Summary of Related Research Work at North Carolina A&T State University

Abdollah (Abbie) Homaifar (Ph.D.)

Duke Energy Eminent Professor and

Director, Autonomous Control and

Information Technology Center (ACIT)

Office: 336-285-3709

Fax: 336-334-7716

Website: http://acitcenter.ncat.edu

North Carolina Agricultural and Technical State University

Explore. Discover. Become.

08/16/2013

ACIT Center Research Team Members/Collaborators

Dr. Abdollah Homaifar (NCAT)

Dr. Ruben Buaba (NCAT)

Dr. Gholamreza Fetanat (NCAT)

Chaunté Lacewell (Ph.D. student, NCAT)

M. Gorji-Sefidmazgi (Ph.D. student, NCAT)

Muhammad Sohail (Undergrad, NCAT)

Dr. Yuh-Lang Lin (NCAT)

Dr. Manoj Jha (NCAT)

Sayemuzzaman Mohammad (Ph.D. student, NCAT)

Dr. Eric Kihn (NOAA/NGDC)

Dr. Ken Knapp (NOAA/NCDC)

Dr. Alok Choudhary (NU)

Dr. William Hendrix (NU)

Dr. Stefan Liess (UMN)

Website: http://acitcenter.ncat.edu

Overview

Selected Research Areas

- Data mining
- Machine Learning
- Search Algorithms
- Modeling/Prediction

Outline

- Tracking of cloud clusters developing into tropical cyclones
- Objective tropical cyclone intensity estimation using satellite Images
- Nearest Neighbor Search in Large High-Dimensional Dataset
- Analyzing climatic time series, low frequency variability of climate

Tracking of Cloud Clusters Developing Into Tropical Cyclones

Chaunté Lacewell
(Ph.D. student)

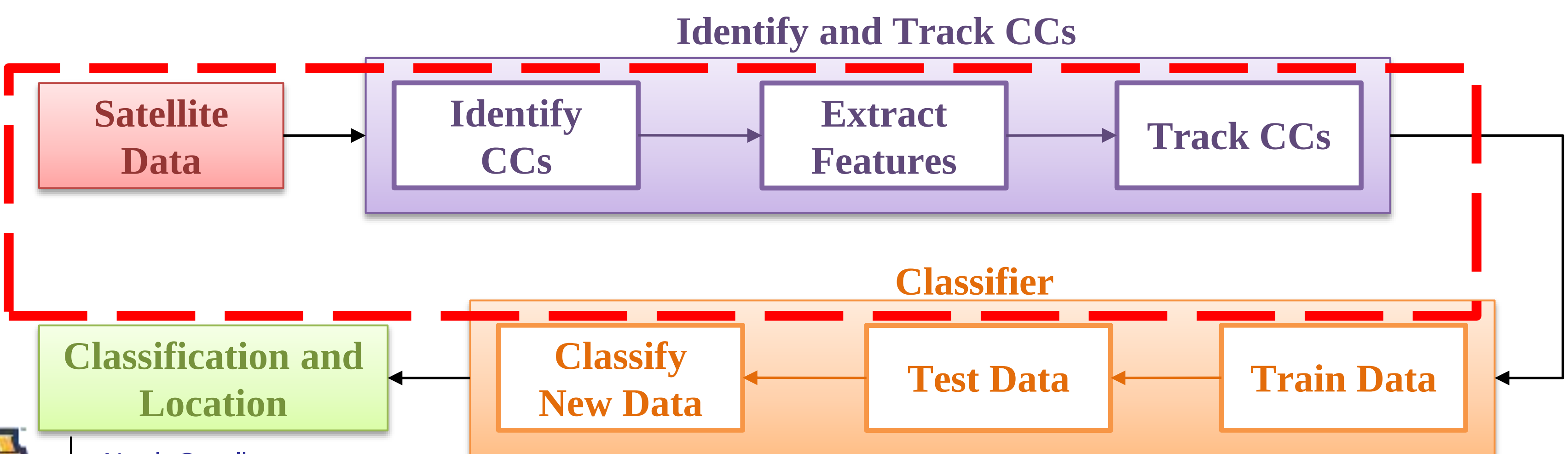
Introduction

- Forecasters need new techniques using pattern recognition to determine whether a tropical cyclone (TC) will develop from a loosely organized cluster of clouds.
- Refined observational data and forecasting techniques are not always available or accurate in areas in which data is sparse such as western North Africa.
- Prior studies have attempted to predict tropical cyclogenesis (TCG) using numerical weather prediction models and radar data.
- Global gridded satellite data are used and are readily available when advanced data are not available.

Methodology

- Extract information
  - » Objectively identify and track individual CCs
  - » Determine relevant features
- Classify information
  - » Evaluate data
  - » Classify developing (D) and non-developing (ND) CCs

