

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.

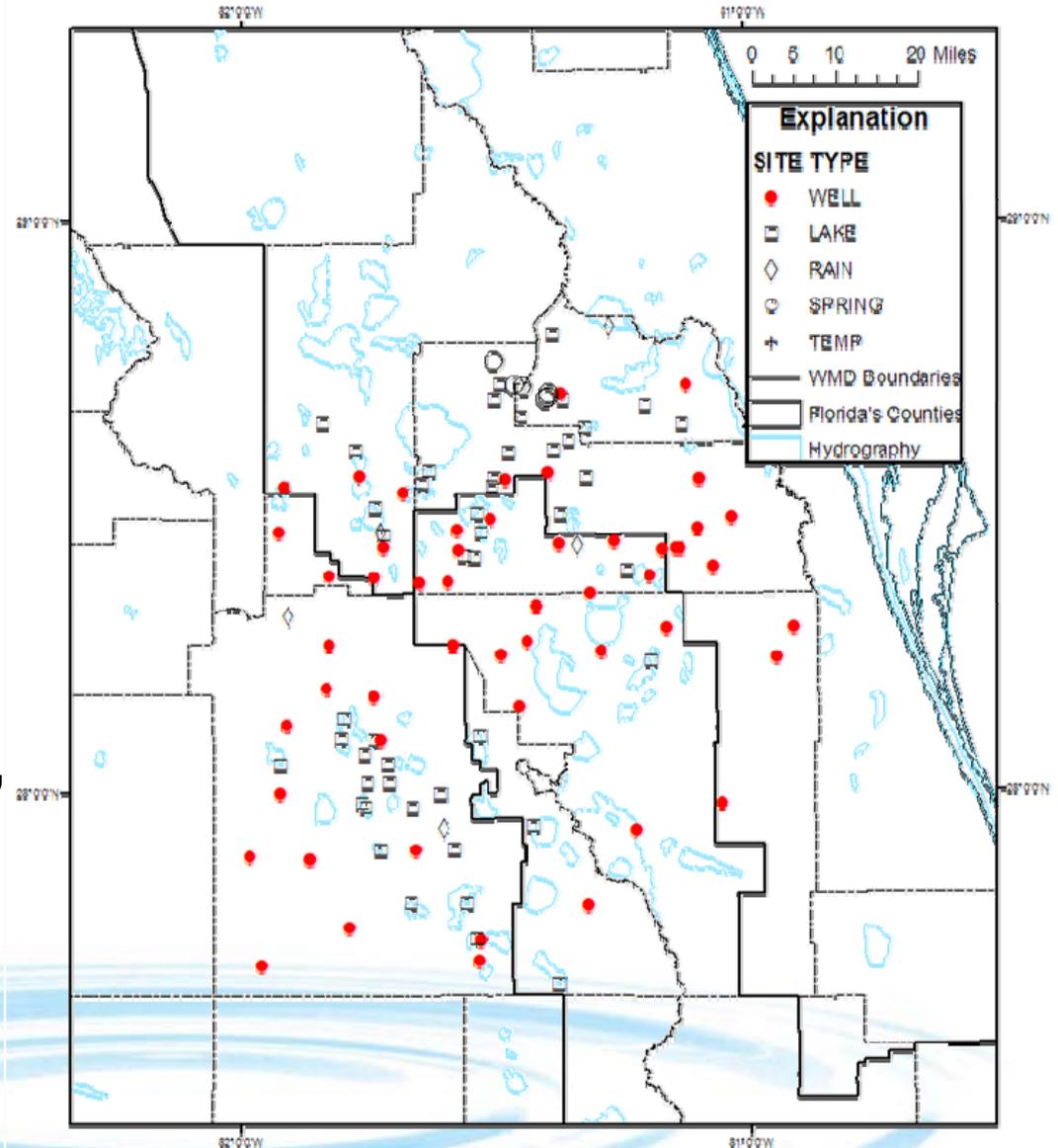
Statistical Trends in Hydrologic Data

- 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



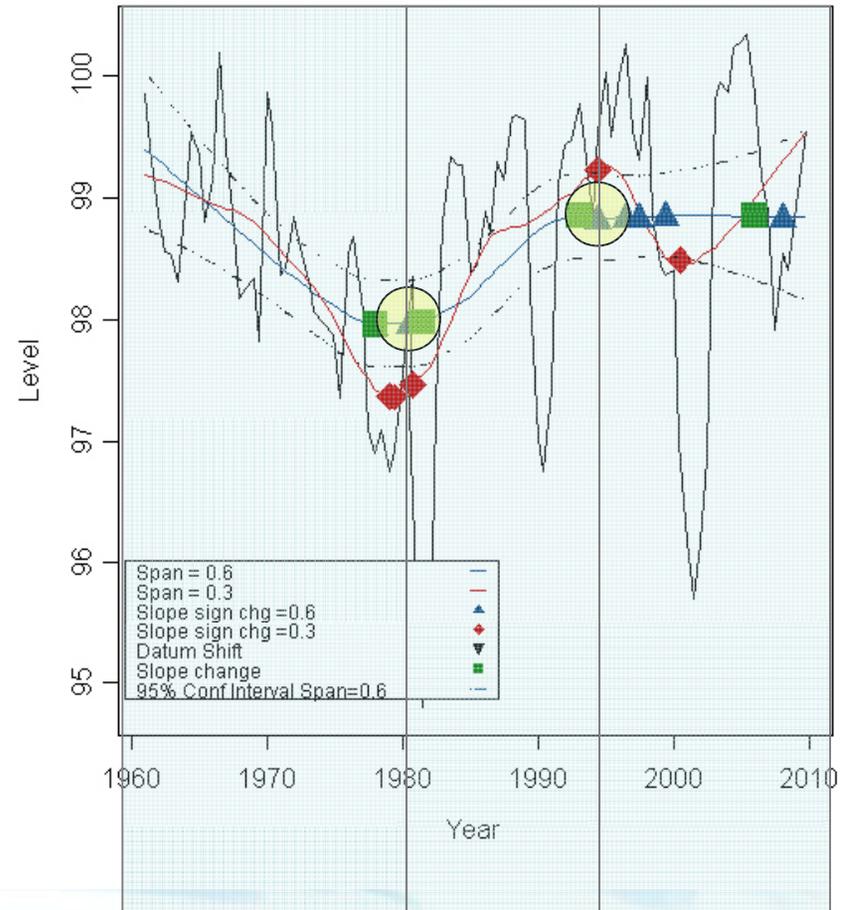
Trend Analysis of Entire Time Series vs. Segments of Time

