SPECIAL PUBLICATION SJ2007-SP12

STATISTICAL MODELING OF SPRING DISCHARGE AT ROCK AND WEKIVA SPRINGS IN ORANGE COUNTY, FLORIDA

FINAL REPORT

ST. JOHNS RIVER WATER MANAGEMENT DISTRICT

Statistical Modeling of Spring Discharge at Rock and Wekiva Springs in Orange County, Florida

FINAL REPORT

Prepared for:

St. Johns River Water Management District
4049 Reid Street
Palatka, Florida 32177

Prepared by:

1541 N. Dale Mabry Highway
Suite 202
Lutz, Florida 33548

June 22, 2007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The St. Johns River Water Management District (District) is engaged in hydrologic modeling and data analysis in support of ongoing Minimum Flows and Levels (MFLs) and Water Supply Development projects. MFLs define the frequency and duration of high, average, and low water events necessary to prevent significant ecological harm to aquatic habitats and wetlands from permitted water withdrawals. An integral component of the District's MFL program is the development of long-term daily discharge predictions at various springs in the District. This report describes the development of statistical models for predicting daily spring discharge time series for Rock and Wekiva springs from an assortment of auxiliary data including: (a) previously recorded spring discharge rates at the spring of interest and at adjacent springs, (b) groundwater level measurements from adjacent monitoring wells, and (c) rainfall data from nearby gauging stations.

The presented regression models are based on the statistical correlation between the explanatory and response variables. For example, spring discharge is correlated with aquifer water levels, perhaps with a lead time. This correlation explains some of the variability in the observed spring discharge rates. Further, the correlation is improved using the average water level values rather than the individual measurements which are known to display higher variances.

Stepwise regression analysis was used to build multivariate linear input-output models between the response variable (spring discharge) and the independent variables (spring discharge from nearby springs, water level measurements, lake levels and precipitation) at the springs of interest. Piecewise multiple regression models that incorporated linear temporal trends were constructed due to the relationship between spring discharge and the independent variables changing with time and significant temporal trends in spring discharge. The period of interest (1959-2005) was broken into two time periods, each with a different regression model:

- The period 1959-1997 for Rock Springs, or 1959-2003 for Wekiva Springs, used the water levels in a nearby well (USGS well 283253081283401, Orlo Vista), the moving average discharge at Rock Springs, the rainfall at Orlando, and the year as independent variables. The year was used as an independent variable because there was a linear trend with time that was not accounted for by the other model variables.

- The period 1998-2005 for Rock Springs used two nearby wells (283253081283401 Orlo Vista and 283204081544902 Mascotte Shallow Well).
- Gaps in the Wekiva Springs discharge record over 2003-2005 were filled using linear interpolation for gaps less than 30 days. Gaps in Wekiva Springs discharge larger than 30 days were filled using linear regression on Rock Springs discharge.

The observed daily time-series of Rock Springs from 1998-2002 was used to quantify the error associated with filling data gaps with linear interpolation or regression. For gaps less than 30 days, linear interpolation gave lower errors than regression. Accordingly, gaps less than 30 days were filled with linear interpolation, while for gaps larger than 30 days, the regression was used.

The flow duration curves at both Rock and Wekiva springs changed over the two calibration periods, and were generally lower for the more recent time-period. Though no mechanism is identified for this decrease, it highlights the potential for future flow duration curves to be significantly different from the historically observed flow duration curve.

This report incorporates comments provided by peer review of the first report in this Statistical Modeling of Spring Discharge series. The peer review comments and their resolution as they apply to this report are in Appendix B.

1.0 INTRODUCTION

The Minimum Flows and Levels (MFLs) Program of the St. Johns River Water Management District (District), mandated by state water policy (section 373.042, *F.S.*), establishes MFLs for lakes, streams and rivers, wetlands, and groundwater aquifers. MFLs define the frequency and duration of high, average, and low water events necessary to prevent significant ecological harm to aquatic habitats and wetlands from permitted water withdrawals. The MFLs Program is subject to rule (Chapter 40C-8, F.A.C.) and provides technical support to the District's regional water supply planning process and the consumptive use permitting (CUP) program.

MFLs designate hydrologic conditions that prevent significant harm and above which water is available for reasonable beneficial use. The determinations of MFLs consider the protection of non-consumptive uses of water, including navigation, recreation, fish and wildlife habitat, and other natural resources. MFLs take into account the ability of wetlands and aquatic communities to adjust to changes in hydrologic conditions. Therefore, MFLs allow for an acceptable level of change to occur relative to the existing hydrologic conditions. However, when use of water resources shifts the hydrologic conditions below those defined by the MFLs, significant ecological harm occurs. As it applies to wetland and aquatic communities, significant harm is a function of changes in the frequencies and durations of water level and/or flow events, causing impairment or destruction of ecological structures and functions.

Currently, the District is engaged in hydrologic modeling and hydrologic data analysis in support of the ongoing MFLs and Water Supply Development projects. An integral component of the District's MFL program is the development of long-term daily discharge models at various springs in the District (Osburn et al., 2002). MFLs for two springs in Orange County, Florida, namely, Rock and Wekiva springs, are currently needed. Though both springs have long historical records, there are significant gaps ranging from several days to several years. This study evaluates data availability and applies statistical models to fill in the data gaps and to generate long-term daily discharge simulations and flow duration curves for these two springs.

2.0 OBJECTIVE OF STUDY

The objective of this study is to develop a historic daily spring discharge time series for Rock and Wekiva springs from an assortment of auxiliary data such as: (a) previously recorded spring discharge at the spring of interest and at adjacent springs, (b) groundwater level measurements from adjacent monitoring wells, (c) lake levels from nearby lake-level gages and (c) rainfall data from nearby gauging stations. The study will investigate the correlation structure between various data types and test the applicability of simple multivariate linear models to generate daily discharge records based on these other variables for the common period of record.

This report presents the results of data screening and preliminary statistical analysis for rainfall, groundwater level, and spring discharge data for Rock and Wekiva springs. It also explores the use of empirical models to provide estimates of daily discharge at these springs. These statistical models will take advantage of all available data to try to provide the most accurate estimates. In general, early time records are sparse and often not available for a number of locations. This will require the use of different models ranging in sophistication from simple correlation based models to multivariate regression models which can only be constructed when enough supporting data (e.g., rainfall and groundwater levels) are available at a sufficient number of nearby locations. These models will be used to run a continuous simulation model covering the period of record referenced by the constituent data. From the results of statistical modeling, standard flow-duration analysis for the system (discharge versus percent exceedance for the long-term simulation) will be conducted and standard high- and low flow frequency analyses for the system (frequency of spring discharge for set durations) will be carried out.

This report is organized as follows: Data screening and preliminary statistical analysis are described in Section 3. Section 4 contains the regression modeling methodology and the regression models developed for Rock and Wekiva springs. In section 5, daily discharge predictions are presented along with flow duration curves and frequency analyses for each of these springs. Section 6 contains conclusions and recommendations from this study.

3.0 DATA SCREENING AND PRELIMINARY ANALYSIS

This section summarizes the available data and shows the results of data screening and preliminary statistical analyses conducted for the available time series. The objective of these analyses is to identify the correlation structure between the spring discharge at the springs of interest and the other time series. Results from these analyses will be used to guide the construction of explanatory models which will predict daily discharge values at each spring.

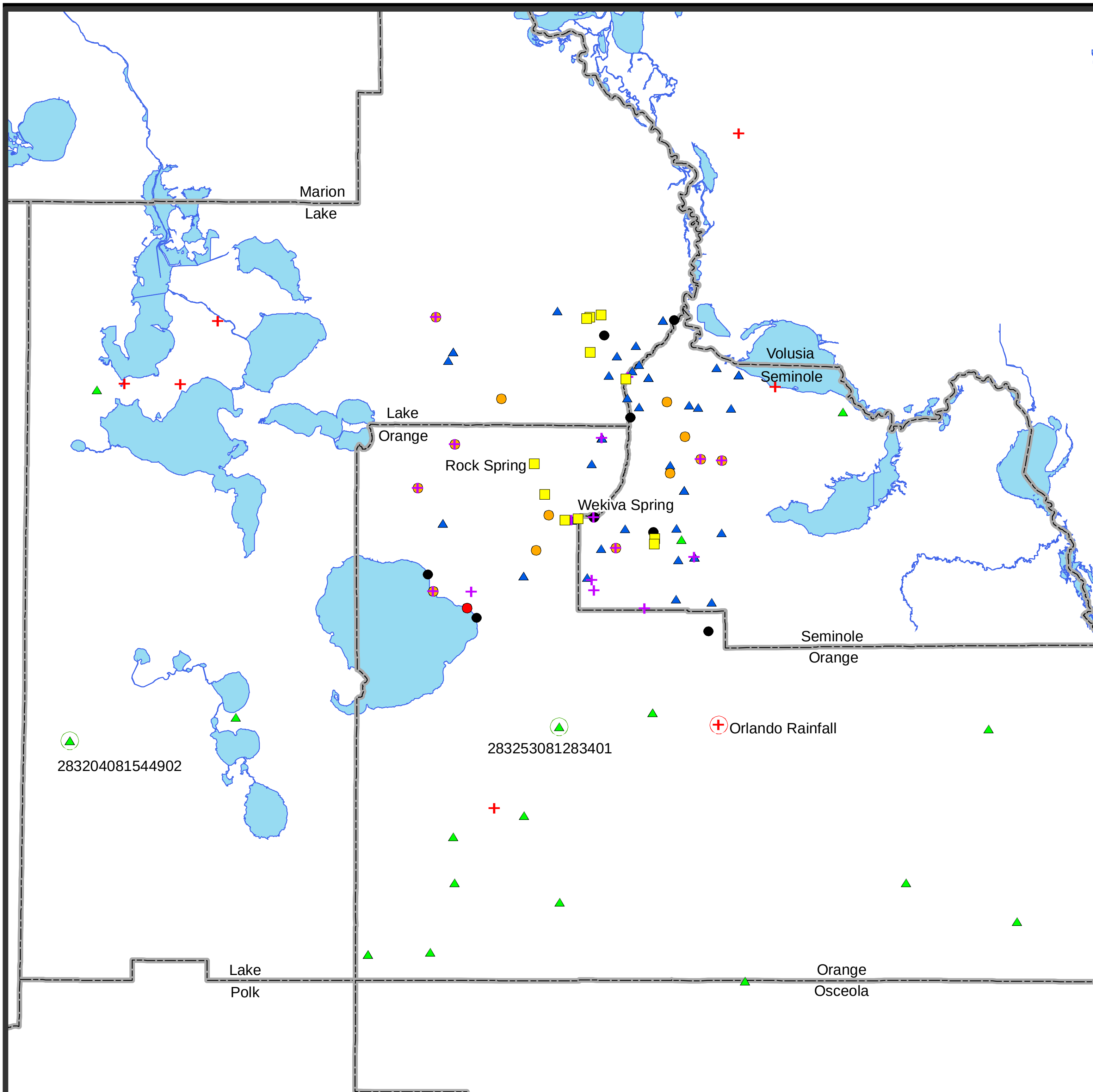
3.1 Data sources

Figure 1 shows a map of the study area and highlights the location of various data sources. Although the map shows numerous groundwater wells around the springs of interest, very few wells have data records with consistent frequency and a long enough period of record to be considered for statistical modeling. The selected groundwater wells with a reasonable data frequency and period of record have been highlighted in the map. Also, one long term NOAA rainfall gage has been selected which is discussed below. The following data sources were used in estimating daily flow duration curves for each spring (Figure 1):

- Measured discharge at Rock and Wekiva springs
- Groundwater level measurements at monitoring wells:
    - 283250381283401, USGS W. 0R47 at Orlo Vista (hereafter w283401)
    - 283204081544902, USGS shallow well near Mascotte (hereafter w544902)
- Precipitation measurements at rain gages:
    - Orlando Rainfall

The above list of data sources includes only the data used in the final regression models. Other water level data from nearby monitoring wells were used to fill in gaps (via regression) in the water levels at wells w283401 and w544902. Those locations were:

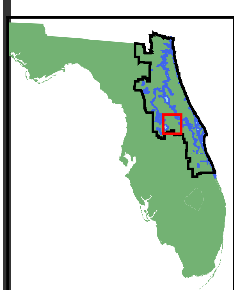
- Groundwater level measurements at monitoring wells:
    - 283249081053201, OR-0007 Bithlo 1 Well at Bithlo, FL (hereafter w053201)
    - 283204081544901, USGS deep well near Mascotte (hereafter w544901)



N

0 2.5 5 10 15 20 Miles

Projection: Universal Transverse Mercator (Zone 17N) NAD 1983 HARN



Spring	NOAA Rain Gage
SJRWMD Lake Gage	SJRWMD Rain Gage
SJRWMD Marsh Gage	SJRWMD Well
SJRWMD Surface Water Gage	USGS Well

Station Data Used in Statistical Model

NOAA Rain Gage USGS Well

Data Sources used in this Study

Filename: SJRWMD_Location.mxd
Last Update: September 14, 2006
For: St. Johns River Water Management District

Figure 1

Variables were selected from the large initial dataset by performing a correlation analysis on all variables versus discharge for each spring. The variable with the highest correlation coefficient was included in the model; other variables were added in a step-wise fashion and only those that increased the R^2 the most were included in the final model (see Section 4.0 below). In order to conduct exploratory data analysis and select the final model variables, a database was compiled of spring discharge (response variable), groundwater levels (explanatory variable) and precipitation (explanatory variable) with a common time basis. Table 1 shows summary statistics (i.e., minimum, maximum, average and standard deviation) for these various data types as well as data for Rock and Wekiva springs.

With respect to the selection of the rainfall station to be used in the analysis, both the Orlando and Sanford rainfall stations were considered. However, the Orlando rainfall station yielded a higher R^2 and more statistically significant regression parameters than the other rainfall stations when included in the final multiple regressions. While the Sanford Rainfall station is closer to the springs of interest, the Orlando station was selected because it falls within the watershed boundary of the springs and gave a slightly higher R^2 than the Sanford rainfall station.

The frequency of observation for each data type was subsequently calculated. This is useful for determining appropriate lag and moving average windows. Moving averages were calculated for recorded water levels, precipitation and spring discharge at adjacent springs at selected lag times: 1, 2, 3, 4, 6, 8, 12, 24, 48, and 52 weeks for use in the regression modeling discussed below. These moving averages act as independent variables and carry useful information regarding the physical state of the system prior to the time of interest.

Table 1 Basic statistics for data at Rock and Wekiva springs, 1931-2005.

Data type	Date Range	N obs	Min	Max	Average	Std Dev
Rock Springs (cfs)	2/5/1931 - 9/30/2005	2426	34.1	83.2	53.4	8.0
Wekiva Springs (cfs)	3/8/1932 - 9/30/2005	666	38.6	91.7	66.8	5.8
Orlando Rainfall (in)	1/1/1942 - 12/31/2004	22643	0	8.4	0.1	0.4
w283401 (ft)	8/1/1943 - 9/30/2005	19963	48.3	80.8	61.8	5.2
w544902 (ft)	1/28/1959 - 9/30/2005	15546	94.9	103.5	100.4	1.4
w544901 (ft)	1/27/1959 - 9/30/2005	15828	93.9	102.7	99.8	1.4
w053201 (ft)	6/2/1961 - 9/30/2005	15892	28.7	43.2	35.9	2.0

3.2 *Frequency analysis*

Table 2 shows the mean, minimum, and maximum of frequency of observation for each data type for Rock and Wekiva springs. Rock Springs had three periods with different frequencies: from 1931-1959, one to three measurements were made per year in 15 different years. From 1960-1997, average measurement frequency was once every 75 days, and daily measurements were made from 10/1/1998 to 9/30/2005. Wekiva Springs also had three periods with different measurement frequencies: from 1932-1959, one or two measurements were made per year in 10 different years; from 1960-2003, average measurement frequency was once every 74 days, and daily measurements began from 4/30/2003, but with frequent data gaps.

Table 2 Data frequency and gap analysis.

Data type	Observation frequency (days)		N gaps	Gap length (days)			Isolated measurements
	Mean	Min		Mean	Min	Max	
Rock Springs (cfs)	143	Daily	250	99	1	3639	1931-1962
Wekiva Springs (cfs)	129	Daily	364	77	1	3641	1931-1967
Orlando Rainfall	Daily	Daily	0	-	-	-	-
WL @ w283401 (ft)	Daily	Daily	203	15	1	101	None
WL @ w544902 (ft)	Daily	Daily	61	26	1	420	None
WL @ w544901	Daily	Daily	34	36	1	419	None
WL @ w053201	Daily	Daily	40	16	1	110	None

Daily data were available for the wells for most of the period of record but with several gaps ranging from a day to several months or a year (Table 2). For gaps less than 15 days, linear interpolation between observed points was used. For well w544902 and gaps longer than 15 days, a regression of the water level in well w544902 on the water level in well w544901 was calculated for data 30 days prior to and after the data gap, and the regression used to interpolate values at w544902 (average R^2 = 0.84). Both wells w544901 and w544902 were missing data for 10/1/2003 to 11/22/2004; a linear regression of well 544902 on well w053201 for 2000-2005 (w544902 = 0.6641*w053201+77.096, R^2 =0.71) was used to predict the levels for w544902 for that time period.

3.3 Analysis of overlap

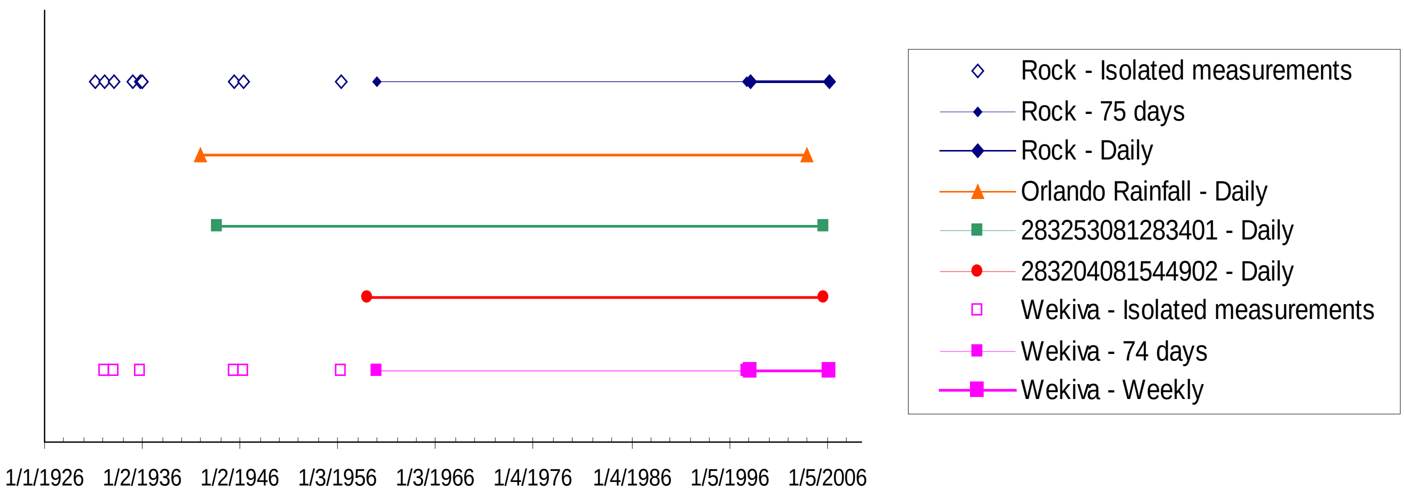
Periods of overlap between different data types were analyzed for each of the springs of interest (Figure 2). This is useful for determining how the period of record can be split up into sub-periods with common sets of explanatory variables. The frequency of observation for each data type was subsequently calculated. Moving averages were calculated for recorded water levels, precipitation and spring discharge at adjacent springs at selected lag times: 1, 2, 3, 4, 6, 8, 12, 24, 48, and 52 weeks for use in the regression modeling discussed below.

Figure 2 shows the overlap between various data types for the Rock and Wekiva springs. Shown here are the periods of record for: (a) Rock and Wekiva spring-discharge, (b) groundwater levels at monitoring wells w283401 and w544902 and (c) precipitation measurement at Orlando. Also indicated therein is the average frequency of observation for each data type (as was discussed in detail in the previous section). As previously mentioned, Rock Springs had three periods with different frequencies: from 1931-1959, one to three measurements were made per year in 15 different years. From 1960-1997, average measurement frequency was once every 75 days, and daily measurements were made from 10/1/1998 to 9/30/2005. Wekiva Springs also had three periods with different measurement frequencies: from 1932-1959, one or two measurements were made per year in 10 different years; from 1960-2003, average measurement frequency was once every 74 days, and daily measurements began from 4/30/2003, but with frequent data gaps.

For the regression modeling, the time-period 1959-2006 was divided into two periods according to the temporal frequency of the measurements of spring discharge, and due to a linear trend in discharge over 1959-1998 (Rock Springs) or 1959-2003 (Wekiva Springs). Using only one time period resulted in poor regression predictions, since the high density of data points over 1998-2005 (Rock Springs) or 2003-2005 (Wekiva Springs) resulted in biased regression parameters.

From 1959, several time series are available which could be used to estimate daily discharge at Rock and Wekiva springs. Several USGS observation wells (N = 19) had daily data for some period over 1959-2005, but only three wells (w283401, w544901, w544901 and w183401) had data from 1959. Data for most of the other wells started in the 1960s. Rainfall

Data Range and Frequency: Rock and Wekiva springs



Date: August 21, 2006	Data Coverage and Frequency for Rock and Wekiva Springs.
File: Fig 2.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida — Figure 2

was available at four rain gage stations; only two stations close to the springs (Orlando and Sanford) showed statistical significance in the regression modeling.

Based on the above discussion of overlap analysis for Rock and Wekiva springs, the following two datasets are used for Partial Correlation Coefficient and Stepwise Analysis to build a regression model:

- Dataset for Rock Springs regression model to predict pre-1998 Rock Springs discharge values:
  - Dependent variable: Rock Springs (203 discharge values from 11/24/1959-12/31/1997)
  - Independent variables:
    - 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of Rock Springs
    - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w283401
    - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w544901
    - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w544902
    - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of rainfall at Orlando
    - Year
- Dataset for Rock Springs regression model to predict post-1997 Rock Springs discharge values:
  - Dependent variable: Rock Springs (2213 discharge values from 1/1/1998)
  - Independent variables:
    - 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of Rock Springs

        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w283401
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w544901
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w544902
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of rainfall at Orlando
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of rainfall at Sanford

- Dataset for Wekiva Springs regression model to predict 1959-2002 Wekiva Springs discharge values:
    - Dependent variable: Wekiva Springs (251 discharge values from 11/25/1959)
    - Independent variables:
        - Daily discharge predictions and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of Rock Springs
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w283401
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of w544902
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48- and 52-week moving averages of w544901
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of rainfall at Orlando
        - Daily observations and 1-, 2-, 3-, 4-, 6-, 8-, 12-, 24-, 48-, and 52-week moving averages of rainfall at Sanford

- Dataset for Wekiva Springs regression model to predict 2003-2005 Wekiva Springs discharge values:
  - Dependent variable: Wekiva Springs (409 discharge values from 1/15/2003)
  - Independent variables:
    - Daily discharge at Rock Springs, 30-day window on either side of gaps.

3.4 Partial correlation coefficient and stepwise analysis

Partial Correlation Coefficient (PCC) is the degree of correlation between any two variables, all others being kept constant. PCCs can be used to find which variables are responsible for multicollinearity. Thus PCCs can be used to drop the explanatory variable(s) which causes multicollinearity. Another option is to include all the variables in the stepwise regression analysis, where variables are added or removed one at a time until no additional variables can be found that improve the goodness-of-fit of the input output model. Stepwise procedures select the most correlated independent variable first, remove the variance in the dependent, then select the second independent which most correlates with the remaining variance in the dependent, and so on until selection of an additional independent does not increase the R-squared by a significant amount (significance = .05). In other words, stepwise regression chooses the variables with the highest partial correlations and includes variables until the partial correlation of all remaining excluded variables with the dependent variable is below some limit. This selection process in a way ensures that no variables with high multicollinearity are picked in the regression model using stepwise regression.

The PCCs and Pearson correlation coefficients for Rock springs is presented in Tables 3 and 4 for the time periods 1959 to 1997 and from 1998 to 2005. For both datasets, some variables had high Pearson correlation coefficients but low PCCs, and vice-versa. This can happen if the independent variable correlates strongly with the other variables in the dataset, which is likely given the number of variables tested (45). In order to obtain a regression with the fewest independent variables and highest adjusted R^2, the variable with the highest Pearson r was selected first for the stepwise analysis. This may result in some variables with a low PCC in the context of the entire dataset being included in the final variable selection, for example the Rock

Table 3 PCCs, Pearson correlation coefficients (r), and variables selected in stepwise regression for the Rock Springs dataset for predicting discharge from 1959-1997.