Flowchart

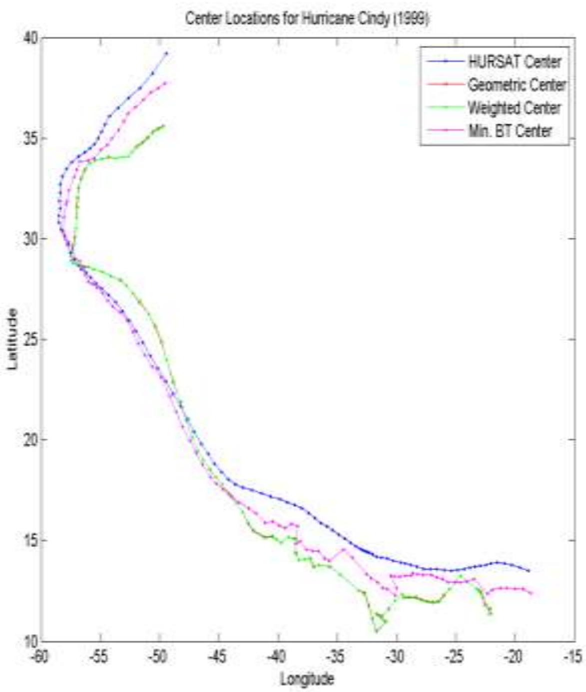


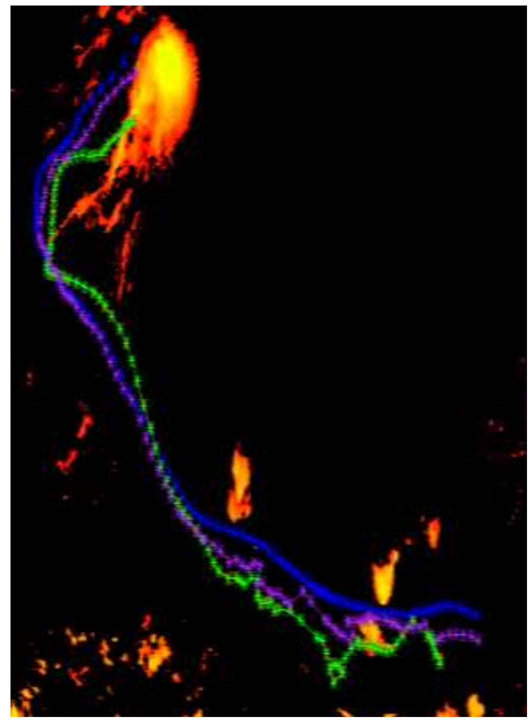
Methodology…

- Identification of CCs
  - » Brightness Temperature (BT) threshold (245 K)
  - » Radius of at least 1^{o} (111 km)
  - » CC area of at least 2,400 km^2
  - » Other features extracted such as geometric center, weighted center, min. BT location, average sea surface temperature, eccentricity, etc.

- Tracking of CCs
  - » Incorporates both area overlap and detect and spread methods
  - » Method is able to track splitting, merging, and continuance of CCs
  - » Uses a reward incorporating distance between CCs and overlapping area to determine track decisions

Results/Conclusion





- No “ground truth” data to verify identified and tracked CCs
  - » Use Hurricane Satellite (HURSAT) data to analyze the usefulness of tracking technique
  - » Method provides reasonable results when compared to actual tracks, which are completed by experts

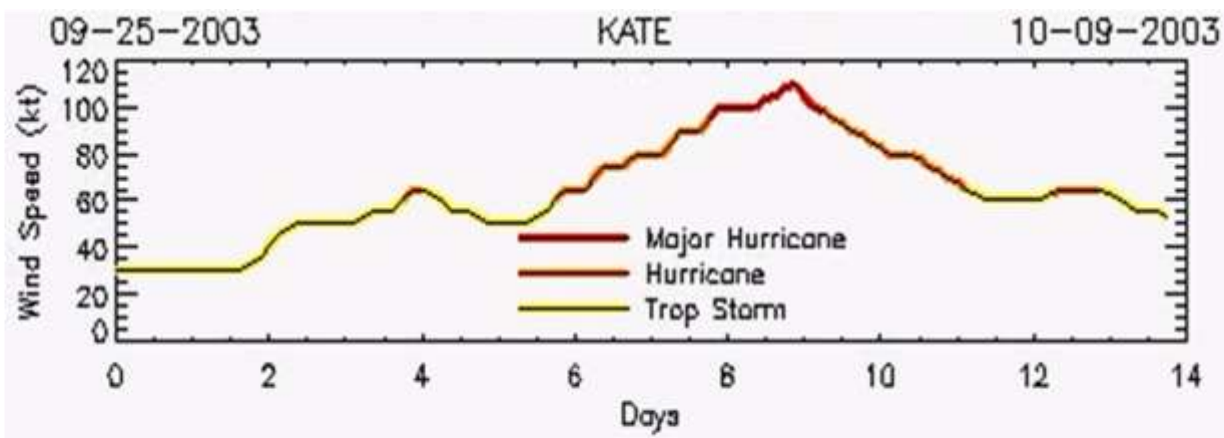
HURSAT center (blue), geometric center (red), weighted center (green), and minimum Brightness temperature location (magenta) of Hurricane Cindy (1999)

Objective Tropical Cyclone Intensity Estimation using Satellite Images

Gholamreza Fetanat (Ph.D.)

Introduction

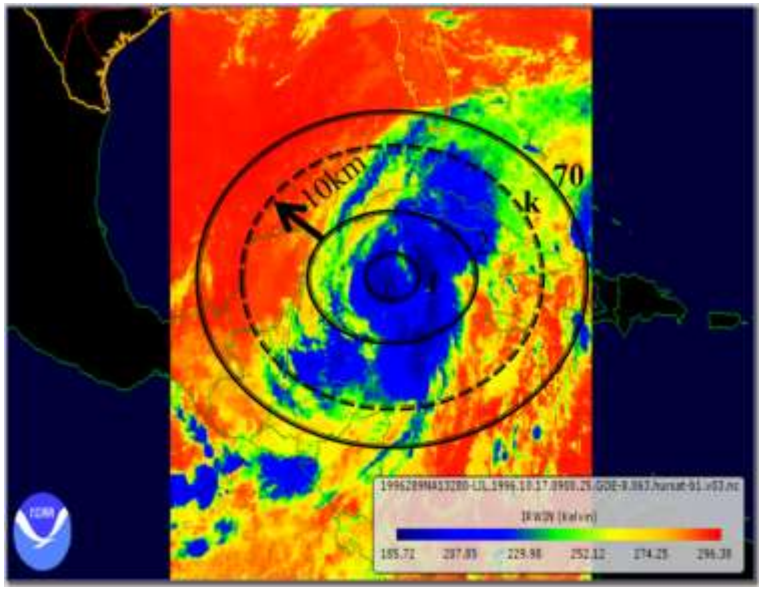
- Developing an objective method that provides quick and accurate Tropical Cyclone intensity estimation using satellite images
- A TC is a storm system characterized by large air masses circulating clockwise or counterclockwise
- The intensity of a TC is measured by the minimum sea level pressure (MSLP-mb) or surface maximum sustained wind speed (MSW-kt), which is defined as the one-minute wind speed average



Ref: Knapp, K. R. and J. P. Kossin, 2007: New global tropical cyclone data from ISCCP B1 geostationary satellite observations, *Journal of Applied Remote Sensing*, 1, 013505.

