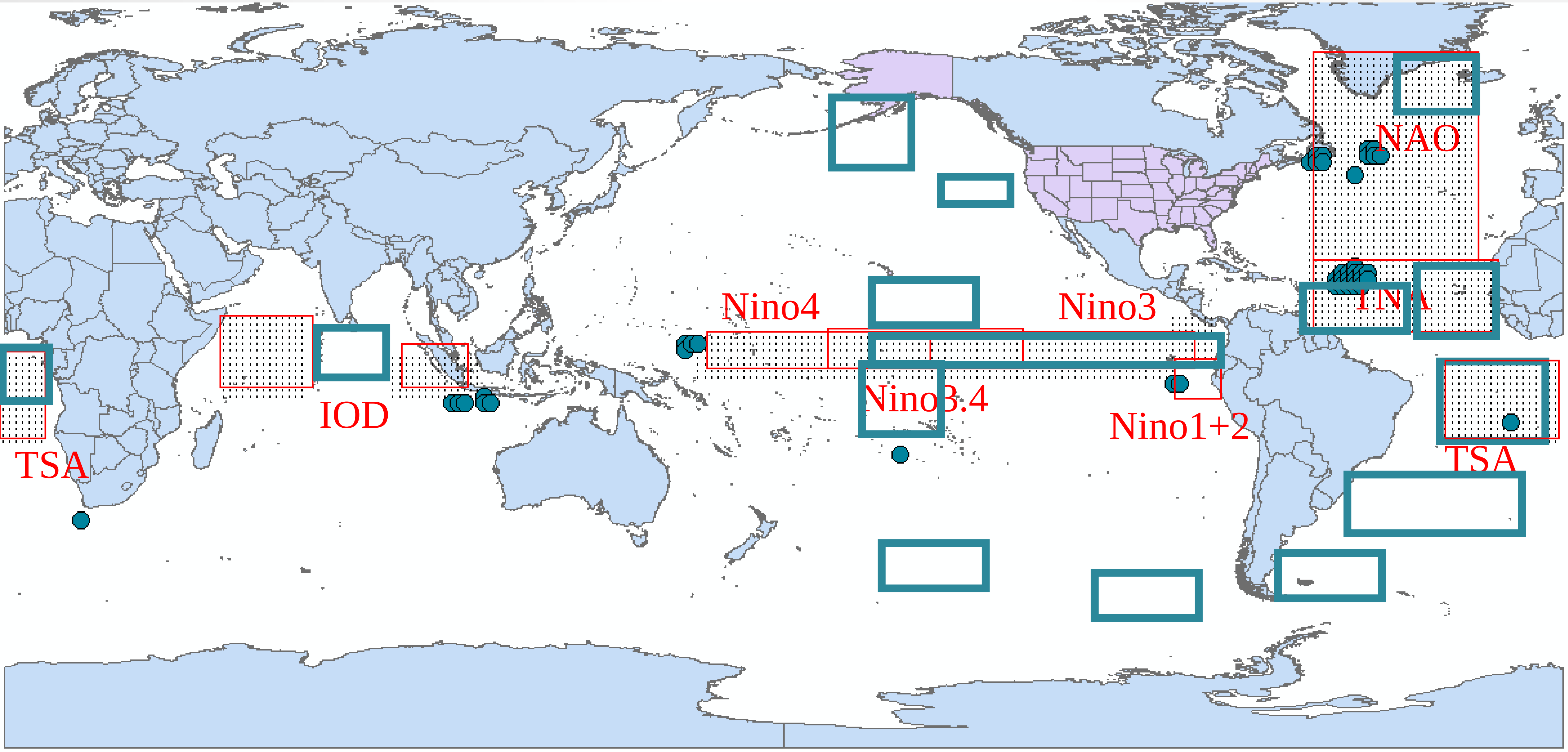
UCF

NOAA - Existing Teleconnection Patterns

Summer 0.005

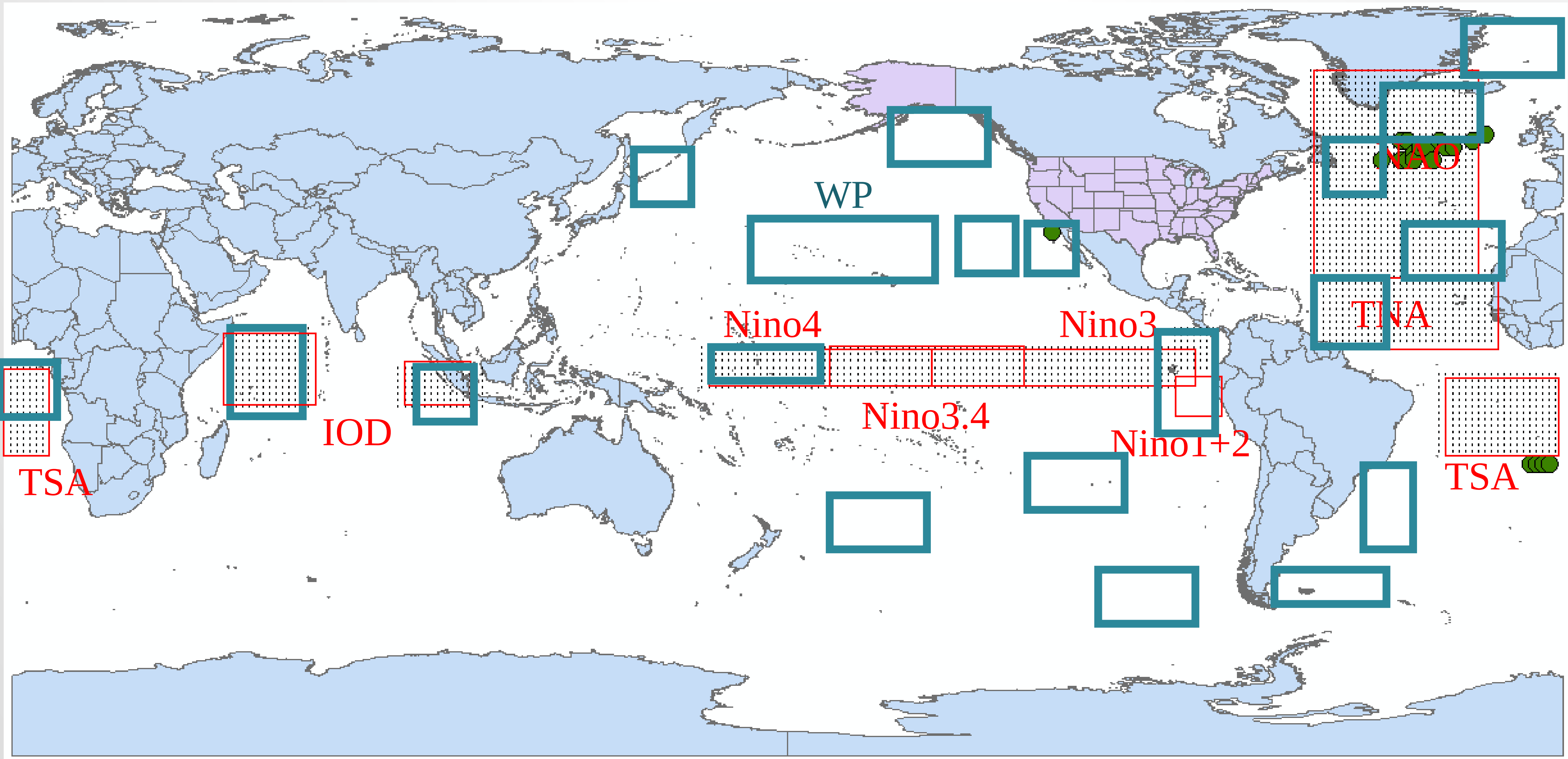

NAO
TNA
Nino4
Nino3
Nino3.4
Nino1+2
IOD
TSA
TSA
UMASS-BOSTON
UCF
NOAA - Existing Teleconnection Patterns

Fall 0.005



UMASS-BOSTON
UCF
NOAA - Existing Teleconnection Patterns

Winter 0.005



UMASS-BOSTON

UCF

NOAA - Existing Teleconnection Patterns

Thank You !

Ni-Bin Chang

nchang@ucf.edu