Variable	PCC	r
Year	-0.32	0.00
W544901	-0.17	0.30
W544902	0.18	0.26
W283401	0.11	0.65
W384601	-0.05	0.51
W384602	0.13	0.43
W040101	0.15	0.51
Apopka Rainfall	0.05	0.11
Deland Rainfall	0.09	-0.05
Leesburg Rainfall	0.21	0.13
Lisbon Rainfall	-0.23	0.02
Orlando Rainfall	-0.28	-0.09
Sanford Rainfall	0.14	0.04
RockSpring.48wk	0.27	0.57
W283401.1wk	-0.09	0.63
W283401.2wk	-0.13	0.63
W283401.4wk	-0.13	0.63
W283401.6wk	0.05	0.62
W283401.8wk	-0.08	0.61
W283401.12wk	0.05	0.59
W283401.24wk	0.15	0.50
W283401.48wk	-0.20	0.34
W544902.1wk	-0.06	0.28
W544902.2wk	0.04	0.25
W544902.4wk	-0.09	0.23
W544902.6wk	0.11	0.24
W544902.8wk	0.01	0.25
W544902.12wk	-0.12	0.22
W544902.24wk	0.00	0.22
W544902.48wk	-0.07	0.08
SANFORD.1wk	0.19	0.27
SANFORD.2wk	-0.03	0.28
SANFORD.4wk	-0.05	0.25
SANFORD.6wk	0.08	0.28
SANFORD.8wk	-0.08	0.34
SANFORD.12wk	0.12	0.39
SANFORD.24wk	0.09	0.45
SANFORD.48wk	0.10	0.59
ORLANDO.1wk	-0.08	0.14
ORLANDO.2wk	0.06	0.14
ORLANDO.4wk	0.02	0.16
ORLANDO.6wk	-0.12	0.11
ORLANDO.8wk	0.13	0.21
ORLANDO.12wk	-0.11	0.29
ORLANDO.24wk	0.16	0.41
ORLANDO.48wk	0.04	0.59

Rock Springs 1959-1997 Selected variables	
Variable	PCC
Year	-0.13
RockSpring.48wk	0.57
W283401.1wk	0.36
ORLANDO.48wk	0.57

Table 4 **PCCs, Pearson correlation coefficients (r), and variables selected in stepwise regression for the Rock Springs dataset for predicting discharge from 1998-2005.**

Variable	PCC	r
Year	-0.17	-0.23
W544901	0.00	0.69
W544902	0.02	0.65
W283401	0.04	0.86
W384601	0.22	0.81
W384602	-0.22	0.78
W040101	-0.10	0.83
Apopka Rainfall	0.03	0.00
Deland Rainfall	0.00	-0.06
Lisbon Rainfall	-0.02	-0.04
Orlando Rainfall	-0.06	-0.10
Sanford Rainfall	0.03	-0.05
RS.48wk Rainfall	0.16	0.40
W283401.1wk	0.03	0.87
W283401.2wk	0.03	0.87
W283401.4wk	0.11	0.87
W283401.6wk	0.01	0.86
W283401.8wk	-0.11	0.84
W283401.12wk	0.02	0.80
W283401.24wk	-0.18	0.59
W283401.48wk	-0.24	0.34
W544902.1wk	-0.06	0.66
W544902.2wk	-0.05	0.66
W544902.4wk	0.02	0.66
W544902.6wk	-0.12	0.65
W544902.8wk	0.02	0.65
W544902.12wk	-0.12	0.62
W544902.24wk	0.23	0.50
W544902.48wk	-0.08	0.44
SANFORD.1wk	-0.01	-0.08
SANFORD.2wk	0.02	-0.05
SANFORD.4wk	0.13	-0.01
SANFORD.6wk	-0.07	0.04
SANFORD.8wk	-0.04	0.09
SANFORD.12wk	-0.14	0.20
SANFORD.24wk	0.04	0.52
SANFORD.48wk	-0.08	0.35
ORLANDO.1wk	0.04	-0.13
ORLANDO.2wk	0.04	-0.09
ORLANDO.4wk	-0.11	-0.05
ORLANDO.6wk	-0.01	0.01
ORLANDO.8wk	-0.19	0.05
ORLANDO.12wk	0.07	0.19
ORLANDO.24wk	0.08	0.57
ORLANDO.48wk	0.19	0.67

Rock Springs 1998-2005 Selected variables	
Variable	PCC
W283401.1wk	0.74
W544902.1wk	0.23

Springs regression for 1998-2005 (Table 4). However, such variables have high and statistically significant PCCs in the final variable dataset (see "Selected Variables" sub-table in Tables 3 and 4). After selection of the first variable, each variable in the list of variables was then added one at a time to test for the PCC of the new variable. At each step, only the variable with the largest PCC value was retained, and variables were added until none had statistically significant PCCs. The PCCs and Pearson correlation coefficients for Wekiva Springs is presented in Table 5 for the time period 1959 to 2003. As with the Rock Springs datasets, Wekiva Springs has some variables had high Pearson correlation coefficients but low PCCs, and vice-versa.

Table 5 PCCs, Pearson correlation coefficients (r), and variables selected in stepwise regression for the Wekiva Springs dataset for predicting discharge from 1959-2003.

Variable	PCC	r
Year	-0.16	0.01
W544901	0.17	0.51
W544902	-0.07	0.46
W283401	-0.07	0.64
W384601	0.36	0.56
W384602	-0.42	0.50
W040101	-0.24	0.52
Apopka Rainfall	0.08	-0.02
Deland Rainfall	-0.02	-0.17
Leesburg Rainfall	-0.26	-0.08
Lisbon Rainfall	-0.07	-0.12
Orlando Rainfall	0.28	0.03
Sanford Rainfall	-0.31	-0.03
Rock.spring	0.44	0.68
RS.48wk	-0.18	0.30
W283401.1wk	-0.17	0.64
W283401.2wk	0.24	0.65
W283401.4wk	0.04	0.64
W283401.6wk	-0.16	0.63
W283401.8wk	0.11	0.62
W283401.12wk	-0.13	0.61
continued on following page		

Table 5 (cont.)

Variable	PCC	r
W283401.24wk	0.17	0.57
W283401.48wk	-0.05	0.37
W544902.1wk	0.13	0.52
W544902.2wk	-0.22	0.48
W544902.4wk	0.09	0.47
W544902.6wk	-0.14	0.48
W544902.8wk	-0.01	0.46
W544902.12wk	0.17	0.42
W544902.24wk	-0.22	0.48
W544902.48wk	0.24	0.38
SANFORD.1wk	0.30	0.33
SANFORD.2wk	-0.02	0.32
SANFORD.4wk	0.08	0.33
SANFORD.6wk	-0.27	0.27
SANFORD.8wk	0.35	0.31
SANFORD.12wk	-0.06	0.27
SANFORD.24wk	0.07	0.44
SANFORD.48wk	-0.25	0.55
ORLANDO.1wk	-0.26	0.00
ORLANDO.2wk	0.23	0.03
ORLANDO.4wk	-0.38	0.03
ORLANDO.6wk	0.24	0.03
ORLANDO.8wk	0.02	0.05
ORLANDO.12wk	-0.22	0.07

Wekiva Springs 1959-2003 Selected variables	
Variable	PCC
Rock.springs	0.49
W283401.8wk	0.10

4.0 REGRESSION MODELING

4.1 Methodology

The objective of regression modeling is to build a multivariate linear input-output model between the response variable (spring discharge) and the surrogate predictor variables (measured values and moving averages of spring discharge, groundwater and precipitation) at the spring of interest (Montgomery and Peck, 1992). Such a relationship can be expressed by:

$$q_t = \beta_0 + \beta_1 q_{t-i} + \beta_2 h_{t-j} + \beta_3 r_{t-k} + \varepsilon \quad (1)$$

where q is spring discharge; h is groundwater level; r is precipitation; ε is a random error term; β_0, β_1, β_2, and β_3 are regression coefficients; t is time, and i, j, and k denote lags that maximize the correlation between the response and predictor variable pair of interest. Since spring discharge may depend on the average groundwater condition or precipitation over some time window, moving averages may also be included in the regression model:

[Spring discharge] = ***f*** {[same spring MA] + [groundwater level MA] +

[precipitation MA] + [adjacent spring MA]} (2)

Depending on the information available for the spring of interest, the regression model can contain all four terms in Eq. (2). This is especially true for periods when detailed measurements of groundwater levels are available.

As described earlier, the model building process can be carried out using stepwise regression, where variables are added or removed one at a time until no additional variables can be found that improve the goodness-of-fit of the input-output model. At each successive step in the regression modeling process, the variable that explains the largest fraction of unexplained variance is included.

The model generated at every step is tested to ensure that the each of the regression coefficients is significantly different from zero. A partial **F**-test, or, an equivalent **t**-test, is used to reject the hypothesis that a regression coefficient is zero, at a $100(1 - \alpha)\%$ confidence level (Montgomery and Peck 1992). The stepwise regression process continues until the input-output model contains all of the input variables that explain statistically significant amounts of variance in the output, i.e., no more variables can be found with a statistically significant regression coefficient.

Even when all variables have statistically significant regression parameters, multicollinearity may inflate the variance of the regression parameter values, causing uncertainty in model predictions, particularly outside the calibration domain. The variance inflation factor quantifies the degree of multicollinearity for each independent variable, and is calculated as:

$$VIFj = \frac{1}{1 - R^2} \tag{3}$$

where the VIFj is the variance inflation factor for independent variable j, and R^2 is the R^2 of the multiple regression of variable j on all other independent variables (Montgomery and Peck 1992). A VIF larger than 10 indicates severe multicollinearity problems.

The workflow for modeling the spring discharge can be summarized as follows:

- Split the period of record into a late-time period, where spring discharge measurements are available, and an early time period where only limited spring discharge measurements are available.
- For each period, organize the spring discharge data (response variable) and the corresponding daily and moving averages of groundwater levels, precipitation, discharge at same spring and discharge at adjacent springs (predictors).
- Retain only those predictor variables for which the number of data points is at least 80% of the number of spring discharge measurements. This threshold has been applied to ensure that the characteristics of the spring discharge time series can be captured as much as possible by the regression model.
- Build a stepwise regression model between spring discharge (response) and some or all of the following predictors: discharge at same spring, discharge at adjacent springs, precipitation, and groundwater levels.

An important point to note here is that these regression models are being built with the "best available data." The quality of the model therefore depends on data coverage, presence of groundwater monitoring wells and lake levels in the immediate vicinity, and availability of discharge measurements at nearby springs that can be used as ancillary data sources.

4.2 Regression models for Rock Springs

High data frequency from 1998-2005 complicated the use of a single regression equation (2) to predict daily flow at Rock Springs. Calibration of a linear model to the entire time-series gave strong weighting to the 1998-2005 period which yielded poor predictions in earlier periods. In order to reduce prediction errors, piecewise regression was used, including two time-periods that reflected data availability: 1959-1997 and 1998-2005.

Step-wise regression was used to identify the optimal model. The first variable added was the one with the highest correlation coefficient with Rock Springs discharge. Additional variables were added one at a time, and only variables with statistically significant regression parameters ($p<0.01$) were included in the final regression. Variance inflation factors (*VIFs*) were computed for each independent variable; a value greater than 10 indicates potentially severe multicollinearity. All of the independent variables in the Rock and Wekiva springs models had VIFs below 10 (1.8-5.0) indicating minimal multicollinearity.

The resulting regression model for 1959-1997 is shown in Table 6 and a plot of predicted versus observed 1959-1997 is shown in Figure 3. The residuals were normally distributed but with some evidence of heavy-tails (Figure 4) which may generate outliers that influence the regression parameters (Montgomery and Peck 1992). The departure from normality is relatively minor, and the regression model is assumed to give unbiased estimates of regression parameters and spring discharge. The resulting regression model for 1998-2005 is shown in Table 7, and the fit and normality of residuals are shown in Figures 5 and 6.

Table 6 Rock Springs - 1959-1997 – Regression Statistics.

Regression Summary for Rock Springs 1959-1997:
R2 = 0.758. F statistic 145.5 on 4 and 186 degrees of freedom. p-value <1E-15. Residual standard error: 3.501 cfs).
β is the regression slope for each variable as in Equation 1, and the p-value is the statistical significance of β.

N = 190	*B*	Std dev of β	p-value	VIF_j
Intercept	210.8	67.7	0.0022	-
Orlando.48wk	46.47	14.9	0.0022	1.8
RS.48wk	0.505	0.070	10^{-10}	2.6
w283401.1wk	0.627	0.102	10^{-8}	2.8
Year	-0.114	0.033	0.0008	2.0

Table 7 Rock Springs - 1998-2005 – Regression Statistics.

Regression Summary for Rock Springs 1998-2005: R2 = 0.9105. F statistic 8882 on 2 and 1803 degrees of freedom. p-value <1E-15. Residual standard error: 2.578 cfs).				
N = 1805	β	Std dev of β	p-value	VIF_j
Intercept	-101.9	3.6	10^{-15}	-
w283401.1wk	1.489	0.023	10^{-15}	2.9
w544902.1wk	0.714	0.046	10^{-15}	2.9

To compare observed versus predicted discharges, it is also useful to consider the variance values for the two records. The F-test for variance equality is often employed for this purpose. This test makes a statistical comparison between the variances of two data sets through the calculation of three values (Ott, 2006):

- Calculated F-value: depends on the variance values for the observed and predicted discharge values and the two sample sizes,
- Critical F-value: depends on the two sample sizes and the desired significance level for the test, and
- P-value: calculated based on the difference between the calculated and critical F-values.

If the Calculated F-value is greater than the Critical F-value then, reject H_0 (the null hypothesis which states that the standard deviations of two normally distributed populations are equal, and thus that they have similar spreads) at the chosen level of confidence (alpha = 0.05). If this is the case then look at the P-value to evaluate the chances of observing an F-value that is greater than the calculated value.

In general, it is expected that regression-predicted values are generally smoother than actual observed discharge values. To quantify the effects of this smoothing on the generated period of record, two tools are used, a quantitative evaluation and visual comparison. The quantitative evaluation is the Kolmogorov-Smirnov (K-S) test which evaluates the differences between the empirical distribution functions for the observed and predicted time-series (D'Agostino and Stephens, 1986). Under the null hypothesis that the two cumulative distribution functions are identical, the test statistic D for this test is the greatest absolute vertical distance between the two empirical distribution functions. This test statistic is not dependent on the two underlying distributions. Therefore the p-value for this test is only dependent on the two sample sizes, which can be different.

The K-S D statistic can be used to evaluate if the two cumulative distributions functions (CDFs) are statistically similar. Another qualitative tool often employed to compare two data sets is the box-whisker plot (also known in the literature as the box plot, Ott, 2006). This plot is a convenient way of graphically depicting the location and spread of the two (or more) data sets. The plot shows the smallest observation, lower quartile (Q1), median, upper quartile (Q3), and largest observation. Furthermore, the plots show which observations, if any, are considered to be outliers. These plots visually show different types of populations, without any assumptions of the statistical distribution or requirements about the sample sizes. The box size (difference between Q3 and Q1) helps indicate variance. Skew is also graphically shown through (1) the location of the median in relation to Q1 and Q3, (2) the maximum and minimum values, and (3) the number of value of outliers.

Table 8 shows the F-test and K-S test between observed Rock spring time-series and predicted Rock Springs time-series on days corresponding to observed data. Results for the F-test indicate that there is no significant difference between the two variances; with a 22% chance of observing the calculated F-value under the equal variance hypothesis for this sample size. However, the K-S D statistic shows a significant difference between the two empirical CDFs.

Figure 7 shows the box-whisker plots for three data sets:

(1) observed discharge values at Rock Springs for the time period 1959-2005,

(2) regression-predicted values for the same dates at which observed discharge value are available. These predicted values come from two different regression models as described above, and

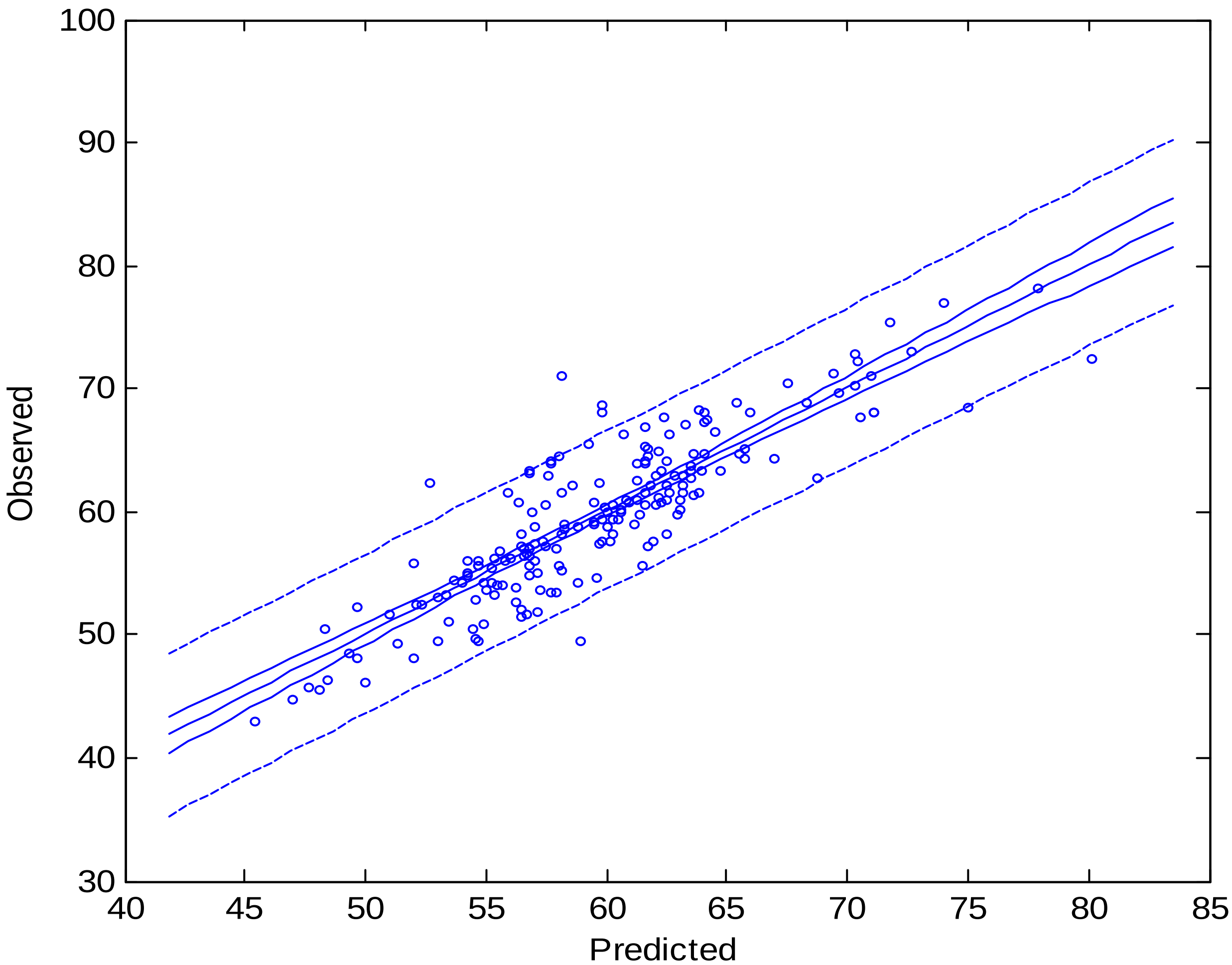
(3) regression-predicted values from the two regression models for each day in the time period 1959-2005.

The plots show that the observed discharge values at Rock Springs show slightly higher variability than the regression-predicted values (data sets 1 and 2). However, data set 3 (which shows a complete record of pooled model predictions) shows slightly higher variability than data set 2. This shows that the regression predictions show slightly higher variability than the observed values. It is expected, however, that more variance would have been observed if more observations had been made in the same time period. In conclusion, the regression-predicted values show a similar range of variability as the observed discharge values with the complete daily predicted record showing plausible variability.

Table 8 Rock Springs - 1959-2005 – Observed and Regression-Predicted Variance Statistics.

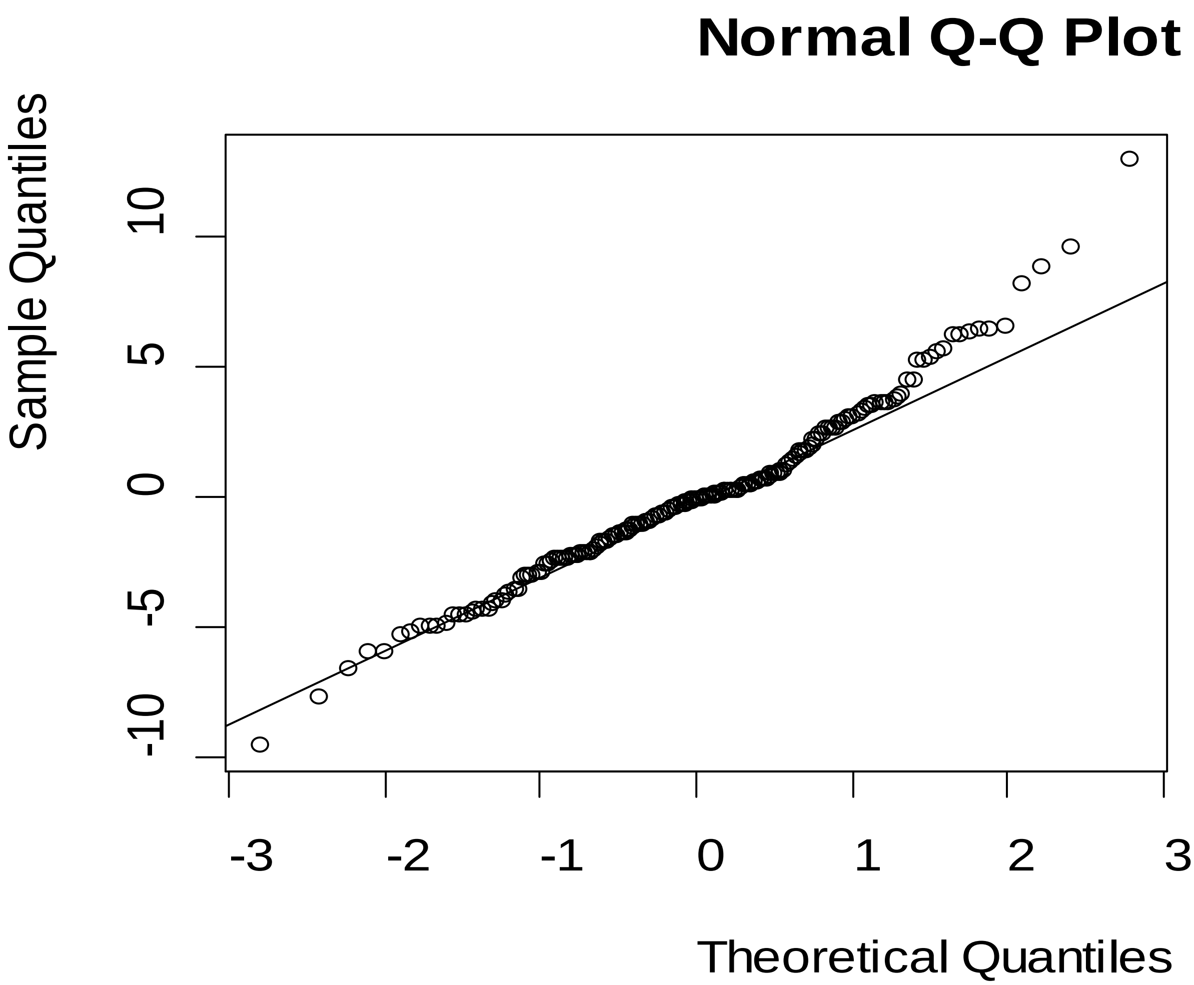
	Rock (predicted)	*Rock (observed)*
Mean	53.95	53.37
Variance	66.00	63.99
Observations	2411	2411
df*	2410	2410
F	1.03	
P(F<=f) one-tail	0.22	
F Critical one-tail	1.07	
K-S D statistic	0.07	
p-value for K-S test	0.00	

* df are the degrees of freedom which are equal to the sample size minus 1 for the F-test.

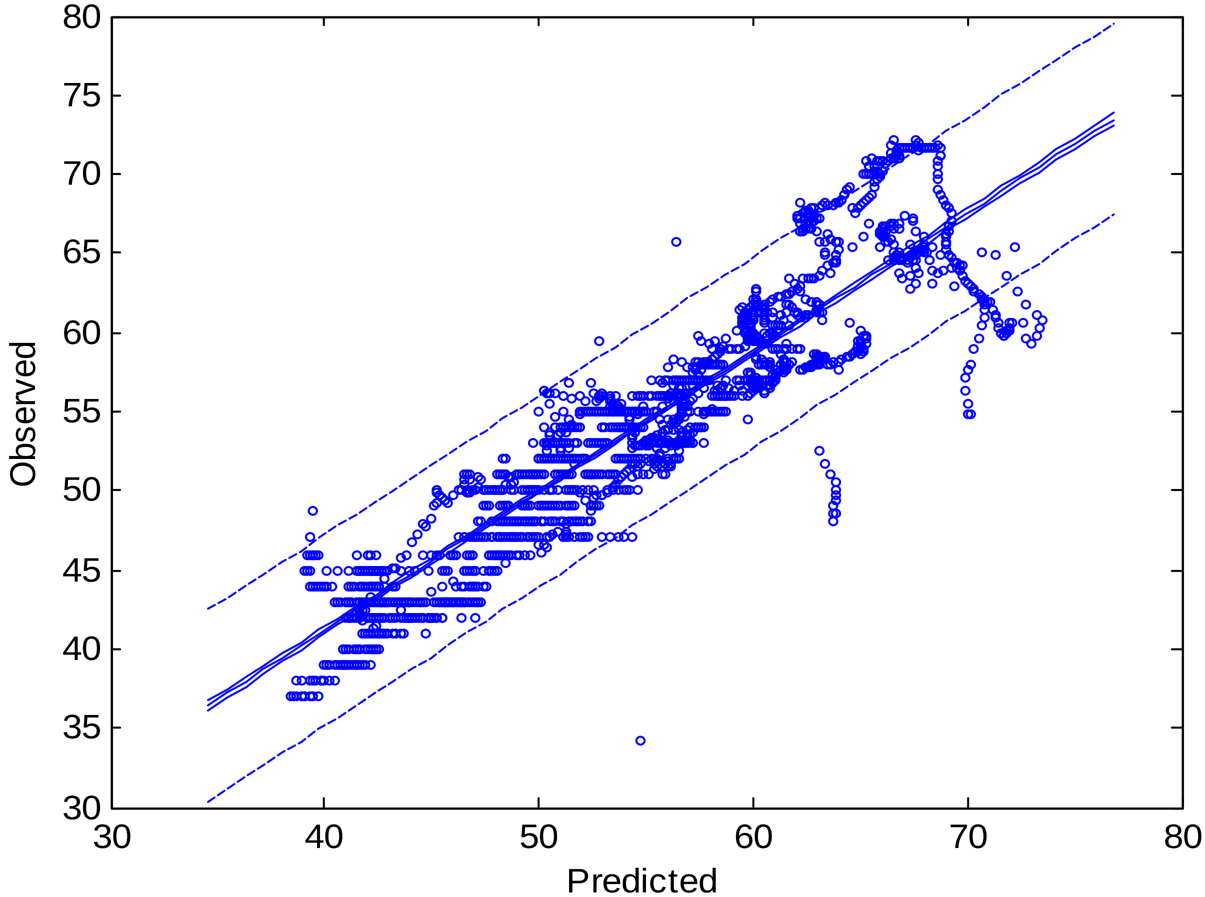


Solid Lines are the Pointwise Confidence Interval (95%) and Dashed Lines Indicate 95% Confidence Intervals for Predictions.

Date: August 22, 2006	Predicted versus Observed Discharge (cfs) for Rock Springs, 1959-1997.
File: Fig 3.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida — Figure 3



Date: August 22, 2006	Normal Probability Plot of Residuals for the Rock Springs Regression, 1959-1997.
File: Fig 4.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida
	Figure 4



Solid Lines are the Pointwise Confidence Band (95%), and Dashed Lines Indicate 95% Confidence Intervals for predictions.