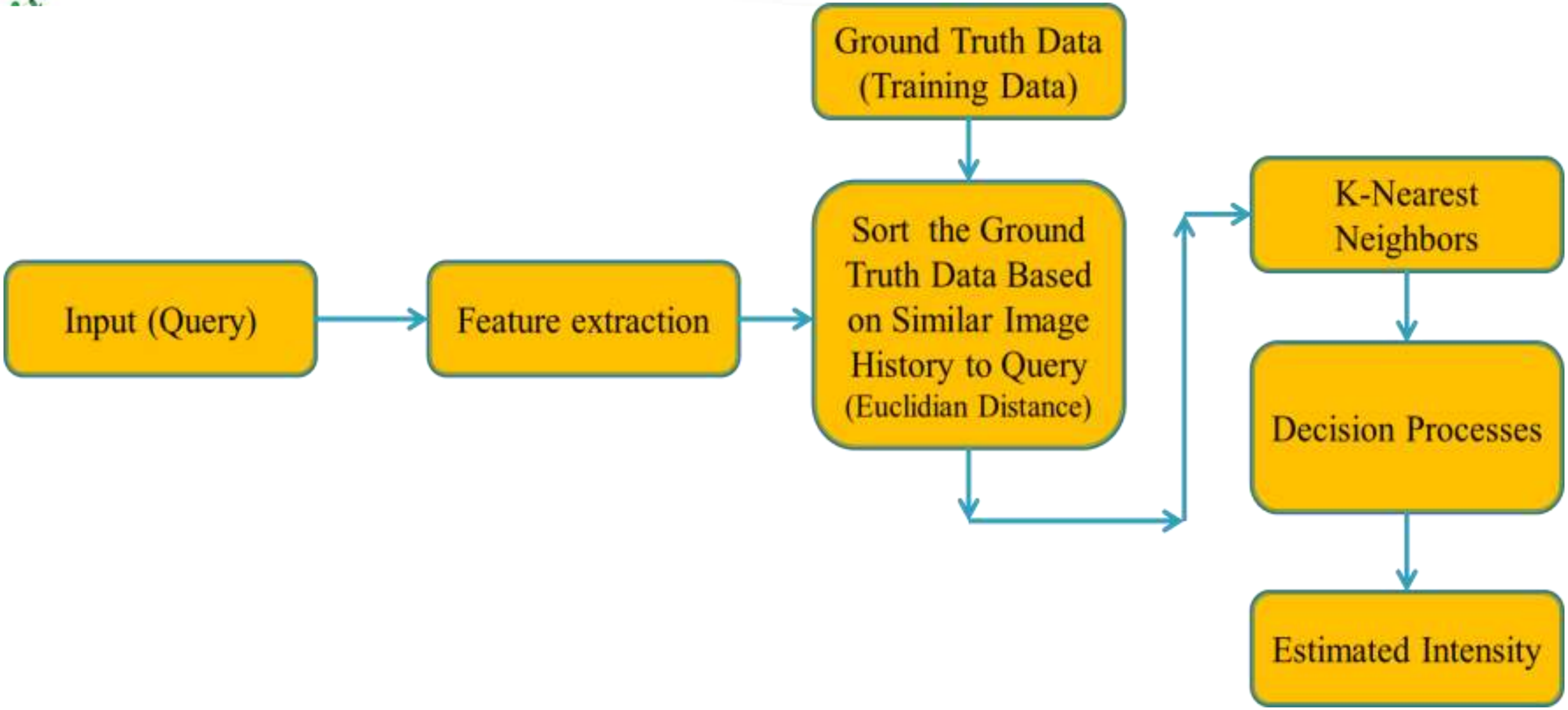
Approach Description
(Image Analysis)

- Features: (*current* and the preceding 6, 12, 24 hours’ images)
  - Images are described based on BTs mean and SD of the selected image rings (14 consecutive rings from the center of the storm) are used for comparison
  - Dimension: (4*2*14=112) attributes
  - (301X301) image described by 112 features



$$INT = f(g(x,y)_t, g(x,y)_{t-6}, g(x,y)_{t-12}, g(x,y)_{t-24})$$

Methodology



Feature Analogs in Satellite Imagery (FASI)

Results/ Conclusion

- RMSE: 4.6 knots compared to 11.7 knots by DT
- 50% of the estimates have MAE less than 2.4 knots, 75% are within 4.4 knots and 90% are within 7.5 knots
- The accuracy is competitive when compared to other objective methods (e.g. advanced Dvorak technique)
- Simplicity, objectivity and consistency of the proposed approach makes it an important tool for estimating the intensity of TCs compare to subjective DT
- Proposed technique very suitable for Hurricane Category 1-3
- Overall, 30% to 55% improvement achieved compared to the DT

Nearest Neighbor Search in Large High-Dimensional Dataset

Ruben Buaba (Ph.D.)

Motivation

- **“Big Data”**

- Data is growing rapidly!!!
- 2.5 quintillion (2.5×10^{18}) bytes of technological data created per day worldwide (per-capita)
- 90% of the world’s data has been created in the last two years alone
- Computers are not catching up
- Faster computing technique is crucial
- Algorithms must be able to run on PC
- Faster Information Retrieval is crucial
  - Saves lives, property, money, time, etc.

