Statistical Evaluations

Central Florida Coordination Area

October 27, 2010

Presented by Patrick Tara, P.E., Senior Engineer
Renee Murch, P.E., Engineer
INTERA

Statistical Evaluations

Objectives:

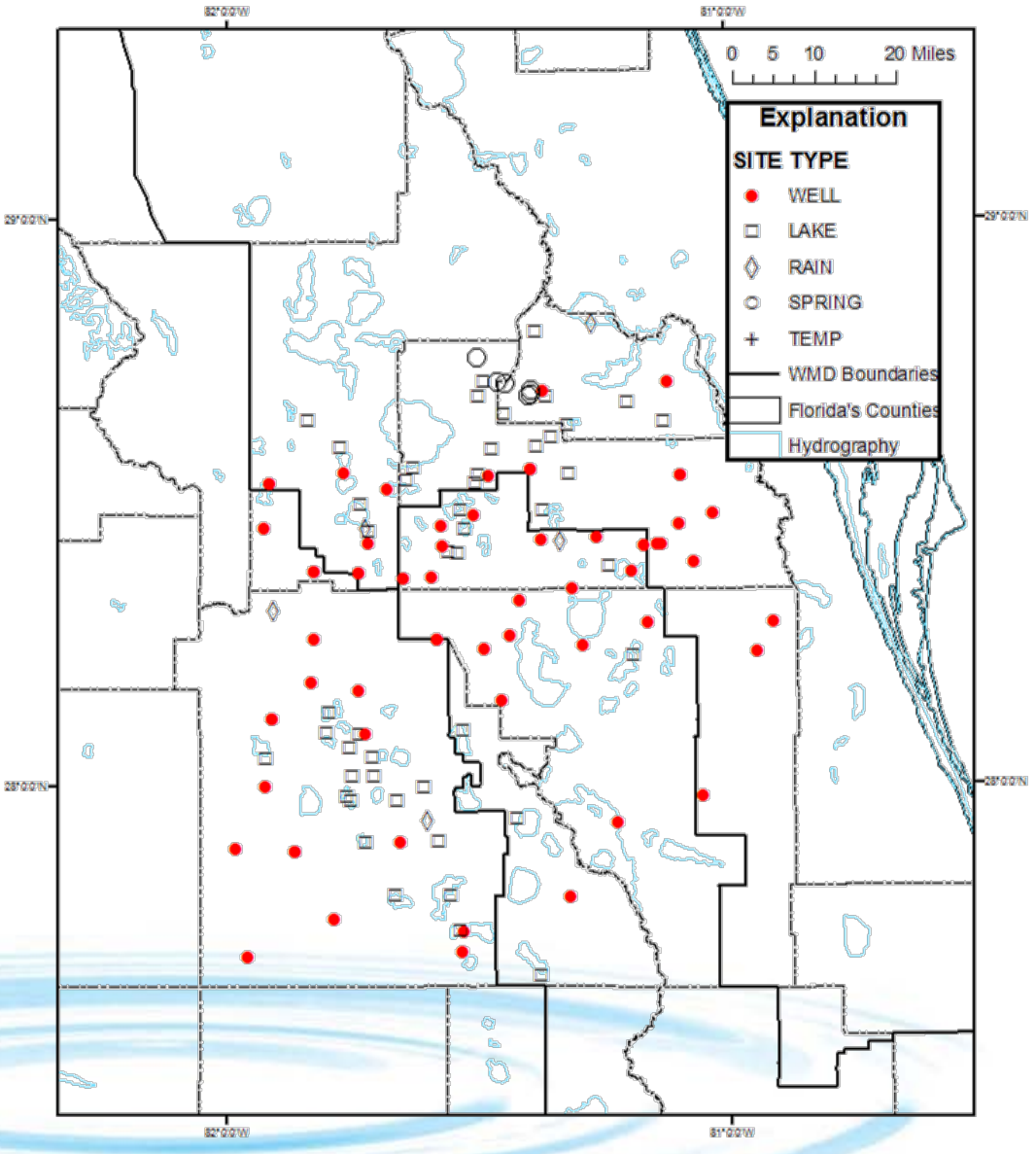
- Statistical and numerical models will provide separate lines of evidence to corroborate and improve confidence in each model's results.
- Multiple lines of evidence will yield a more robust characterization of the hydrologic system in the CFCA, ultimately leading to better management decisions.

Statistical Evaluations

- **Statistical Trends in Hydrologic Data** – evaluating trends at single sites and correlations between sites.
  - Exploratory Data Analysis
  - Trend Analysis
  - Cluster Analysis
- SPLUS scripts previously developed by INTERA were utilized for several phases.

Hydrologic Database

- Compiled from SJRWMD, SWFWMD, and USGS databases
- 120 sites: 62 wells, 6 springs, 47 lakes, and 5 rain gages

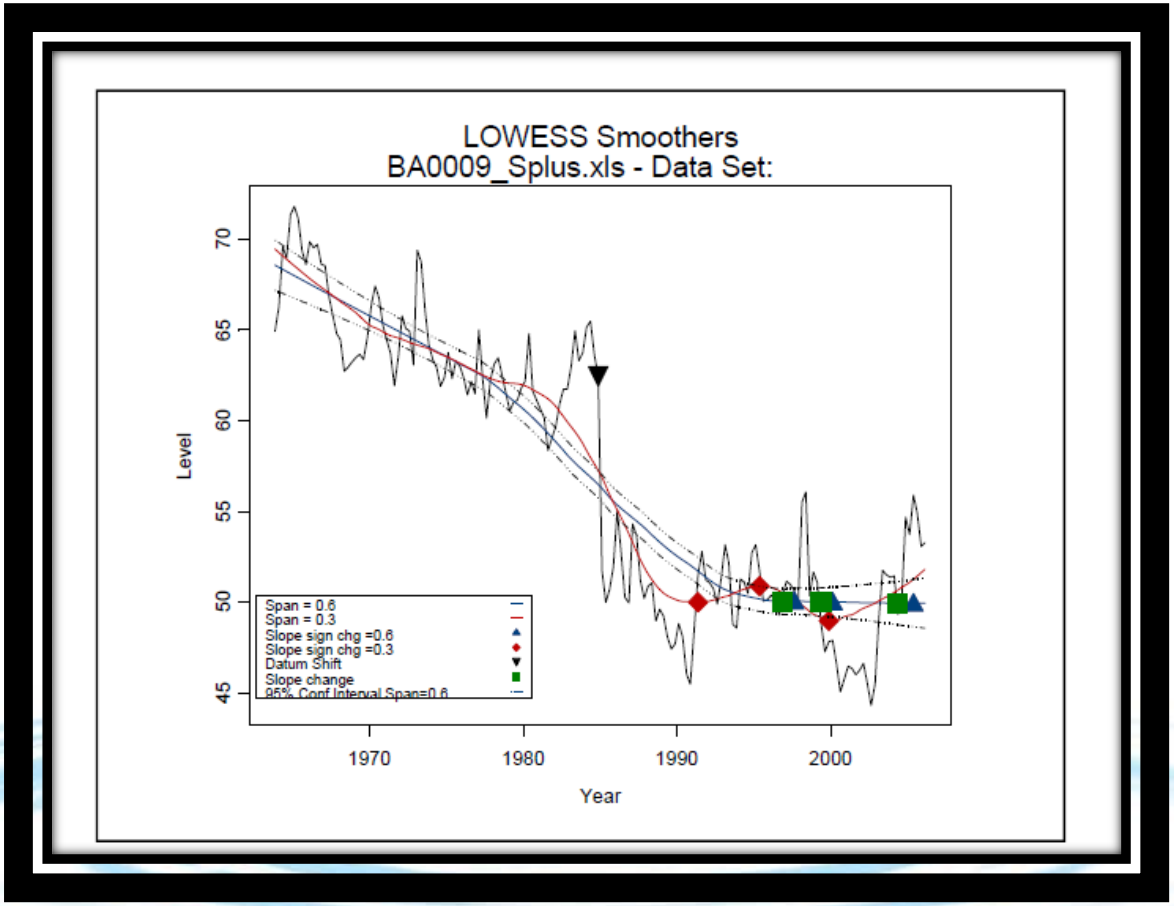


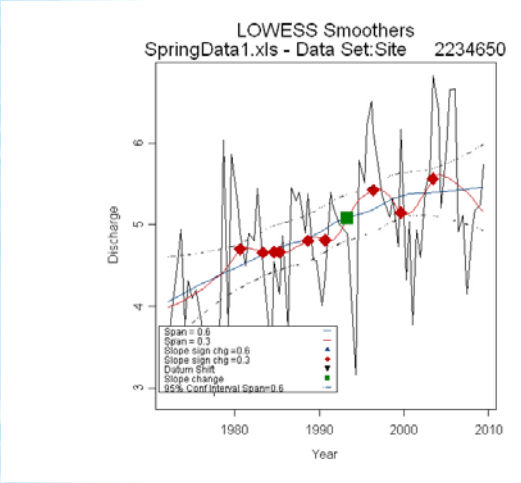
Exploratory Data Analysis

- To describe and summarize the data,
- To compile summary statistics for each data set,
- To develop locally weighted scatter-plots (LOWESS plots) for each data set, and
- To determine the appropriate LOWESS breakpoints to utilize for further analysis.