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.

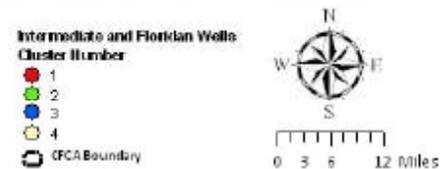
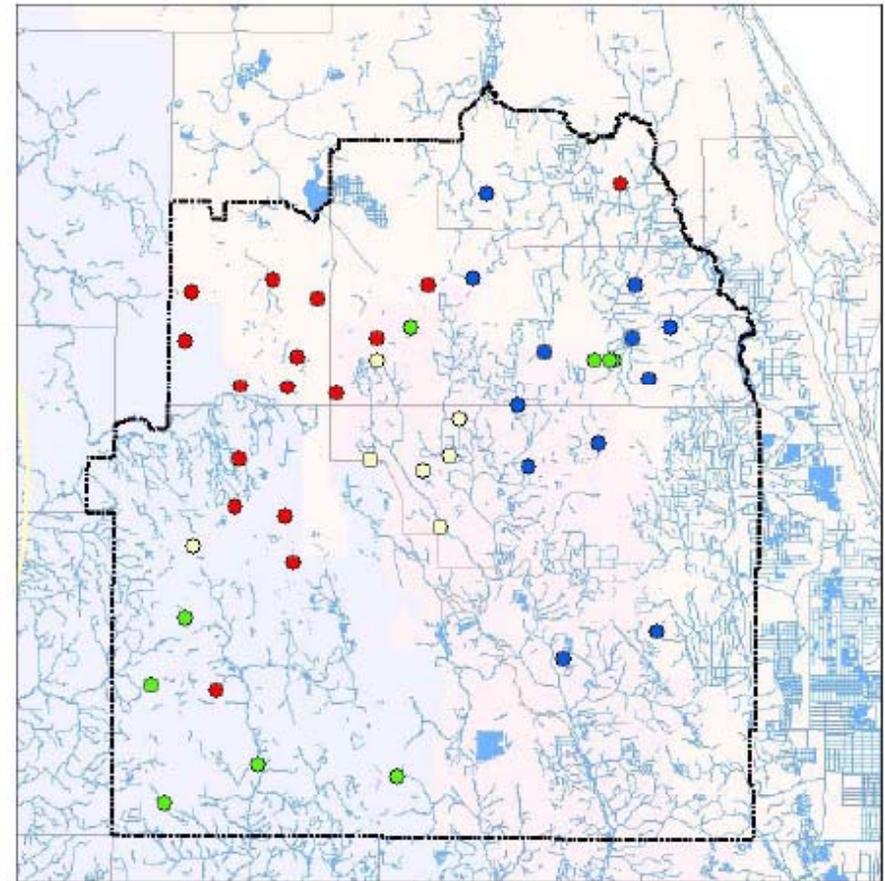
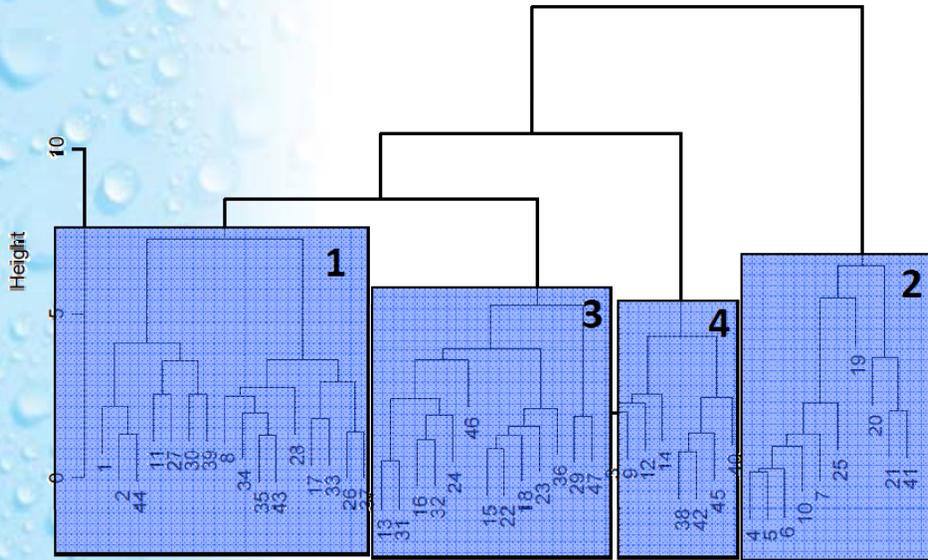
Note: Elimination of BP2 would have caused no change in conclusions.