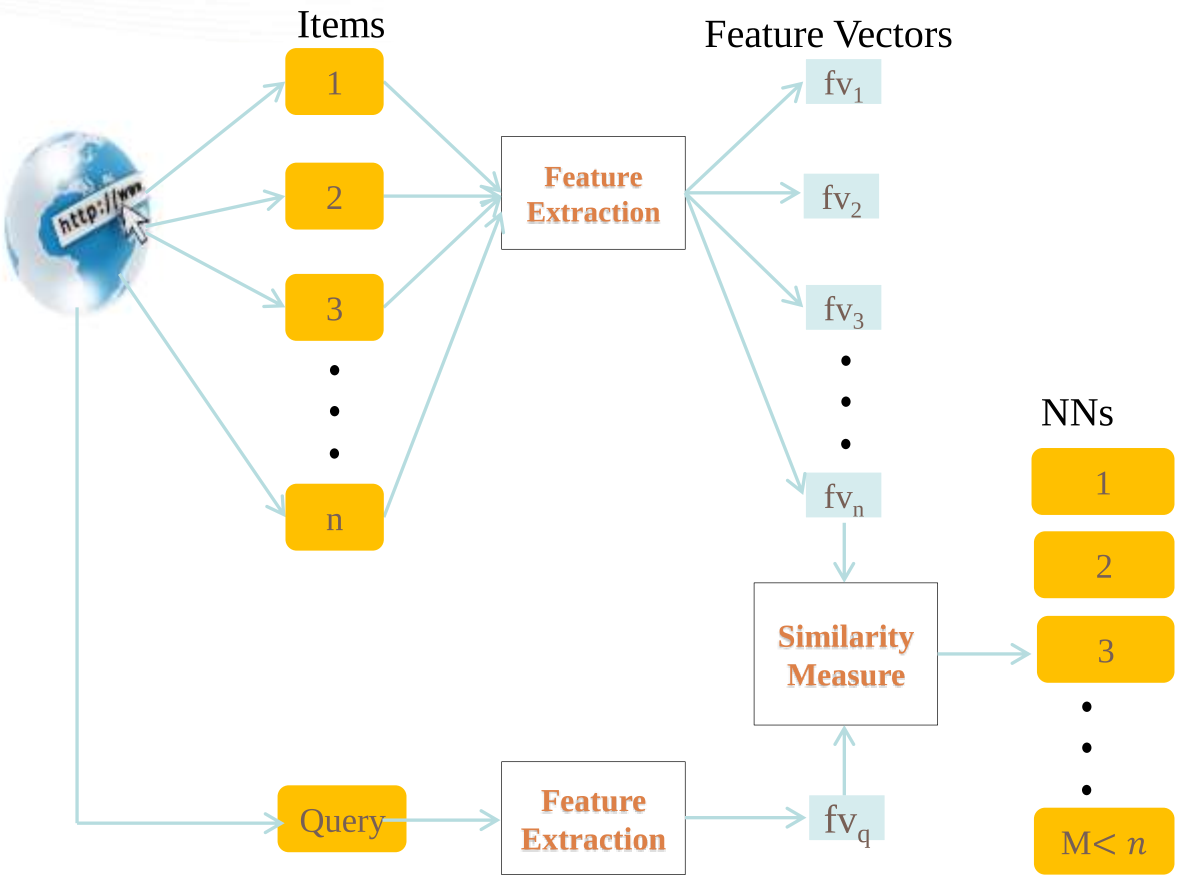
- **Areas Affected**

- Government
  - » In 2012, the Obama administration announced the Big Data Research and Development Initiative
- Science and Research
  - » Decoding the human genome originally took 10 years to process; now it can be achieved in one week
- Private Sector
  - » Walmart: more than 1 million customer transactions every hour (~2.5 petabytes of data)
  - » Facebook handles about 40 billion photos from its user database
  - » Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide

Problem Formulation

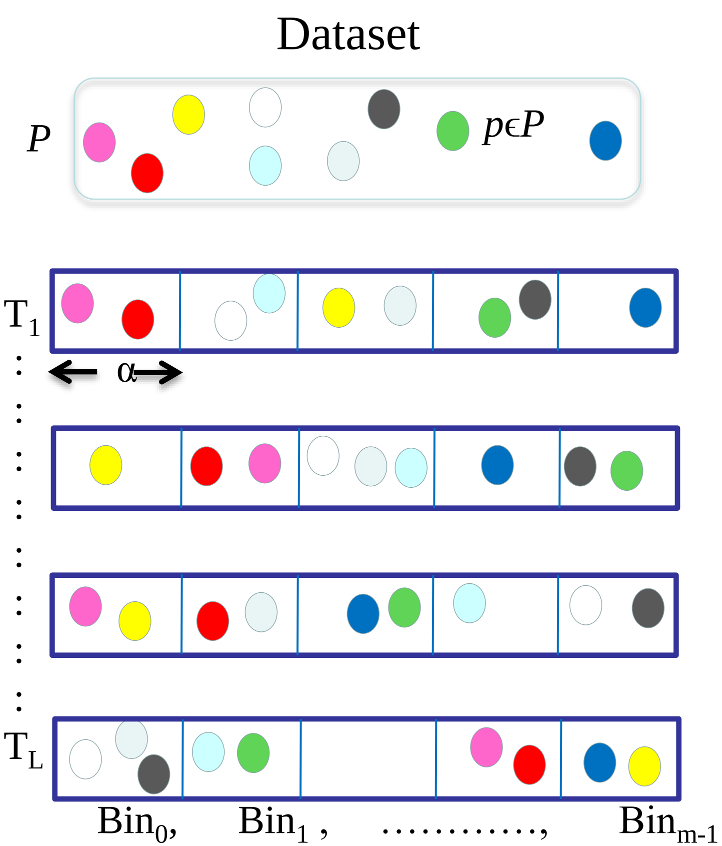
Techniques

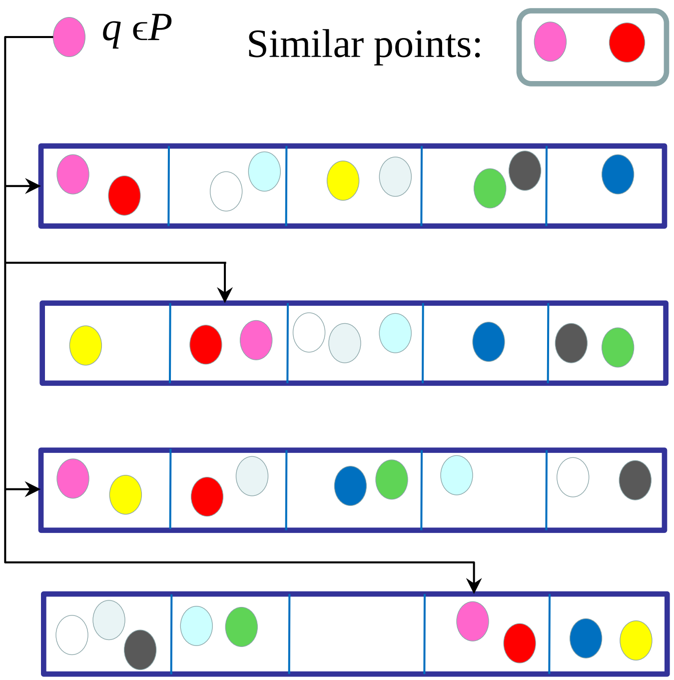
- Data structures
  - Nearest Neighbor Search
  - Scalable Algorithms
  - Trees
  - Locality Sensitive Hashing (LSH)
  - Small Binary Codes



Fast Locality Sensitive Hashing

Basic Concept



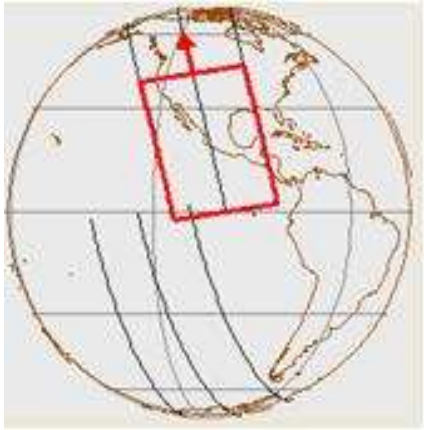


$U\{T_1Bin_0, T_2Bin_1, T_3Bin_0, T_4Bin_3\} = \{$ ● ● ● $\}$

Sample Dataset (1.6 million)

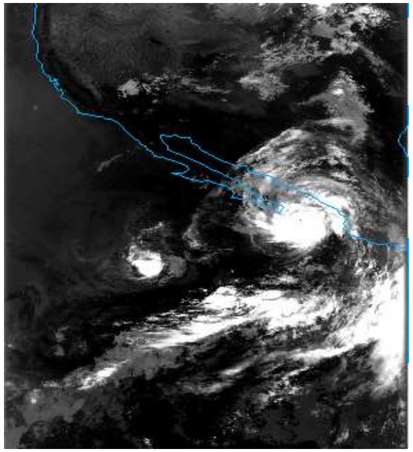
DMSP Satellite images

- Location



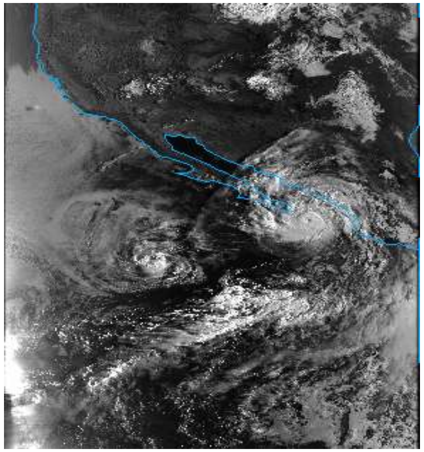
2.7 km ground sample distance

- Visual image



363 x 293 pixels

- Thermal image



363 x 293 pixels

» "Curse of dimensionality"

