Statistical Evaluations

Central Florida Coordination Area
Tools Development Team Workshop
July 29, 2010

Presented by Jill Hood, P.G.; Senior Professional Geologist
Southwest Florida Water Management District

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

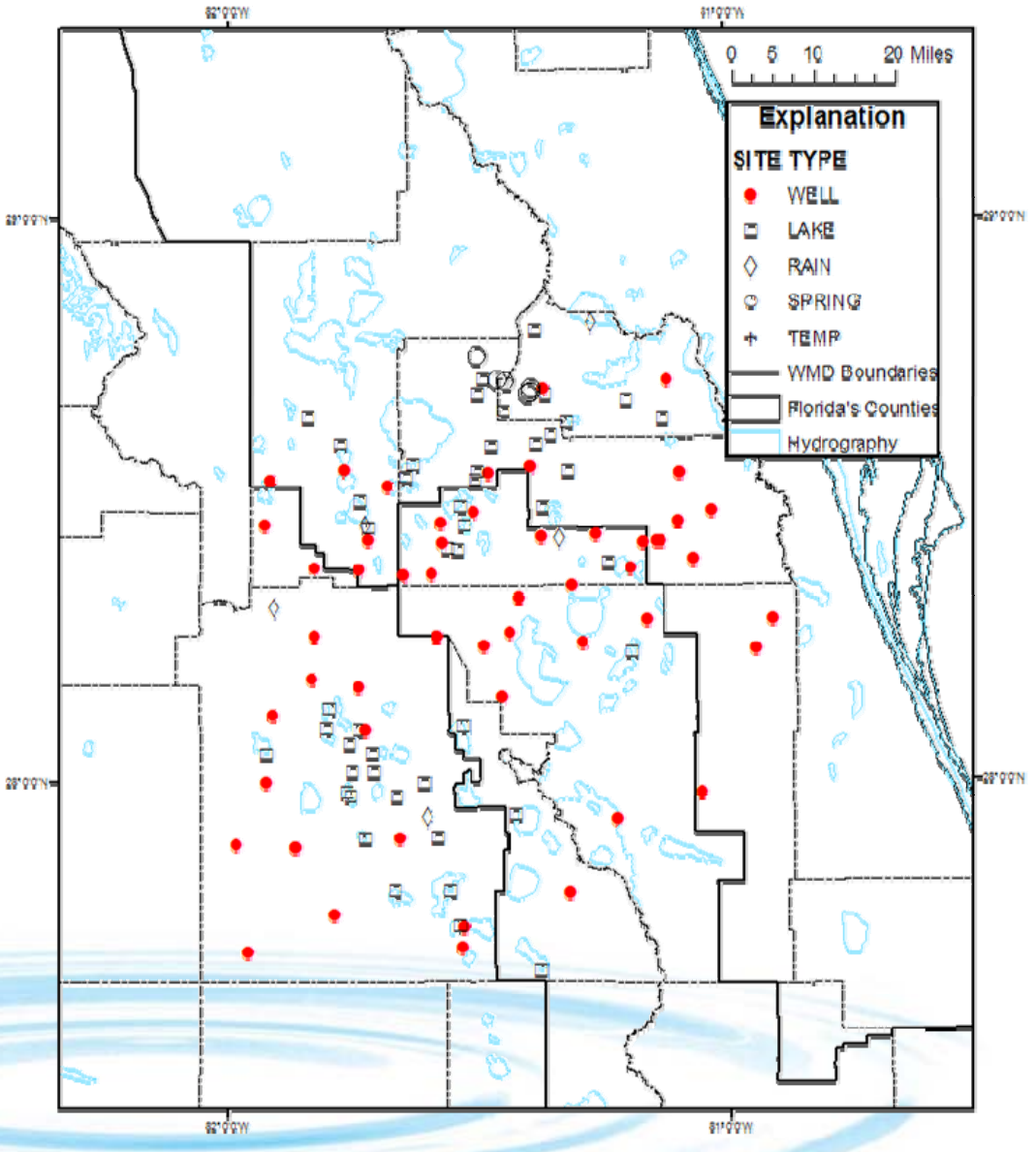
- **Statistical Trends in Hydrologic Data** – evaluating trends at single sites and correlations between sites.
- **Data Mining** – evaluating cause and effect relationships.