LOWESS Smoothers
LakeData2.xls - Data Set: Site TIBET-BU

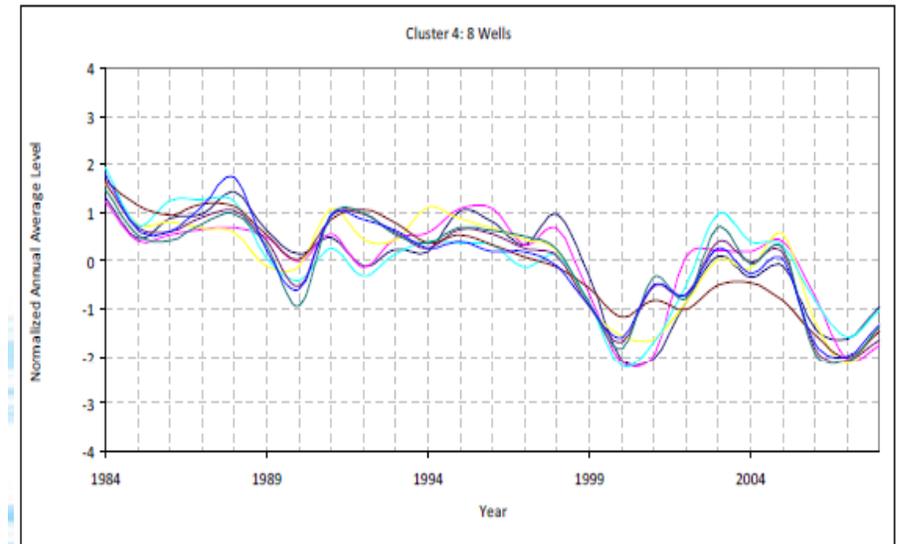
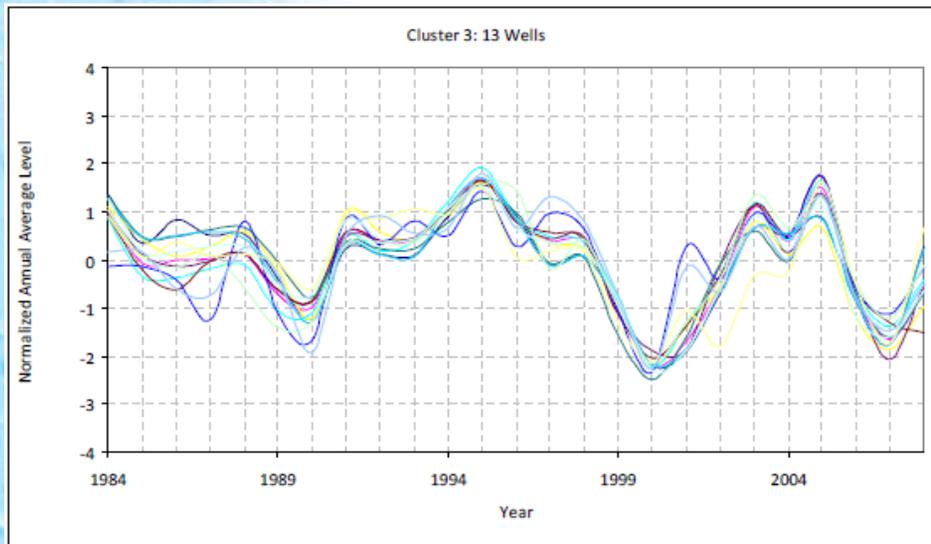