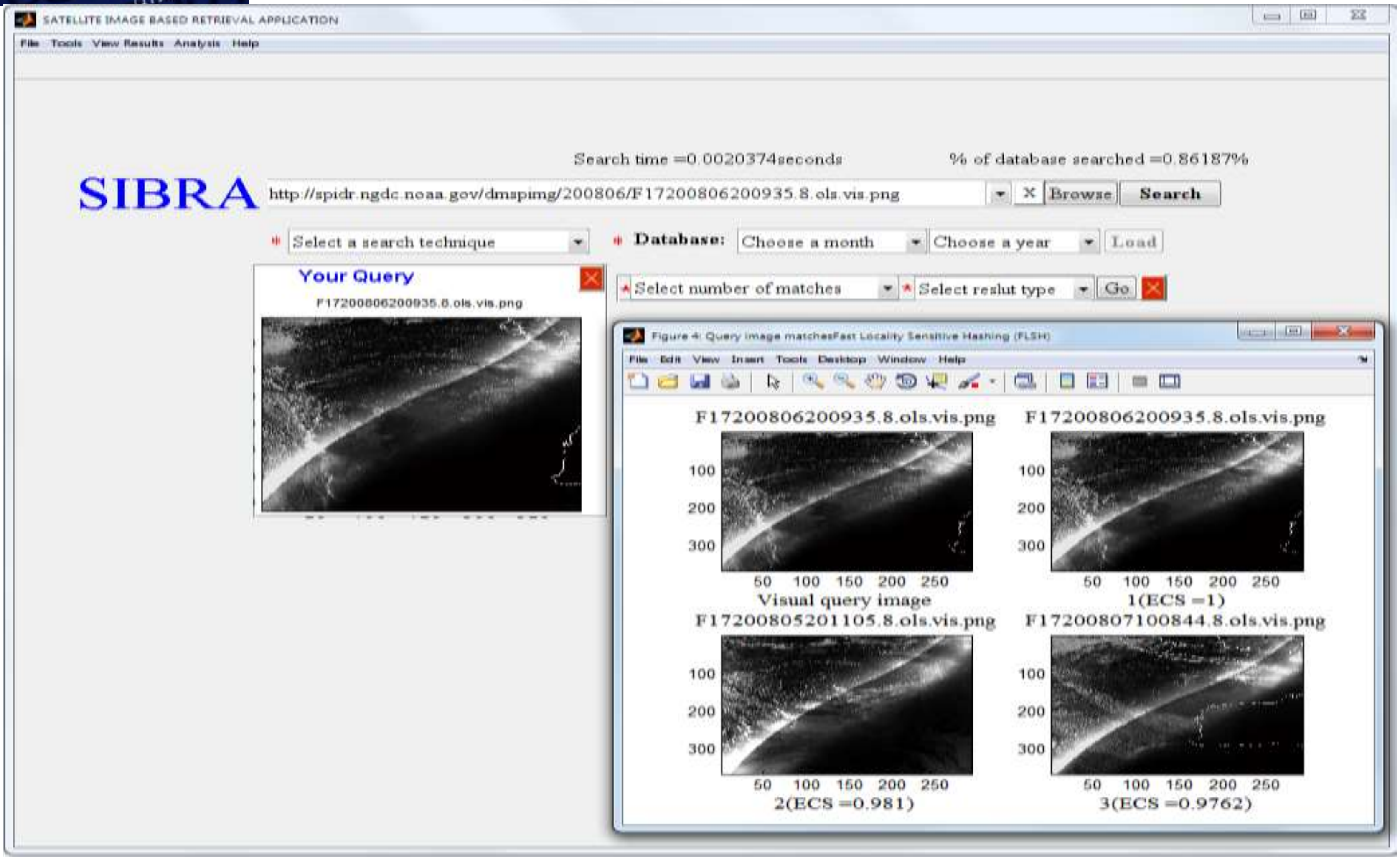
» 17k/month, approx. 200k/yr , (24 GB/yr)

» About 3.8 million images available (about 0.5 TB)

DMSP: **D**efense **M**eteorological **S**atellite **P**rogram

Developed Simulation Interface

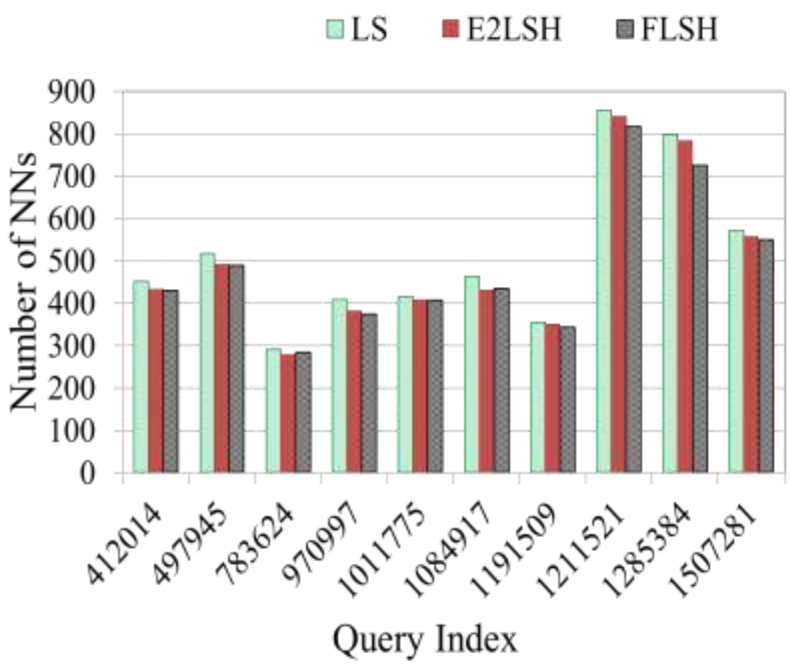
Download @:
http://acitcenter.ncat.edu/resources.html



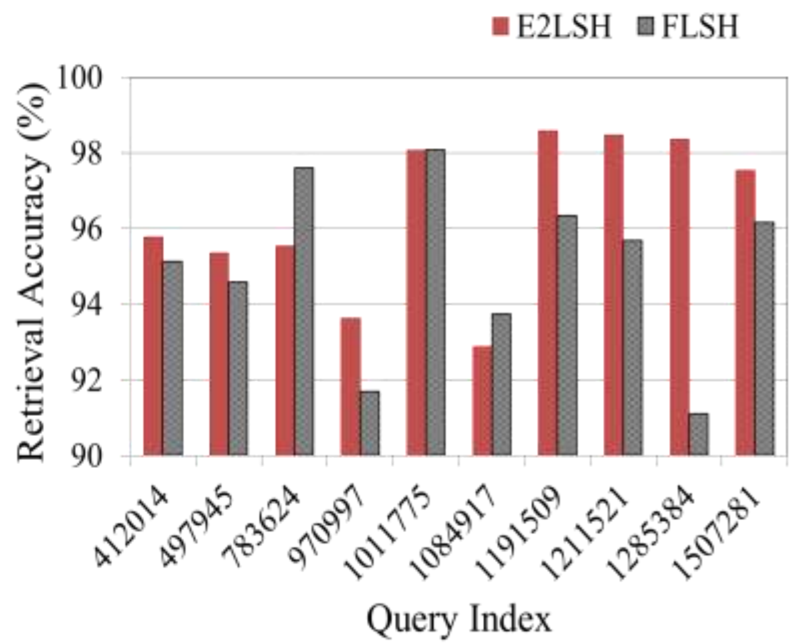
SIBRA: **S**atellite **I**mage **B**ased **R**etrieval **A**pplication

Results/Conclusion

NUMERICAL



PERCENTILE



LS: Linear Search
E2LSH: Exact Euclidean LSH (MIT)
LSH: Locality Sensitive Hashing