Statistical Trends in Hydrologic Data

- Exploratory Data Analysis.
- Identify Breakpoints in LOWESS scatterplots.
- Examine for Seasonal and Serial Correlation.
- Trend Analysis for Individual Sites.
- Cluster Analysis for Evaluating Spatial Groupings of Trends.

Hydrologic Database

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

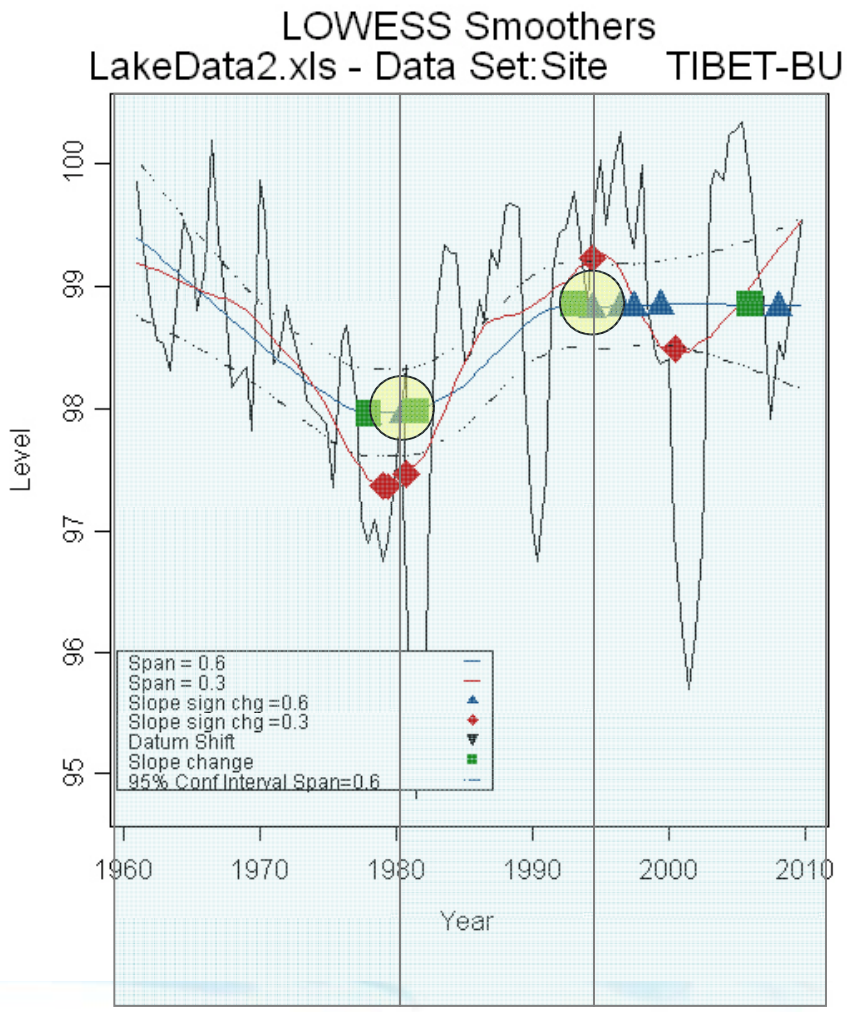


Trend Analysis of Entire Time Series vs. Segments of Time

Segment	Mann-Kendall p-value	Result (Reject Ho if p < 0.1)
Entire Time Series	0.128	Fail to Reject Ho
Segment 1	0.0002	Reject Ho; Man-Kendall Regression Sen Slope: -0.146 ft/yr
Segment 2	0.125	Fail to Reject Ho
Segment 3	0.5526	Fail to Reject Ho
Segment 2 & 3	0.1282	Fail to Reject Ho