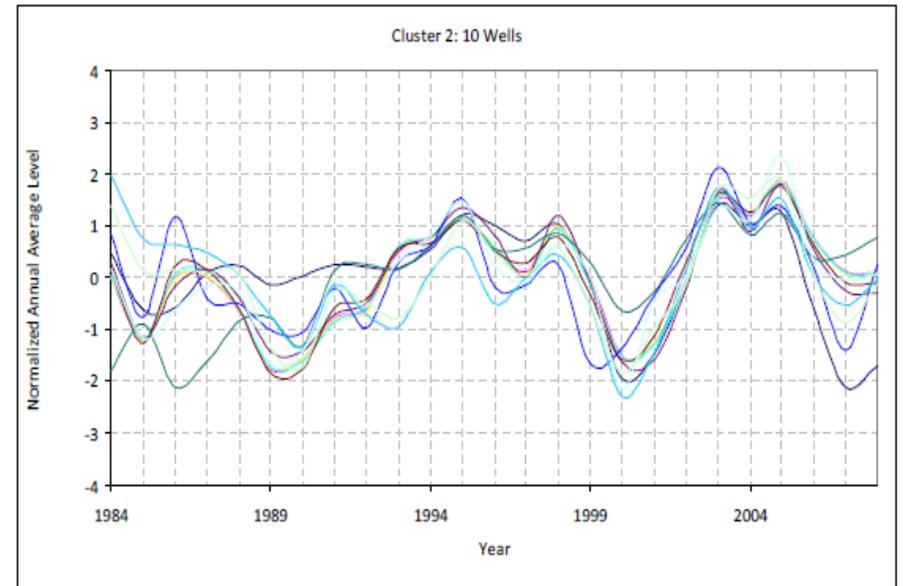
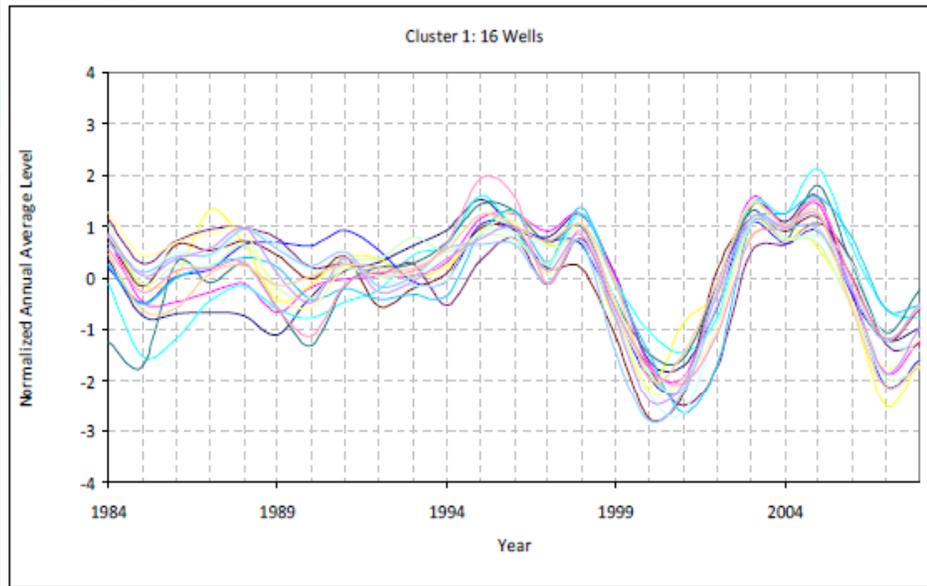


•Segment 1 •Segment 2 •Segment 3

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”