Pros:

- FLSH is twice as fast
- FLSH cuts computational complexity by 50%
- FLSH cuts memory complexity by 50%

Analyzing Temperature Regime/Trends During 1950-2010 in North Carolina

Mohammad Gorji-Sefidmazgi
(Ph.D. student)

Introduction

- Historical climate analysis is important to analyze and has societal, environmental and economical impact
- Analyzing spatiotemporal pattern in climate change across NC

Physical Process (Climate)

Time Series

a very complex and unknown dynamical model
$\dot{x} = f(x, u)$

Understand some information about the climate

Regime Analysis for Climatic Time Series

Regimes in the time series are hidden

Time series clustering →

low-frequency variability of climate

Time series is stationary in each regime (Local Stationarity)

→

Conventional clustering based on Hidden Markov Model and Gaussian Mixture Model

Time series is not locally stationary

→

Problem is more challenging
Developed new clustering method based on Finite Element

Finite Elements Methods for Time Series Clustering

Linear trend analysis is common in climate literature → Find climatic variables are rising or falling

Combining trend analysis and time series clustering → Find period of times when climatic variables are rising or falling

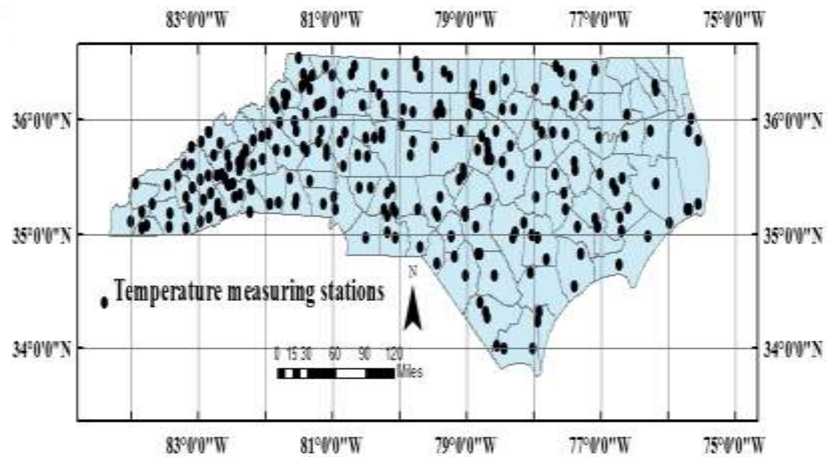
For each regime, assume the model of the time series

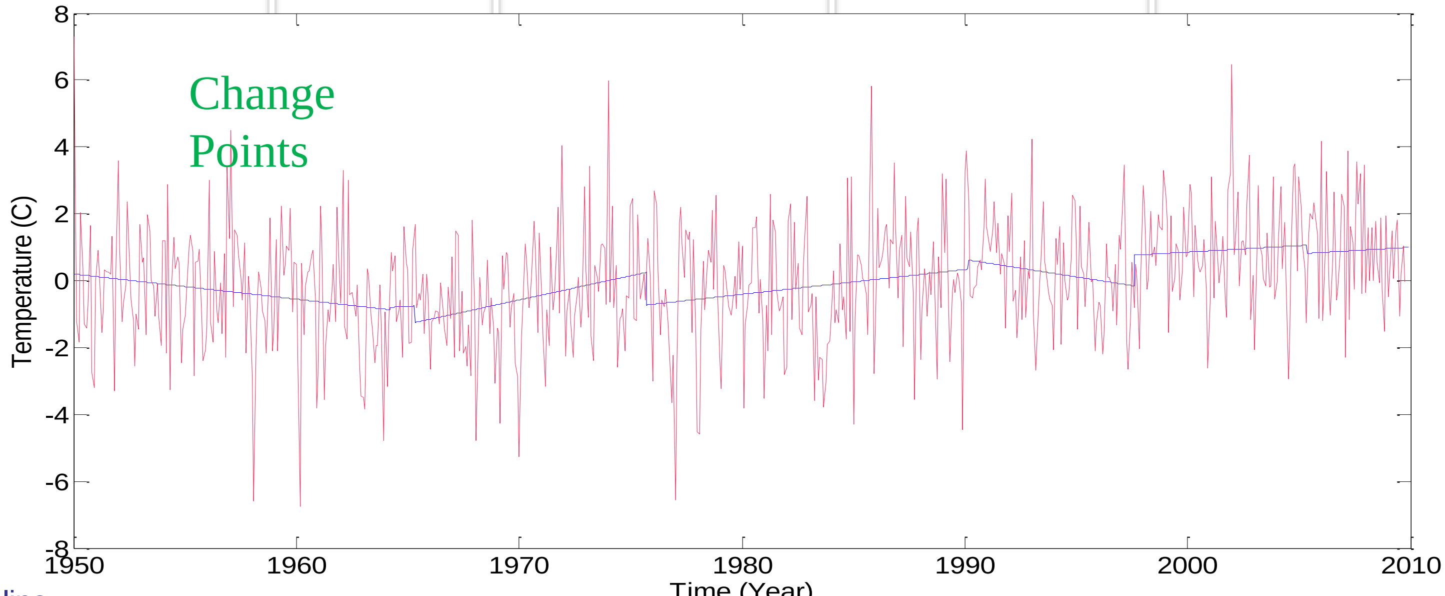
$$x(t) = \theta_0 + \theta_1 . t + \text{noise}$$

Is not locally stationary → Goal is to find switching times between unknown number of regimes