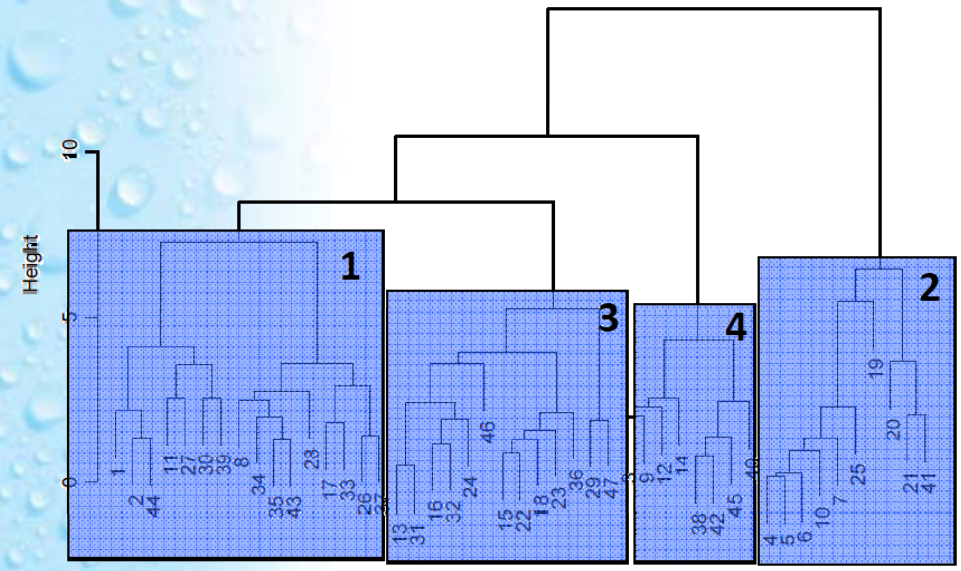
Therefore, a statistically significant trend exists through 7/1/81.

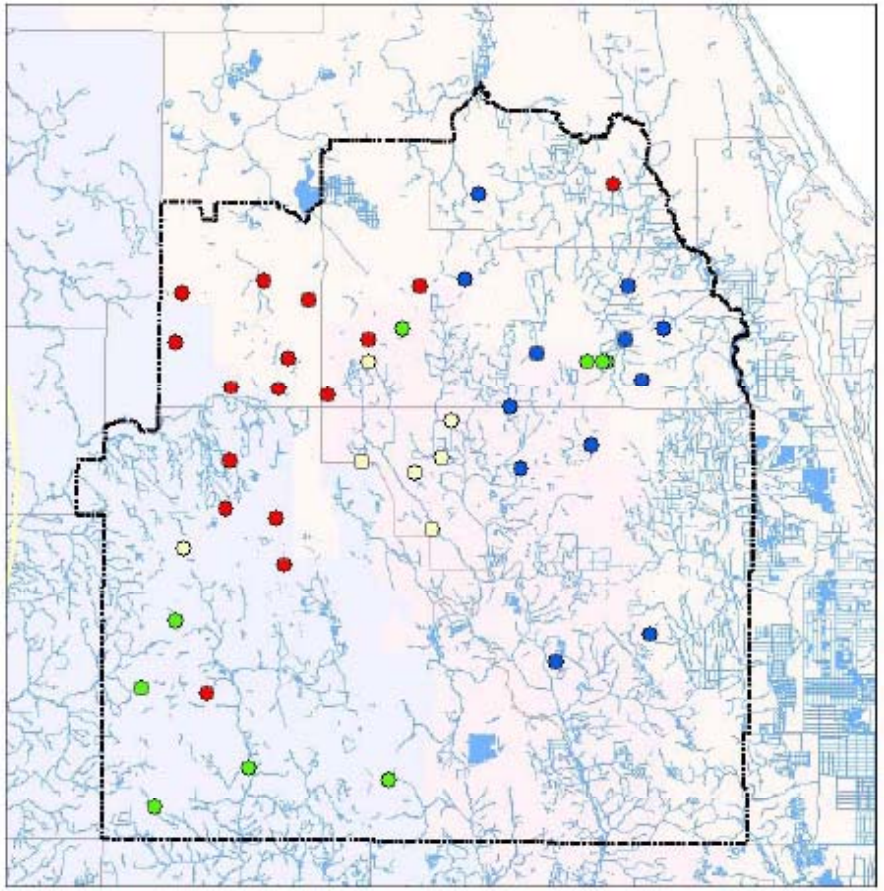
Note: Elimination of BP2 would have caused no change in conclusions.