Date: August 22, 2006

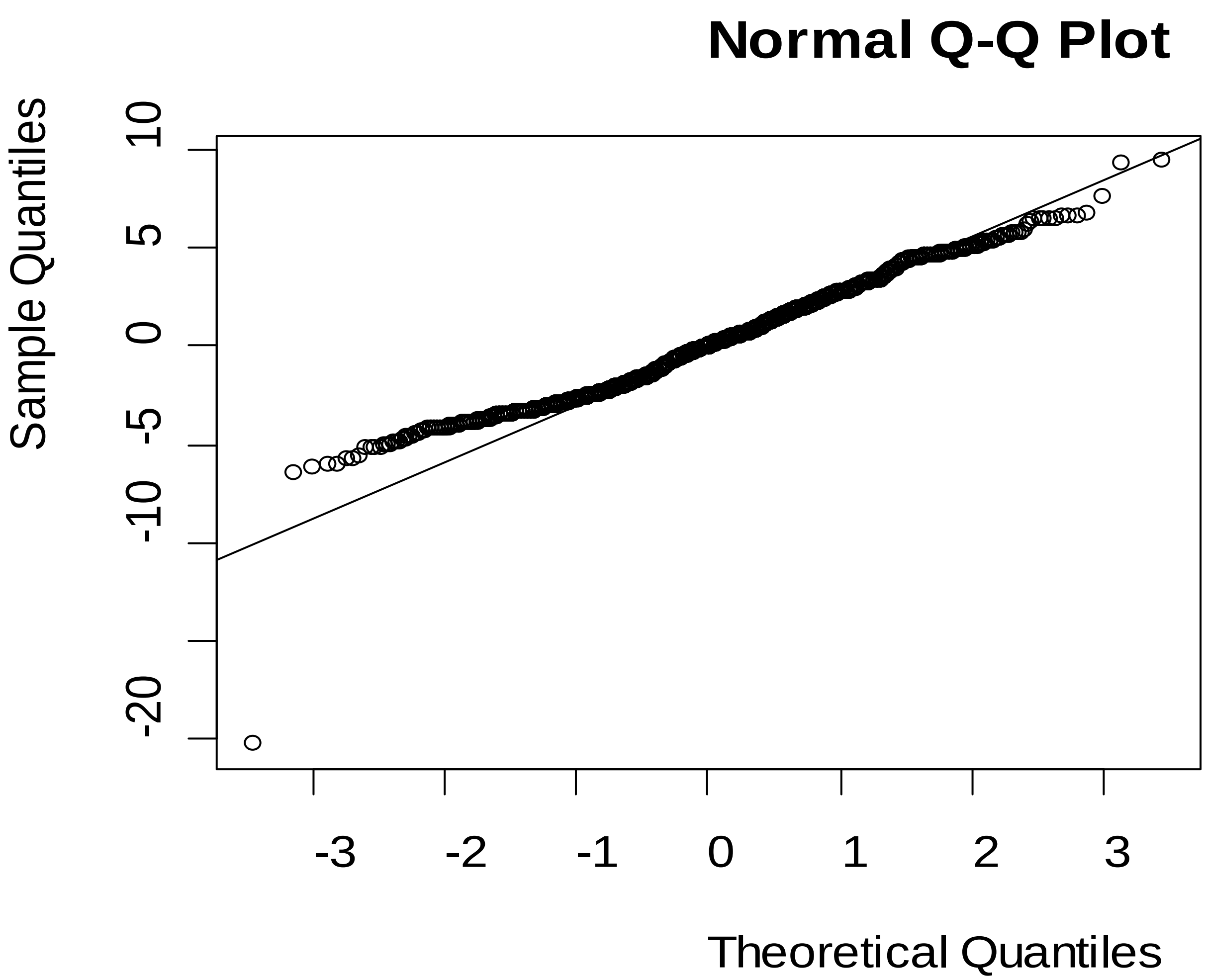
File: Fig 5.pdf

INTERA

Predicted versus Observed Discharge (cfs) for Rock Springs, 1998-2005.

St. Johns River Water Management District
Palatka, Florida

Figure 5



Date: August 22, 2006

File: Fig 6.pdf

Normal Probability Plot of Residuals for the Rock Springs Regression, 1998-2005.

St. Johns River Water Management District
Palatka, Florida

Figure 6

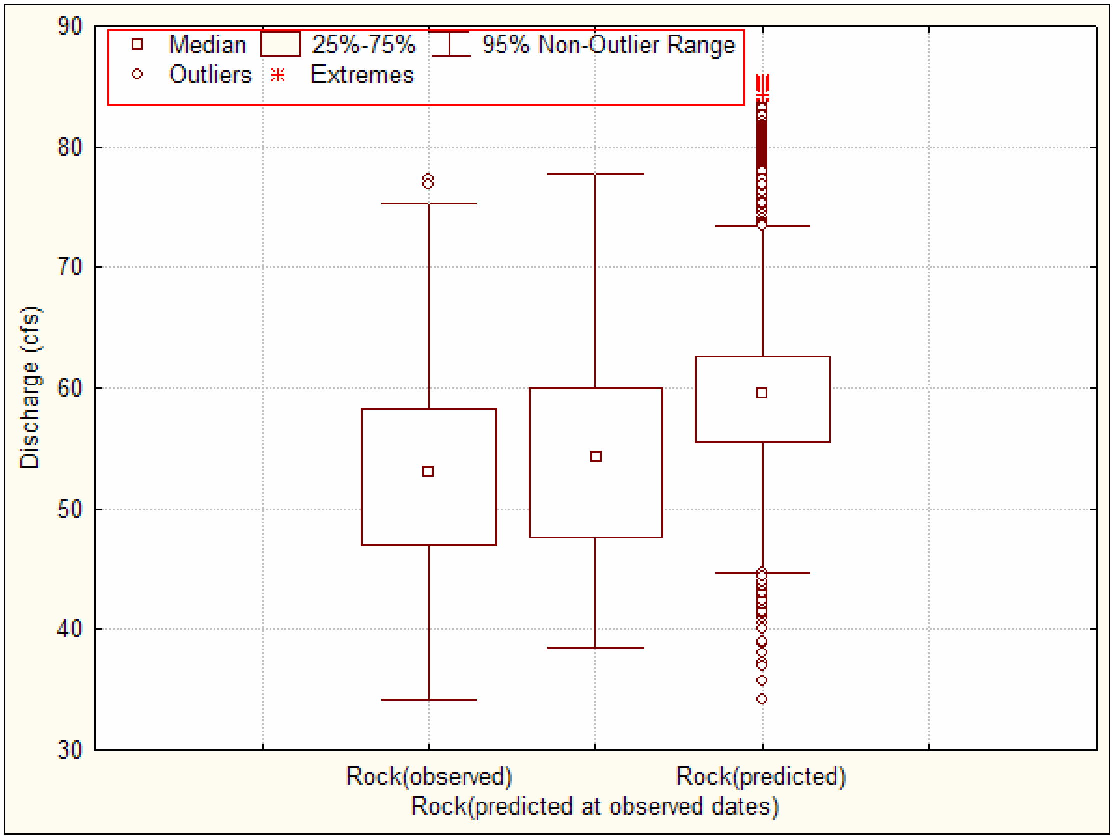


Figure 7 Box-Whisker Plots for Observed and Regression-Predicted Discharge Value for Rock Springs Regression, 1959-2005.

4.3 Regression models for Wekiva Springs

Wekiva Springs had much less data than Rock Springs, so Rock Springs was used as an independent variable in the regression. Wekiva Springs discharge correlated well with Rock Springs discharge (R^2 = 0.65) until 2003-2005 (Figure 8). The time series of Wekiva Springs was divided into two periods, 1959-2002 and 2003-2005, since daily data with gaps was available from 2003-2005 and monthly or tri-monthly data was available from 1959-2002. As with Rock Springs, separating the data was important for preventing dominance of the period with daily records, which reduces model fit and predictive capability in the earlier period with less data. Table 9 lists the resulting regression model and the fit and normality of residuals are shown in Figure 9 and Figure 10, respectively.

Starting in April, 2003, Wekiva Springs had a near daily time-series with gaps of less than 15 days, so the small gaps were filled using linear interpolation. For the 2003-2005 period, gaps larger than 30 days were filled using linear regression on Rock Springs discharge. The linear regression was computed using observed discharge at Wekiva Springs as the dependent variable and observed discharge at Rock Springs as the independent variable over a moving window 30 days before and after the gap.

Table 9 Wekiva Springs - 1959-2002 – Regression Statistics.

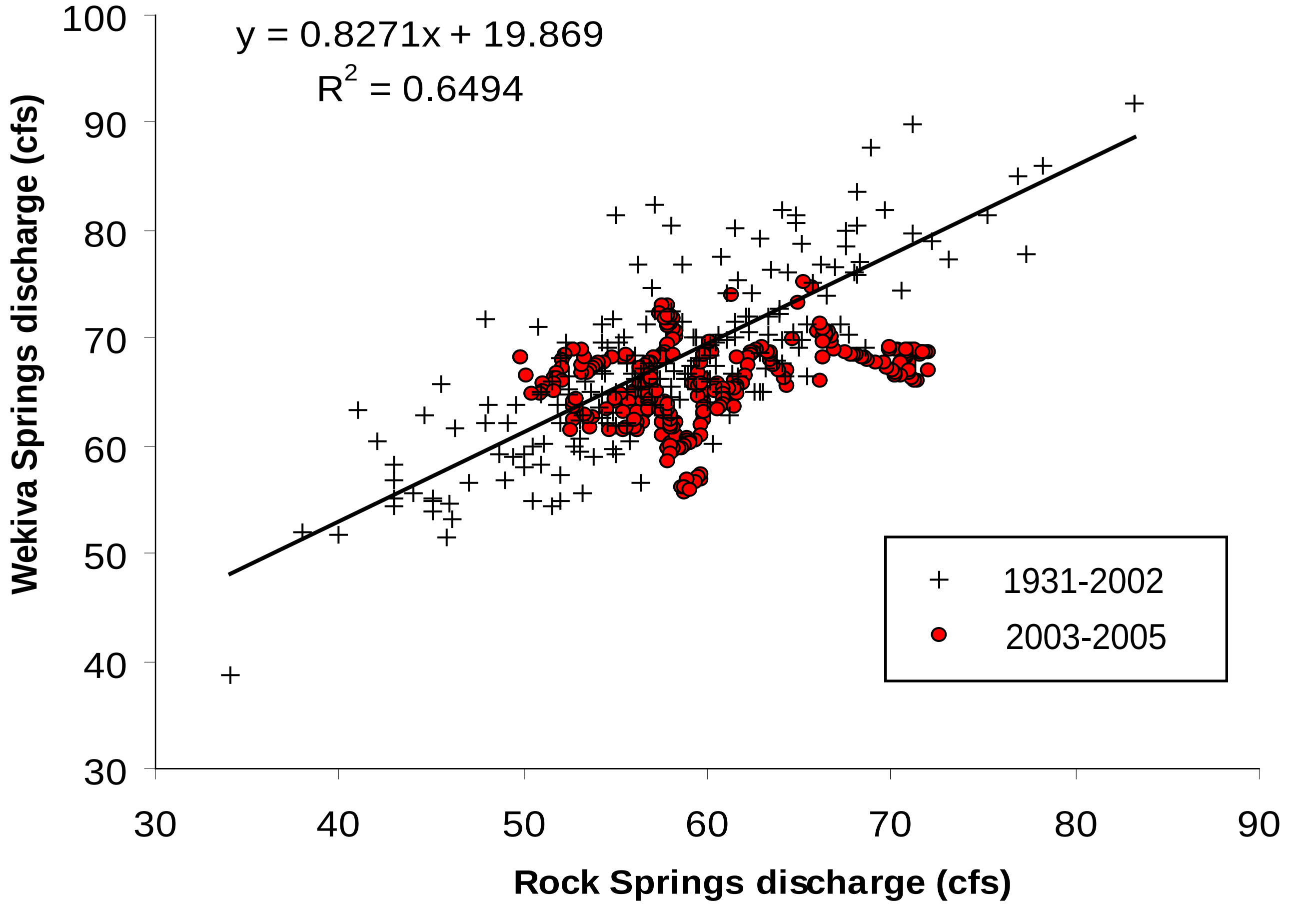
Regression Summary for Wekiva Springs 1959-2002: R2 = 0.6849. F statistic 207.6 on 4 and 146 degrees of freedom. p-value <1E-15. Residual standard error: 4.107 cfs				
N = 148	β	Std dev of β	p-value	VIF_j
Intercept	2.727	5.304	0.608	-
Rock_spring_daily	0.60443	0.07448	10^{-13}	5.0
w283401.8wk	0.50994	0.14112	0.0004	5.0

To compare observed versus predicted discharges, the same methods described before for Rock Springs are used for Wekiva Springs. Results for the F-test and K-S D statistic are shown in Table 10. Results for the F-test indicate that there is a statistically significant difference between the two variances; with values of 32.26 and 21.63 for the observed and regression-predicted values, respectively. The K-S D statistic shows a similar significant difference between the two empirical CDFs.

As mentioned before for Rock Springs, the F-test and the K-S D statistic do not show the nature of the difference between the two time series. To provide some insight into these differences, Figure 11 shows the box-whisker plots for the observed and regression-predicted discharge values (along with the complete regression-predicted period of record). The plots show that the differences between the observed and predicted values are largely due to the existence of more outliers and extreme values in the observed time series. The non-outlier range is almost identical for the two time series. As with Rock Springs, data set 3 (which shows a complete record of pooled model predictions) shows much more variability than data set 2, with an overall variability that is slightly higher for the observed record. It is expected, however, that more variance would have been observed if more observations had been made in the same time period. In conclusion, the regression-predicted values show a reasonably similar range of variability as the observed discharge values with the complete daily predicted record showing plausible variability.

Table 10 Wekiva Springs - 1959-2005 – Observed and Regression-Predicted Variance Statistics.

	Wekiva(observed)	*Wekiva(predicted)*
Mean	66.54	65.30
Variance	32.26	21.63
Observations	633	633
df	632	632
F	1.49	
P(F<=f) one-tail	0.00	
F Critical one-tail	1.14	
K-S D statistic	0.20	
p-value for K-S test	0.00	

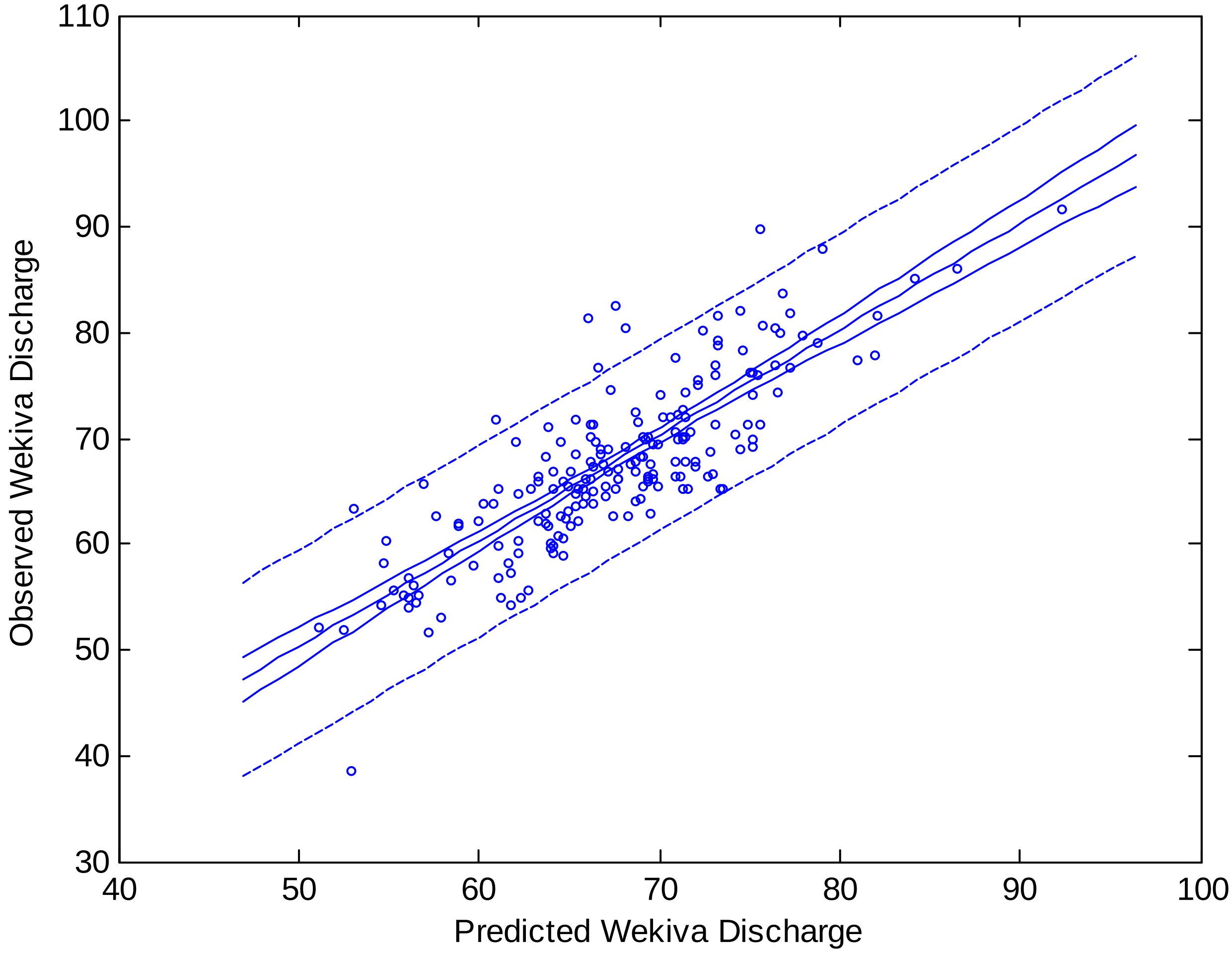


Date: August 22, 2006	Wekiva Springs Discharge versus Rock Springs Discharge over 1931-2002 (+) and from 2003-2005 (red dots).
File: Fig 8.pdf	

INTERA

St. Johns River Water Management District
Palatka, Florida

Figure 8

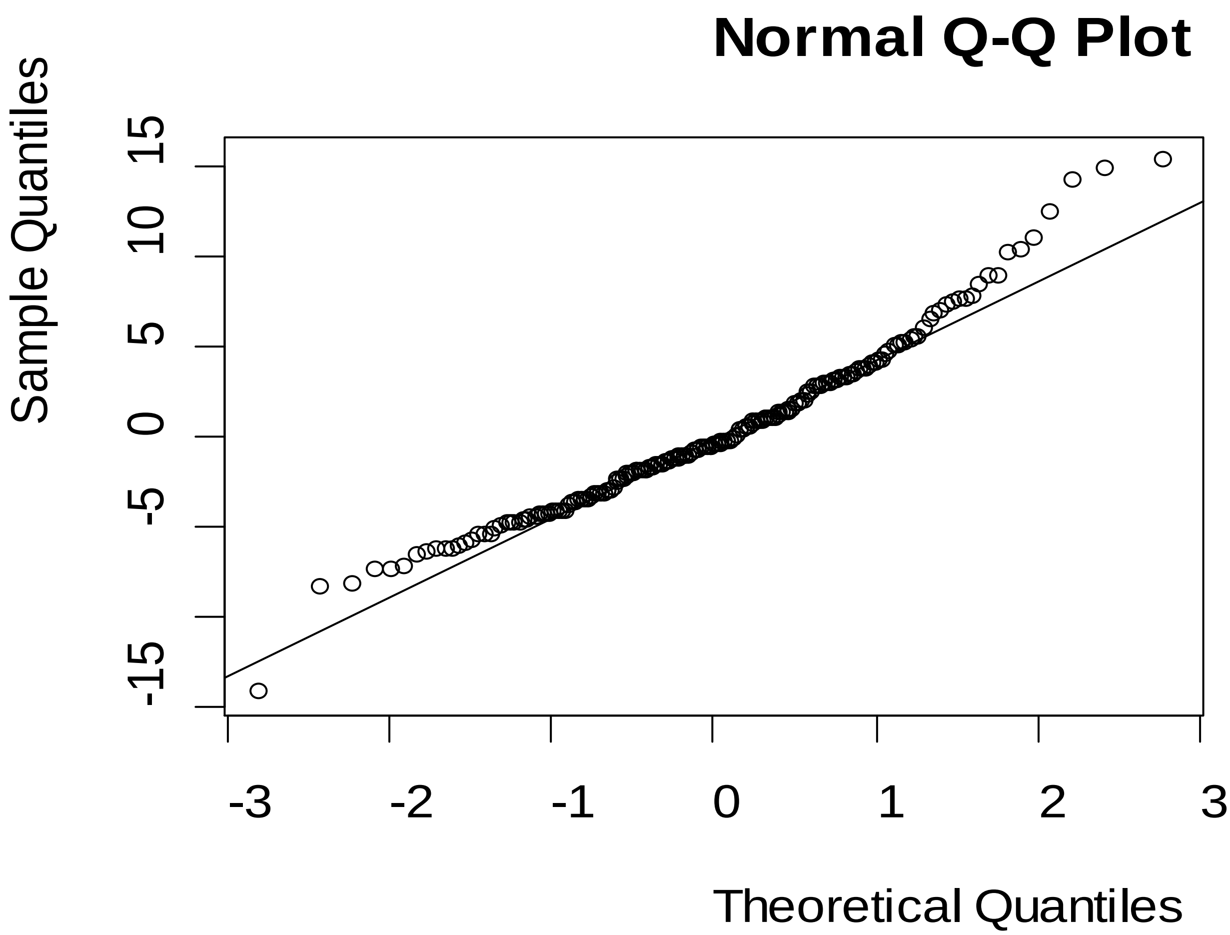


Date: August 22, 2006	Wekiva Springs Regression Model, Predicted versus Observed, 1959-2002.
File: Fig 9.pdf	

INTERA

St. Johns River Water Management District
Palatka, Florida

Figure 9



Date: August 22, 2006	Normal Probability Plot for Residuals for the Wekiva Springs Regression, 1959-2002.
File: Fig 10.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida — Figure 10

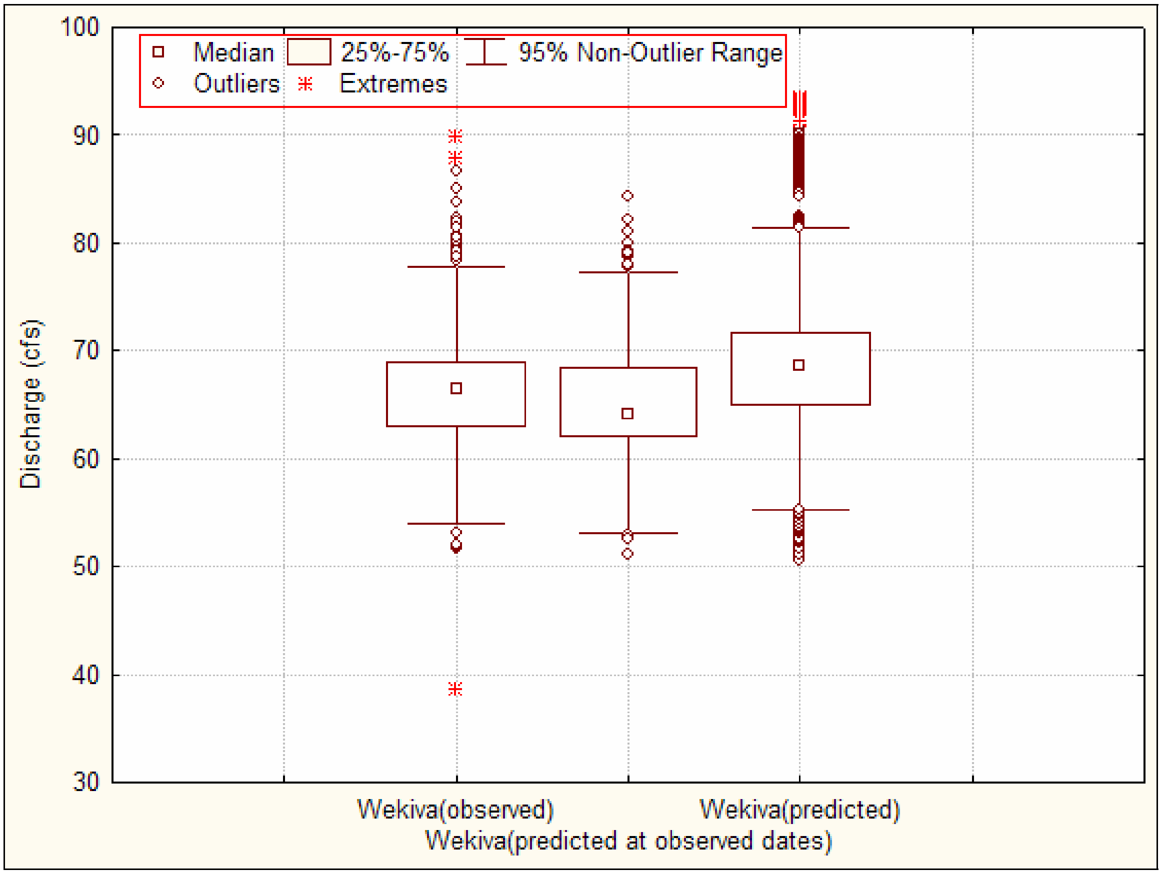


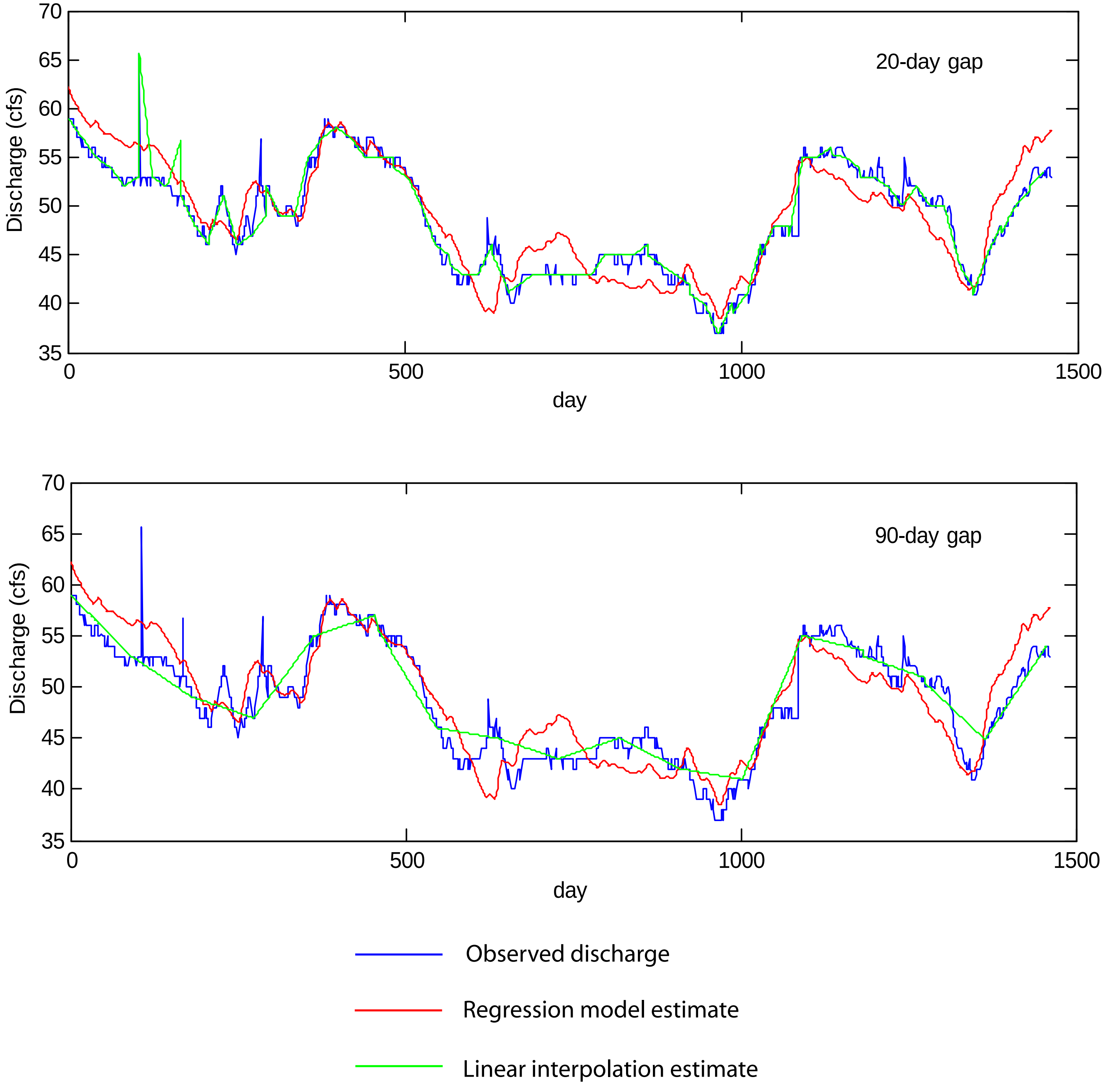
Figure 11 Box-Whisker Plots for Observed and Regression-Predicted Discharge Value for Wekiva Springs Regression, 1959-2005.