Results/Conclusion

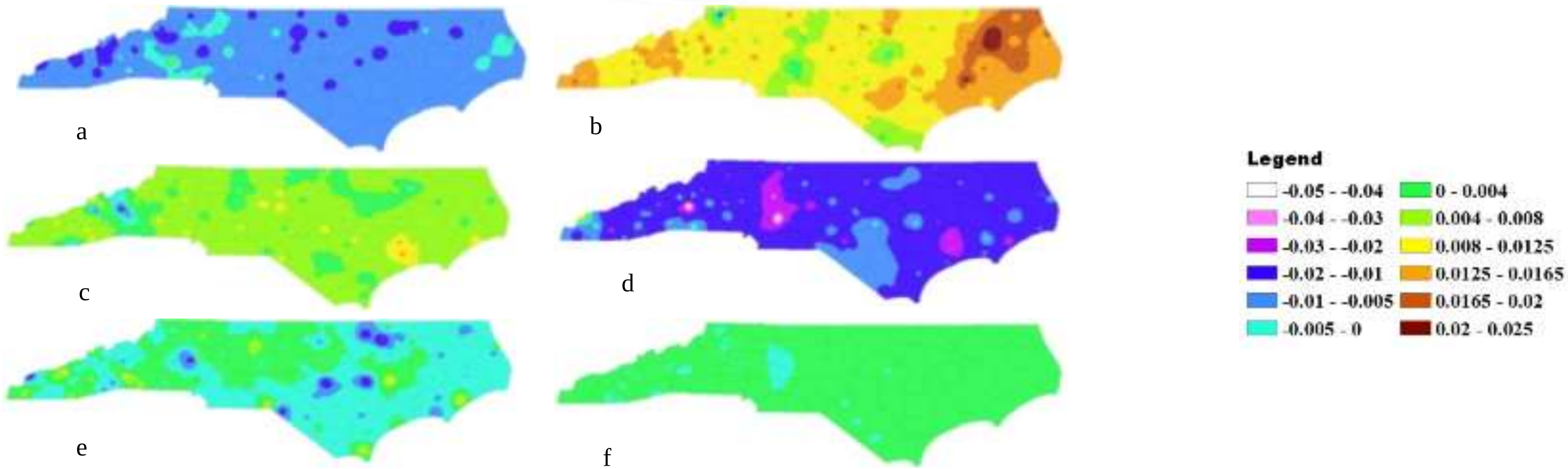
- Average temperature data gathered from 249 stations across North Carolina during 1950-2010
- Needs some pre-processing includes filling missing data and remove seasonality
- FEM finds regimes and trends at the same times. It also help us estimate optimal number of regimes.
- The trends are compared with important climatic indices, it





Results/Conclusion

Spatial distribution of trends in six regimes



Regime	Length	Average trend in NC (ºC per month)	Average temperature change in NC (ºC)	% change
Regime 1	1950 – 1965	-0.0077	-1.42	-9.80
Regime 2	1965 –1976	0.0107	1.34	9.10
Regime 3	1976 – 1990	0.0047	0.83	5.81
Regime 4	1990-1998	-0.0133	-1.20	-8.03
Regime 5	1998-2005	-0.0005	-0.05	-0.32
Regime 6	2005-2010	0.0011	0.06	0.39

Future Work

- Develop the image search algorithm into a web application
- Improve the algorithm for tracking the loosely cloud clusters that develop into TCs and develop CC classification algorithms
- FASI is in the process of transitioning to the operational phase (John Knaff will assist)
- Use data from the entire US or N America to investigate the usefulness of the proposed Finite Elements Methods for Time Series Clustering

Posters

- Tracking of Cloud Clusters Developing Into Tropical Cyclones
- Objective Tropical Cyclone Intensity Estimation using Satellite Images
- Nearest Neighbor Search in Large High-Dimensional Dataset
- Analyzing Temperature Regime/Trends During 1950-2010 in North Carolina
- Tracking Hurricane Eye Using Rotational Invariant Features


Abdollah Homaifar (PI.)


Gholamreza Fetanat (Ph.D.)


Ruben Buaba (Ph.D.)


Mohammad Gorji (Ph.D. student)


Muhammad Sohail (Undergrad)

Recent Publications (2013)

1. G. Fetanat, A. Homaifar and K. Knapp, “Objective Tropical Cyclone Intensity Estimation using Analogs of Spatial Features in Satellite Data”, Weather and Forecasting, June 2013

2. R. Buaba, A. Homaifar and E. Kihn, “Optimal Load Factor for Approximate Nearest Neighbor Search under Exact Euclidean Locality Sensitive Hashing”, International Journal of Computer Applications 69(21):22-31, May 2013. Published by Foundation of Computer Science, New York, USA.

3. G. Fetanat, A. Homaifar and K. Knapp , “Tropical cyclone intensity estimation from satellite images” , 12th AMS Annual Student Conference, 5–6 Jan. 2013, Austin, TX.

4. C.W., Lacewell, A. Homaifar, and Y.-L. Lin, “Tracing the origins and propagation of pre-tropical storm Debby (2006) mesoscale convective systems using pattern recognition and image fusion”, Meteorology and Atmospheric Physics, Volume 119,I ssue 1-2, pp 43-58, January 2013.

Recent Publications (2013) (submitted)…

5. C. W. Lacewell, A. Homaifar, Y.L. Lin, “Tracing the Origins and Propagation of Pre-Tropical Storm Debby Mesoscale Convective Systems using Pattern recognition and Image Fusion”, Meteorology and Atmospheric Physics, 2013.
6. R. Buaba, A. Homaifar, E. Kihn, “Approximate Nearest Neighbor Search-A New Perspective to Exact Euclidean Locality Sensitive Hashing”, ACM, Journal of Experimental Algorithmics, 2012.
7. M. Gorji-Sefidmazgi, M. Sayemuzzaman,A. Homaifar, M. Jha and S. Liess , “Finite Element Method Based Non-Stationary Time Series Clustering Method for Trend Analysis”, Elsevier Environment International, 2013
8. M. Gorji-Sefidmazgi, M. Sayemuzzaman, , A. Homaifar, M. Jha and S. Liess , “Quantifying Climate Trend using Non-Stationary Time Series Clustering Method in North Carolina”, Journal of Geophysical Research-Atmospheres, 2013.

Thank You!!!