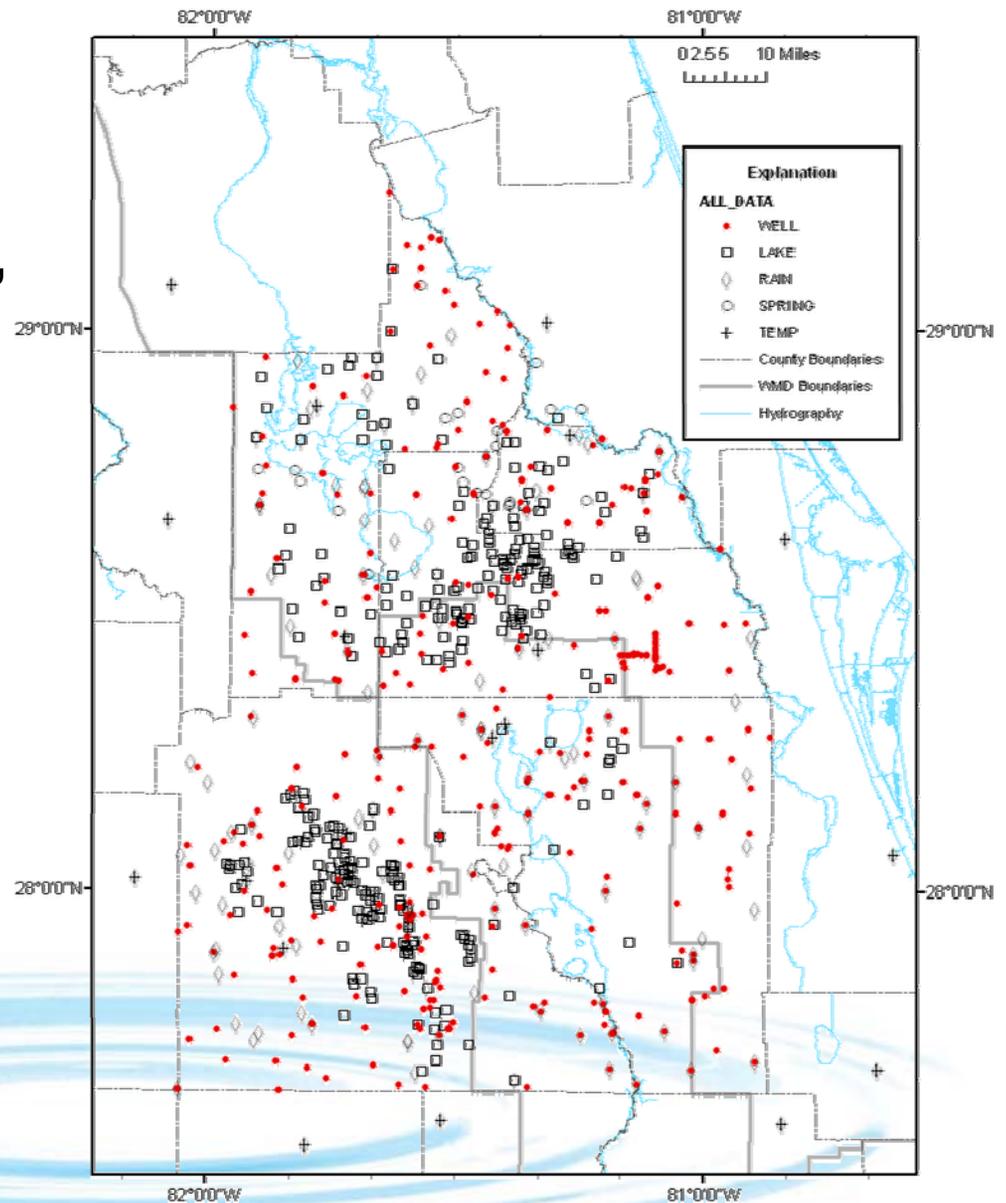
Data Mining and MODFLOW

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

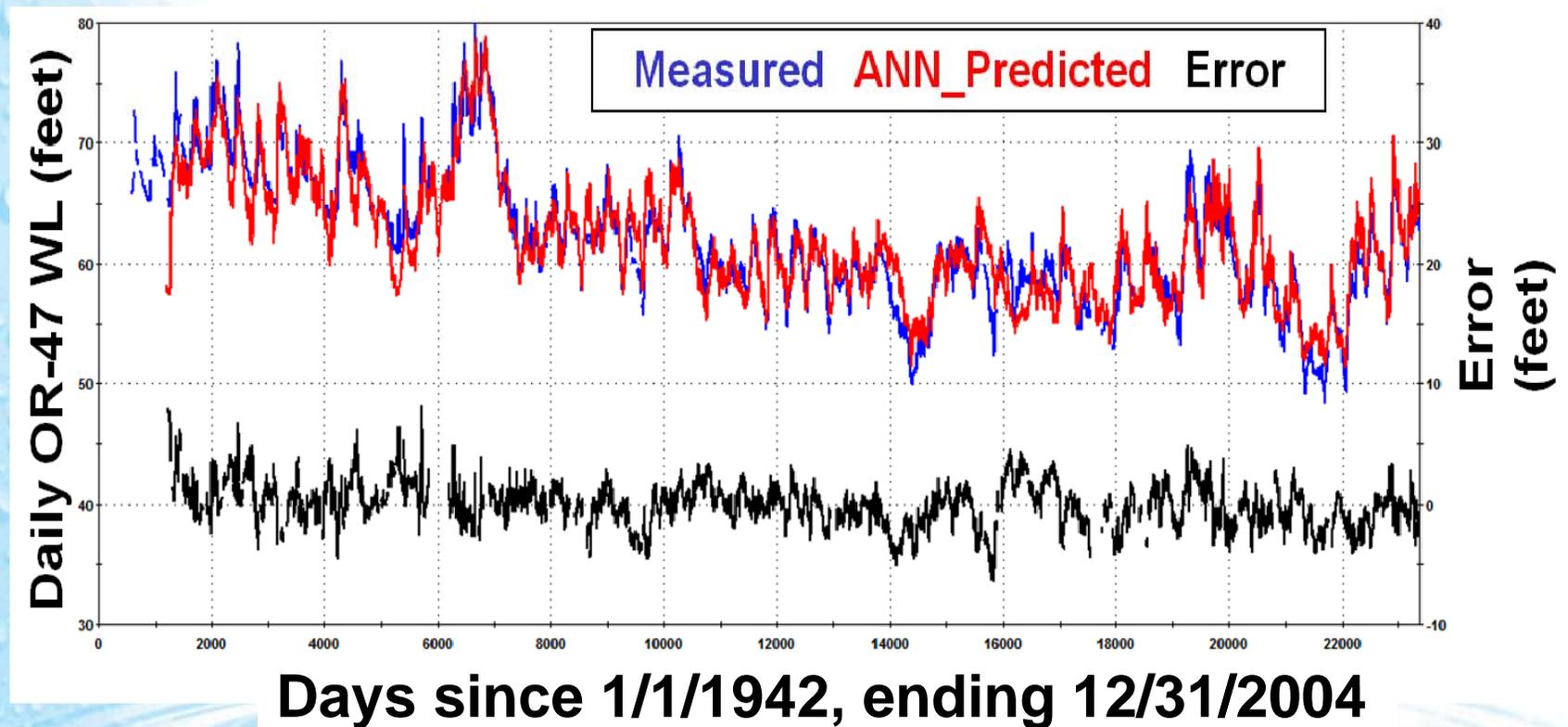


Data Mining