5.0 PREDICTION OF DAILY DISCHARGE AND FLOW DURATION

5.1 Methodology: Regression models and linear gap filling

The objective of regression modeling and gap filling is to provide a historic daily time series of spring discharge, using both observed discharge data and other observed data such as well levels and rainfall. The Rock and Wekiva springs had several years of daily data from 1998-2005 and several years with data observation frequencies of 45 days or better. In order to not discard observed data, the final data series included the observed data points, and gaps filled using one of two methods 1) multiple regressions, which predicted the discharge on unmeasured days given the well levels and rainfall and 2) linear interpolation between observed discharge points. Linear interpolation was most valid for small data gaps (1-30 days) and regression was more valid for longer gaps, where the discharge between two observed days may not be linear. The daily time-series for Rock Springs from 10/1/1998-9/30/2002 had no gaps, and was used to calculate the errors generated by using either linear interpolation or regression. For example, Figure 12 shows how linear interpolation fits the observed time series better for 20-day gaps, but regression performs better for 90-day gaps.

Whether regression or linear interpolation is used to fill gaps depends on the gap length and the R^2 of the regression model, with longer gaps requiring regression. In order to determine the threshold gap length where linear interpolation will be used instead of regression, the daily time-series for Rock Springs from 10/1/1998-9/30/2002 was used to generate synthetic daily discharge time series using linear interpolation for a range of gaps (10-120 days). Flow duration curves (FDC) were then generated for the observed data (the "true" FDC), the linearly interpolated data, and the regression time-series (Figure 13). The root mean square error (RMSE) between the observed flow duration curve and the FDCs generated using either linear interpolation or regression for different gap lengths (10–120 days) were calculated by summing the squared difference in fraction exceedance for each discharge amount, and taking the square root of the sum after dividing by the number of discharge values.



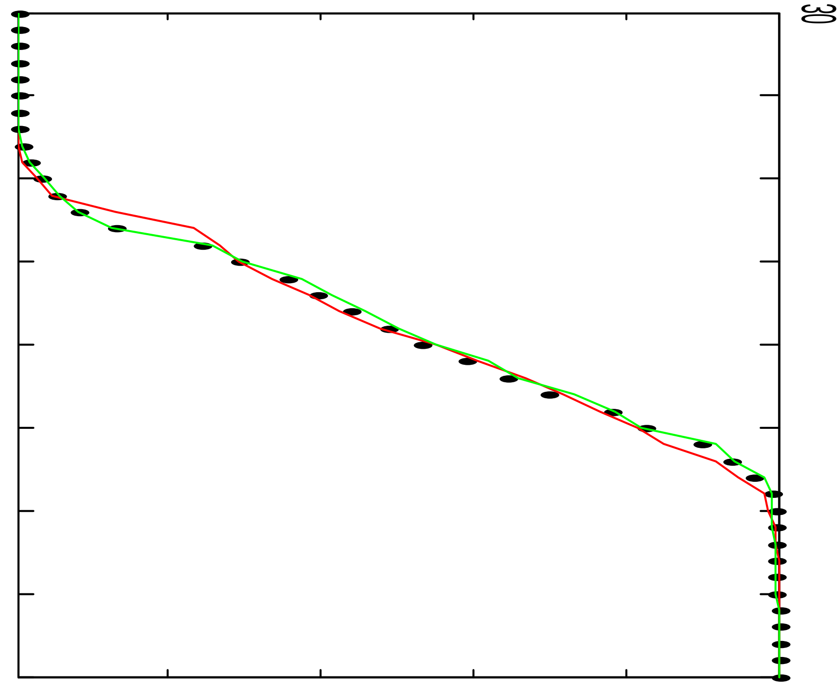
Date: August 22, 2006

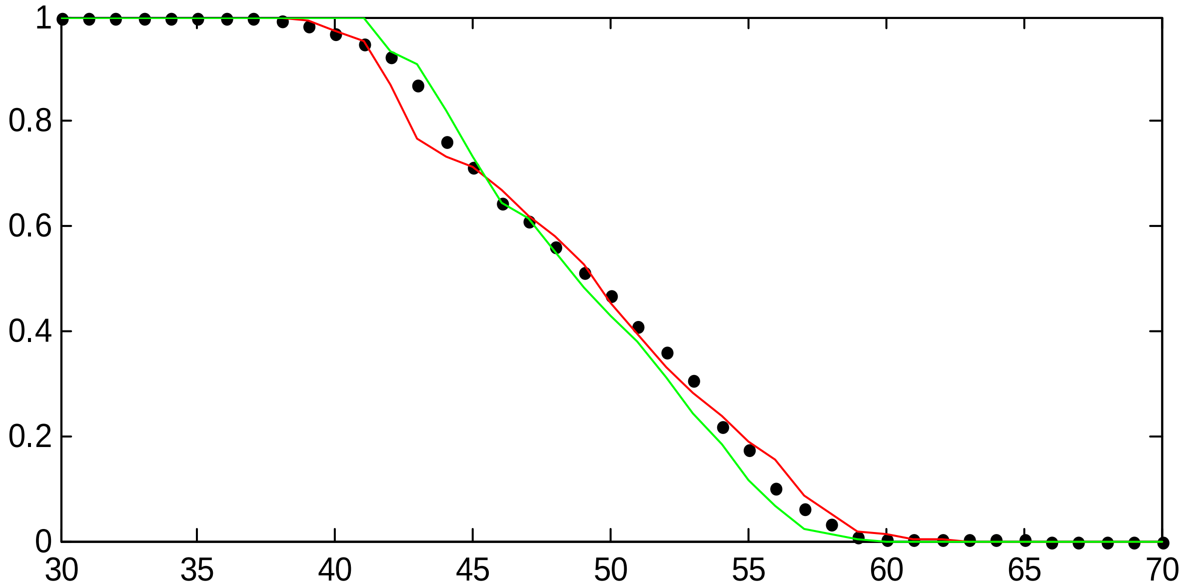
File: Fig 12.pdf

Observed Daily Discharge at Rock Springs and Interpolated Time Series using Linear Interpolation and Regression, 10/1/1998-9/30/2002, for 20-day Gap and 90-day Gap.

St. Johns River Water Management District
Palatka, Florida

Figure 12





Date: August 23, 2006

File: Fig 13.pdf

Flow-Duration Curves for Rock Springs, 10/1/1998 to 9/13/2002 using Observed Data (black dots), Linear Interpolation (green line) and Regression, for 20-day Gap and 90-day Gap.

INTERA

St. Johns River Water Management District
Palatka, Florida

Figure 13

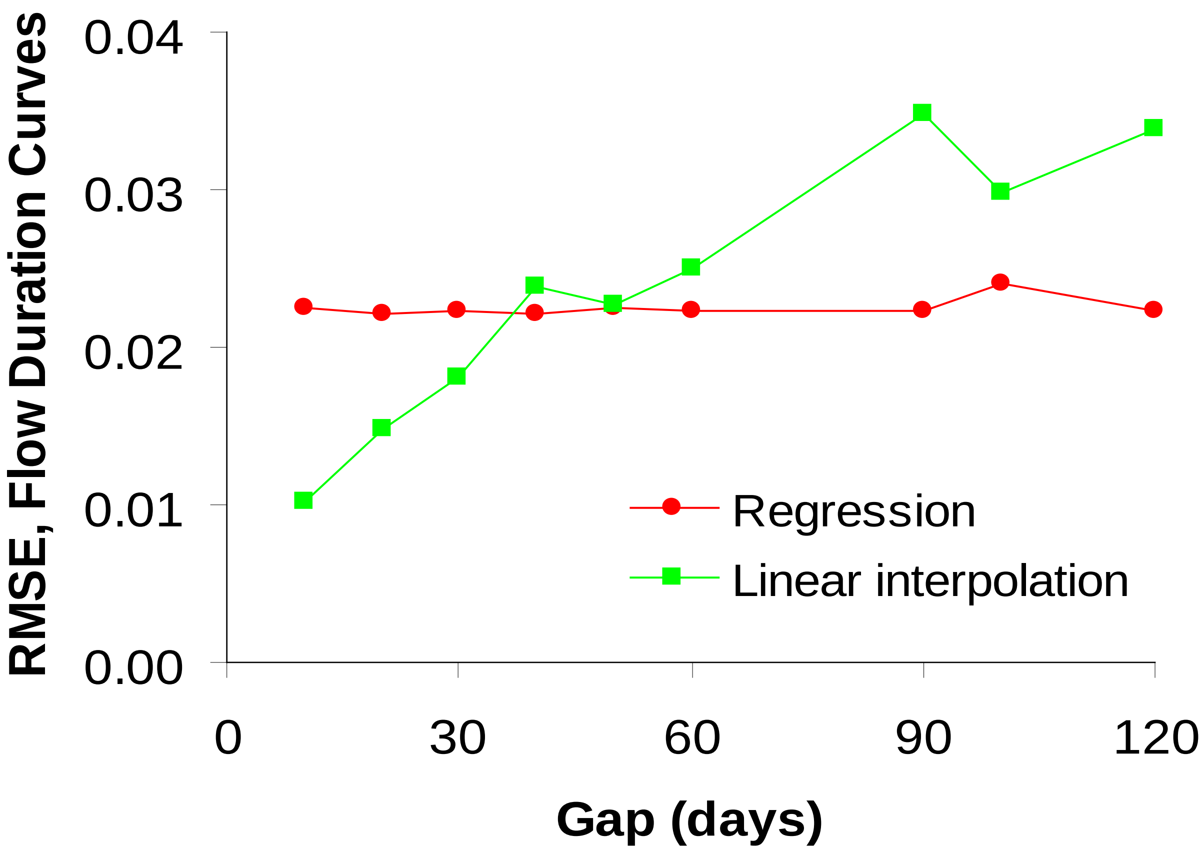
The plot of RMSE for the flow duration curve versus gap length (Figure 14) suggests that below a 30-50 day gap length, linear interpolation gives lower error than regression. This is a minimum threshold for the gap length for linear interpolation, since the R^2 of the regression over 1998-2005 was high ($R^2 = 0.91$). For a lower R^2, a higher gap threshold may provide more accurate flow duration curves. Here we generate two daily discharge time-series and flow duration curves: one set with a gap threshold of 30 days, and another with gap threshold of 60 days. This will provide an estimate of the effects of using each threshold gap length for the overall flow duration curve.

5.2 Daily discharge and flow duration curves for Rock Springs

The Rock Springs daily discharge series was calculated using the estimate from the linear regression model on gaps larger than 30 days, and linear interpolation on gaps smaller than 30 days. The time-series of observed and modeled discharge shows generally good agreement for 1959-1997 (Figure 15) and 1998-2005 (Figure 16). The time-series from 1959-1997 had few gaps smaller than 30 days, so the regression model was used to predict discharge for most days during that time period. The 1998-2005 period, by contrast, had many small gaps that were filled using linear interpolation. In Figure 16, the blue line is the regression prediction; the yellow line is produced using linear interpolation between points for gaps less than 30 days and the regression model for gaps longer than 30 days.

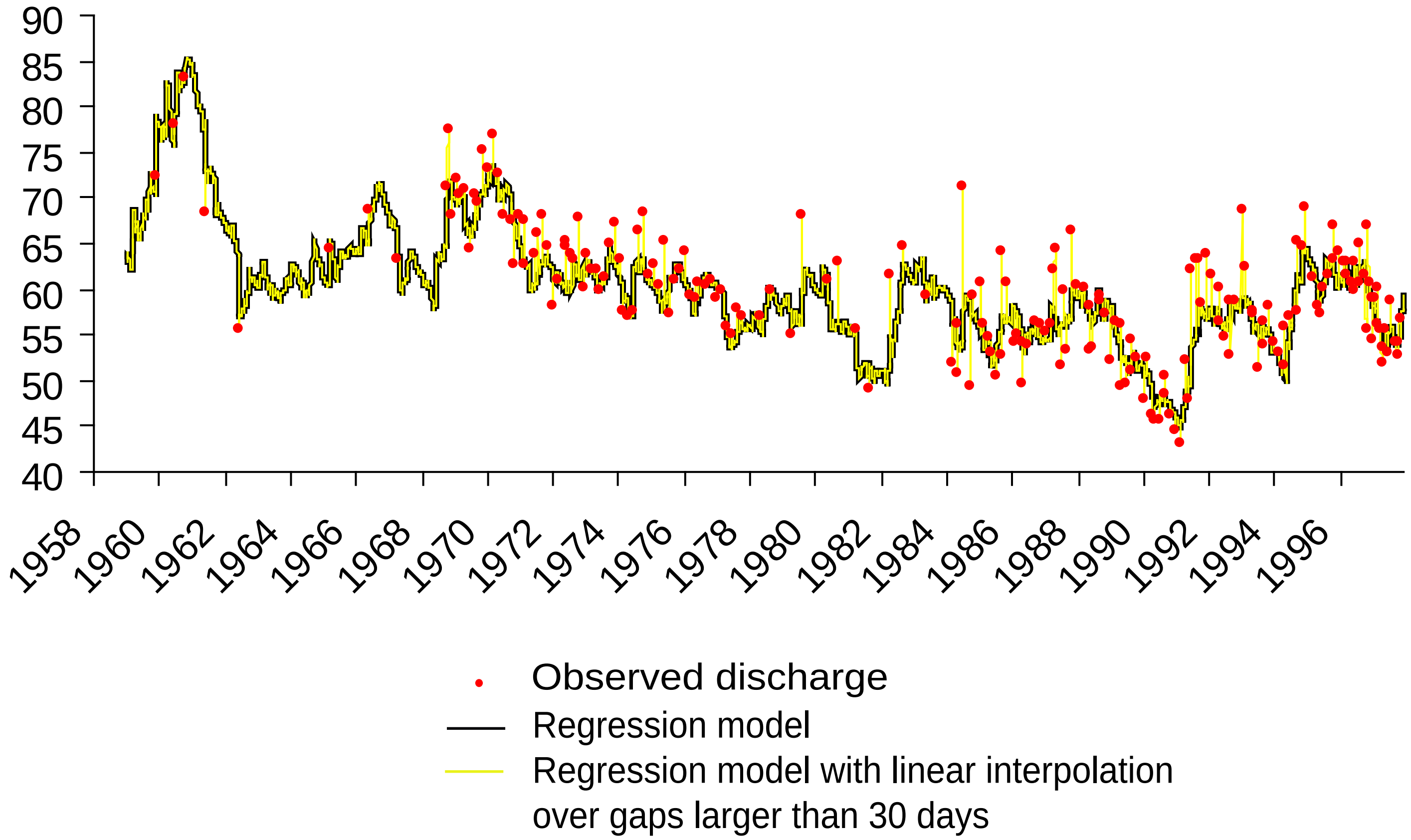
Daily flow duration curves (FDCs) over 1998-2005 for Rock Springs based on the daily discharge time-series are shown in Figure 17. In Figure 17, the red line is the regression prediction, the blue line is the gap-filled time series with a gap threshold of 30 days and the green line is the gap-filled time series with a gap threshold of 90 days. Overall, the linear interpolation over gaps 30 days or smaller predicted the observed time series better than regression alone.

The FDC for Rock Springs changed over time. Discharge at all frequencies decreased from the 1960s through the 1990s, so that the FDC for 1998-2005 is significantly lower than the FDC for 1959-1997 (Figure 18). FDCs may change temporarily and reversibly in response to climate shifts, such as changes in precipitation, or they may change permanently due to groundwater pumping or land use change. The mechanism for the change in the FDC at Rock Springs is not known and is beyond the scope of this current work. The high and low flow frequency analyses for Rock Springs are shown in Figures 19a and 19b.



Date: August 22, 2006	The Root Mean Square Error of the Flow Duration Curve versus gap Length for the Regression Model and Linear Interpolation on Rock Springs, 1998-2002
File: Fig 14.pdf	

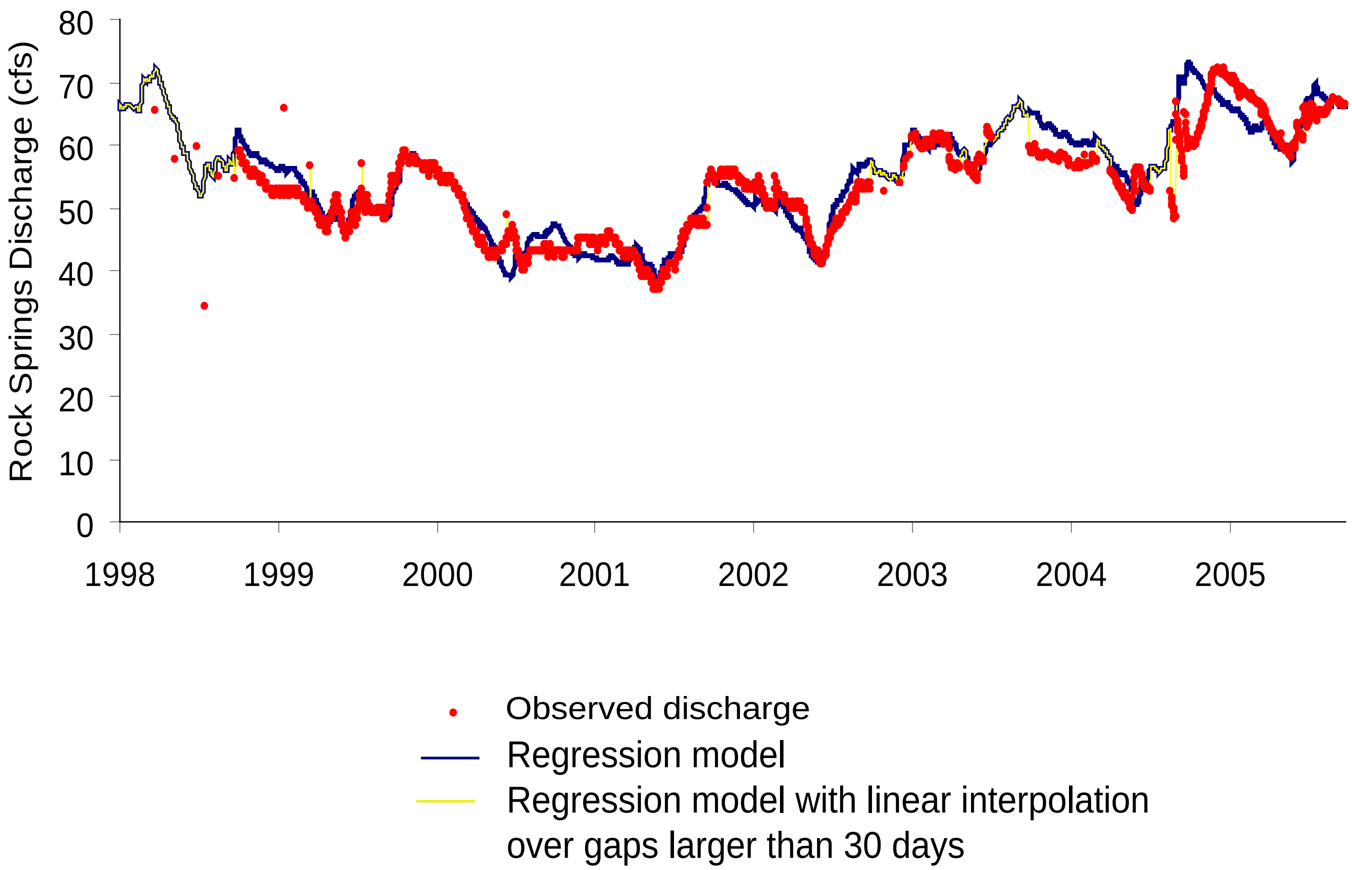
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 14



Date: August 22, 2006	Time Series of Predicted and Observed Discharge for Rock Springs 1959-1997.
File: Fig 15.pdf	

St. Johns River Water Management District
Palatka, Florida

Figure 15

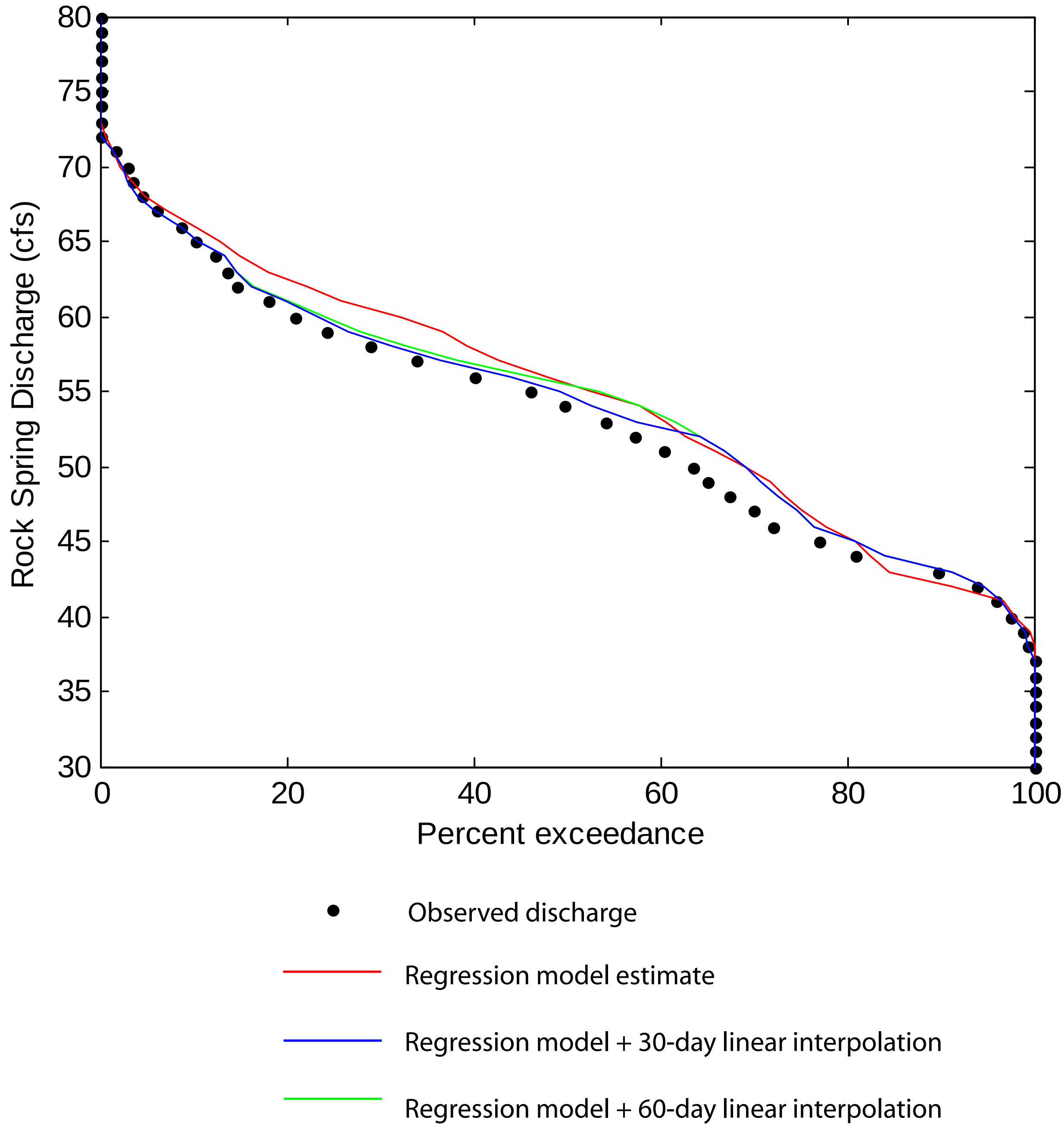


Date: August 22, 2006	Time Series of Rock Springs Discharge, with the Best-Fit Piecewise Regression Model,1998-2005.
File: Fig 16.pdf	

