Introduction
In 2005, groundwater withdrawals averaging 762 million gallons per day (MGD) constituted 95% of the total amount of water withdrawn in the study area. The well-drained karst terrain is the largest component of the water balance for the region's Floridan aquifer system (Jensen, 1994). Consequently, variations in both rainfall and groundwater use can affect water levels and flows in aquifers. Moreover, deterministic models have been developed to quantify cause-effect relationships and to help regulators and other stakeholders manage these regional resources. However, the models have been found to have difficulty simulating the complex interactions between the water system and the surface and subsurface environments in a karst terrain.

The goal of this project was to develop a decision support system (DSS) based on data mining results to complement the deterministic models. A DSS is a powerful, easy-to-use package that combines data, analytical results, predictive models, and supporting graphics that allow resource managers and stakeholders to evaluate alternative management strategies (Roehl and others, 2006).

Description of the Data
Substantial historical hydrologic and climate data were available for data mining (Figure 1). Loss complete groundwater data was also available. They comprised:

- Daily hydrographs for 23 wells (23 Floridan aquifer system, 3 surficial aquifer system), 22 lakes, and 6 springs;
- Daily rainfall, air temperature, and estimated potential evapotranspiration from 18 National Oceanic and Atmospheric Administration (NOAA) sites; and
- Monthly actual and estimated groundwater use representing utility pumping, phosphate mining, agriculture, citrus farming, golf course irrigation, and drainage well recharge.

The completeness (fewer missing data) and quality (more measured and less estimated data) of the data varied significantly. In general, the NOAA meteorological data were the most complete and have the highest quality, followed by the well, lake, spring, and groundwater use data.

Technical Approach
Artificial neural networks (ANNs) are a multivariate, nonlinear curve fitting method from the field of Artificial Intelligence that is commonly used for industrial process modeling and control (Jensen, 1994). Because of delays in availability of groundwater use data, the data mining initially focused on determining the extent to which rainfall, air temperature, and potential evapotranspiration could explain daily variability in the hydrographs from 1942 through 2008. As a first step, an empirical, multi-layer perceptron ANN model was developed for each hydrograph. For inputs to the ANNs, the climate time series were decomposed into deconstructed spectral ranges that had window sizes from 30 days to 6 years to represent the dynamics of the spectral time periods.

The ANN for each site was systematically trained by using sensitivity analyses to culled less predictive inputs. This training sensitivity process revealed that rainfall, temperature, and potential evapotranspiration inputs were removed, resulting in 51 rainfall-only ANNs. For most sites, the data was further reduced with input elimination and tested for independent statistics about model accuracy. This was not possible for some sites because their measurement population was too small.

The groundwater use impacts were subsequently modeled using inputs derived from aggregated data that summed all different types of groundwater use for each month. This approach was necessary because most of the groundwater use data were estimates whose temporal patterns varied slightly, a problem for empirical modeling that cannot be overcome. The aggregated data sites justified the assumption that ground-water conductivity of the Floridan aquifer system that disperses localized impacts, and the one-month time step that dampens transient variability.

The groundwater use data were processed into spectral ranges similar to the rainfall. The 51 groundwater use ANNs simulate the monthly-averaged prediction errors (residuals) of the rainfall ANNs. The residuals represent the portion of the variability in the hydrographs that is not explained by rainfall ANNs. For all few new, testing data were not used because of low measurement populations resulting from the change time step from daily to monthly. Figure 2 shows the outputs of each site's ANN pair are summed to compute a final prediction.

Results and Discussion
Daily Rainfall ANNs - Accuracy statistics (based on testing data when available) for the rainfall ANNs indicate the average coefficient of determination (R²) is highest for the wells, followed by the lakes and then the springs (Table 1). The average percent error is lowest for the wells and highest for lakes and springs. As fitted to the training by least-squares with respect to time (green lines in Figure 3) denote their long-term trends and suggest long-term changes in water use, land use, and other factors. The long-term decreasing trend of Wells A is accurately predicted using rainfall ANNs because rainfall in central Florida was observed to decline. The ANN poorly replicates the long-term trend and more extreme high frequency variability of Well B, which may be caused by pumping and shallow water-table dynamics, respectively.