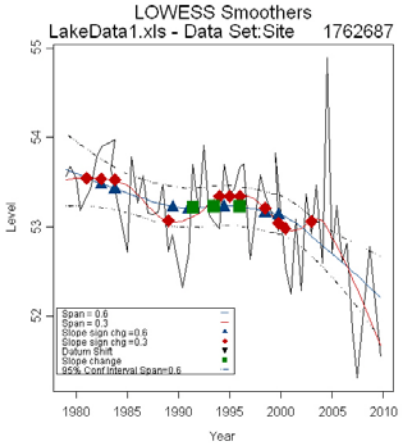
LOWESS

- Identifies break points in the time-series
- Break points - times in the series when the slope of the trend changes sign (e.g., from increasing to decreasing, from mild to steep, etc.).

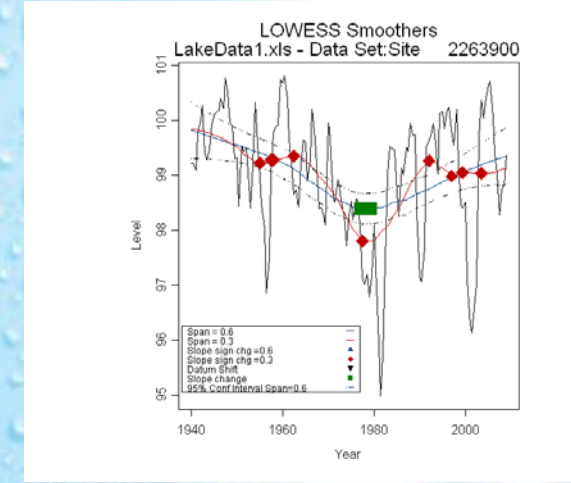




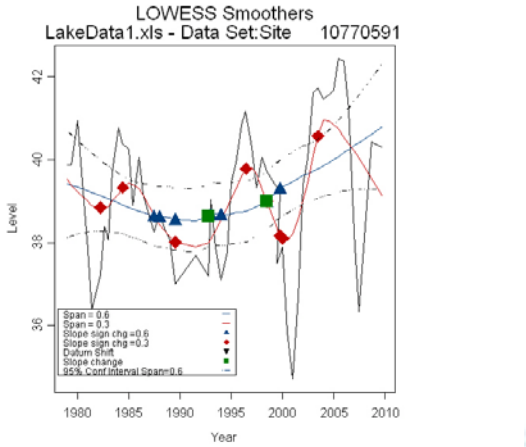
Monotonic (M)



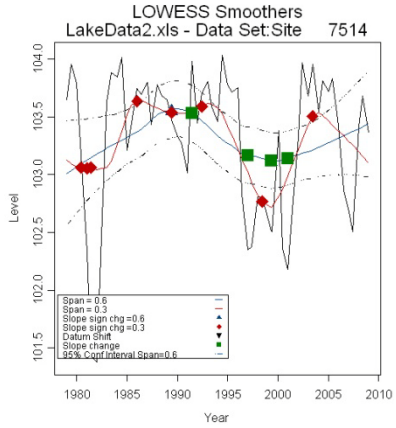
Monotonic with Slope Change (MS)



Piecewise (P)



Piecewise (P)



Double Piecewise (2P)

EDA Results

Trend Type	Abbreviation	Number of Stations
Monotonic	M	37
Monotonic with Slope Change	MS	20
Piecewise	P	43
Double Piecewise	2P	20
Total		120

Trend Analysis: Hypothesis Testing

1. Choose statistical test,

2. Setup null and alternative hypotheses,
   ("There is no trend in the data with time.")

3. Select the appropriate significance level and critical p-value (α),
   (80% significance level, 2-tailed critical p-value of 0.1)

4. Compute the test statistic and the p-value,
   (Compare p-value to 0.1)

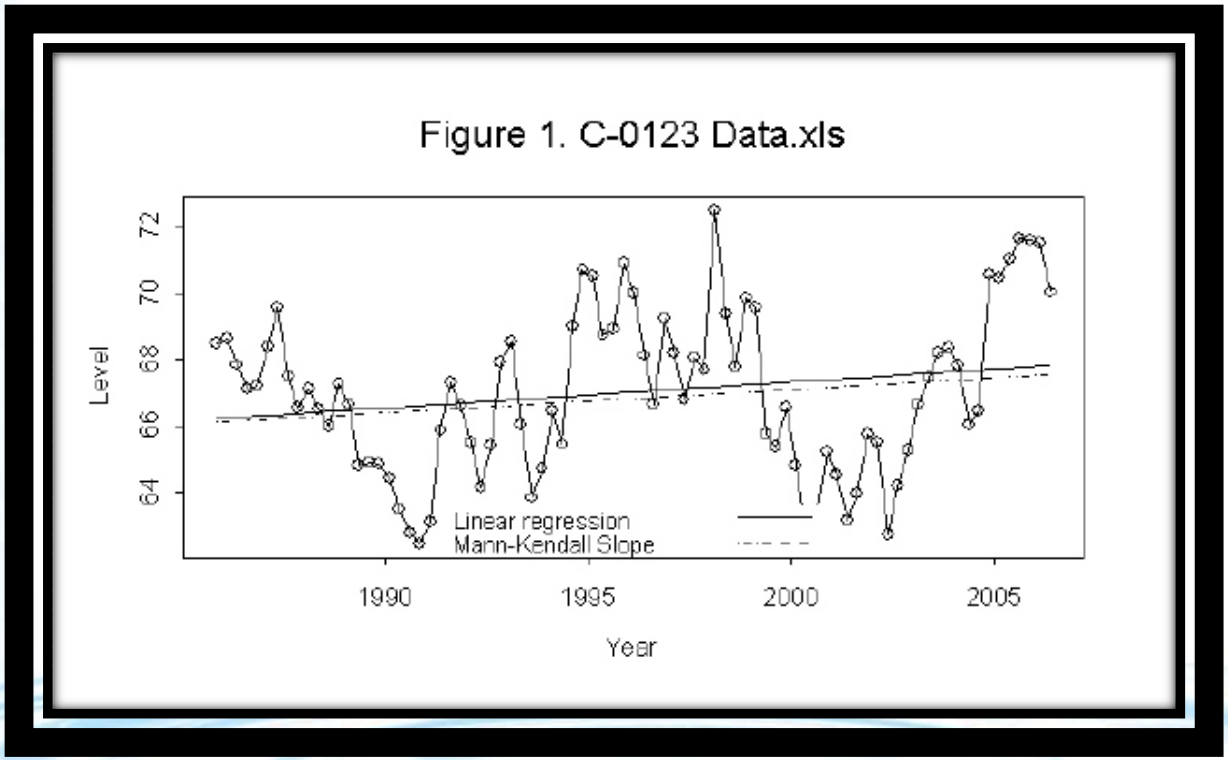
5. Determine test conclusion:
   - if p < α: Reject Ho => "There is a trend in the data."
   - otherwise: Fail to reject Ho