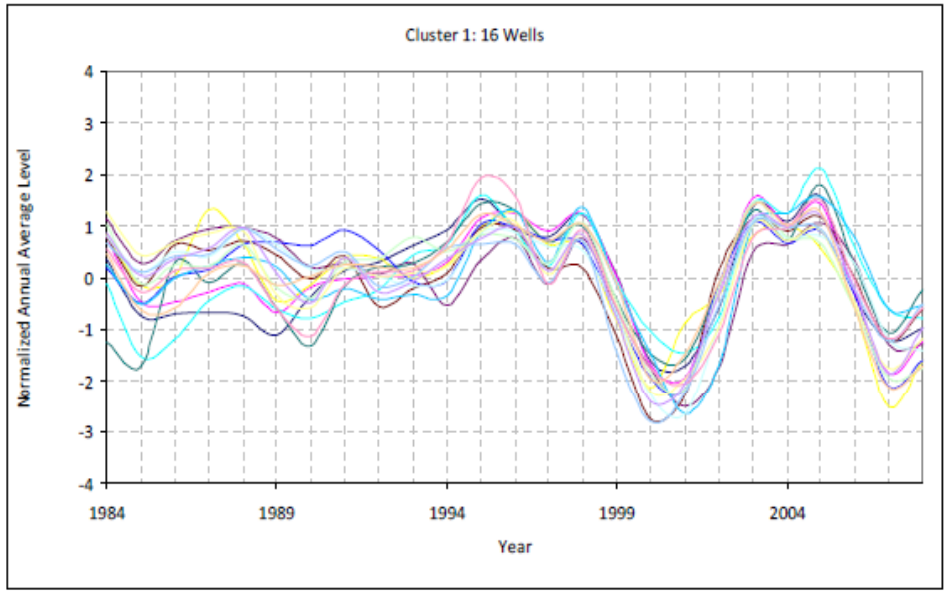
•Segment 1 •Segment 2 •Segment 3

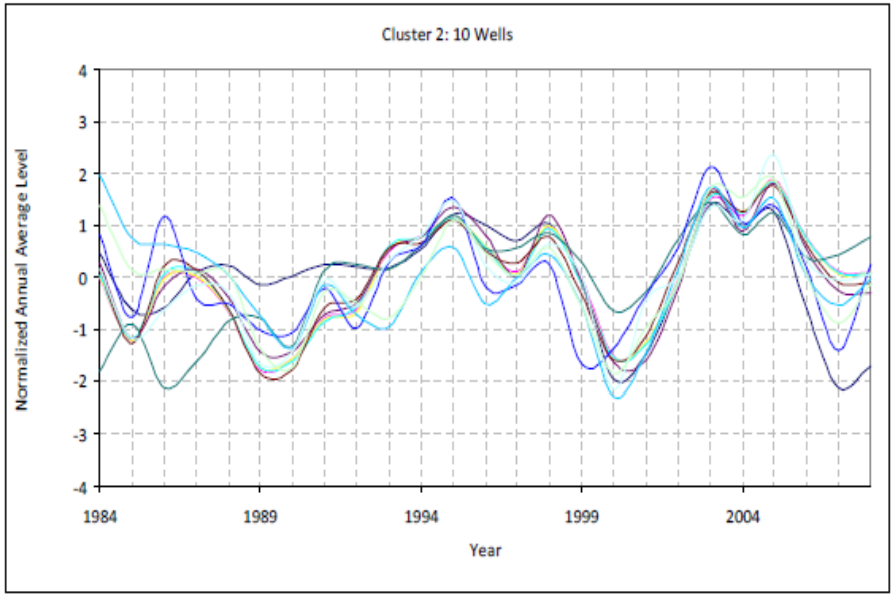
Statistical Trends in Hydrologic Data

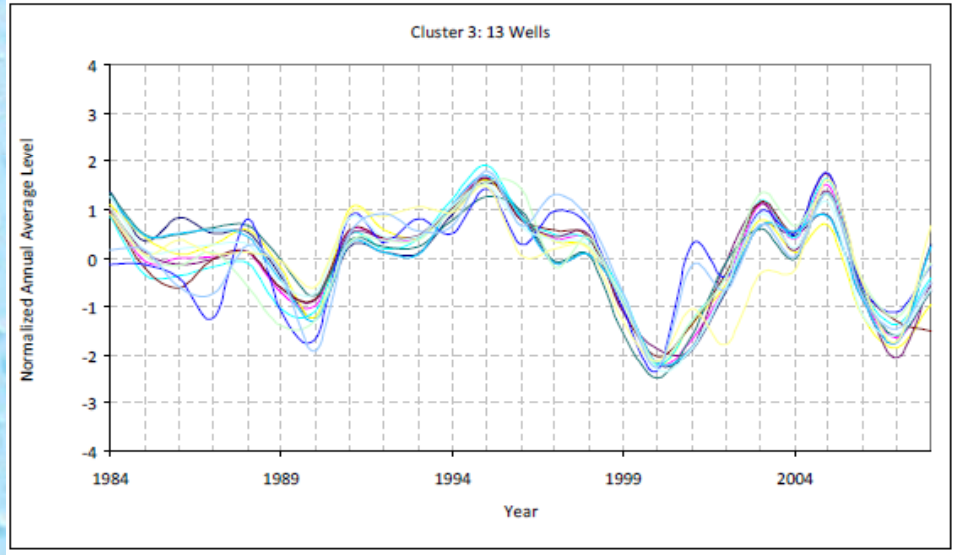


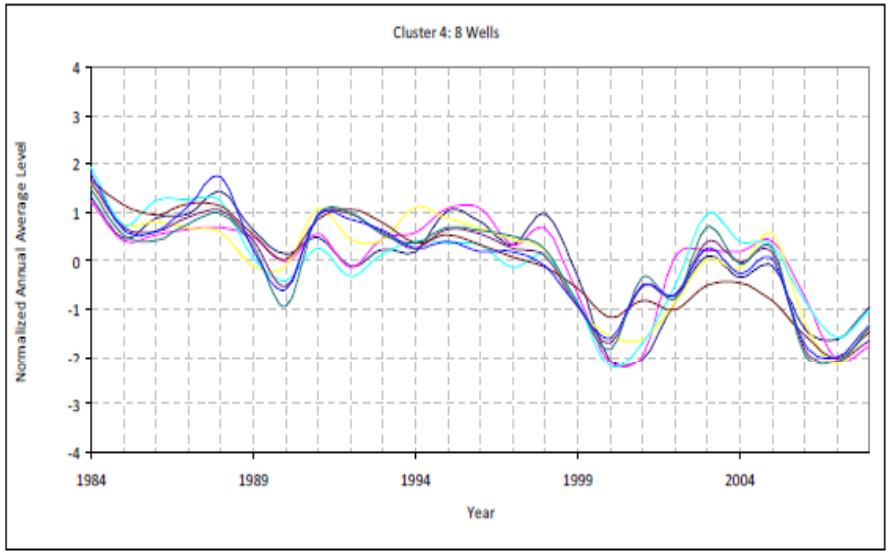


Statistical Trends in Hydrologic Data









Statistical Trends in Hydrologic Data

Next Steps:

- Explain Data Clusters
  - Hydrogeologic Setting
  - Climate
  - Pumping
  - Land Use Changes

What is Data Mining?

- *Data Mining*: the search for valuable knowledge in massive volumes of data.
- *Data Mining Tool Box:*
  - Signal processing, statistics, machine learning, chaos theory, advanced visualization.
  - Artificial Neural Networks (ANN) models – one approach to machine learning.

Data ⇨ Information ⇨ Knowledge

"To understand the past is to understand the future"