INTERA

St. Johns River Water Management District
Palatka, Florida

Figure 16

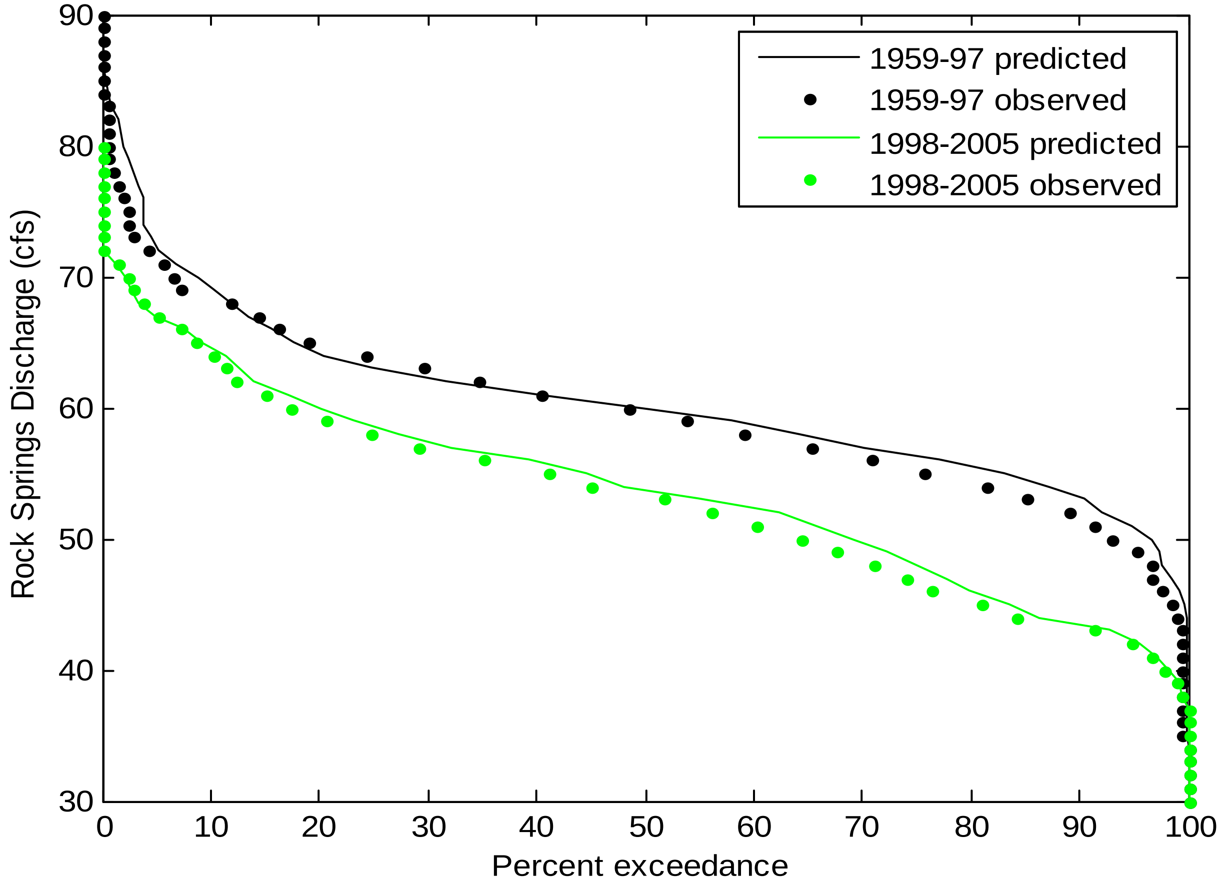


Date: August 23, 2006	Flow Duration Curves for Rock Springs, 1998-2005.
File: Fig 17.pdf	

INTERA

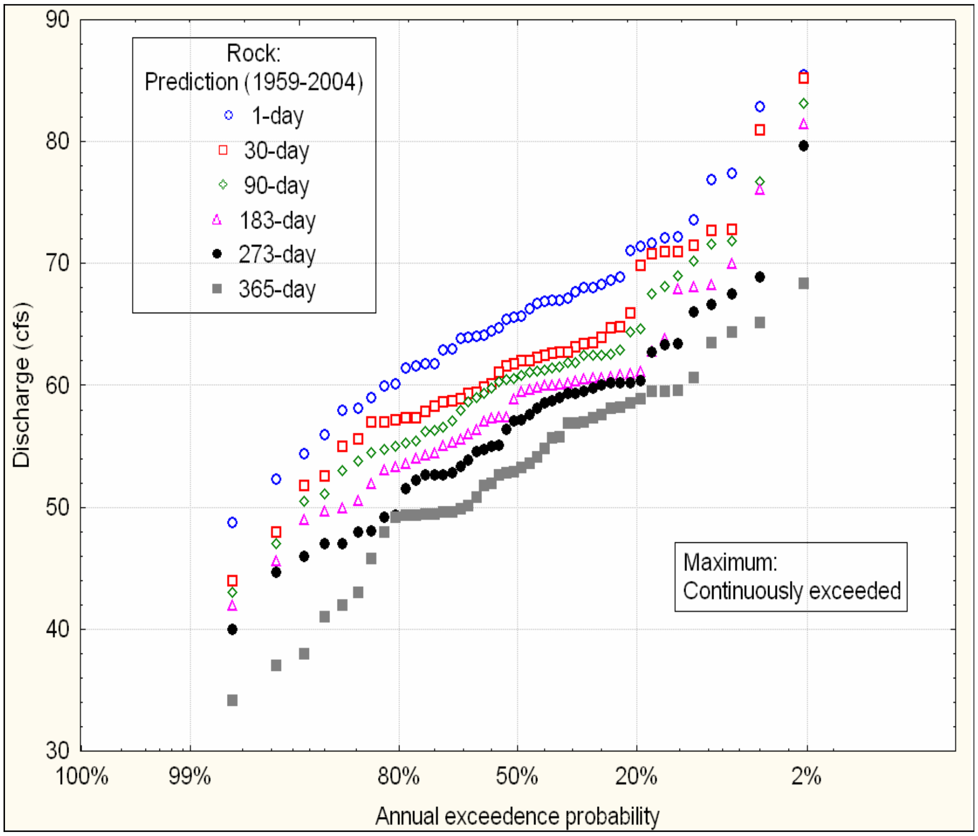
St. Johns River Water Management District
Palatka, Florida

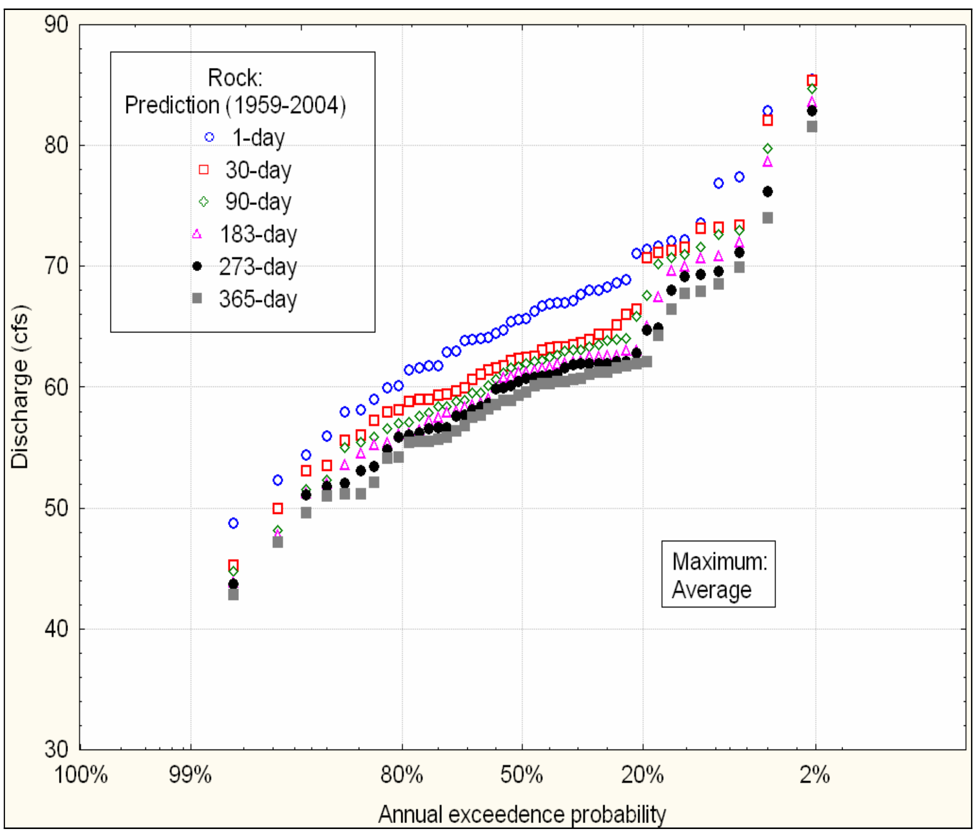
Figure 17



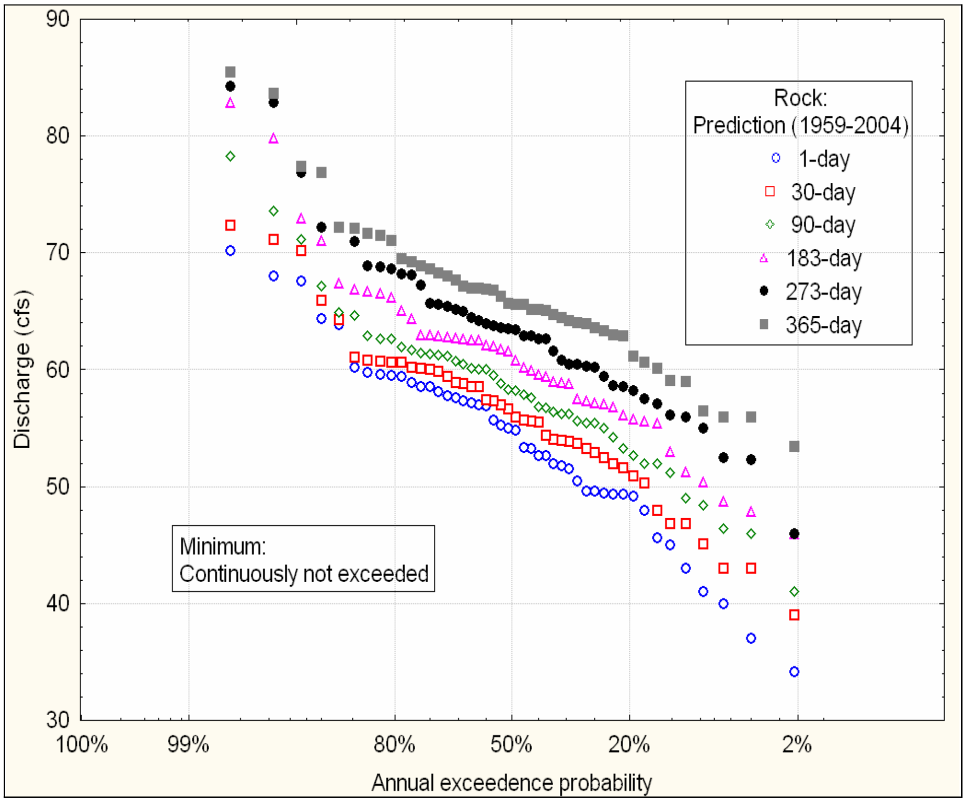
Date: August 23, 2006	Flow Duration Curves for Rock Springs over 1959-1997 and 1998-2005.
File: Fig 18.pdf	

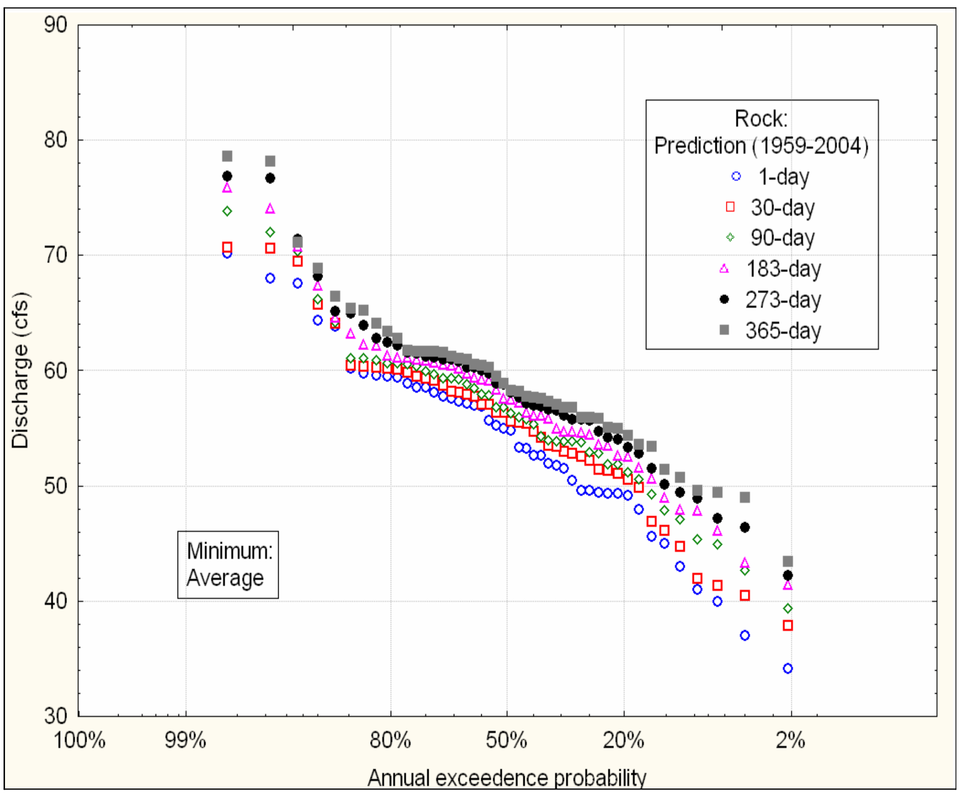
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 18





Date: August 22, 2006	High Flow Frequency Plots, Rock Springs	
File: Fig 19a.pdf		
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 19a





Date: August 22, 2006	Low Flow Frequency Plots, Rock Springs
File: Fig 19b.pdf	

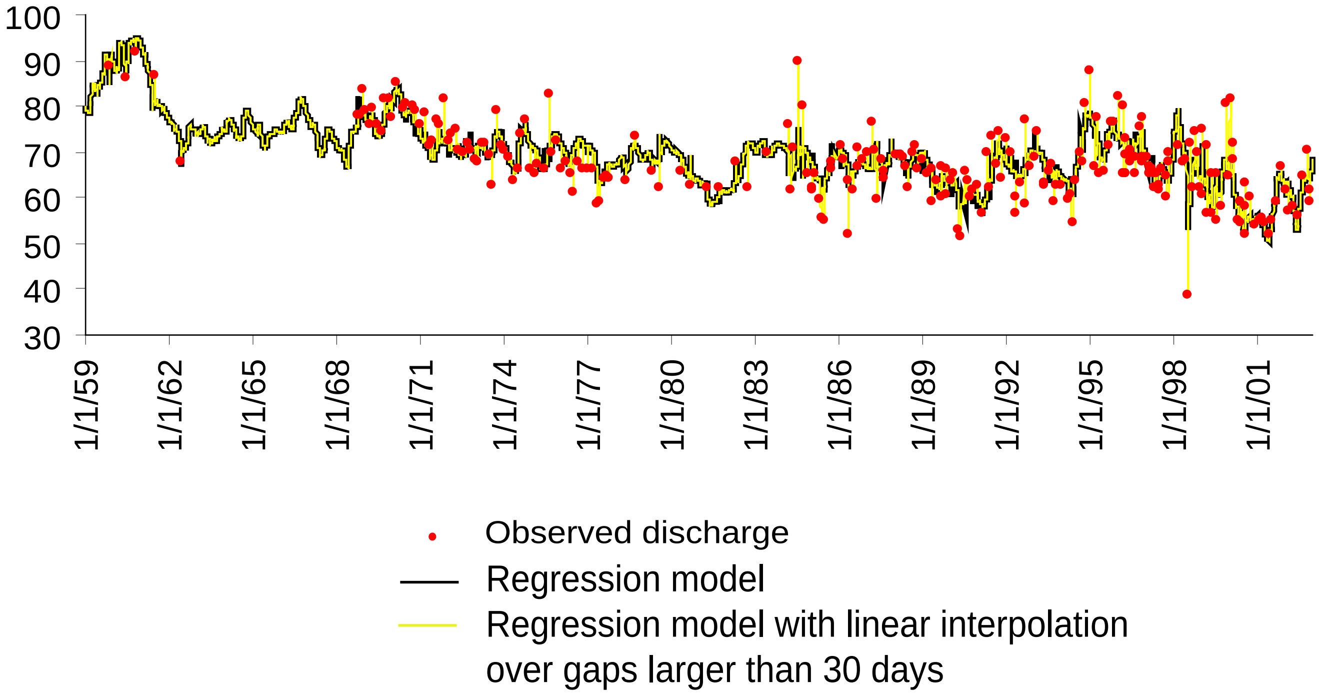
INTERA

St. Johns River Water Management District
Palatka, Florida

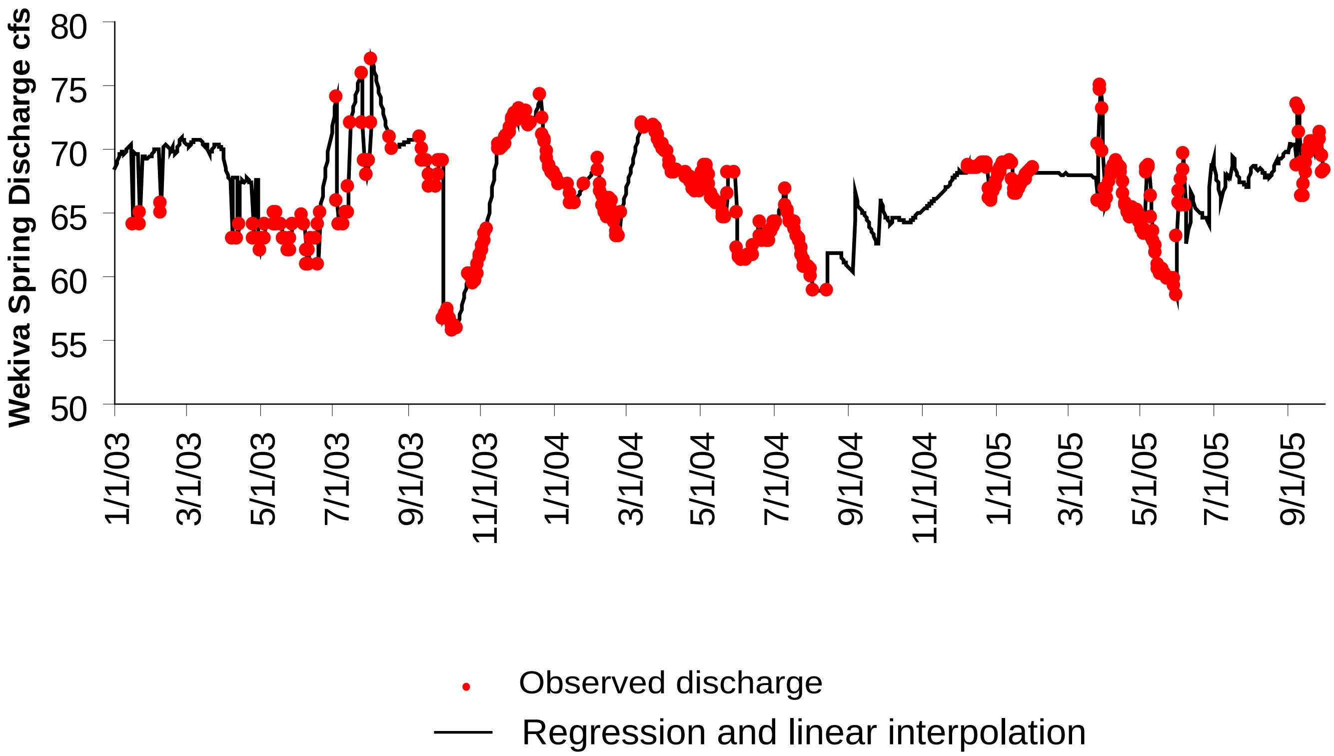
Figure 19b

5.3 Daily discharge and flow duration curves for Wekiva Springs

The daily discharge time-series for Wekiva matches the observed time series well, though some highs and lows were missed as expected in regression predictions (Figure 20). In Figure 20, the black line is the regression prediction; the yellow line is produced using linear interpolation between points for gaps less than 30 days and the regression model for gaps longer than 30 days. Linear interpolation and regression on the 60-day moving window gave good matches between observed and predicted flow over 2003-2005 (Figure 21). Daily FDCs for Wekiva Springs show good fit to the observed FDC for 2003-2005 (Figure 22 top). As with Rock Springs, the FDC for Wekiva Springs was significantly lower in 2003-05 compared with 1959-2002 (Figure 22 bottom). As with Rock Springs, the reasons for this decline are not documented in this report. High and low flow frequency analyses are presented in Figure 23 and Figure 24.



Date: August 22, 2006	Time Series of Daily Discharge at Wekiva Springs, 1959-2002.
File: Fig 20.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida
	Figure 20

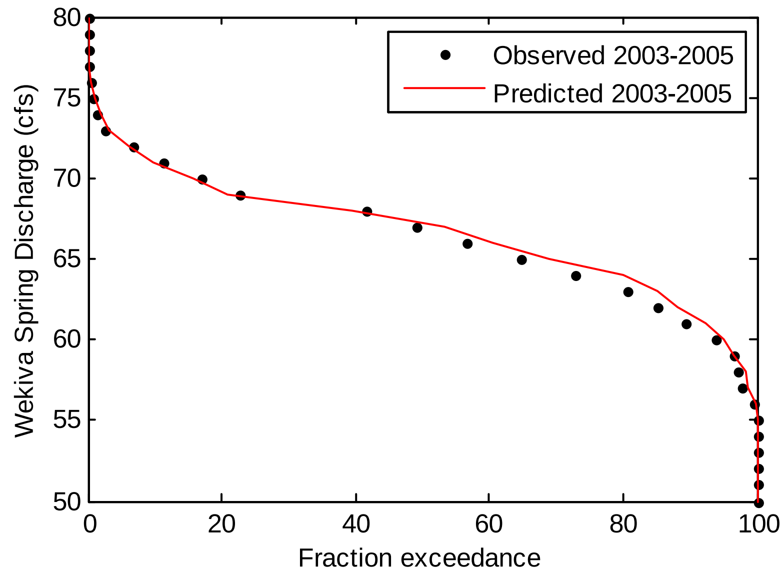


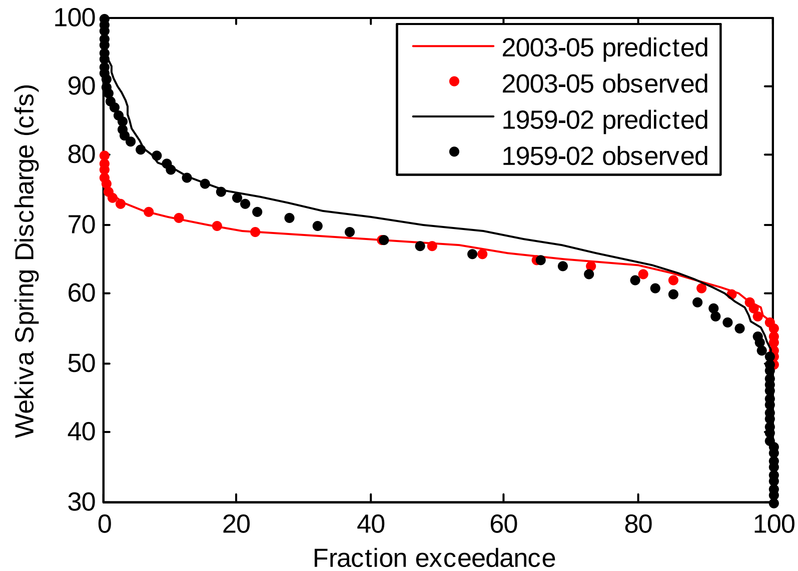
Date: August 22, 2006	Wekiva Springs Time Series and Linear Gap Filling, 2003-2005.
File: Fig 21.pdf	

INTERA

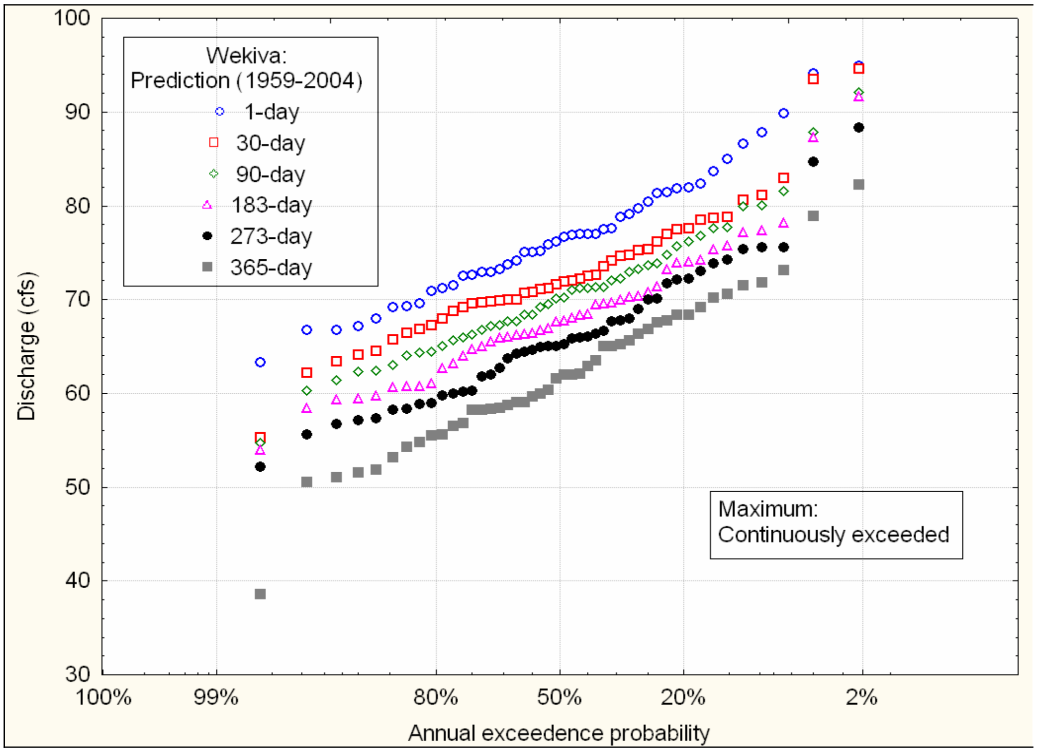
St. Johns River Water Management District
Palatka, Florida