Trend Analysis: Mann-Kendall

Non-Parametric Test

Versus linear regression: Weights outliers less,

and detects all monotonic trends, not just linear trends



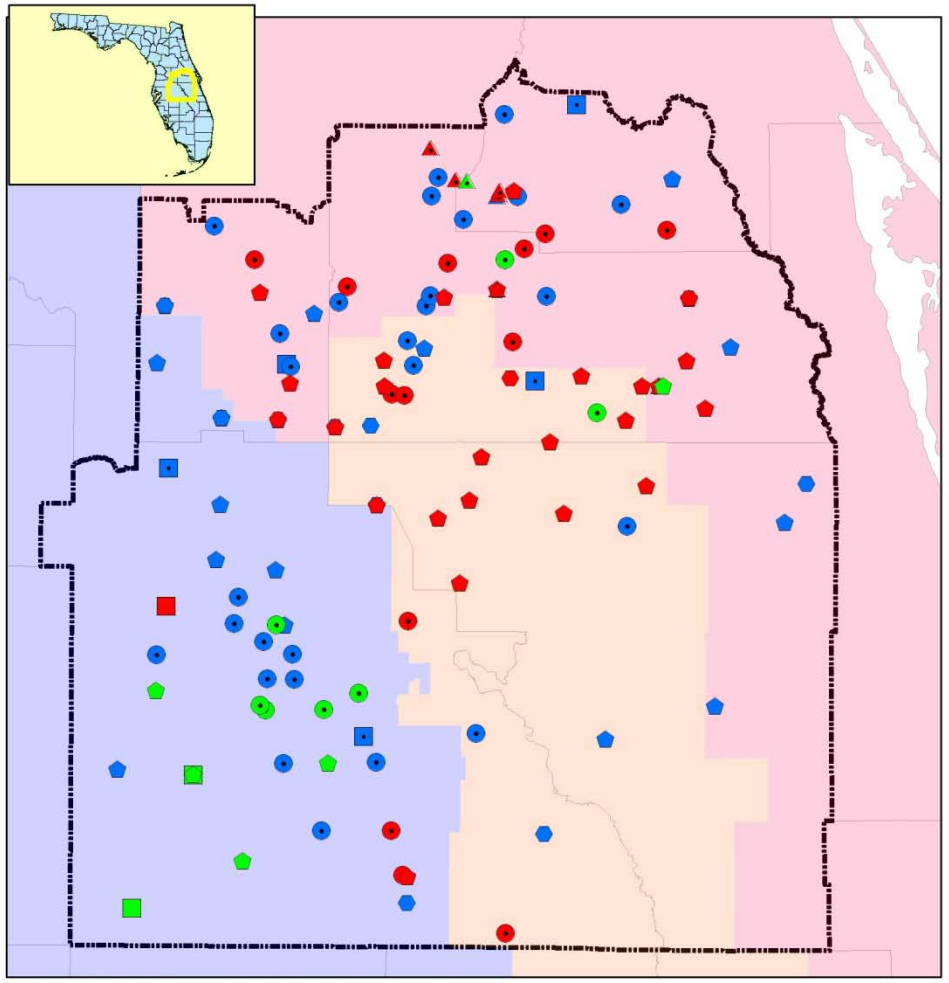
Trend Analysis Tests

Trend Type	Number of Break Points	Trend Tests Performed
M		Trend single period Trend seasonal single period
MS and P	1	Trend single period Trend seasonal single period Trend piecewise Trend seasonal piecewise CDF Compare
2P	2	Trend single period Trend seasonal single period Trend single period for each segment

Period of Record Trends

(Statistically Significant)

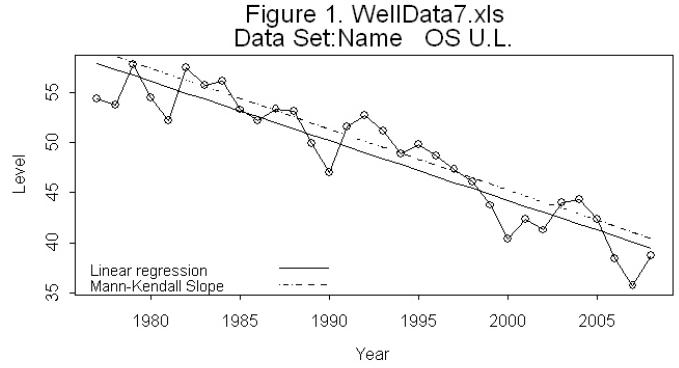
Site Type	Increasing Trends	Decreasing Trends
GW_IAS	2	1
GW_LFA	0	2
GW_SAS	0	4
GW_UFA	5	24
LK	7	13
SP	1	4
Total	15	48

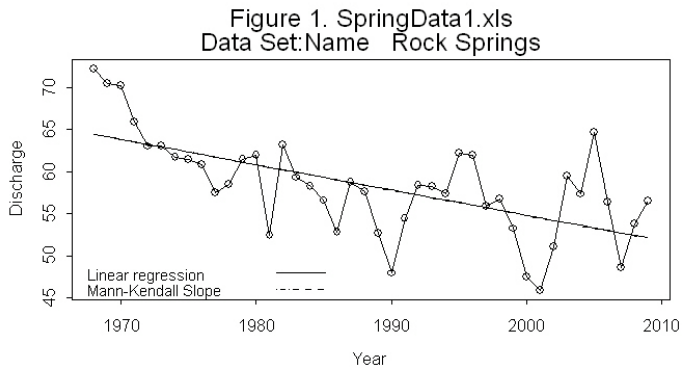


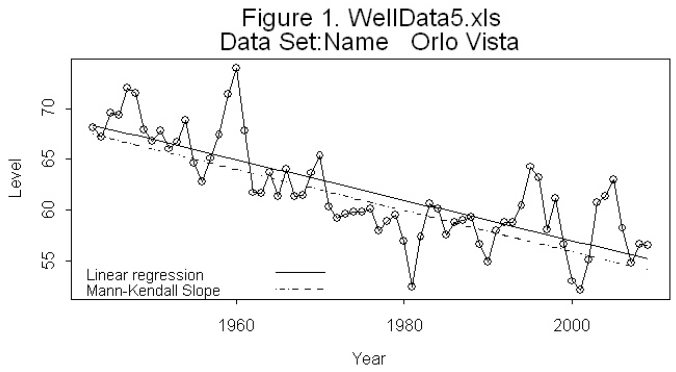
Decreasing Trends

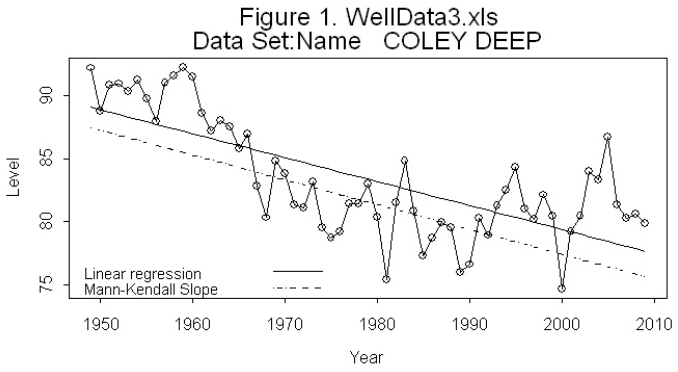
Site Name	Start Date	EndDate	MK p-value	Magnitude of Change	Unit
OS U.L.	5/4/1977	9/16/2008	0.0000	-19.060	ft
Orlo Vista	8/1/1943	4/28/2009	0.0000	-13.318	ft
Rock Springs	10/11/1968	8/17/2009	0.0000	-12.343	cfs
COLEY DEEP	11/18/1949	11/4/2009	0.0000	-11.783	ft
Wekiwa Springs	10/16/1968	8/20/2009	0.0013	-10.326	cfs
Longwood	10/25/1951	5/11/2009	0.0000	-9.958	ft
Cocoa D	7/31/1968	5/11/2009	0.0763	-9.232	ft
Bay Lake nr Windermere	3/1/1966	5/11/2009	0.0000	-8.311	ft
Cocoa C - Zone 1	2/24/1967	2/2/2009	0.0000	-8.053	ft
CROOKED LAKE NR BABSON PARK (R)	4/29/1945	10/27/2009	0.0000	-7.994	ft
Shingle Creek nr Kissimmee	5/3/1978	3/27/2009	0.0000	-7.841	ft
Horsehead Pond - SAS	1/8/1984	1/29/2009	0.0000	-6.671	ft
Cocoa P	3/5/1971	5/12/2009	0.0000	-5.689	ft
Mercantile Lane nr Kissimmee	5/7/1977	3/27/2009	0.0000	-5.669	ft