Data Mining and MODFLOW

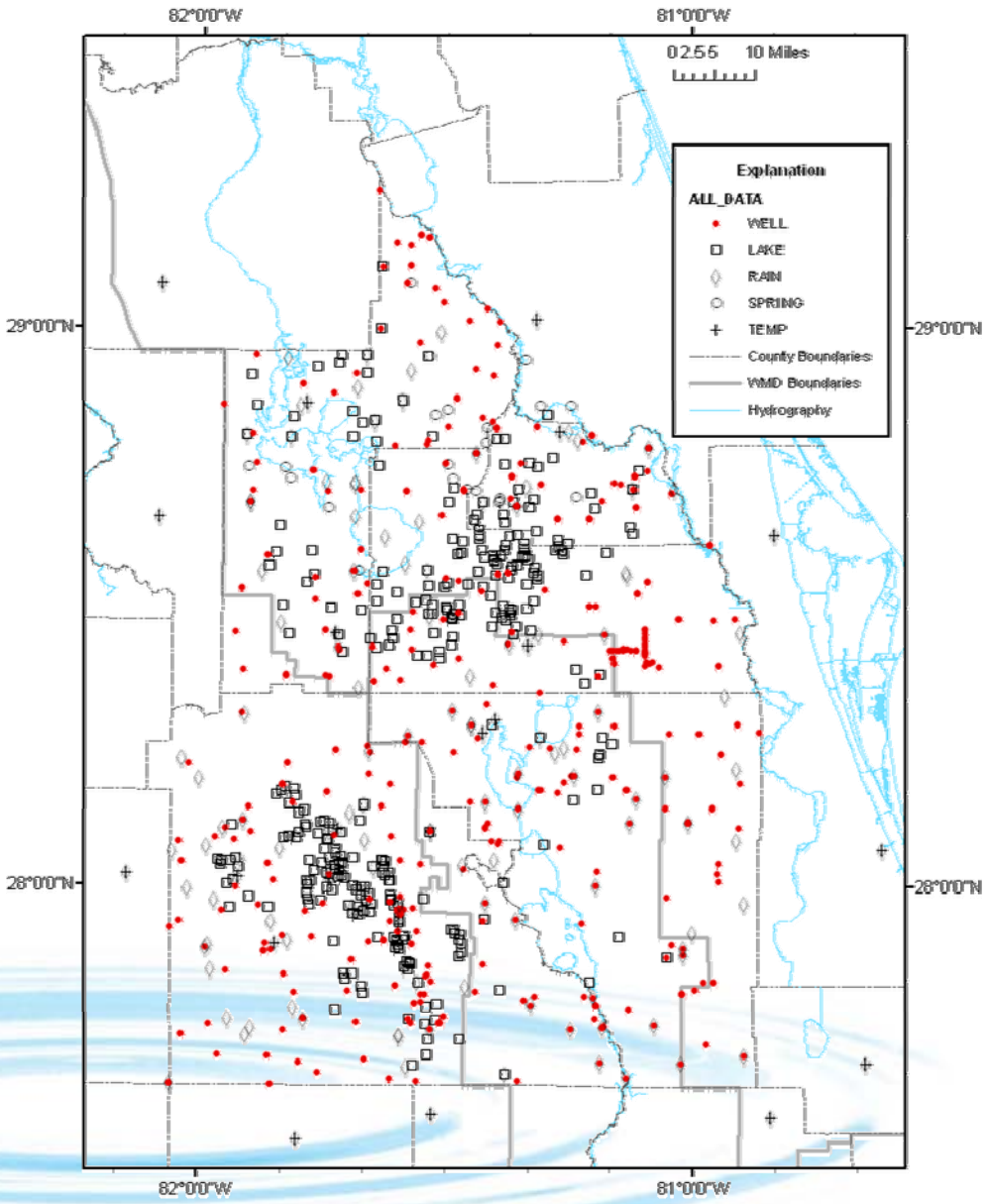
Two Approaches – Similar Goals

- Can be complementary.
- Data mining can provide great insight into hydrologic system behavior.
- Insights can be used to guide development and evaluation of MODFLOW model.

Hydrologic Database

- Compiled from SJRWMD, SWFWMD, SFWMD, USGS, NOAA, Orange County, and Seminole County databases

- 963 sites: 470 wells, 22 springs, 307 lakes, 143 rain gages, and 21 air temperature sites