- 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.*

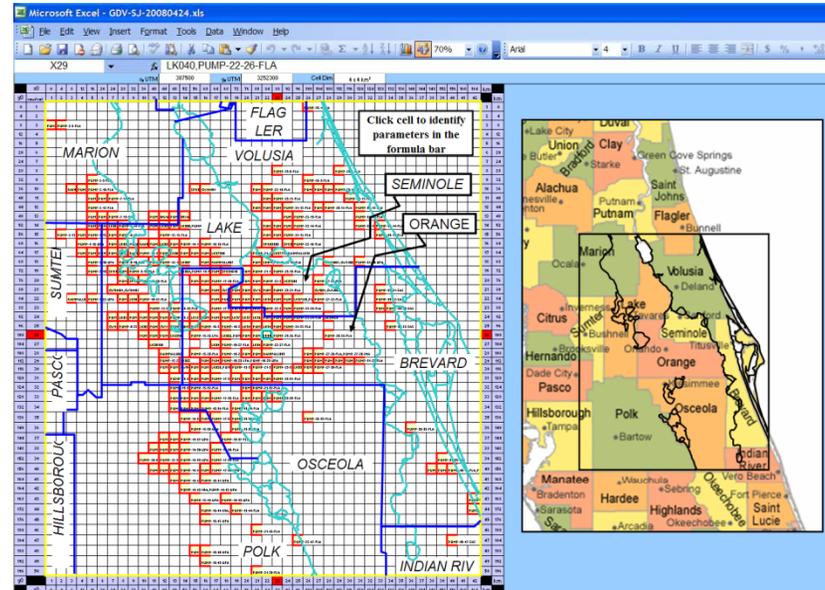
Artificial Neural Networks

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



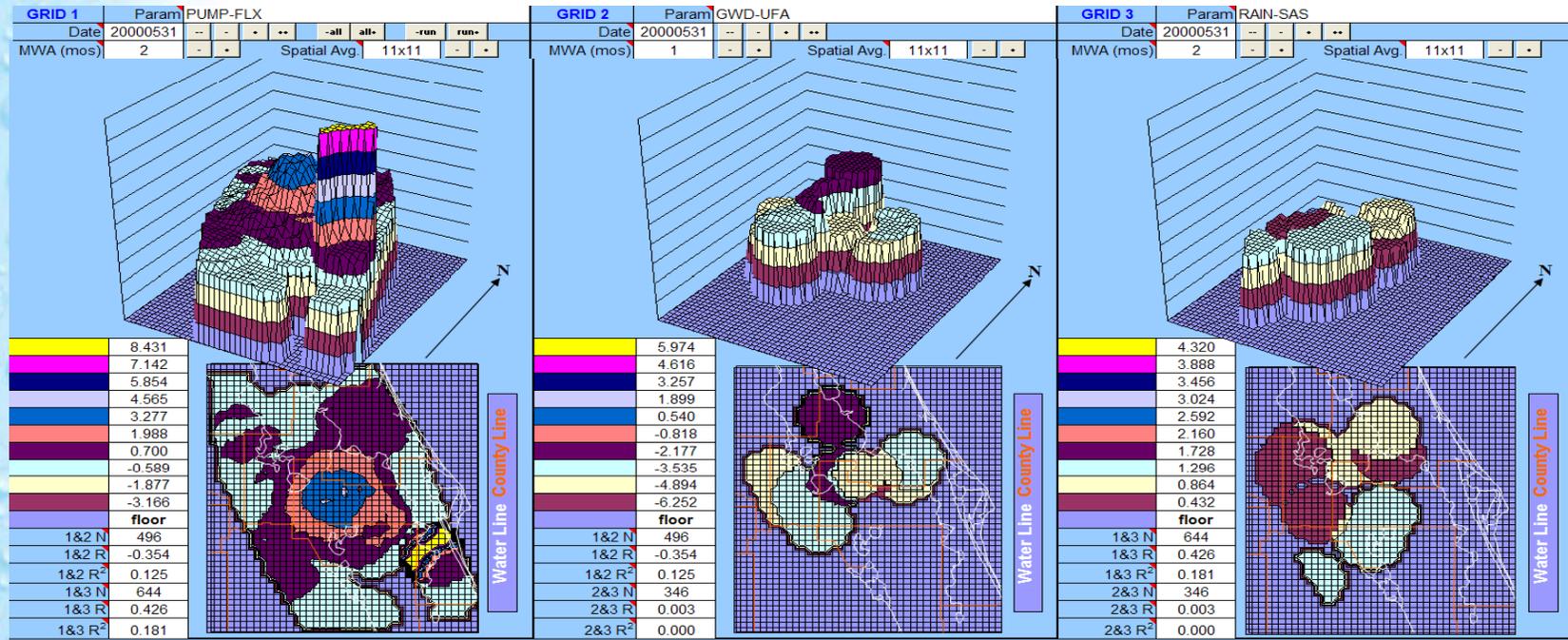