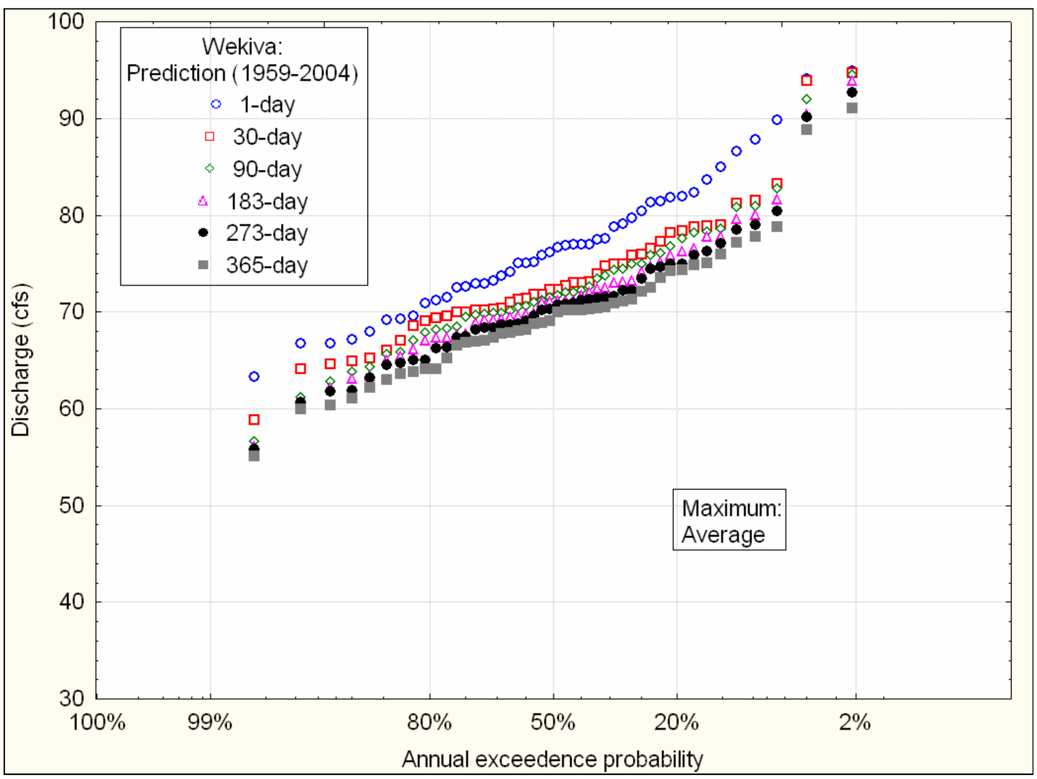
Figure 21



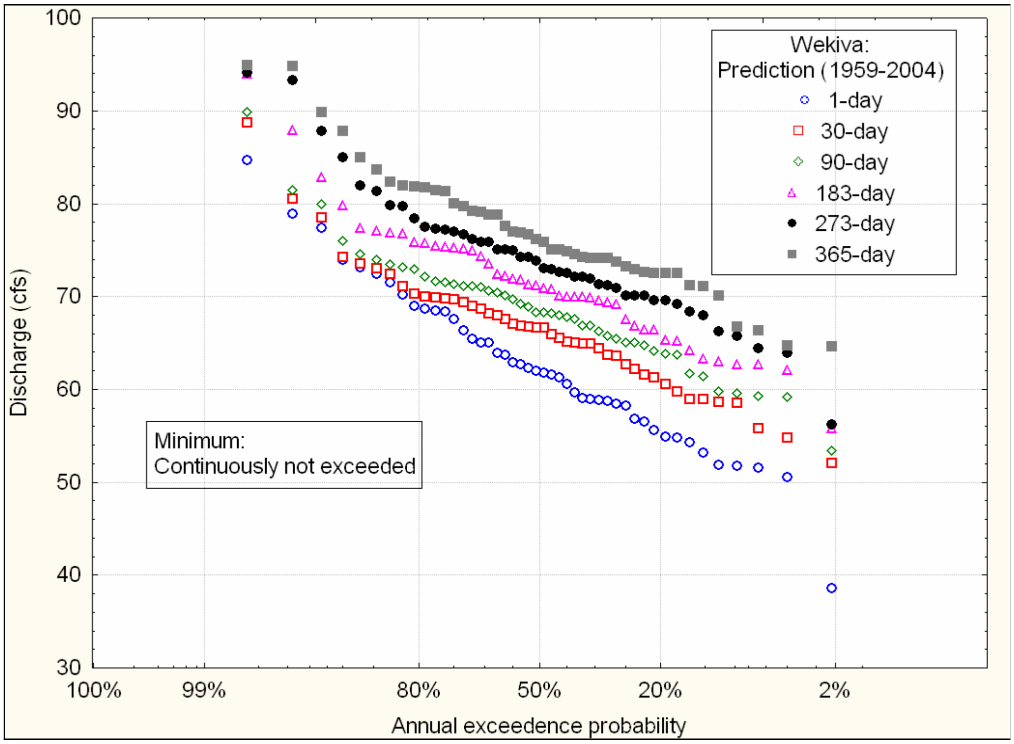


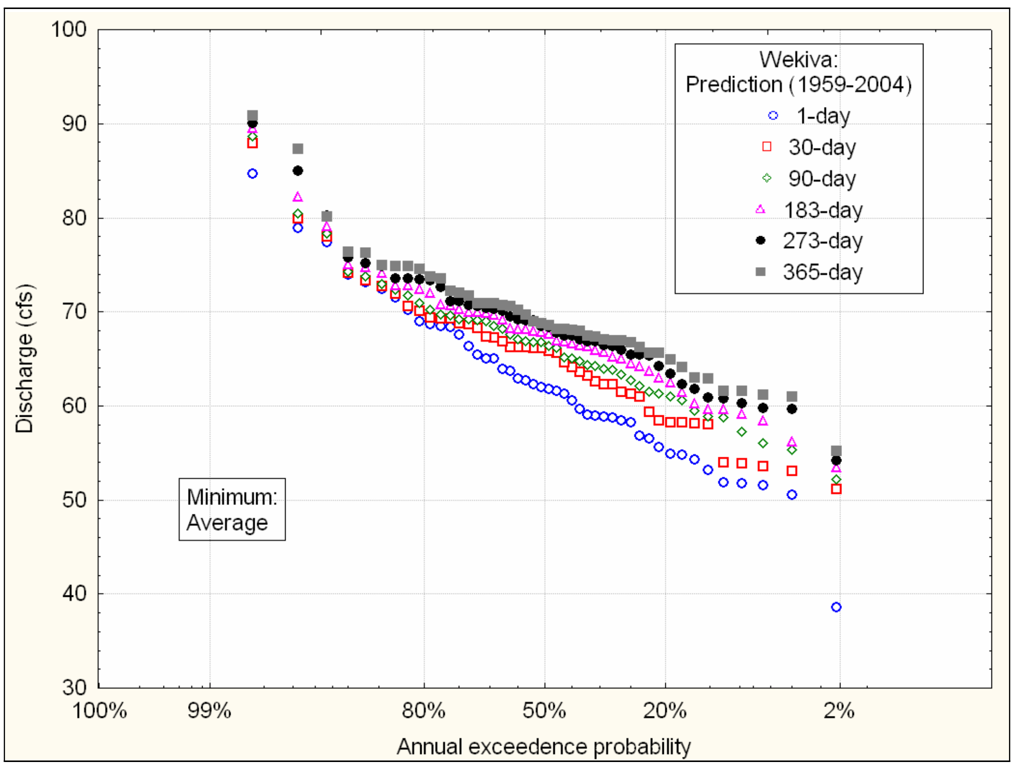
Date: August 23, 2006	Flow Duration Curves for Wekiva Springs, showing Observed and Predicted Curves for 2003-2005 and 1959-2002.
File: Fig 22.pdf	
INTERA	St. Johns River Water Management District Palatka, Florida
	Figure 22





Date: August 22, 2006	High Flow Frequency Plots, Wekiva Springs	
File: Fig 23.pdf		
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 23





Date: August 22, 2006	Low Flow Frequency Plots, Wekiva Springs	
File: Fig 24.pdf		
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 24

6.0 CONCLUSION AND RECOMMENDATIONS

This document presents an evaluation of the spring discharge data for Rock and Wekiva springs; groundwater levels at adjacent monitoring wells, and precipitation measurements at nearby rain gage stations. Based on this evaluation, a regression modeling methodology is developed and applied for generating historic daily spring discharge records at Rock and Wekiva springs. Flow duration curves are also generated along with high- and low-frequency analyses for set durations from the simulated daily spring discharge. The following general conclusions can be made based on this study.

- Measurements of well levels were available at a daily time step for the wells used for flow prediction.
- Two regression models were required for each spring: one where daily discharge measurements were available with some gaps, and a second where discharge measurements were at longer intervals. Separation into two time periods was required in order to get representative regression parameters for each time period, and to avoid giving too much weight to recent periods with high data frequency.
- The flow duration curves for both Rock and Wekiva springs changed over the study period: discharge decreased at all probability levels. Flow duration curves based on the historical time-series may not correctly represent current or future flow duration curves due to changes in the relationships between precipitation, groundwater levels, and spring discharge.
- The statistical modeling could be complemented by a more process-based approach that includes the effects of pumping and land use change on spring discharge. Such an exercise would help explain the causes of the decreased spring discharge over 1960-1990, and project whether such decreases are temporary and due to random climate fluctuations, or permanent and due to land use change or groundwater pumping.

The daily period of record generated by the multiple regression models provides an estimate for the historic time series of spring discharge values. These estimated discharge values are developed for uses where such a time series is required, such as a frequency analysis of historic flows for MFL determinations. It must be explicitly stated that the presented multiple regression models are not physical and should not be used for predictive purposes or to interpret the relationships between spring discharge values and explanatory variables such as groundwater levels, recorded rainfall, or recorded discharges at nearby springs. A specific caution is made that predictions achieved by altering the explanatory variables from their observed values and re-generating the spring discharge time series entail assumptions not supported here.

7.0 REFERENCES

D'Agostino, R.B. and M.A. Stephens, 1987. Goodness-of-Fit Techniques, *Journal of Educational Statistics*, Vol. 12, No. 4, pp. 412-416.

Montgomery, D.C., and E.A. Peck, 1992. I*ntroduction to Linear Regression Analysis.* John Wiley and Sons, New York.

Osburn, W., D. Toth, and D. Boniol, 2002. Springs of the St. Johns River Water Management District. Technical Publication SJ2002-5, St. Johns River Water Management District, Palatka, FL.

Ott, R.L., 2006. *Introduction to Statistical Methods and Data Analysis (6th Edition).* PWS-Kent Publishing Company, Boston, MA.

APPENDIX A
Model Usage Notes

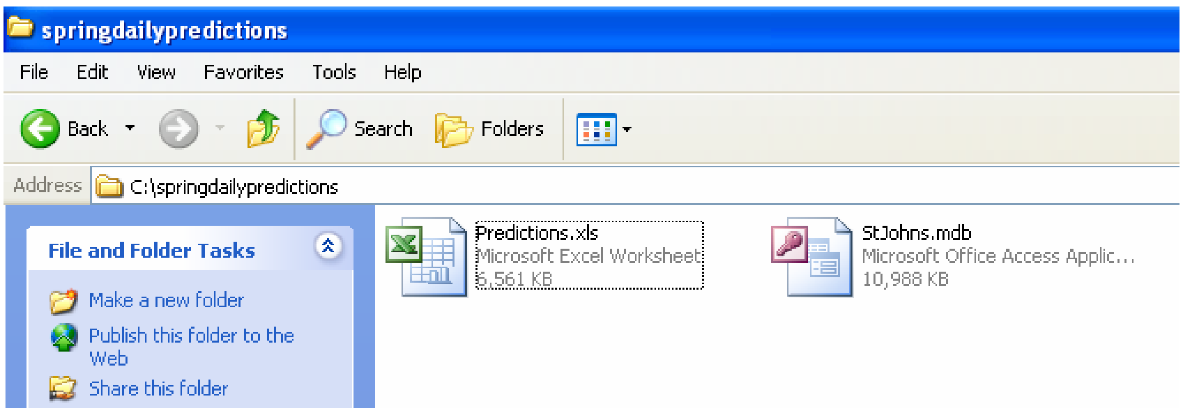
Model Usage Notes

This Appendix describes the structure and operation of an ACCESS database created to facilitate predictive applications of the statistical spring discharge models described earlier in Section 4. An example using Rock Springs data is also presented.

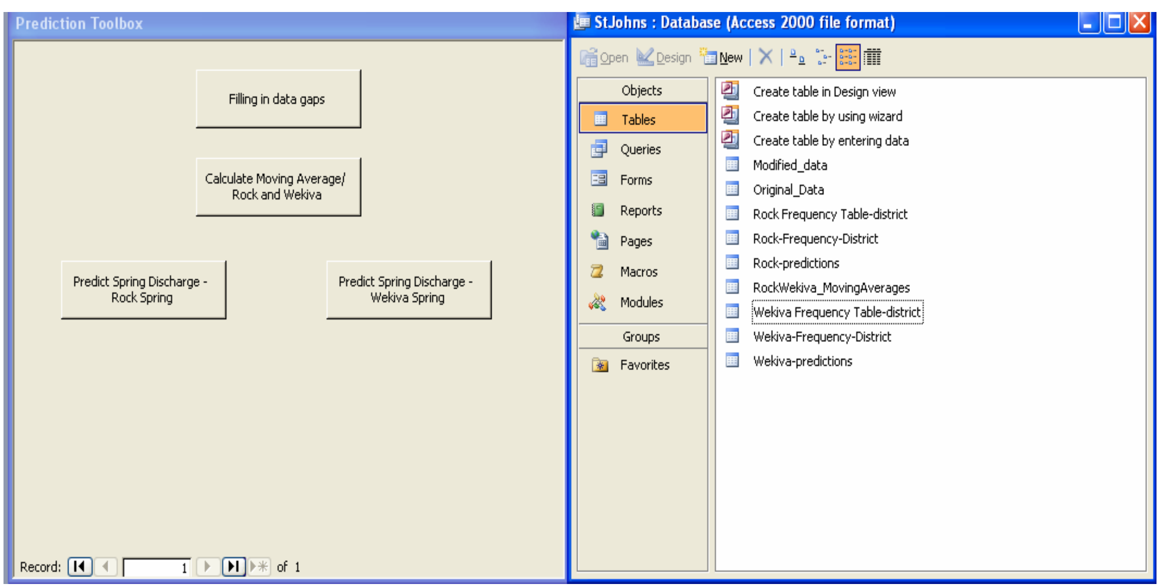
1. *Folder*: <u>springdailypredictions</u> –

The folder **<u>springdailypredictions</u>** has two files as shown below:

- **<u>St.Johns.mdb</u>**
- **<u>Predictions.xls</u>**



After building the statistical models, **<u>St.Johns.mdb – an ACCESS database</u>** was built for applying the statistical models to generate daily predictions for the two springs. A screenshot of the database is shown below.

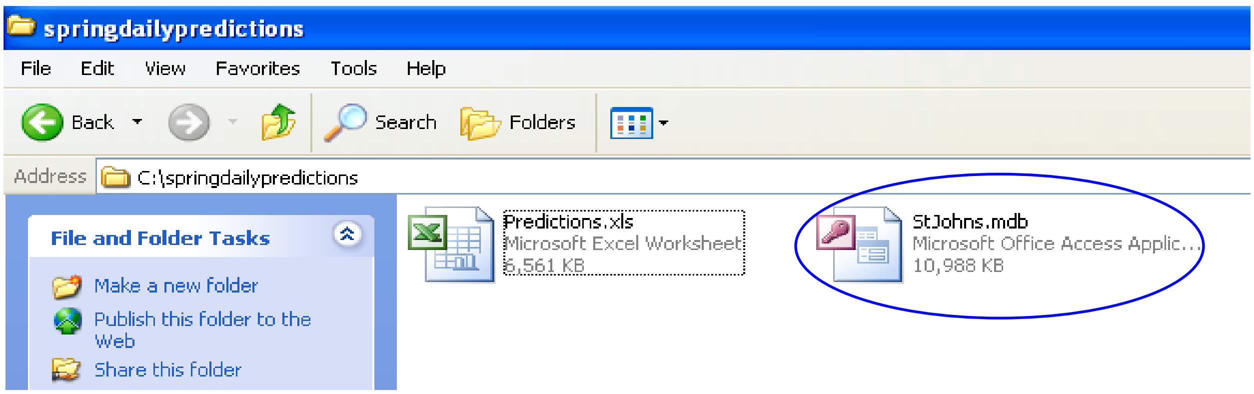


On the left, are the different tables present in the database and on the right is a prediction toolbox. The prediction toolbox executes ACCESS queries and/or VISUAL BASIC APPLICATION Modules, on the click of different buttons. **<u>Predictions.xls – EXCEL file</u>** is used to graphically display the daily predictions and frequency analysis generated in **St.Johns.mdb**. The next few pages will walk the user through using the toolbox for generating daily predictions and frequency analysis with the help of an example. It will also guide the user on how to save the results for different cases.

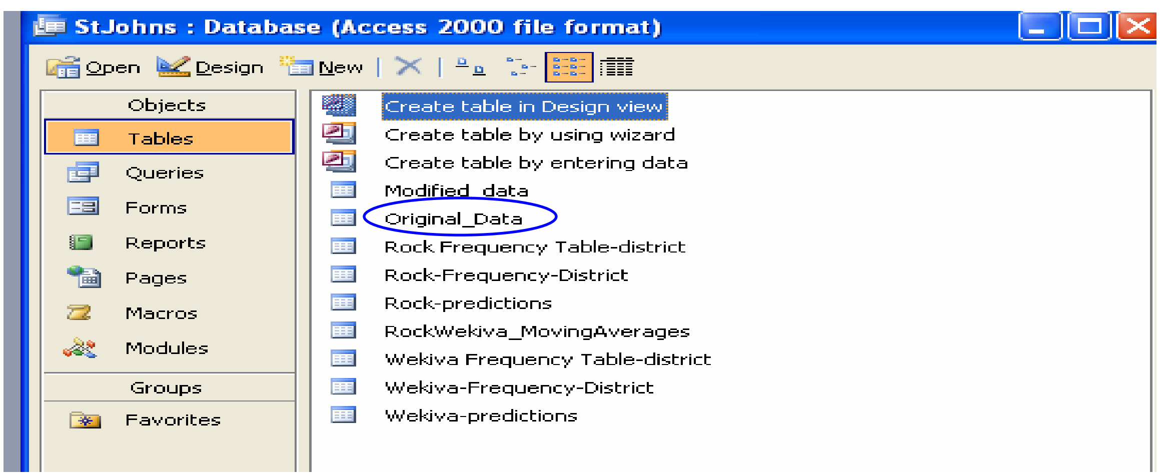
In the example below, our primary task would be to get Rock Springs daily predictions from 1/1/1959 to 9/30/2005.

2. Open St.Johns.mdb

Open **St.Johns.mdb** (highlighted below) by double clicking the file.



The original spring discharge, groundwater elevation and precipitation data reside in the "**Original_Data**" ACCESS data table. The screenshot below indicates the **Original Data** table within the database.



Double-clicking this table would open the **Original Data** table as shown below.

Original_Data : Table

Date	Rock Spring	Orlando Rainfall	w283401	w544902R	Wekiva Spring
1/1/1958		0.06			
1/2/1958		0.02			
1/3/1958		0			
1/4/1958		0			
1/5/1958		0	65.87		
1/6/1958		0.49			
1/7/1958		0.86			
1/8/1958		0			
1/9/1958		0			
1/10/1958		0	66.31		
1/11/1958		0			
1/12/1958		0			
1/13/1958		0.7			
1/14/1958		0			
1/15/1958		0.05	66.56		
1/16/1958		0.01			
1/17/1958		0			
1/18/1958		0			
1/19/1958		0			
1/20/1958		0	66.22		

Record: 1 of 17440

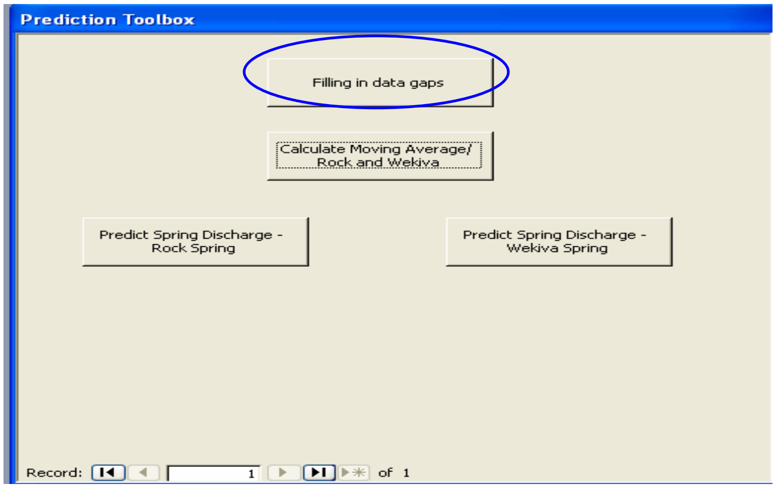
The table has 17440 records for dates ranging from 1/1/1958 to 9/30/2005. If the user wants to change a particular data time series, pasting the new time series (with dates from 1/1/1958 to 9/30/2005) over the old one is one of the ways to do it.

If the user has another ACCESS database with new time series data, it can be added to the **Original Data** table using an *Append Query*. *Append Query* allows the user to append one or more columns to the **Original Data** table. For example, if a new time series for w283401 becomes available, append the new data column as *w283401 (new)* using the *Append Query*. Then delete the old *w283401* column from **Original Data** table and rename *w283401 (new)* as *w283401*. If data is not available for a particular date, the user can leave it blank as seen in **Original Data** table for different variables.

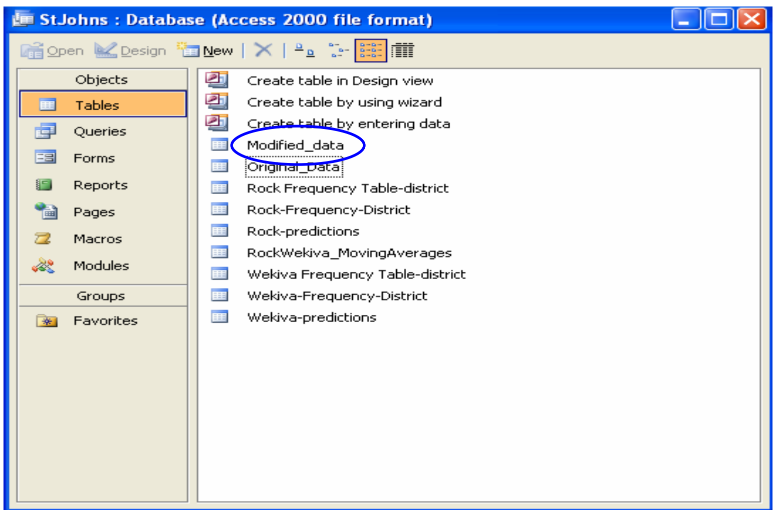
3. Data Gap Filling to create "Modified Data" Table

Gaps in the data which are less than 30 days are filled by linear interpolation. The need to fill data gaps for wells arises during the calculation of moving averages. Also, as indicated in the report, spring predictions for Rock and Wekiva, for gaps less than 30 days, perform better than predictions for Rock and Wekiva from regression.

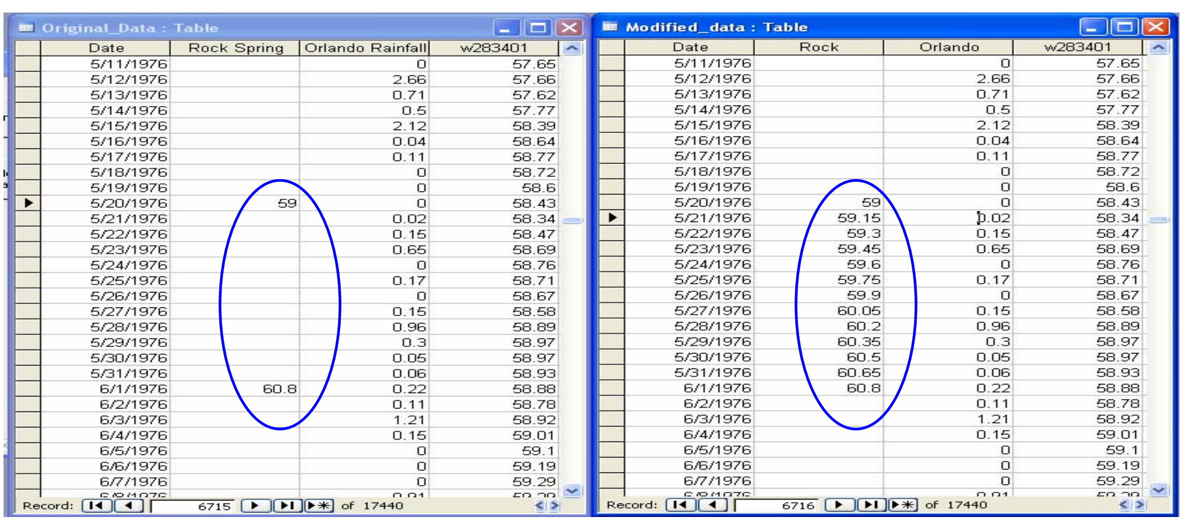
Therefore the next step is clicking the "Filling in data gaps" button on the prediction toolbox.