Decreasing Trend Examples





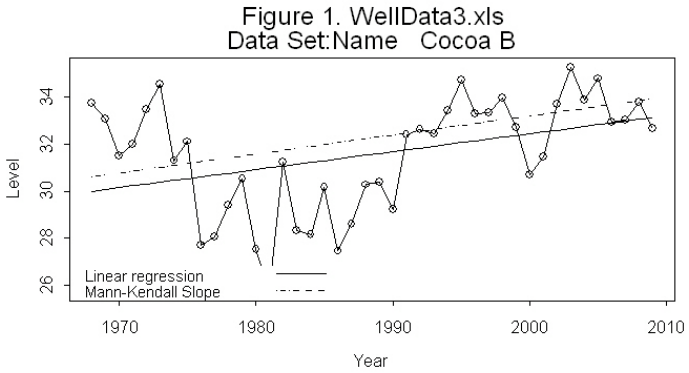


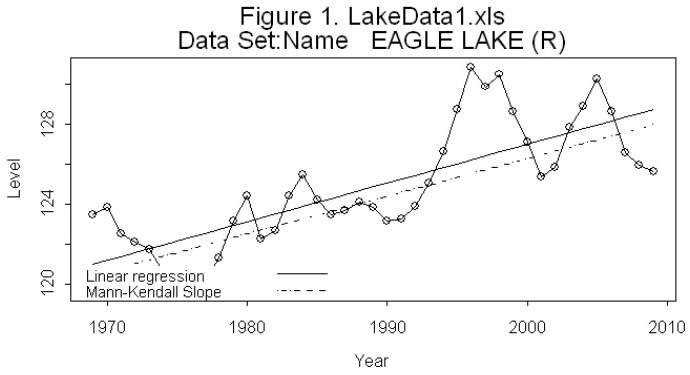


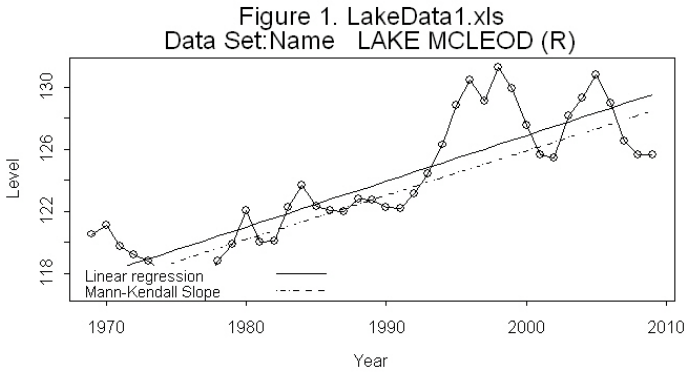
Increasing Trends

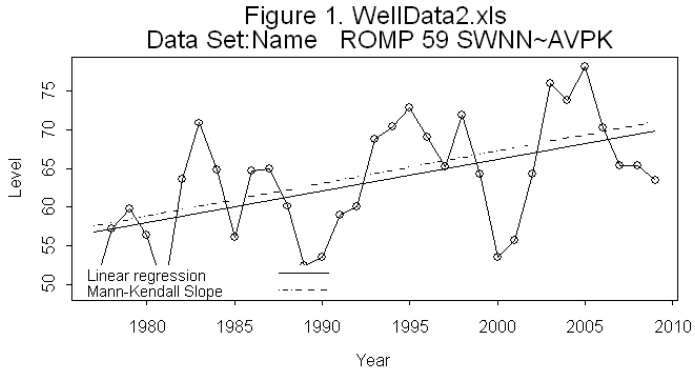
Site Name	Start Date	EndDate	MK p-value	Magnitude of Change	Unit
Killarney	7/1/1959	10/3/2008	0.015	0.376	ft
Whip-Por-Will	8/1/1960	1/6/2009	0.020	0.529	ft
Miami Springs	3/28/1972	8/17/2009	0.000	1.571	cfs
LAKE ANNIE (R)	8/21/1970	10/29/2009	0.007	3.097	ft
Cocoa B	7/31/1968	5/11/2009	0.007	3.302	ft
LAKE RUBY (R)	10/2/1971	10/29/2009	0.000	3.518	ft
LAKE ALFRED (R)	3/30/1961	10/19/2009	0.005	4.341	ft
STATE ROAD 60 DEEP NR LAKE WALES	9/18/1975	9/18/2008	0.042	5.871	ft
ROMP 59 HTRN	2/2/1977	10/27/2009	0.016	7.911	ft
EAGLE LAKE (R)	3/10/1965	10/29/2009	0.000	8.416	ft
ROMP 45 AVPK	8/21/1980	11/4/2009	0.020	11.233	ft
FORT GREEN SPRINGS INT	8/31/1964	10/3/2008	0.006	11.637	ft
SANLON RANCH FLDN	1/10/1970	10/27/2009	0.000	11.915	ft
LAKE MCLEOD (R)	3/13/1965	10/29/2009	0.000	12.708	ft
ROMP 59 SWNN~AVPK	9/10/1976	10/27/2009	0.004	13.975	ft