Decision Support System

- 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:**

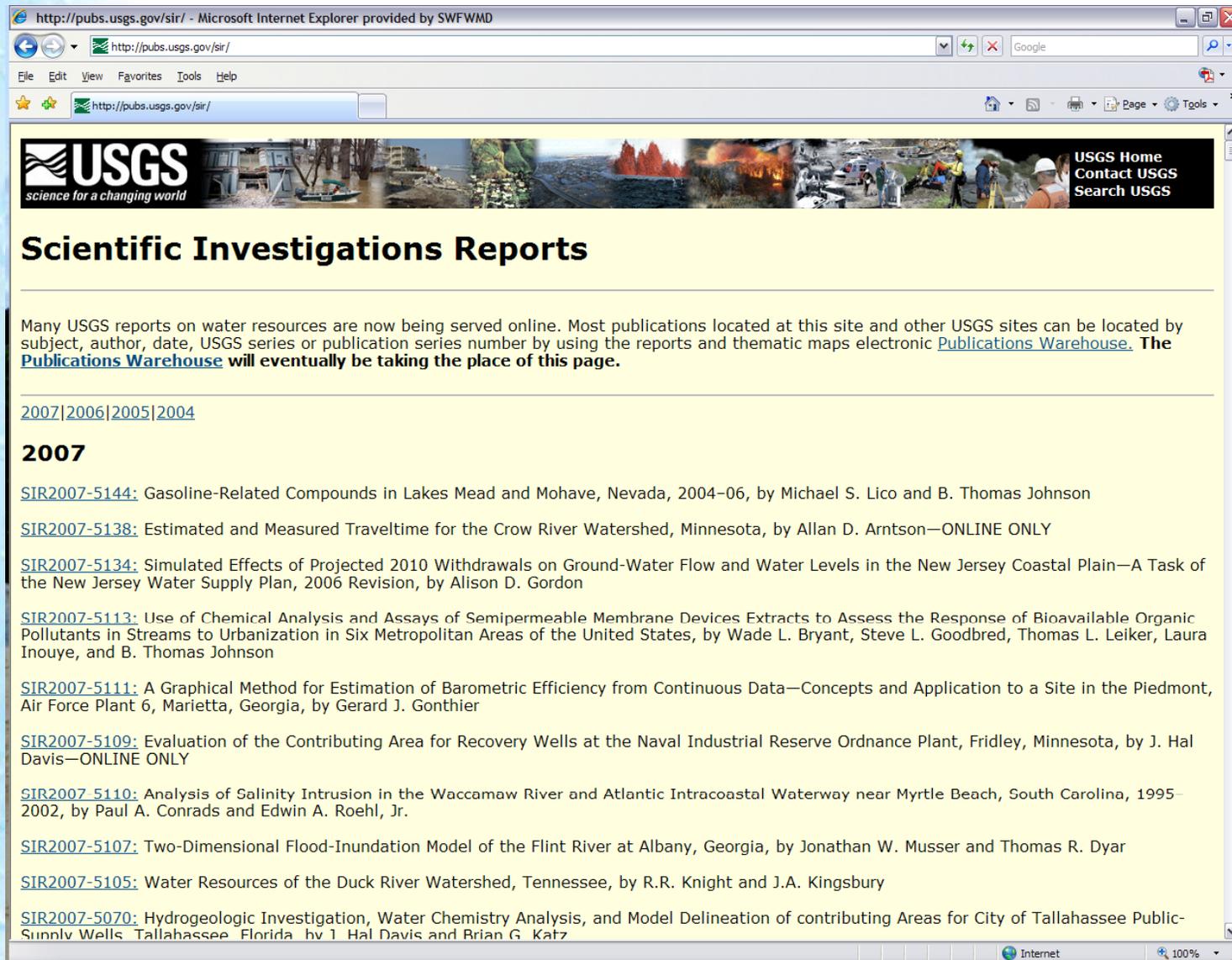
- Finalizing Database of Stressors (Water Use)
- Completing Site-Specific ANNs