Data Mining

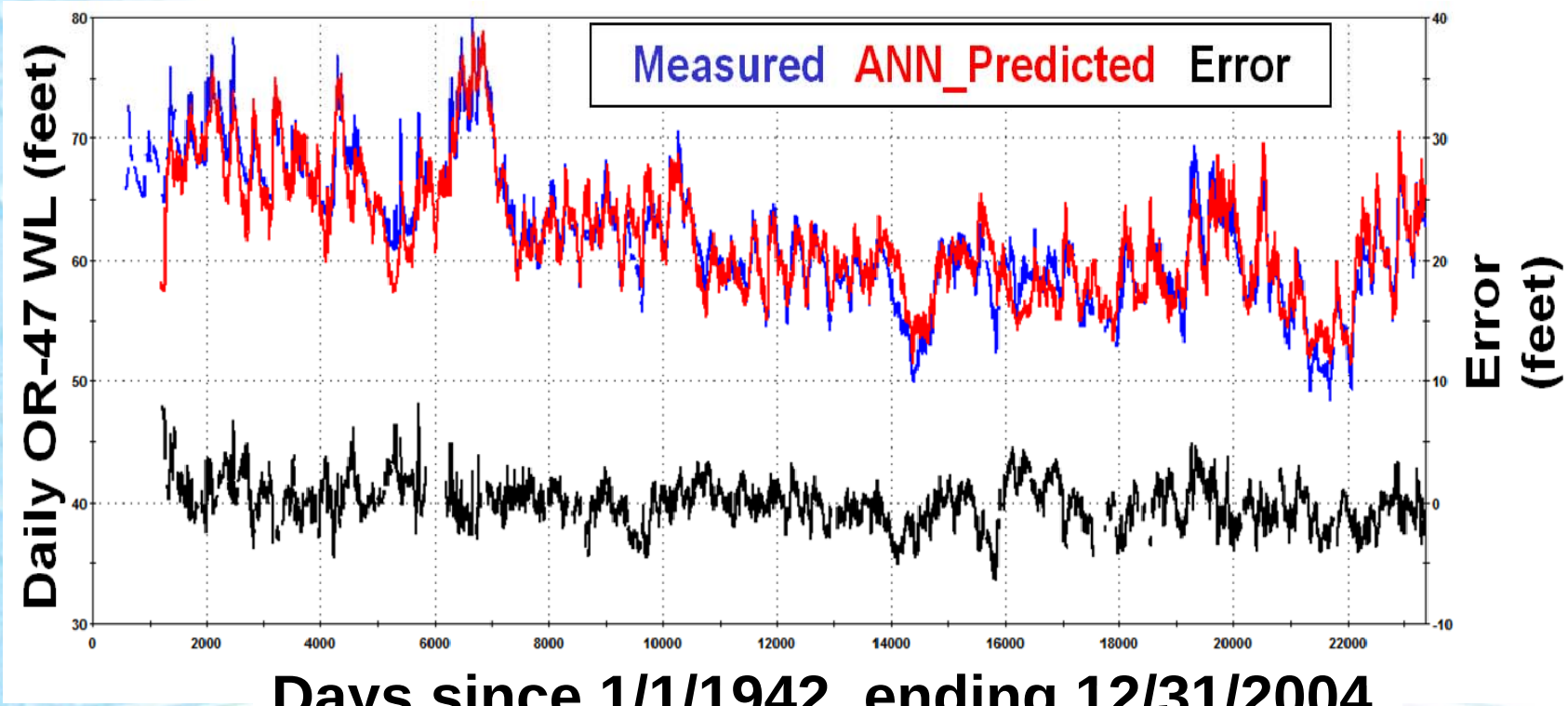
I. **Historical data analysis -** *Database compilation and QA; Exploratory Data Analysis; Develop site-specific Artificial Neural Network (ANN) models.*

II. **Decision support system (DSS) incorporating a regional data mining model -** *Spatial interpolation of multiple ANN models to quantify regional groundwater system behavior.*

III. **Comparison with physics-based groundwater model -** *Compare ANN results with MODFLOW model assessments to strengthen model predictive capabilities based on system behavior inferred from historical data.*

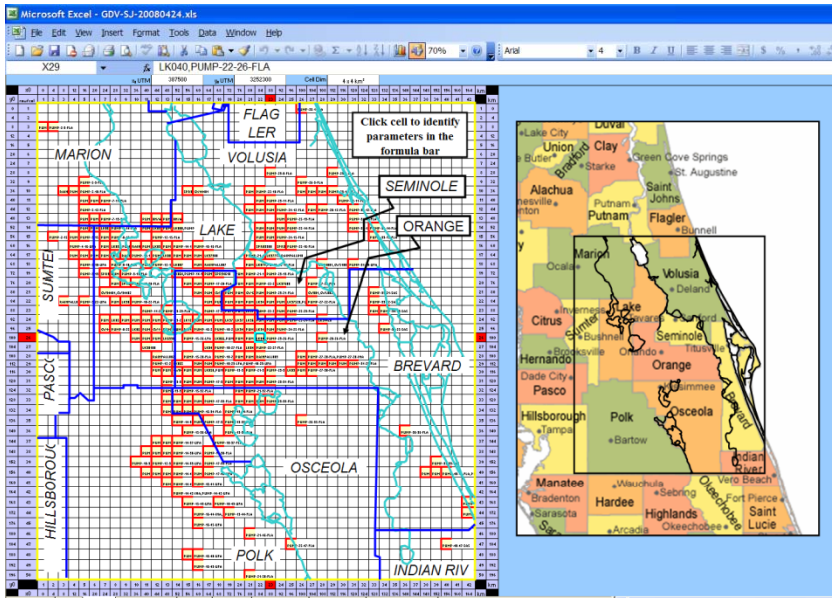
Artificial Neural Networks

- Results from pilot study indicate the effectiveness of ANNs for hydrologic simulation.