Increasing Trend Examples

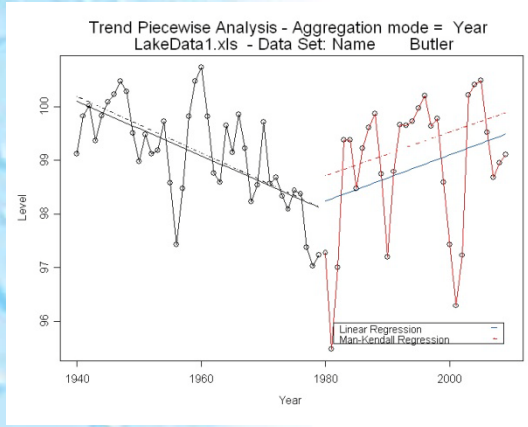




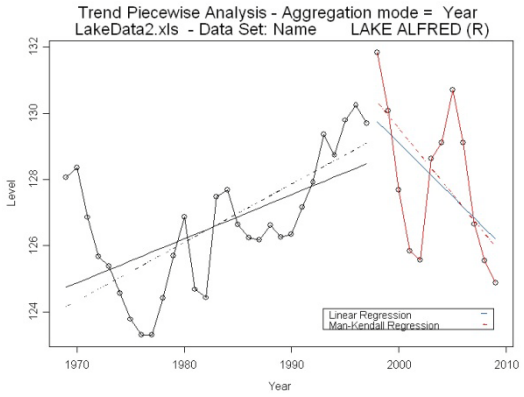




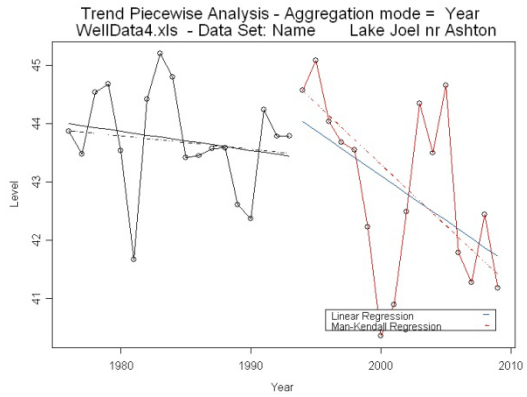
Non-Monotonic Trend Analysis



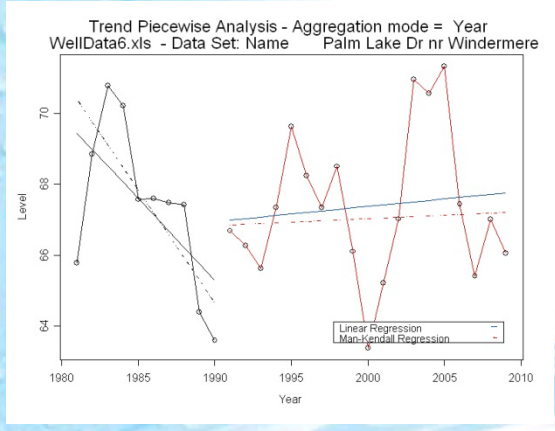
Trend, Trend



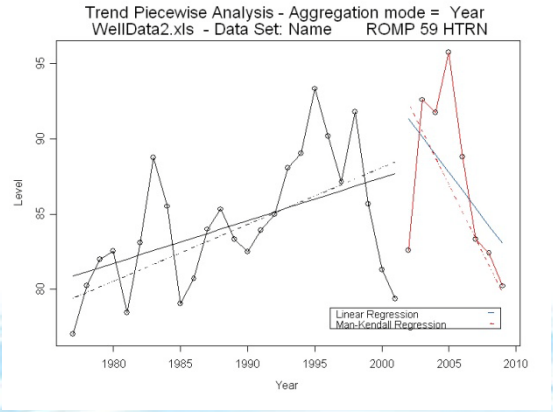
Trend, Trend



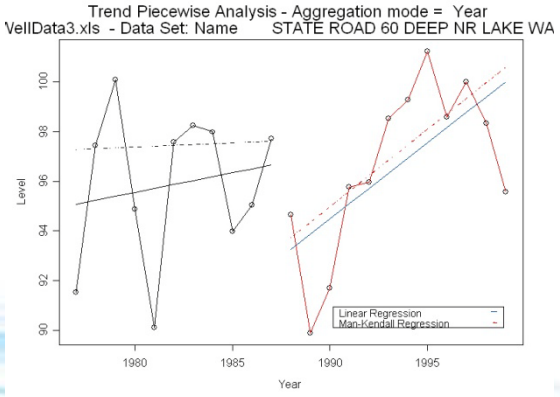
No Trend, Trend



Trend, No Trend

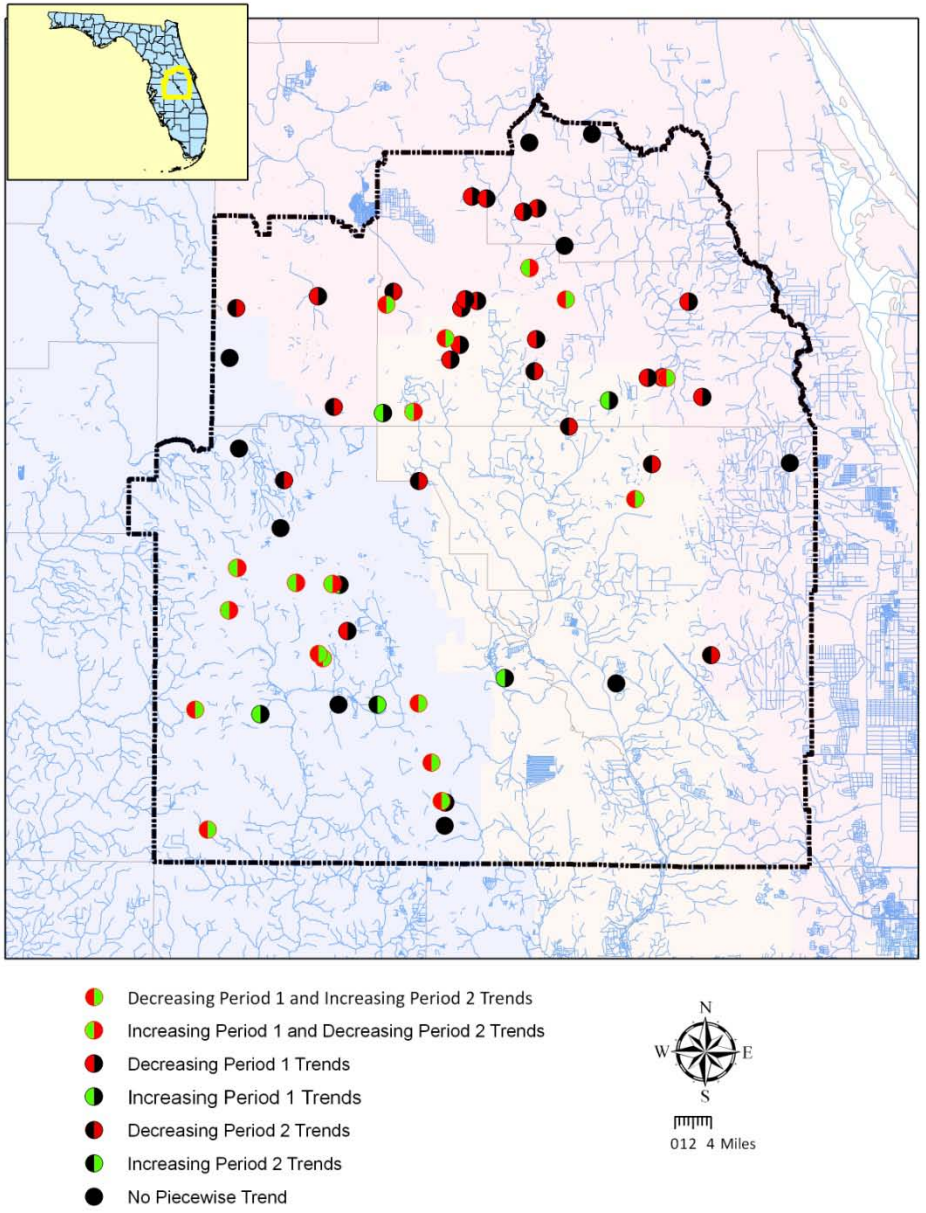


Trend, No Trend
(Segment 2 p = 0.1078)



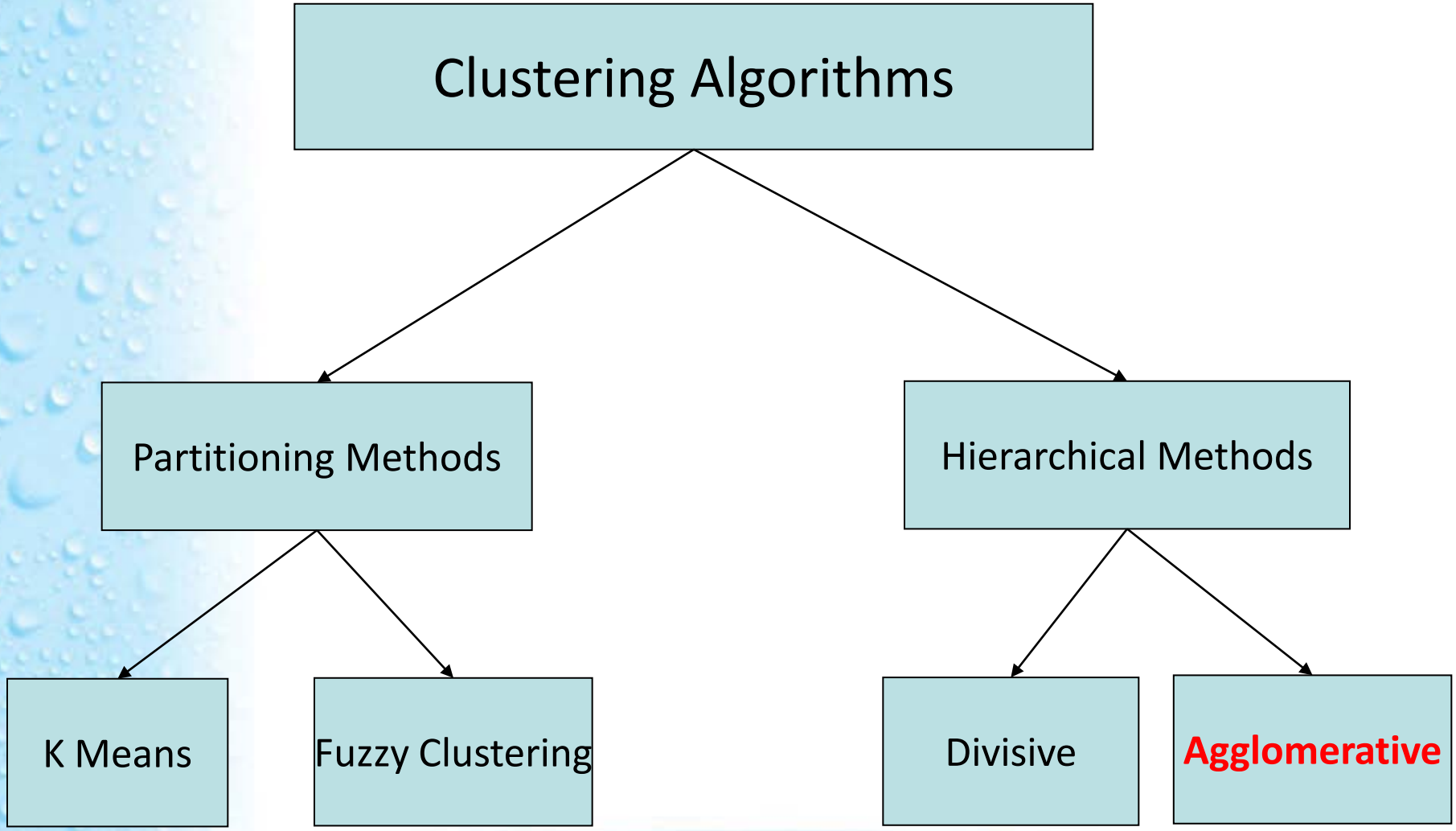
No Trend, Trend

Non-Monotonic Station Results



Cluster Analysis

- In cluster analysis, we search for groups (clusters) in the data in such a way that objects belonging to the same cluster resemble each other.
- A data set for clustering can consist of either rows of observations, or a dissimilarity object (a measure of dissimilarities between observations).
- Location of stations is examined in conjunction with results.


Clustering Algorithms
Partitioning Methods
Hierarchical Methods
K Means
Fuzzy Clustering
Divisive
Agglomerative

- For CFCA- data was clustered based on normalized average over pre-defined time periods.