Springs and lakes are clustered at the northern center of the study area (Figure 1). The springs were sporadically measured for most of the study period, but more frequent measurements were made in the last decade. Spring discharge “flat lining,” or consecutive days of identical flow, is possible due to procedures used to estimate daily data from direct measurement. The late summer and early fall (Figure 2) shows lower frequency variability than the late spring. ANNs were not used on other sites during this period and not predicted by the ANNs. Spring B’s high frequency variability during the last decade is not accurately predicted possibly due to more localized rainfall events not observed in any of the 18 NCRAA rainfall gauges.

The longer-term up and down trending at Lake A is accurately predicted using rainfall ANNs. At Lake B, the minimum water levels in 1973 and 2000 were 22 ft. and were not pumping during a sustained drought.

Monthly Groundwater Use ANNs - Limited improvement in prediction accuracy was gained by incorporation of groundwater use. The R² values for summed rainfall and usage ANNs (Table 2) are similar to those for rainfall ANNs (Table 1), but for monthly time steps. The trend of ANNs can most likely be attributed to the general fact that most of the time at most sites, actual usages are not accurately represented in the largely estimated data, and (or) the variability in the rainfall ANNs manifest forcing that is not represented in the usage data or the ANNs. However, R² values tended to be higher for the springs, suggesting larger usage effects. Limited measurement population produced using testing data for usage ANNs.

Table 1. Statistics for rainfall ANNs. %Error = 100 * root mean square error/historical range.

<table>
<thead>
<tr>
<th>Site</th>
<th>#Sites</th>
<th>Max</th>
<th>Min</th>
<th>Avg</th>
<th>Max</th>
<th>Min</th>
<th>Avg</th>
<th>R²</th>
<th>%Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wells</td>
<td>23</td>
<td>0.31</td>
<td>0.00</td>
<td>0.08</td>
<td>0.91</td>
<td>0.67</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Springs</td>
<td>6</td>
<td>0.56</td>
<td>0.04</td>
<td>0.27</td>
<td>0.74</td>
<td>0.37</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lakes</td>
<td>22</td>
<td>0.35</td>
<td>0.00</td>
<td>0.12</td>
<td>0.90</td>
<td>0.32</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Statistics for usage ANNS (Training R²), and summed rainfall and usage ANNs (Sum R²).

<table>
<thead>
<tr>
<th>Site</th>
<th>#Sites</th>
<th>R²</th>
<th>%Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wells</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Springs</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lakes</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 (above). Measured and predicted hydrographs with residuals for example well, springs, and lakes based on daily rainfall ANNs. The well and lake examples are those having the highest and lowest R². The spring examples are those having the highest and lowest R², and they are shown in Figure 1.

Decision Support System (DSS)
A DSS was developed using Microsoft Excel. It integrates the 102 rainfall and usage ANNs with the historical database, and provides user controls (Figure 4) and streaming graphics to allow users to run simulations having alternative rainfall and groundwater use scenarios (Figure 5). The DSS executes at a monthly time step from 1965 through 2008.

Conclusions
For nearly all sites, groundwater use was found to explain much less of the observed variability in hydrographs than climatic forcing, although relative groundwater use impacts are greater during droughts. These results may be affected by the relatively poor completeness and quality of the groundwater use data. Nevertheless, results indicate that consideration of both climate variability and groundwater use predictions of future hydrologic system behavior would benefit the sustainable management of the resource. The ANN models were embedded in a DSS that will be distributed to resource managers and other stakeholders.

Acknowledgments
The authors thank the St. Johns River Water Management District, South Florida Water Management District, and the Southwest Florida Water Management District for supporting this project. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References


Figure 4. DSS rainfall set point controls. Rainfall data are modulated as either a percentage of historical values or using a constant bias. As shown in the map at right, the sites were grouped based on k-mean clustering of 1,440 daily moving window averages.

Figure 5. DSS simulation and streaming graphics controls. Predicted hydrographs and actual groundwater usage scenarios (green dashed curves) are visualized with the historical hydrograph (blue curve). Black curve indicates the difference between scenario and historical hydrographs.