# Explaining and Predicting Lake Levels using Multivariate Statistical Modeling of Temporally-weighted Rainfall and Spatially-explicit Groundwater Production

Dan Schmutz, Chief Environmental Scientist, Greenman-Pederson, Inc (GPI), Orlando, FL and Devon Moore, Environmental Scientist, City of Winter Haven Natural Resources Division, Winter Haven, FL

## Abstract

Management of lake levels may be improved with an understanding of the contribution of various factors contributing to water level fluctuations. Greenman-Pedersen, Inc. (GPI) performed multivariate statistical analyses of existing data to provide the City of Winter Haven with an improved understanding of the statistical relationship between groundwater production and monitored Environmental Management Plan (EMP) lake levels at five sites under various rainfall scenarios. Groundwater production was represented by the Production Weighted by Distance (PWD) Index, a spatially-referenced index of groundwater production intensity, considering locations and production rates of all active wells for the 12 months preceding the sampling events. Rainfall was represented using a rainfall decay index, calculated as a weighted average of the preceding 60 months of rainfall. A mean-centered multivariate linear regression model was developed using nine years of in-sample data for the rainfall index, PWD, and their interaction which was found to explain 66% of the variability in lake levels. The root mean square error (RMSE) of the model was 1.12 feet, giving an idea of the average expected error for predicting end of the dry (May) and wet season (September) mean water levels. Out-of-sample testing was performed using data collected after statistical model development from the five selected EMP sites for an additional four events: September 2017, May 2018, September 2018, and May 2019. The RMSE value for these four out-of-sample events was 1.13 feet, which is almost identical to the development dataset RMSE of 1.12 feet. The implication of the out-of-sample testing is that the statistical model is continuing to successfully predict lake levels at these five EMP sites on average within about 1 foot of the actual values. We believe this level of accuracy will be useful for water supply management planning purposes.