$$z_{annual} = \frac{x_{annual} - \bar{x}_{poa}}{s_{poa}}$$

- No missing data
- Cluster analysis period: 1984 – 2008

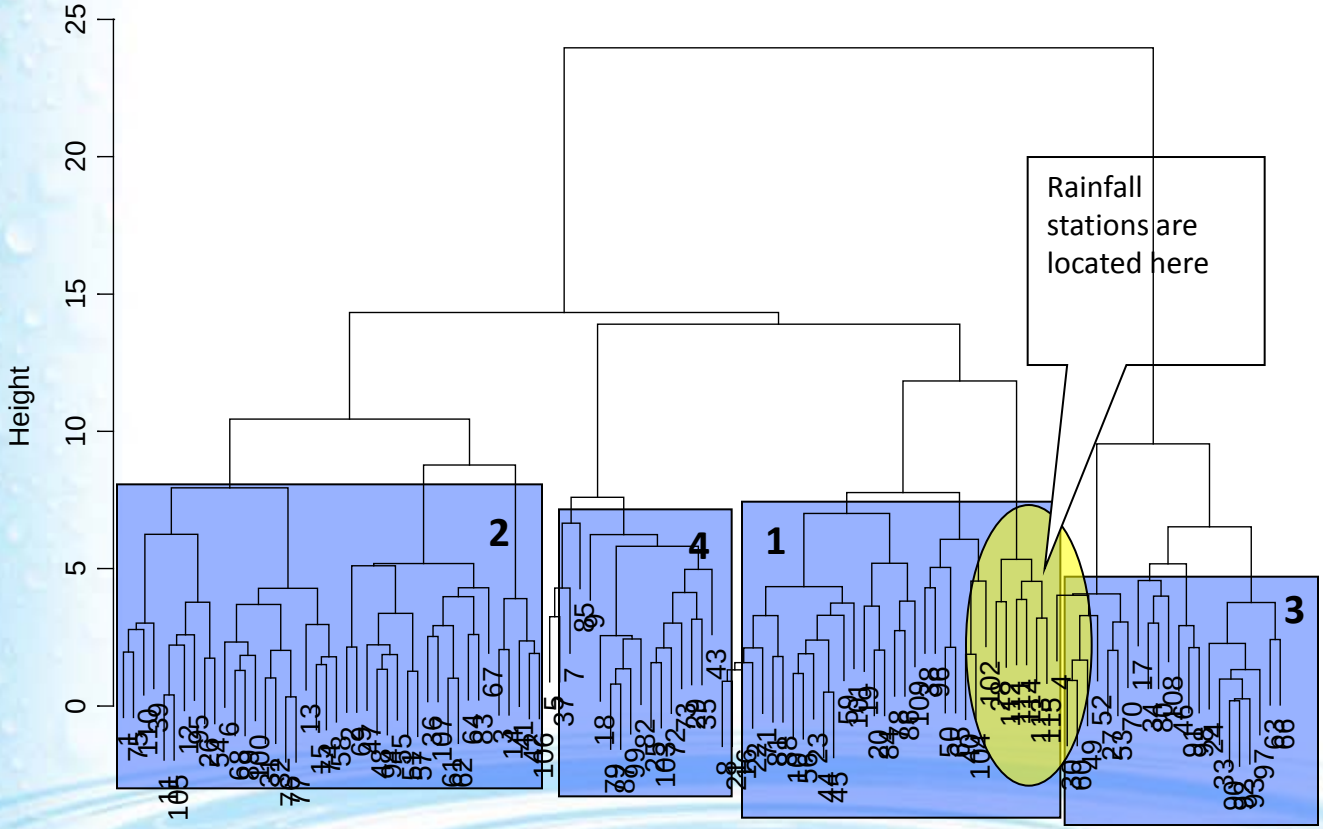
Need for Data Normalization

Cluster Analysis: CFCA Application

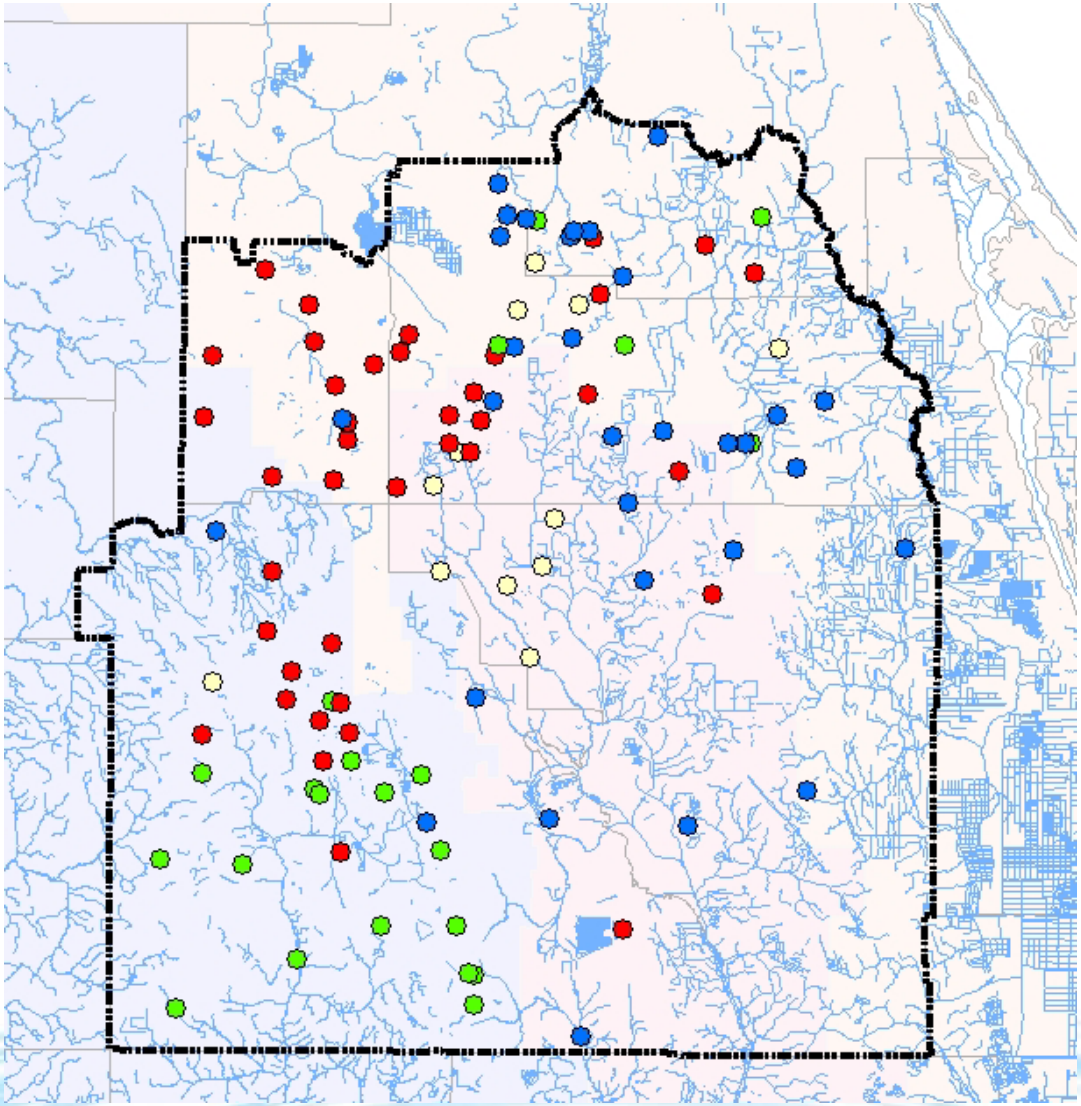
- Total of 115 stations
- Period of study: 1984 - 2008

Number	Site_Name	Type	1984	1985	…	2007	2008
1	Alligator	LK	0.19	-0.19		-0.21	0.38
2	Apopka	LK	0.51	0.05		-1.491	-1.37
3	Apshaw	LK	1.10	-0.05		-1.08	-1.62
4	Barton Big	LK	0.14	-0.30		0.42	0.58
5	Bay	LK	0.69	-0.08		-0.44	-0.064
6	Bay Lake nr Windermere	GW_UFA	1.93	0.73		-1.58	-1.02

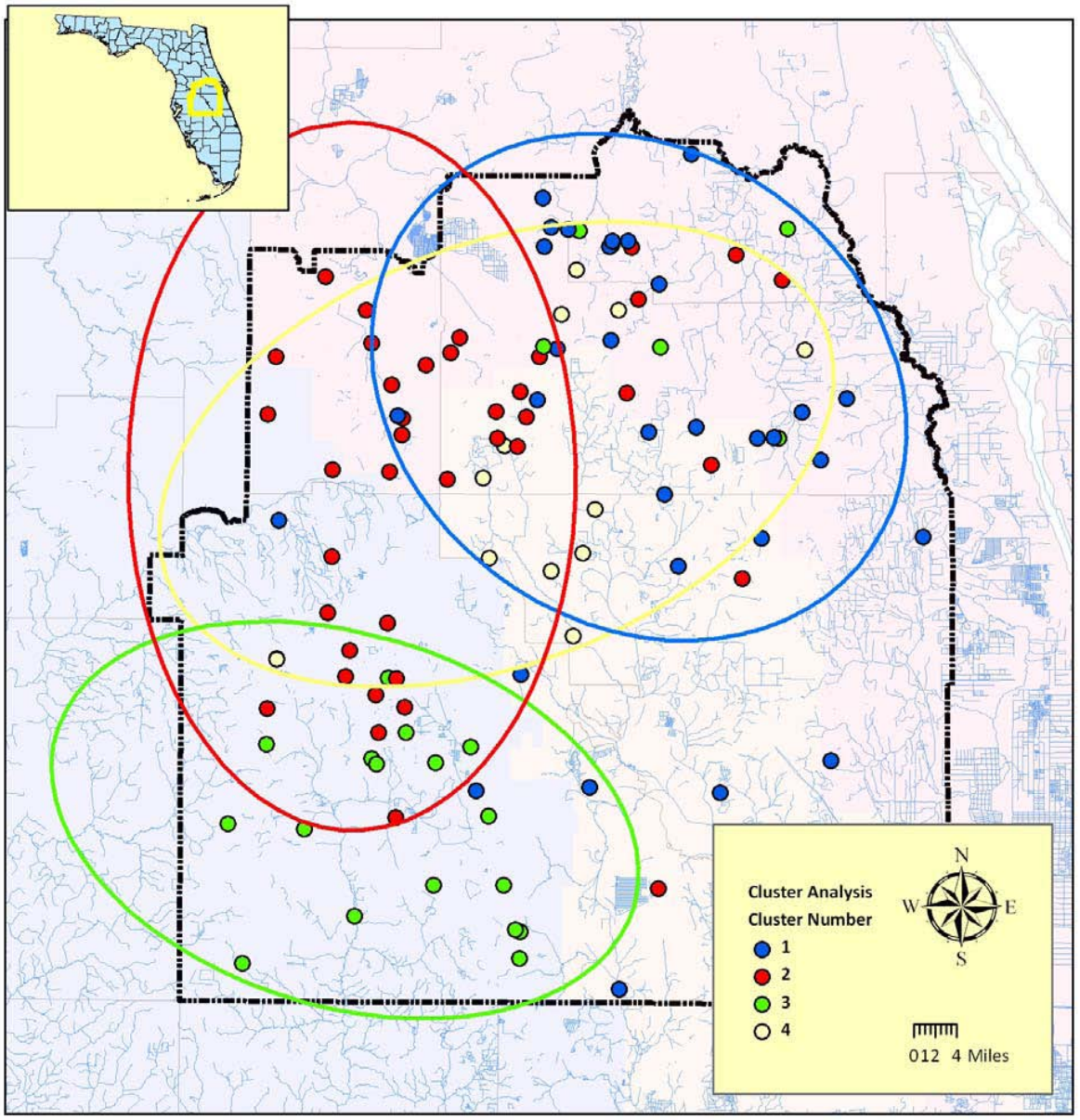
CFCA: 115 Stations, 1984-2008
AHCA Dendrogram