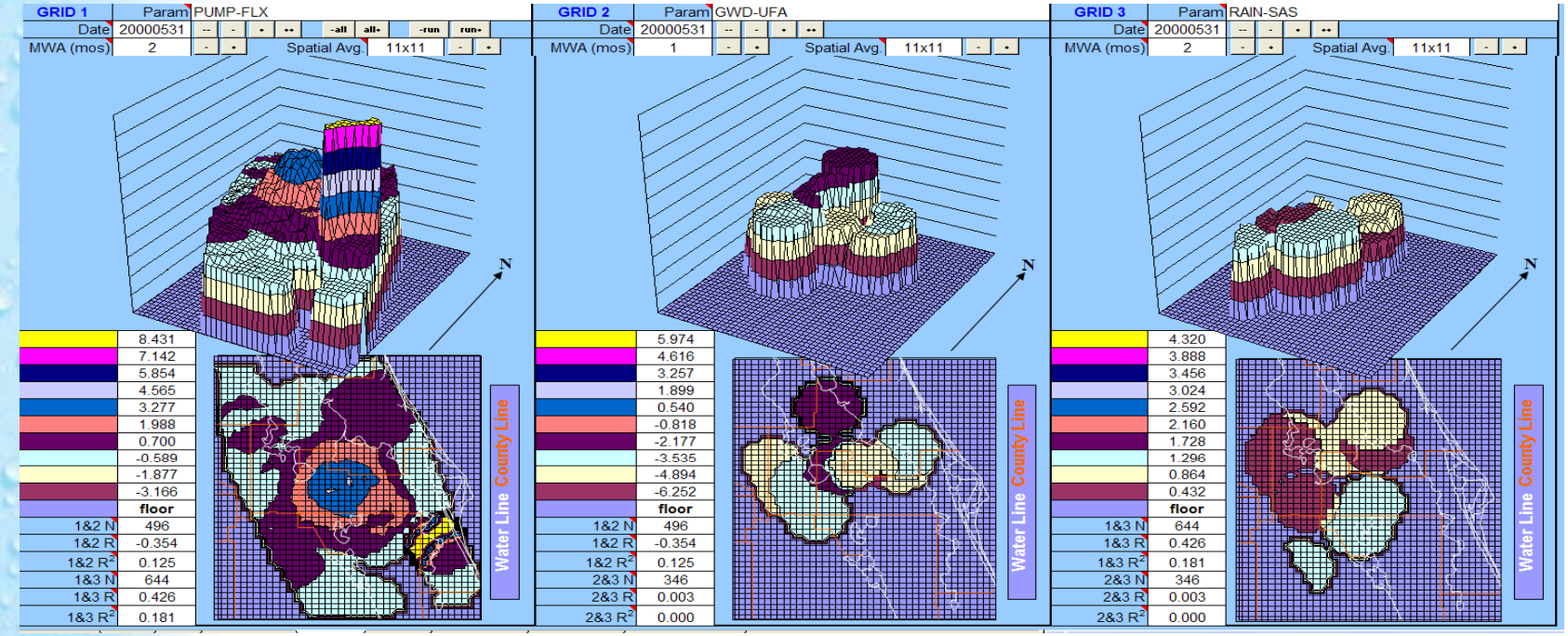
Decision Support System

- An Excel tool built around ANN models.
- Composed of multiple ANN-based "sub-models" to predict regional groundwater system.
- Fast run time and user friendly.



Tabs: *Info*, ***Map***, *Parameters; 3DVis, Output, Database1, Database2, Release Notes*

Decision Support System



1&2 N	496
1&2 R	-0.354
1&2 R^2	0.125
1&3 N	644
1&3 R	0.426
1&3 R^2	0.181

1&3 N	644
1&3 R	0.426
1&3 R^2	0.181
2&3 N	346
2&3 R	0.003
2&3 R^2	0.000

Comparison of ANNs and Models

- Compare site-specific ANN results (key lakes, wells, and springs) to same sites simulated by numerical model.
- ANN and numerical models will provide separate lines of evidence, providing opportunity 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.

Data Mining

- Status:
  - Finalizing Database of Stressors (Water Use)
  - Completing Site-Specific ANNs
- Next Steps:
  - Compare ANN's and Numerical Models

Data Mining

Report complete by September 2011.