- **Next Steps:**

- Compare ANN's and Numerical Models

Data Mining

Report complete by September 2011.



The screenshot shows a Microsoft Internet Explorer browser window displaying the USGS Scientific Investigations Reports page. The browser's address bar shows the URL <http://pubs.usgs.gov/sir/>. The page features a banner with the USGS logo and the tagline "science for a changing world" on the left, and navigation links for "USGS Home", "Contact USGS", and "Search USGS" on the right. The main heading is "Scientific Investigations Reports". Below this, a paragraph states: "Many USGS reports on water resources are now being served online. Most publications located at this site and other USGS sites can be located by subject, author, date, USGS series or publication series number by using the reports and thematic maps electronic [Publications Warehouse](#). The [Publications Warehouse](#) will eventually be taking the place of this page." Below the paragraph are navigation links for the years 2007, 2006, 2005, and 2004. The "2007" section is expanded, listing several reports with their titles and authors:

- [SIR2007-5144](#): Gasoline-Related Compounds in Lakes Mead and Mohave, Nevada, 2004–06, by Michael S. Lico and B. Thomas Johnson
- [SIR2007-5138](#): Estimated and Measured Traveltime for the Crow River Watershed, Minnesota, by Allan D. Arntson—ONLINE ONLY
- [SIR2007-5134](#): Simulated Effects of Projected 2010 Withdrawals on Ground-Water Flow and Water Levels in the New Jersey Coastal Plain—A Task of the New Jersey Water Supply Plan, 2006 Revision, by Alison D. Gordon
- [SIR2007-5113](#): Use of Chemical Analysis and Assays of Semipermeable Membrane Devices Extracts to Assess the Response of Bioavailable Organic Pollutants in Streams to Urbanization in Six Metropolitan Areas of the United States, by Wade L. Bryant, Steve L. Goodbred, Thomas L. Leiker, Laura Inouye, and B. Thomas Johnson
- [SIR2007-5111](#): A Graphical Method for Estimation of Barometric Efficiency from Continuous Data—Concepts and Application to a Site in the Piedmont, Air Force Plant 6, Marietta, Georgia, by Gerard J. Gonther
- [SIR2007-5109](#): Evaluation of the Contributing Area for Recovery Wells at the Naval Industrial Reserve Ordnance Plant, Fridley, Minnesota, by J. Hal Davis—ONLINE ONLY
- [SIR2007-5110](#): Analysis of Salinity Intrusion in the Waccamaw River and Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002, by Paul A. Conrads and Edwin A. Roehl, Jr.
- [SIR2007-5107](#): Two-Dimensional Flood-Inundation Model of the Flint River at Albany, Georgia, by Jonathan W. Musser and Thomas R. Dyar
- [SIR2007-5105](#): Water Resources of the Duck River Watershed, Tennessee, by R.R. Knight and J.A. Kingsbury
- [SIR2007-5070](#): Hydrogeologic Investigation, Water Chemistry Analysis, and Model Delineation of contributing Areas for City of Tallahassee Public-Supply Wells, Tallahassee, Florida, by J. Hal Davis and Brian G. Katz