CENTRAL FLORIDA COORDINATION AREA



CENTRAL FLORIDA COORDINATION AREA

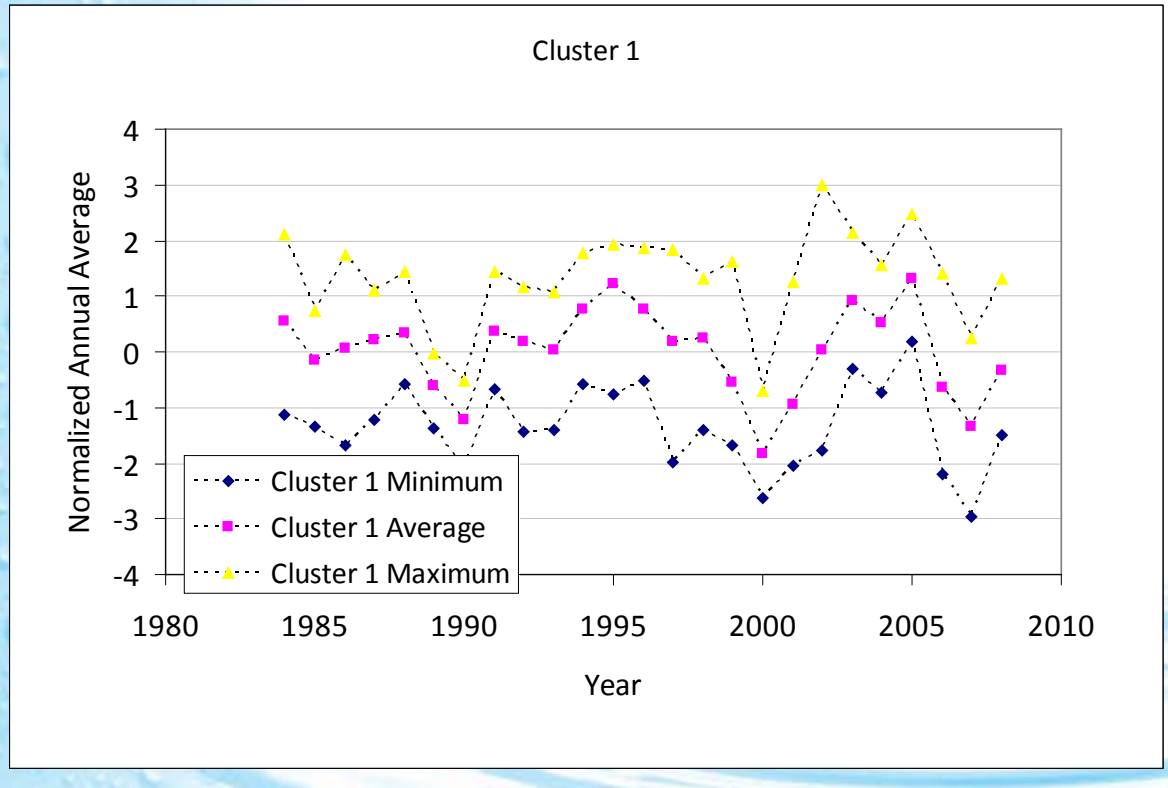


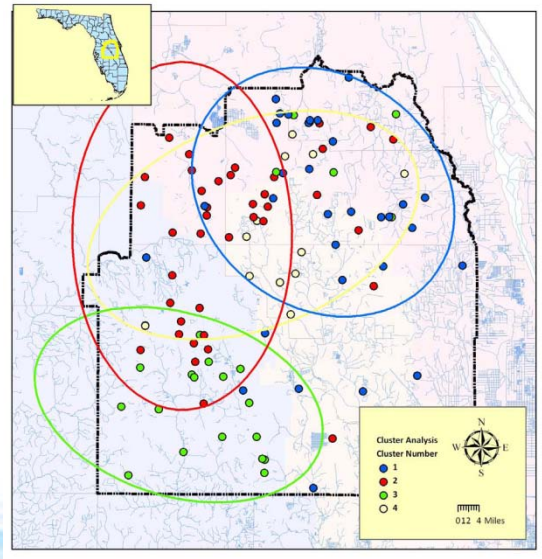
Cluster Membership and Slopes

	Cluster Number			
Site Type	1	2	3	4
GW_IAS			2	1
GW_LFA	1			1
GW_SAS	1	4	1	5
GW_UFA	16	14	7	5
LK	6	24	12	4
RF	5			
SP	4		1	1
Total	33	42	23	17

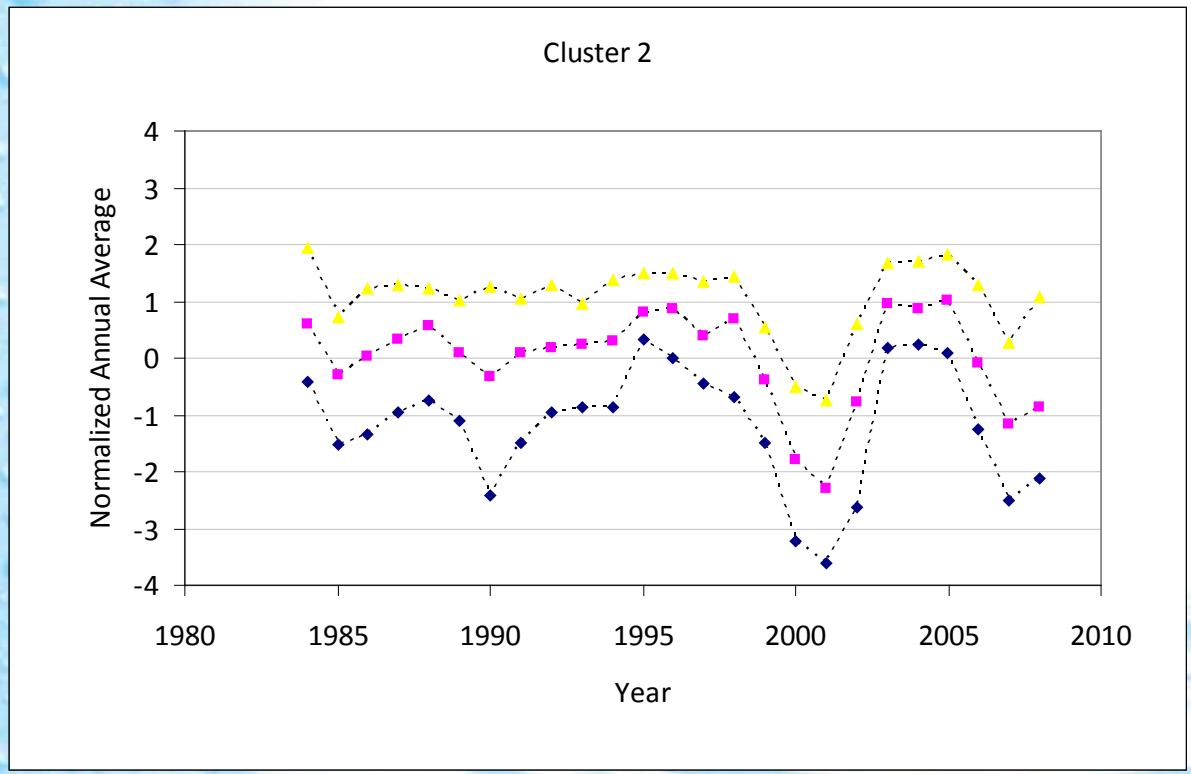
	Cluster Number			
Site Type	1	2	3	4
GW_IAS			0.33016	-0.0536
GW_LFA	-0.1131			-0.1617
GW_SAS	0.02079	-0.0293	0.07761	-0.0946
GW_UFA	-0.0396	-0.0489	0.2779	-0.2856
LK	-0.0098	0.00198	0.19025	-0.0407
RF	0.2717			
SP	-0.0798		0.04057	-0.0277

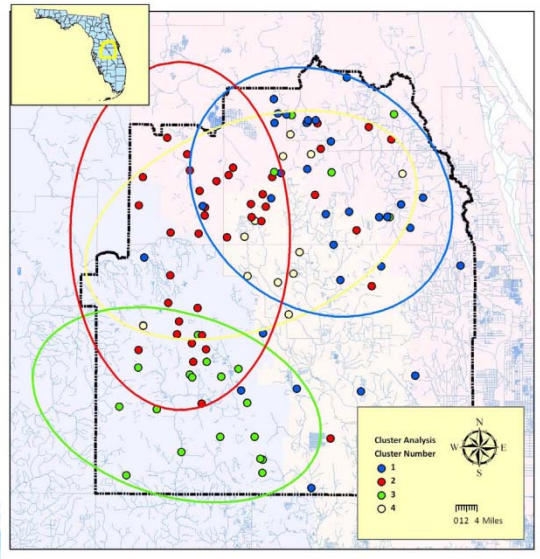
- Reduced levels in 1990, followed by several years of rebound,
- More pronounced reduced levels in 2000, followed by a rebound period, and
- Negative Mann Kendall slopes, with the exception of the rainfall stations and the surficial well, which exhibit positive slopes over the period of analysis.