Clicking this button creates a **<u>Modified_data</u>** table as highlighted below:



Open the **<u>Modified_data</u>** table by double-clicking on it. Below is the screenshot:



Original_Data : Table

Date	Rock Spring	Orlando Rainfall	w283401
5/11/1976		0	57.65
5/12/1976		2.66	57.66
5/13/1976		0.71	57.62
5/14/1976		0.5	57.77
5/15/1976		2.12	58.39
5/16/1976		0.04	58.64
5/17/1976		0.11	58.77
5/18/1976		0	58.72
5/19/1976		0	58.6
5/20/1976	59	0	58.43
5/21/1976		0.02	58.34
5/22/1976		0.15	58.47
5/23/1976		0.65	58.69
5/24/1976		0	58.76
5/25/1976		0.17	58.71
5/26/1976		0	58.67
5/27/1976		0.15	58.58
5/28/1976		0.96	58.89
5/29/1976		0.3	58.97
5/30/1976		0.05	58.97
5/31/1976		0.06	58.93
6/1/1976	60.8	0.22	58.88
6/2/1976		0.11	58.78
6/3/1976		1.21	58.92
6/4/1976		0.15	59.01
6/5/1976		0	59.1
6/6/1976		0	59.19
6/7/1976		0	59.29

Record: 6715 of 17440

Modified_data : Table

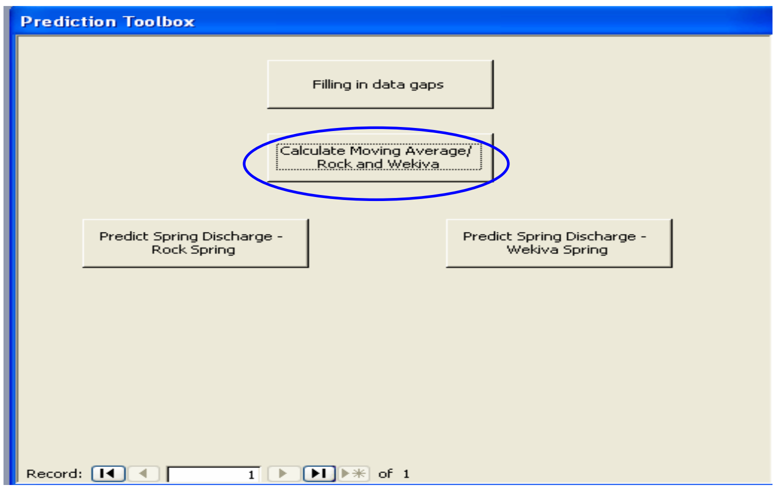
Date	Rock	Orlando	w283401
5/11/1976		0	57.65
5/12/1976		2.66	57.66
5/13/1976		0.71	57.62
5/14/1976		0.5	57.77
5/15/1976		2.12	58.39
5/16/1976		0.04	58.64
5/17/1976		0.11	58.77
5/18/1976		0	58.72
5/19/1976		0	58.6
5/20/1976	59	0	58.43
5/21/1976	59.15	0.02	58.34
5/22/1976	59.3	0.15	58.47
5/23/1976	59.45	0.65	58.69
5/24/1976	59.6	0	58.76
5/25/1976	59.75	0.17	58.71
5/26/1976	59.9	0	58.67
5/27/1976	60.05	0.15	58.58
5/28/1976	60.2	0.96	58.89
5/29/1976	60.35	0.3	58.97
5/30/1976	60.5	0.05	58.97
5/31/1976	60.65	0.06	58.93
6/1/1976	60.8	0.22	58.88
6/2/1976		0.11	58.78
6/3/1976		1.21	58.92
6/4/1976		0.15	59.01
6/5/1976		0	59.1
6/6/1976		0	59.19
6/7/1976		0	59.29

Record: 6716 of 17440

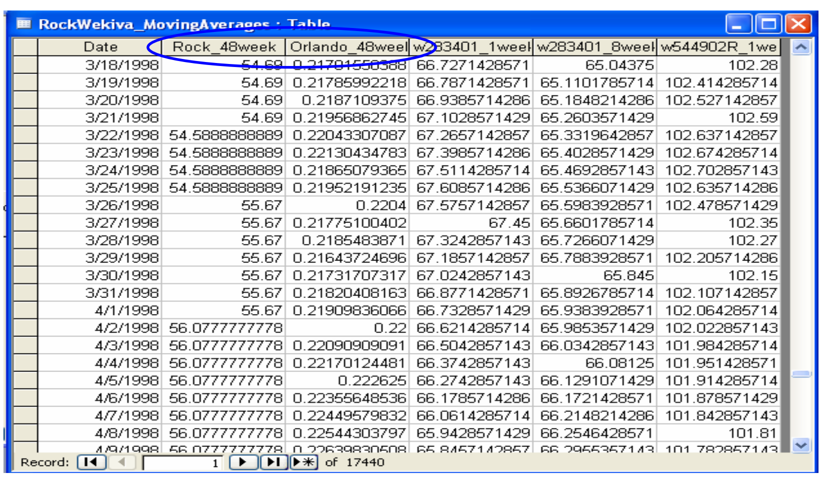
The user would notice linear interpolation values (for data gaps less than 30 days) in the **Modified_Table**. For example, we see two Rock Springs observed values highlighted in **Original_data**. The **Modified_Table** shows linearly interpolated values for Rock Springs between dates 5/20/1976 and 6/1/1976

4. Calculating moving average variables for each spring

The statistical models in the report show the use of moving averages of different variables (spring, groundwater level and rainfall data) for predicting daily discharge for each spring. Computation of these variables, for each spring, is then performed by clicking the button highlighted below.



For example clicking on *Calculate Moving Average/Rock and Wekiva* would fill the table **RockWekiva_MovingAverages** present in the database. The screenshot below shows the table:



RockWekiva_MovingAverages : Table

Date	Rock_48week	Orlando_48weel	w283401_1weel	w283401_8weel	w544902R_1we
3/18/1998	54.69	0.21701550388	66.7271428571	65.04375	102.28
3/19/1998	54.69	0.21785992218	66.7871428571	65.1101785714	102.414285714
3/20/1998	54.69	0.2187109375	66.9385714286	65.1848214286	102.527142857
3/21/1998	54.69	0.21956862745	67.1028571429	65.2603571429	102.59
3/22/1998	54.5888888889	0.22043307087	67.2657142857	65.3319642857	102.637142857
3/23/1998	54.5888888889	0.22130434783	67.3985714286	65.4028571429	102.674285714
3/24/1998	54.5888888889	0.21865079365	67.5114285714	65.4692857143	102.702857143
3/25/1998	54.5888888889	0.21952191235	67.6085714286	65.5366071429	102.635714286
3/26/1998	55.67	0.2204	67.5757142857	65.5983928571	102.478571429
3/27/1998	55.67	0.21775100402	67.45	65.6601785714	102.35
3/28/1998	55.67	0.2185483871	67.3242857143	65.7266071429	102.27
3/29/1998	55.67	0.21643724696	67.1857142857	65.7883928571	102.205714286
3/30/1998	55.67	0.21731707317	67.0242857143	65.845	102.15
3/31/1998	55.67	0.21820408163	66.8771428571	65.8926785714	102.107142857
4/1/1998	55.67	0.21909836066	66.7328571429	65.9383928571	102.064285714
4/2/1998	56.0777777778	0.22	66.6214285714	65.9853571429	102.022857143
4/3/1998	56.0777777778	0.22090909091	66.5042857143	66.0342857143	101.984285714
4/4/1998	56.0777777778	0.22170124481	66.3742857143	66.08125	101.951428571
4/5/1998	56.0777777778	0.222625	66.2742857143	66.1291071429	101.914285714
4/6/1998	56.0777777778	0.22355648536	66.1785714286	66.1721428571	101.878571429
4/7/1998	56.0777777778	0.22449579832	66.0614285714	66.2148214286	101.842857143
4/8/1998	56.0777777778	0.22544303797	65.9428571429	66.2546428571	101.81
4/9/1998	56.0777777778	0.22639830508	65.8457142857	66.2955357143	101.782857143

Record: 1 of 17440

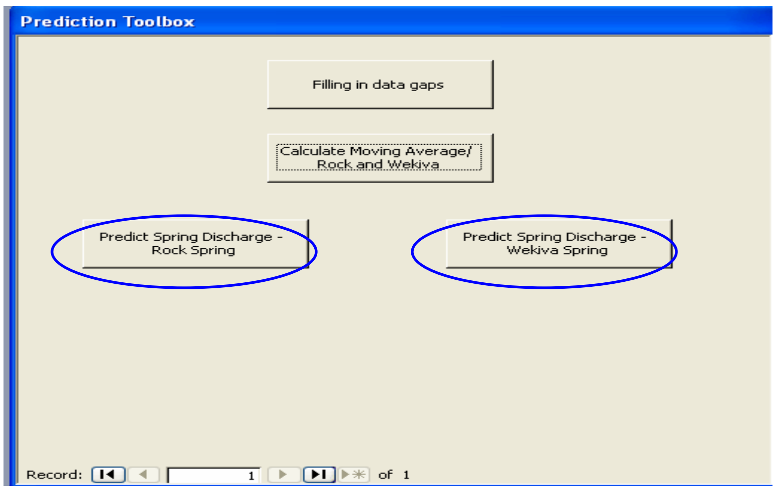
The highlighted columns in the table above show some of the calculated moving averages to be used in the Rock statistical model for daily discharge predictions.

5. Calculate Spring discharge predictions and frequency analysis

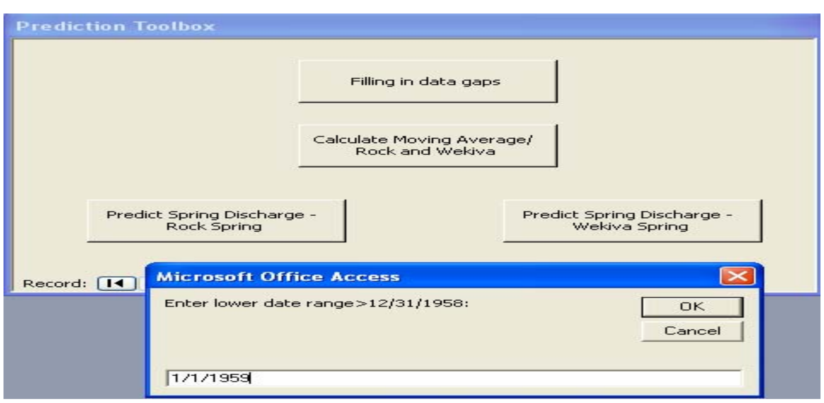
Spring discharge daily predictions are limited by a range of lower and upper date. This is due to limited date range coverage for explanatory variables in the statistical model for a particular spring. The following are the dates for the two springs for which daily discharge predictions can be computed:

Spring	Date Range for discharge predictions
Rock	1/1/1959 to 9/30/2005
Wekiva	1/1/1959 to 9/30/2005

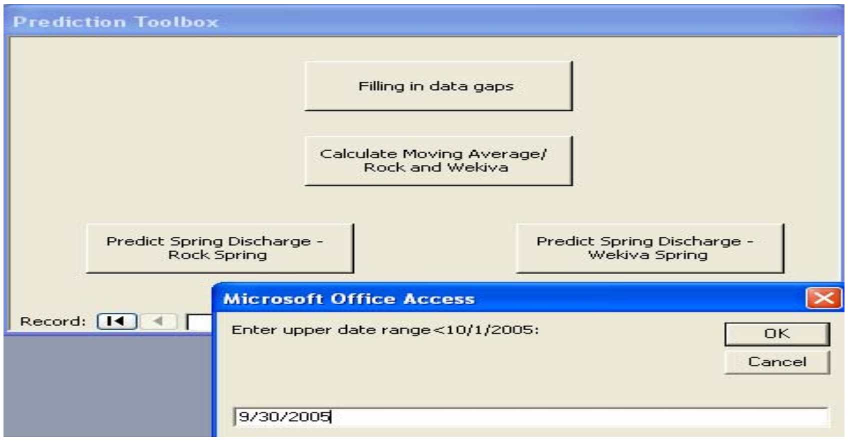
Clicking the buttons highlighted below give daily discharge predictions and maximum and minimum frequencies for date ranges specified by the user. Note that these date ranges have to fall within the ranges mentioned above for a particular spring. Also, since Rock discharge predictions are used in the regression model for predicting Wekiva discharge, it is necessary to first predict Rock discharge values prior to Wekiva predictions



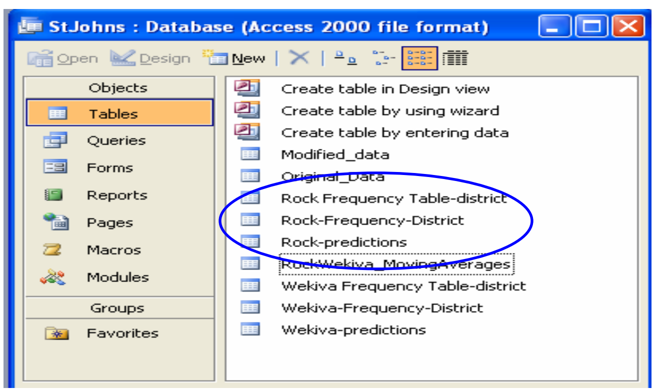
For example, on clicking *Predict Spring Discharge – Rock Springs*, we see a pop-up window asking for the date from which predictions are needed. For our example enter 1/1/1959. As noted earlier, the date entered should be greater than 12/31/1958, since Rock Springs predictions are only available since that date.



Press OK. Another window asking for the date till which predictions are needed. For our example enter 9/30/2005. Again the date entered should be less than 10/1/2005, since Rock Springs predictions are only available till 9/30/2005.



On pressing OK, tables called **Rock-predictions, Rock-Frequency-district and Rock Frequency table-District** are added to the ACCESS database as shown below:



Double click Rock-predictions table to view. The screenshot shows the observed Rock discharge data and the predicted Rock discharge data, between the lower and upper date ranges we entered.

Rock-predictions : Table

Date	Rock(observed)	Rock(predicted)
3/11/1997		56.4733333333
3/12/1997	56.3	56.3
3/13/1997		56.7333333333
3/14/1997		57.1666666667
3/15/1997		57.6
3/16/1997		58.0333333333
3/17/1997		58.4666666667
3/18/1997		58.9
3/19/1997		59.3333333333
3/20/1997		59.7666666667
3/21/1997	60.2	60.2
3/22/1997		56.1486734175
3/23/1997		56.1196405914
3/24/1997		56.0906077653
3/25/1997		56.0615749392
3/26/1997		56.0325421131
3/27/1997		56.0049006044
3/28/1997		55.9758677783
3/29/1997		55.9468349522
3/30/1997		55.9859766770
3/31/1997		55.8832040306
4/1/1997		55.8541712045
4/2/1997		55.8251383784
4/3/1997		55.7961055523
4/4/1997		55.7670727262

Record: 1 of 17075

Double-click table **Rock-Frequency-district** to view. The table has continuously-exceeded and average values for 1-day, 30-day, 90-day, 183-day, 273-day and 365-day periods for each year starting on June 1 of a year and ending on May 31 of the next year. The table also has continuously-not-exceeded and average values for 1-day, 30-day, 90-day, 183-day, 273-day and 365-day periods for each year starting on October 1 of a year and ending on September 30 of the next year. It is important to note that each year range for picking maximums and minimums is assumed to be independent of other years. The screenshot below shows some of the columns present in the table.

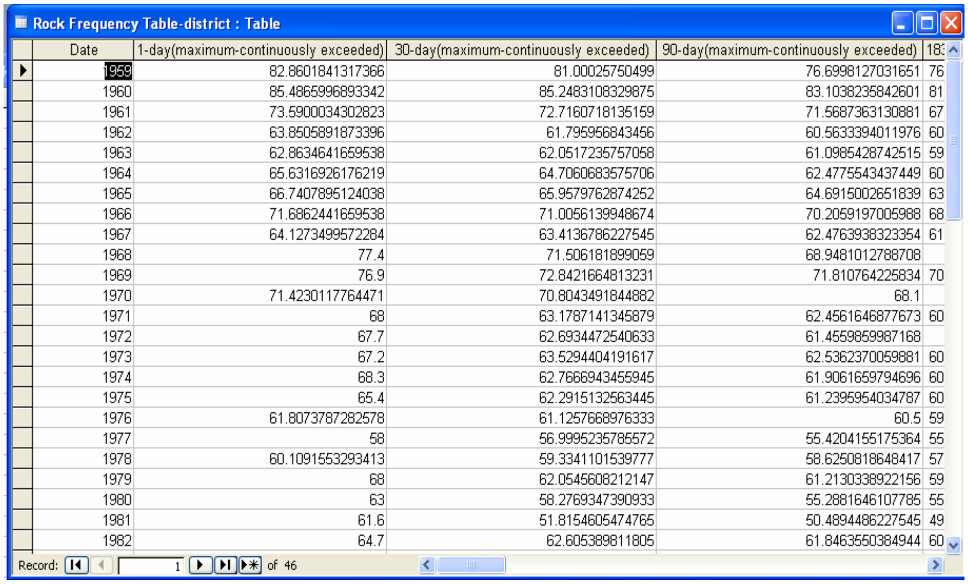
Rock-Frequency-District : Table

Date	Rock	Cont_exceeded_30days	Average_maximum_30days	Cont_not_exceeded_30days	Average_minimum_30days	Cont_exceeded_90days	Average_maxim
11/26/1978	57.725542	57.4936655859709	57.8303706963217	58.4862134559452	57.8303706963217	57.1167599059025	58.1645661163
11/27/1978	57.7138977	57.4936655859709	57.8002478314799	58.4575505988024	57.8002478314799	57.1167599059025	58.1521783325
11/28/1978	57.6771734	57.4936655859709	57.7711166184773	58.395496124893	57.7711166184773	57.1167599059025	58.1404573581
11/29/1978	57.6422405	57.4936655859709	57.7424434165954	58.3172233019675	57.7424434165954	57.1167599059025	58.1346637268
11/30/1978	57.5963681	57.4936655859709	57.7159613088109	58.1746452609067	57.7159613088109	57.1167599059025	58.1296817451
12/1/1978	57.5465127	57.4936655859709	57.6935152309666	58.1316509751925	57.6935152309666	57.1167599059025	58.1252770014
12/2/1978	57.5248065	57.4842443541488	57.6719350102652	58.1029881180496	57.6719350102652	57.1167599059025	58.1216784006
12/3/1978	57.5094838	57.4119550556031	57.6489005748503	58.0689509751925	57.6489005748503	57.1167599059025	58.1151977898
12/4/1978	57.4936656	57.3779179127459	57.6258661394354	58.0268524037639	57.6258661394354	57.1167599059025	58.1083747026
12/5/1978	57.5178499	57.2104097946963	57.5986513857998	57.9151879640718	57.5986513857998	57.1167599059025	58.0998966342
12/6/1978	57.5839828	57.1812513344739	57.5741868314799	57.8748808212147	57.5741868314799	57.1167599059025	58.0916044779
12/7/1978	57.6189157	57.1463184773311	57.5499014200171	57.8381565355004	57.5499014200171	57.1167599059025	58.0832519495
12/8/1978	57.6529528	57.1167599059025	57.5258548656972	57.8139722497862	57.5258548656972	57.1167599059025	58.0770200422
12/9/1978	57.5927623	57.1167599059025	57.5030623113772	57.7488305218135	57.5030623113772	57.1167599059025	58.0710966587
12/10/1978	57.5807361	57.1167599059025	57.4888131650984	57.7255419503849	57.4888131650984	57.1167599059025	58.0664618326
12/11/1978	57.5816318	57.1167599059025	57.4765186749358	57.7138976646706	57.4765186749358	57.1167599059025	58.0639266088
12/12/1978	57.5825275	57.1167599059025	57.4663440419162	57.6771733789564	57.4663440419162	57.1167599059025	58.0626354326
12/13/1978	57.5836142	57.1167599059025	57.4635952112917	57.6529528058169	57.4635952112917	57.1167599059025	58.0622797802
12/14/1978	57.5353001	57.1167599059025	57.4713107998289	57.8737081779299	57.4713107998289	57.1167599059025	58.0632217782
12/15/1978	57.5227601	57.1167599059025	57.4839623661249	57.9759150983747	57.4839623661249	57.1167599059025	58.0644428460
12/16/1978	57.5012629	57.1167599059025	57.5002762070145	58.0359279555175	57.5002762070145	57.1167599059025	58.0659537956
12/17/1978	57.4842444	57.1167599059025	57.5149765902481	58.0359279555175	57.5149765902481	57.1167599059025	58.0677222444
12/18/1978	57.4119551	57.1167599059025	57.5311304242943	58.0359279555175	57.5311304242943	57.1167599059025	58.0672286168
12/19/1978	57.3779179	57.1167599059025	57.5529415183918	58.1479984088965	57.5529415183918	57.1167599059025	58.0660781320
12/20/1978	57.2104098	57.1167599059025	57.5744375209581	58.1627299486741	57.5744375209581	57.1167599059025	58.0644300282
12/21/1978	57.1812513	57.1167599059025	57.5940873781009	58.1734785201027	57.5940873781009	57.1167599059025	58.0636239404
12/22/1978	57.1463185	57.1167599059025	57.6131998066724	58.1922885201027	57.6131998066724	57.1167599059025	58.0628576621
12/23/1978	57.1167599	57.1167599059025	57.6315359495295	58.2030370915313	57.6315359495295	57.1167599059025	58.0619511910

Record: 7270 of 17075

Double-click table **Rock Frequency Table-district** to view. The table contains the maximums from 1-day, 30-day, 90-day, 183-day, 273-day and 365-day continuously-exceeded and average time-series for each year. The table also contains the minimums from 1-day, 30-day, 90-day, 183-day, 273-day and 365-day continuously-not-exceeded and average time-series for each year.

The screenshot below shows a few columns from the table

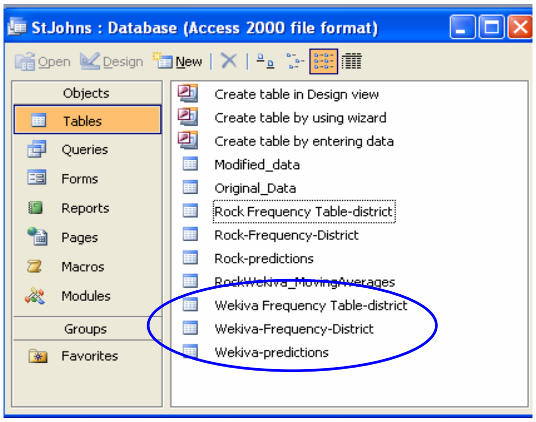


Rock Frequency Table-district : Table

Date	1-day(maximum-continuously exceeded)	30-day(maximum-continuously exceeded)	90-day(maximum-continuously exceeded)	18
1959	82.8601841317366	81.00025750499	76.6998127031651	76
1960	85.4865996893342	85.2483108329875	83.1038235842601	81
1961	73.5900034302823	72.7160718135159	71.5687363130881	67
1962	63.8505891873396	61.795956843456	60.5633394011976	60
1963	62.8634641659538	62.0517235757058	61.0985428742515	59
1964	65.6316926176219	64.7060683575706	62.4775543437449	60
1965	66.7407895124038	65.9579762874252	64.6915002651839	63
1966	71.6862441659538	71.0056139948674	70.2059197005988	68
1967	64.1273499572284	63.4136786227545	62.4763938323354	61
1968	77.4	71.506181899059	68.9481012788708	
1969	76.9	72.8421664813231	71.810764225834	70
1970	71.4230117764471	70.8043491844882	68.1	
1971	68	63.1787141345879	62.4561646877673	60
1972	67.7	62.6934472540633	61.4559859987168	
1973	67.2	63.5294404191617	62.5362370059881	60
1974	68.3	62.7666943455945	61.9061659794696	60
1975	65.4	62.2915132563445	61.2395954034787	60
1976	61.8073787282578	61.1257668976333	60.5	59
1977	58	56.9995235785572	55.4204155175364	55
1978	60.1091553293413	59.3341101539777	58.6250818648417	57
1979	68	62.0545608212147	61.2130338922156	59
1980	63	58.2769347390933	55.2881646107785	55
1981	61.6	51.8154605474765	50.4894486227545	49
1982	64.7	62.605389811805	61.8463550384944	60

Record: 1 of 46

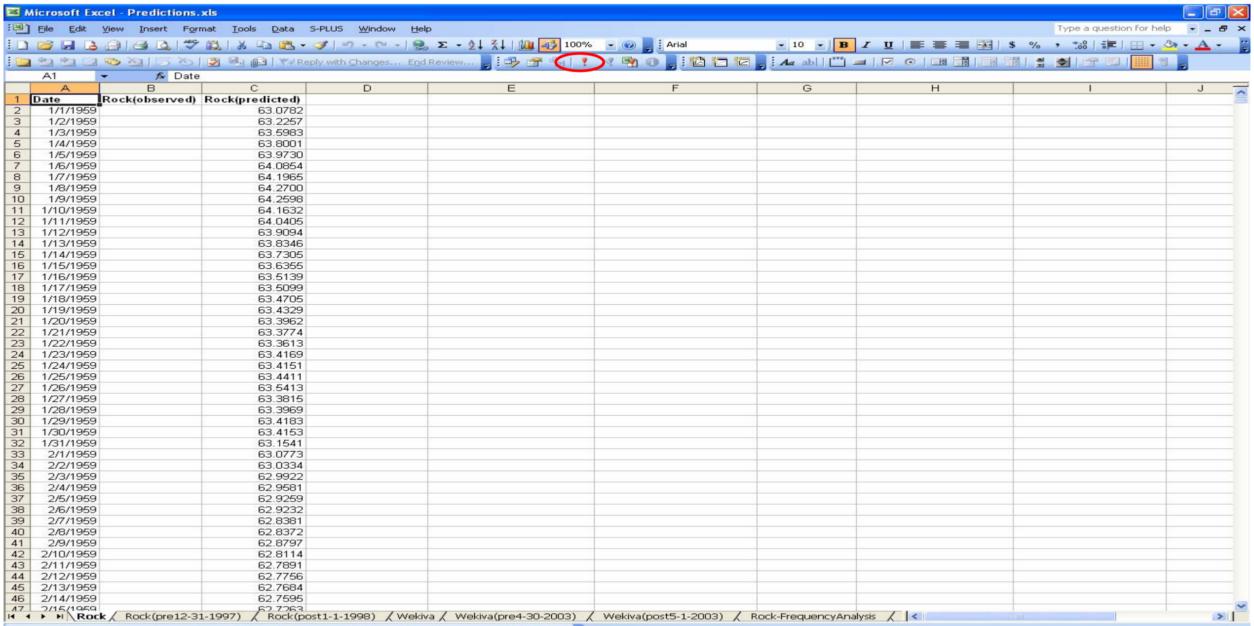
Similarly predictions and, maximum and minimum frequencies, for Wekiva Springs can be obtained for any specified upper and lower date ranges. Tables **Wekiva-predictions, Wekiva-Frequency-District, Wekiva Frequency Table-district** (shown below) are added to the database on clicking *Predict Spring Discharge – Wekiva* and following all the above steps as for Rock Springs.



6. Viewing prediction plots and maximum and minimum frequencies

Plots of observed and predicted daily discharge data can be viewed in the EXCEL file **predictions.xls** which is linked to the prediction tables in ACCESS. The file already has been run to include daily predictions and frequencies for Rock and Wekiva springs for the complete date ranges associated with the two springs.