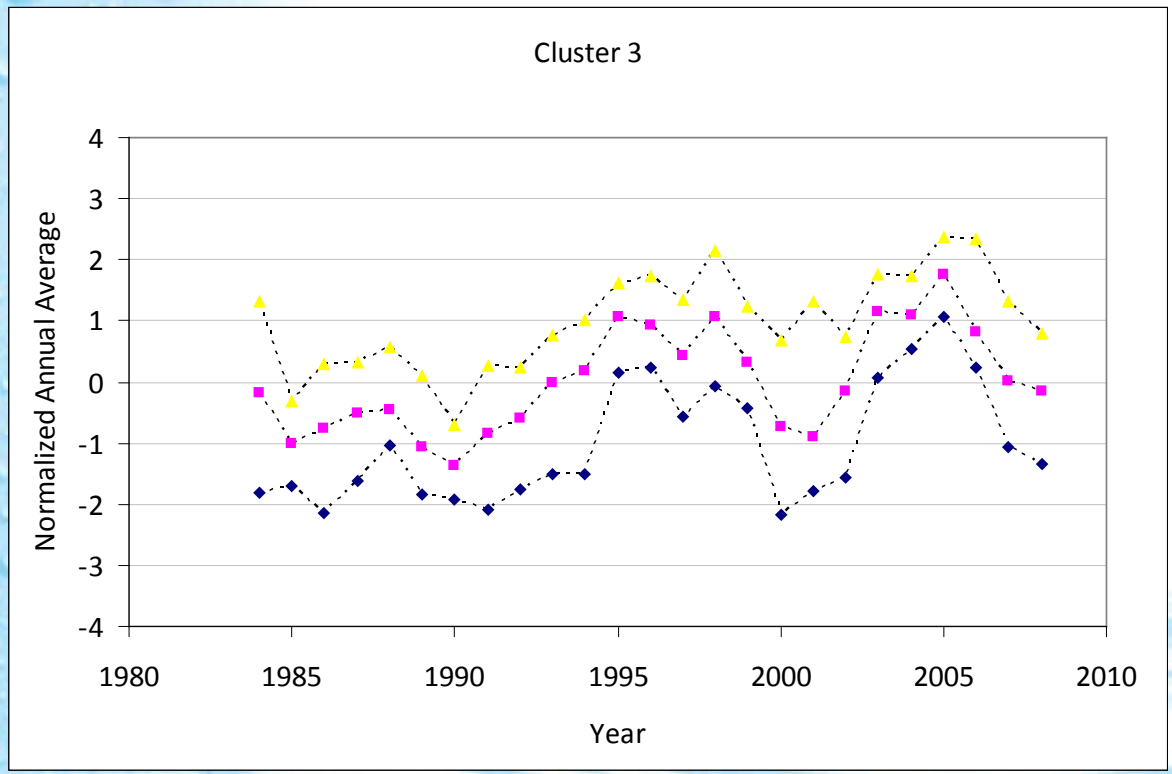


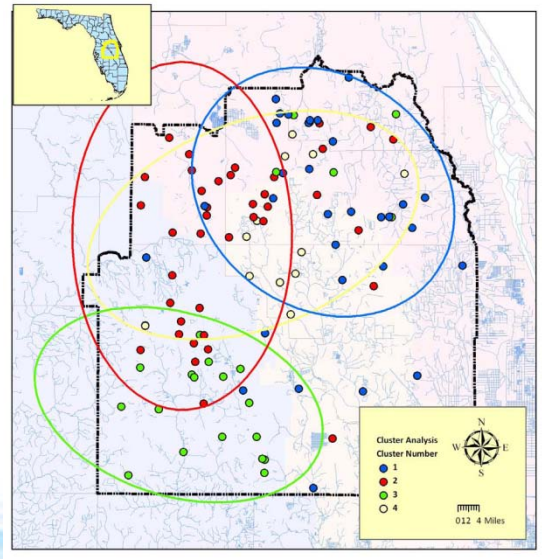
- Low variability at the beginning of the period of analysis (1984 through 1996),
- A cyclic pattern as evidenced by the clear dry period, followed by a wet period, followed by an additional dry period in the latter portion of the analysis period and
- Mann Kendall slopes over the period which are very close to zero.



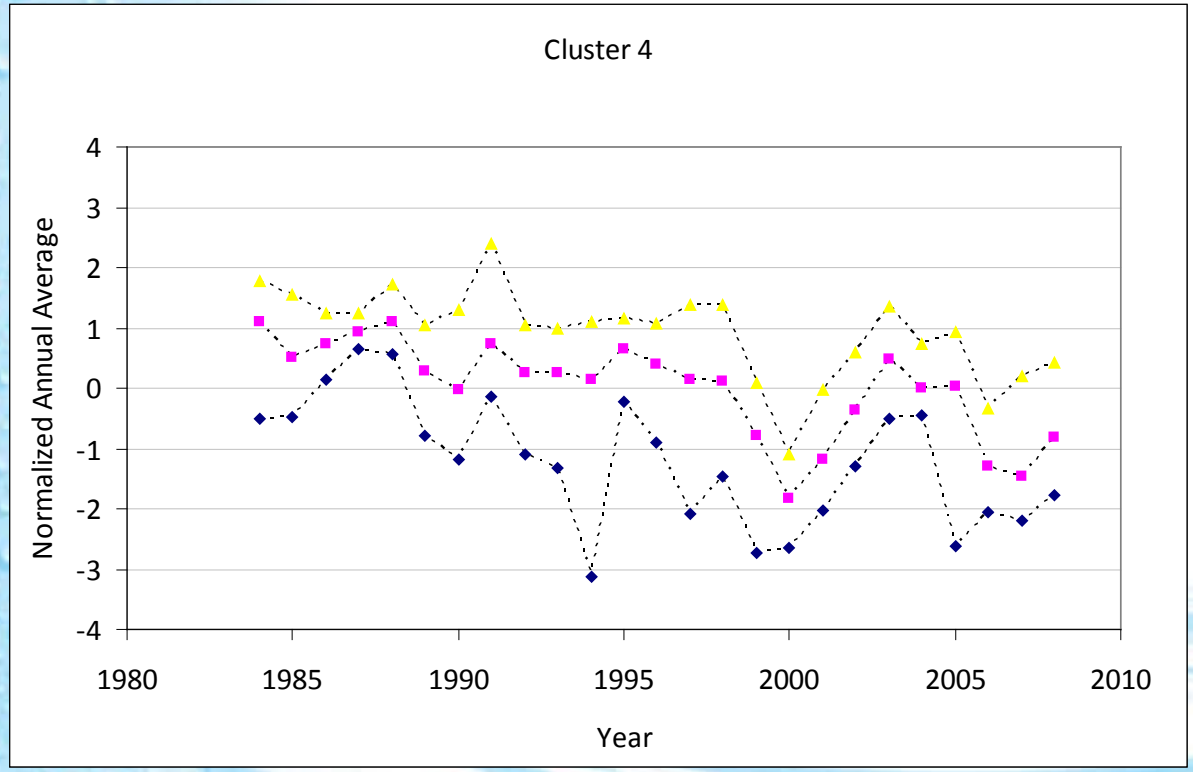


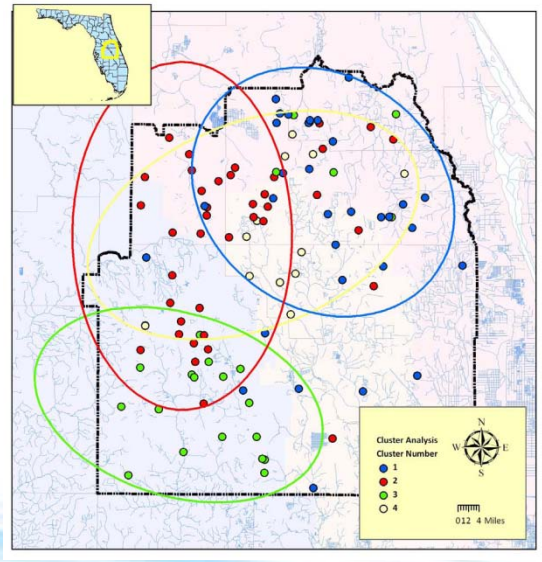
- Positive Mann Kendall slopes over the analysis period, indicating an increasing trend in the data,
- A period of reduced levels in 2000 which is not as pronounced as similar periods experienced by the other clusters.
- Increasing well levels, lake levels, and spring discharges over the analysis period.





- Decreasing levels over the period of analysis,
- The steepest negative Mann Kendall slopes when compared to the remaining clusters,
- A number of stations with statistically significant decreasing trends of high magnitude.
- Monotonic decreasing trends for a majority of stations.





Conclusions

- Trend analysis showed statistically significant trends (both increasing and decreasing) throughout the CFCA.
- Increasing trends were predominately located in Polk County.
- Cluster analysis grouped stations with similar hydrologic behavior.
- Examination of forcing functions would provide more insight into reasons for clustering and individual station response.

Statistical Trends in Hydrologic Data- Future Work

- A more detailed rainfall analysis in the CFCA domain- moving window analysis
- Closer examination of piecewise (non-monotonic) stations
- Analysis of additional stations in Osceola County
- Trends in anthropogenic impacts (land use, pumping)