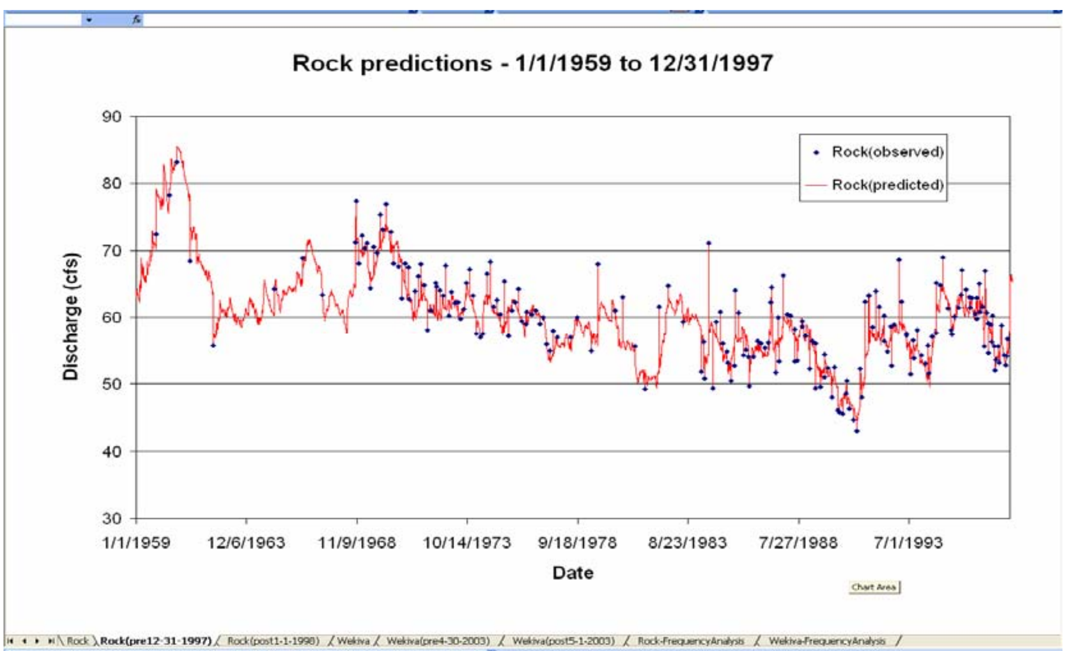
For our example, open **predictions.xls**. The screenshot below shows this file. By default, the *Rock* worksheet opens up; this contains the predictions for the complete range for which daily discharge values can be computed for Rock (1/1/1959 to 9/30/2005?)



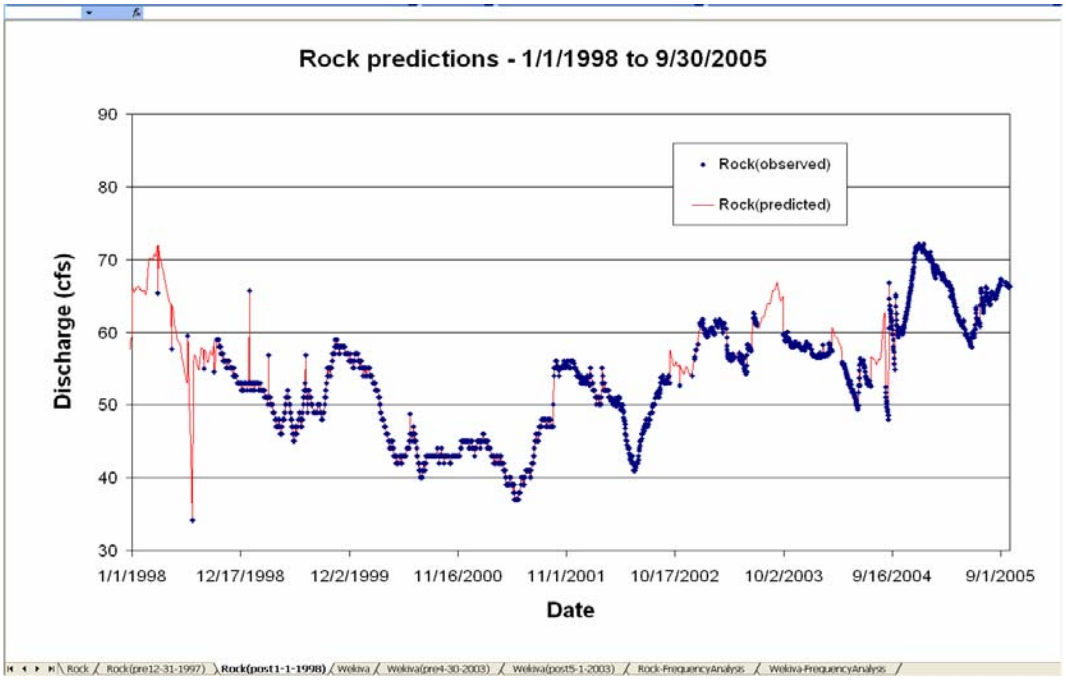
	Date	Rock(observed)	Rock(predicted)
2	1/1/1959		63.0782
3	1/2/1959		63.2257
4	1/3/1959		63.5983
5	1/4/1959		63.8001
6	1/5/1959		63.9730
7	1/6/1959		64.0854
8	1/7/1959		64.1965
9	1/8/1959		64.2700
10	1/9/1959		64.2598
11	1/10/1959		64.1632
12	1/11/1959		64.0405
13	1/12/1959		63.9094
14	1/13/1959		63.8346
15	1/14/1959		63.7305
16	1/15/1959		63.6355
17	1/16/1959		63.5139
18	1/17/1959		63.5099
19	1/18/1959		63.4705
20	1/19/1959		63.4329
21	1/20/1959		63.3962
22	1/21/1959		63.3774
23	1/22/1959		63.3613
24	1/23/1959		63.4169
25	1/24/1959		63.4151
26	1/25/1959		63.4411
27	1/26/1959		63.5413
28	1/27/1959		63.3815
29	1/28/1959		63.3969
30	1/29/1959		63.4183
31	1/30/1959		63.4153
32	1/31/1959		63.1541
33	2/1/1959		63.0773
34	2/2/1959		63.0334
35	2/3/1959		62.9922
36	2/4/1959		62.9581
37	2/5/1959		62.9259
38	2/6/1959		62.9232
39	2/7/1959		62.8381
40	2/8/1959		62.8372
41	2/9/1959		62.8797
42	2/10/1959		62.8114
43	2/11/1959		62.7891
44	2/12/1959		62.7756
45	2/13/1959		62.7684
46	2/14/1959		62.7595
47	2/15/1959		62.7263

The next step is pressing the red exclamation button to refresh the predictions for the date range which the user requested for this example, i.e. 1/1/1959 to 9/30/2005. The exclamation mark is highlighted by a red ellipse in the above figure.

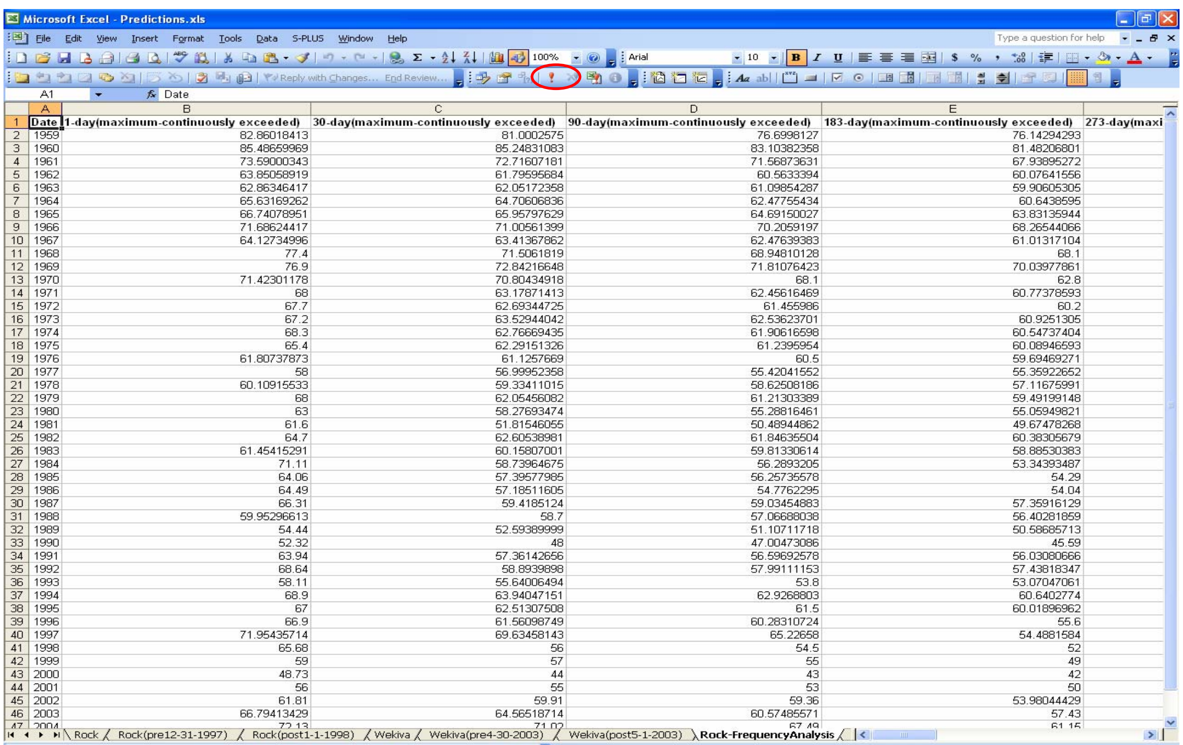
To view the plots for the above data, click on worksheet *Rock (pre12-31-97)* for predictions before 12/31/1997 and worksheet *Rock (post1-1-98)* for predictions from 1/1/1998. The screenshot below shows worksheet *Rock (pre12-31-97)*:



Also, the screenshot below shows worksheet *Rock (post1-1-98)*:



The procedure to view maximum and minimum frequencies is similar to viewing predictions. Click worksheet *Rock-FrequencyAnalysis* as shown below. We see the maximum and minimum frequencies for Rock for the year range 1959-2004



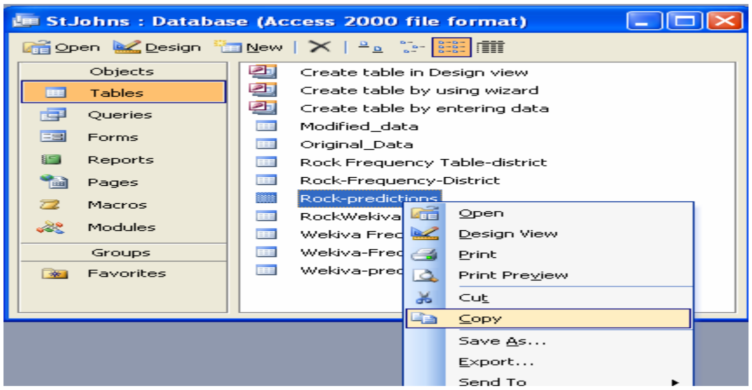
	A	B	C	D	E
1	Date	1-day(maximum-continuously exceeded)	30-day(maximum-continuously exceeded)	90-day(maximum-continuously exceeded)	183-day(maximum-continuously exceeded)
2	1959	82.86018413	81.0002575	76.6998127	76.14294293
3	1960	85.48659969	85.24831083	83.10382358	81.48206801
4	1961	73.59000343	72.71607181	71.56873631	67.93895272
5	1962	63.85058919	61.79595684	60.5633394	60.07641556
6	1963	62.86346417	62.05172358	61.09854287	59.90605305
7	1964	65.63169262	64.70606836	62.47755434	60.6438595
8	1965	66.74078951	65.95797629	64.69150027	63.83135944
9	1966	71.68624417	71.00561399	70.2059197	68.26544066
10	1967	64.12734996	63.41367862	62.47639383	61.01317104
11	1968	77.4	71.5061819	68.94810128	68.1
12	1969	76.9	72.84216648	71.81076423	70.03977861
13	1970	71.42301178	70.80434918	68.1	62.8
14	1971	68	63.17871413	62.45616469	60.77378593
15	1972	67.7	62.69344725	61.455986	60.2
16	1973	67.2	63.52944042	62.53623701	60.9251305
17	1974	68.3	62.76669435	61.90616598	60.54737404
18	1975	65.4	62.29151326	61.2395954	60.08946593
19	1976	61.80737873	61.1257669	60.5	59.69469271
20	1977	58	56.99952358	55.42041552	55.35922652
21	1978	60.10915533	59.33411015	58.62508186	57.11675991
22	1979	68	62.05456082	61.21303389	59.49199148
23	1980	63	58.27693474	55.28816461	55.05949821
24	1981	61.6	51.81546055	50.48944862	49.67478268
25	1982	64.7	62.60538981	61.84635504	60.38305679
26	1983	61.45415291	60.15807001	59.81330614	58.88530383
27	1984	71.11	58.73964675	56.2893205	53.34393487
28	1985	64.06	57.39577985	56.25735578	54.29
29	1986	64.49	57.18511605	54.7762295	54.04
30	1987	66.31	59.4185124	59.03454883	57.35916129
31	1988	59.95296613	58.7	57.06688038	56.40281859
32	1989	54.44	52.59389999	51.10711718	50.58685713
33	1990	52.32	48	47.00473086	45.59
34	1991	63.94	57.36142656	56.59692578	56.03080666
35	1992	68.64	58.8939898	57.99111153	57.43818347
36	1993	58.11	55.64006494	53.8	53.07047061
37	1994	68.9	63.94047151	62.9268803	60.6402774
38	1995	67	62.51307508	61.5	60.01896962
39	1996	66.9	61.56098749	60.28310724	55.6
40	1997	71.95435714	69.63458143	65.22658	54.4881584
41	1998	65.68	56	54.5	52
42	1999	59	57	55	49
43	2000	48.73	44	43	42
44	2001	56	55	53	50
45	2002	61.81	59.91	59.36	53.98044429
46	2003	66.79413429	64.56518714	60.57485571	57.43
47	2004	72.13	71.02	67.49	61.15

The next step is pressing the red exclamation button to refresh the frequencies for the date range which the user requested for this example, i.e. 1/1/1959 to 9/30/2005. The exclamation mark is highlighted by a red ellipse in the above figure.

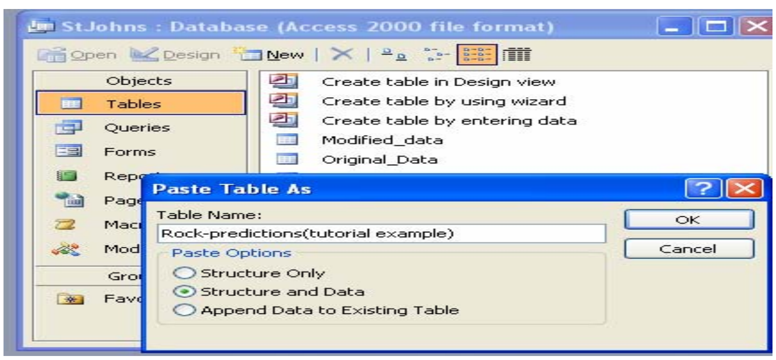
The table above only shows the maximum and minimum frequencies for the years they can be computed.

7. Saving results for different cases

To save the daily discharge predictions and frequencies for a particular set of well or spring data in **Original Data** table, make another copy of the prediction tables in ACCESS and give them a different name. This step is crucial since for a new set of data, the prediction and frequency tables are overwritten. In our example for instance, copy-paste the Rock-predictions table as shown below:



ACCESS prompts for a new name as shown below:



Enter a table name and press OK. The prediction table for our example is created. Similarly create new tables for the Rock-frequency-district and Rock Frequency Table-district.

It is also necessary to save the predictions and frequencies in **predictions.xls** in a different file before the prediction worksheets in EXCEL are refreshed to get predictions for a different case.

APPENDIX B
Resolution of Peer Review Comments

APPENDIX B: Resolution of Peer Review Comments.

Appendix B contains the comments provided by peer review of the first report in this Statistical Modeling of Spring Discharge series and the author's resolution of these comments. This peer review and the subsequent resolution pertain to application of statistical methodology and are, therefore, included in this report as well. The report modifications included some comments on potential use of the presented models as well as a clear statement of the models objectives.

Memorandum

TO: Bob Epting, St. Johns River Water Management District

FROM: Shahrokh Rouhani, Ph.D., P.E., NewFields

SUBJECT: Peer review of "Statistical Modeling of Spring Discharge at Ponce de Leon, Green, and Gemini Springs in Volusia County Florida" by Intera (2005) and "Statistical Modeling of Spring Discharge at Apopka and Bugg Springs in Lake County Florida" by Intera (2006)

DATE: July 16, 2006

****

INTRODUCTION

St. Johns River Water Management District (District) is engaged in ongoing Minimum Flows and Levels (MFLs) and Water Supply Development projects. Such projects require daily discharge time series at a number of springs of interest. Most of these springs suffer from sporadic discharge measurements. Intera (2005 and 2006) utilizes multiple regression models to estimate (hindcast) daily discharges at a number of springs of interest based on a variety of available nearby moving averages of measured spring discharges, groundwater levels, lake levels, and precipitation rates. The estimated daily discharge time series at each spring are then used to generate frequency, duration, discharge curves.

GENERAL COMMENT

In general, I must note that the reports are well written, and easy to follow. Furthermore, from a conceptual point of view, multiple regression of nearby hydrologic data to fill the gaps in times series of daily spring discharges is quite acceptable. The resulting estimated time series and frequency curves also display reasonable patterns consistent with existing, albeit limited, discharge measurements at the investigated springs. However, the review of the reports raises a number of fundamental questions that may warrant further considerations by the authors. These mainly statistical questions are the focus of this memorandum.

SPECIFIC COMMENTS

1. The above reports use multiple regression models that relate moving averages (MA) of nearby hydrologic data to estimate daily spring discharges. Intera (2005) presents the general form of such a model as

   [Spring discharge] = ***f*** {[same spring MA] + [water level MA] + [precipitation MA] + [adjacent spring MA] }

   The authors state that "the use of moving-average-based independent variables is necessitated by the fact that most independent variables are not measured on a daily basis." Although, statistical methods, including multiple regression analysis, are not bound by hydrological principals, it is always desirable to use independent variables that are hydrologically consistent with the dependent variable.

   The independent variable in the above reports is daily spring discharge, i.e. a non-integrated or *differentiated* flow variable. Daily precipitation is also a flow variable, while water levels (either groundwater or lake levels) are storage variables. Within the context of mass balance, the net sum of flows is equal to the rate of change of storage variables. In other words, in a linear model, daily spring discharge is expected to be related to (a) daily values of other flow variables (e.g. precipitation or nearby spring discharges), and (b) daily rates of changes in storage variables (e.g. water levels). This implies that under ideal conditions, non-integrated flow variables and differentiated storage variables should be used in a regression model.

   While I recognize that absence of continuous data may make some of the above differentiations impossible, I am still puzzled about the fact that all dependent variables are uniformly integrated. Integration is the exact opposite of what mass balance suggests. In fact, in cases that continuous daily time series of storage variables (e.g. groundwater or lake levels) are available; their difference values should be explored as an alternative to the current moving averages. For this purpose, continuous or augmented groundwater level time series, such as L-0054 and L-0703, along with other complete daily time series appear to be suitable candidates. I encourage the authors to consider this alternative approach, which is more consistent with the mass balance concept.

2. Intera (2006) notes the issue of multicolinearity, but suggests that computation of partial correlation coefficients (PCC) and stepwise analysis somehow solves this problem. While the use of PCC and stepwise analysis are commendable, they do not address the issue of multicolinearity.

   Multiple regression analysis is based on the fundamental assumption that the variables on the right hand side of the equation are statistically independent, i.e. uncorrelated. Multicolinearity exists when independent variables are highly correlated. Unfortunately, the reports do not contain any systematic information on cross correlations among independent variables. However, statements made in Intera (2006) concerning high correlations among certain groundwater levels (which were used to justify the filling of data gaps in some of the monitoring wells) clearly indicate that at least some of the independent variables are highly correlated. This is especially true for moving averages of the same variables, which are used concurrently as independent variables in the same model. So one can assume that some, if not all of the models used in Intera (2005 and 2006), suffer from multicolinearity.

   A high degree of multicolinearity produces unacceptable uncertainty (large variance) in regression coefficient estimates. Specifically, the coefficients can change drastically depending on which terms are in or out of the model and also the order they are placed in the model. In fact, a typical consequence of multicolinearity is a high regression coefficient, when a number of independent variables have regression coefficients that are deemed as insignificant. For example, Table 8 in Intera (2006) indicates that of the 13 independent variables used to estimate Apopka (post-1990) five variables have statistically insignificant coefficient (i.e. their p values are greater than or equal to 0.05), while R^2 of the same model is nearly 0.80. In other words, the regression results indicate that the collection of selected independent variables has explanatory power but we cannot tell which variable or to what degree the individual variable is explaining the variations of the dependent variable. Generally, such ‘black-box’ models are viewed as undesirable.

   I encourage the authors to consider computing the variance inflation factor (VIF) of each independent variable. VIF associated with the i^{th} independent variable is equal to

$\frac{1}{1-R_i^2}$ where R_i is the regression coefficient of the ith independent variable on all of the other independent variables. A rule of thumb is to treat any VIF in excess of 10 as evidence of multicolinearity. Elimination of collinear independent variables should continue until all VIF are below 10. This approach along with the stepwise analysis would lead to much more defensible models. Other remedies are also discussed in Gujarati (*Basic Econometrics*, 4th Edition, McGraw Hill, 2002, Chapter 10).

3. The results of predicted versus observed time series are visually satisfactory (e.g. Figure 18 in Intera, 2006); however, their corresponding observed versus predicted plots (e.g. Figure 12 in Intera 2006) display poor fits. An explanation of this visual discrepancy would be helpful. I also noticed that the updated frequency curves for Apopka and Bugg springs are much closer to the pattern exhibited by the observed data. However, the addendum dated July 11, 2006 does not describe the reason for this improvement.

4. To compare observed versus predicted discharges, the authors should also consider the comparison of their variances. Results like Figure 12 (Intera, 2006) imply that the predicted values are much less variable that measured discharges. Although, such results are not unexpected (estimated values are generally smoother than actual data), the impacts of such smoothings on the frequency curves must be discussed. Specifically, are extreme discharges adequately estimated?

   Consider the updated frequency curve for Bugg Spring (Intera addendum dated 7/11/06). While observed discharges in the central portion of the curve match their estimated values, extreme values deviate systematically, i.e. biased results. Similar patterns are present in almost all the generated frequency curves. The authors should address this issue, and if deemed significant, appropriate remedies should be considered.

PREPARED FOR:	Bob Epting, St. Johns River Water Management District
PREPARED BY:	Alaa Aly and Srikanta Mishra, INTERA Incorporated
SUBJECT:	Resolution of peer review comments of "Statistical Modeling of Spring Discharge at Ponce de Leon, Green, and Gemini Springs in Volusia County Florida" by Intera (2005) and "Statistical Modeling of Spring Discharge at Apopka and Bugg Springs in Lake County Florida" by Shahrokh Rouhani, NewFields
DATE:	August 8, 2007

INTRODUCTION

St. Johns River Water Management District (District) is engaged in ongoing Minimum Flows and Levels (MFLs) and Water Supply Development projects. Such projects require daily discharge time series at a number of springs of interest. Most of these springs suffer from sporadic discharge measurements. Intera (2005 and 2006) utilizes multiple regression models to estimate (hindcast) daily discharges at a number of springs of interest based on a variety of available nearby moving averages of measured spring discharges, groundwater levels, lake levels, and precipitation rates. The estimated daily discharge time series at each spring are then used to generate frequency, duration, discharge curves.

GENERAL COMMENT

We appreciate the comments from Dr. Rouhani about the validity of the approach and the clarity of the presentation in the report. The following sections address the specific comments in the peer review memorandum.

SPECIFIC COMMENTS

1. …… Within the context of mass balance, the net sum of flows is equal to the rate of change of storage variables. ……. This implies that under ideal conditions, non-integrated flow variables and differentiated storage variables should be used in a regression model. While I recognize that absence of continuous data may make some of the above differentiations impossible, I am still puzzled about the fact that all dependent variables are uniformly integrated. Integration is the exact opposite of what mass balance suggests. ……. I encourage the authors to consider this alternative approach, which is more consistent with the mass balance concept.

While mass balance would suggest exactly what the reviewer points out, the presented models are statistical, not physical. Therefore, they are not intended to be used as mass balance models. The models are based on exploitation of the statistical correlation between the explanatory and response variables. For example, spring discharge is correlated with aquifer water levels, perhaps with a lead time. This correlation explains some of the variability in the observed spring discharge rates. Further, the correlation is improved using the average water level values rather than the individual measurements which always have higher variances. However, as the reviewer notes, spring discharge can also be expected to be correlated to the change in water levels over time. These changes are a function of the "net" change of fluxes to and from the aquifer. In the absence of other significant fluxes such as recharge and pumping, these changes will be closely correlated to the observed spring discharge rates. Unobserved (e.g., pumping) and unobservable (e.g., aquifer recharge) fluxes will complicate this correlation. Further, as noted, this difference is typically very difficult to obtain from real data as data gaps can be a major obstacle for such calculation.

2. Intera (2006) notes the issue of multicolinearity, but suggests that computation of partial correlation coefficients (PCC) and stepwise analysis somehow solves this problem. …… Multiple regression analysis is based on the fundamental assumption that the variables on the right hand side of the equation are statistically independent, i.e. uncorrelated. ….. However, statements made in Intera (2006) concerning high correlations among certain groundwater levels (which were used to justify the filling of data gaps in some of the monitoring wells) clearly indicate that at least some of the independent variables are highly correlated. ...... So one can assume that some, if not all of the models used in Intera (2005 and 2006), suffer from multicolinearity. ….. I encourage the authors to consider computing the variance inflation factor (VIF) of each independent variable.

First, multicolinearity is mainly a problem for the uniqueness and variances for the regression coefficients. That is, when correlated variables are used as explanatory variables, the fitted regression coefficients will not be meaningful and might have very high variances. However, the predicted values from such regression model are still acceptable with the only issue that needs to be addressed is whether adding the correlated variable(s) have resulted in unnecessary inflation of the prediction variance. This variance inflation resulting from adding more variables to the regression equation is exactly what is considered in the stepwise regression algorithm. As detailed below, a variable is only added to the regression equation if it will improve the prediction capability of the final regression equation without adding significantly to the prediction variance. Our experience in applying stepwise regression to outputs of probabilistic risk assessment models confirms this. We have also computed variance inflation factors for the discharge models for Rock and Wekiva springs, and these also indicate that the stepwise regression process has minimized multicolinearity issues. The following description of stepwise regression provides the background information for the procedure showing how multicolinearity is formally dealt with.

In the utilized stepwise approach, a sequence of regression models is constructed starting with the input variable that explains the largest amount of variance in the output, i.e., the variable that has the highest Pearson correlation coefficient with the output. At each successive step in the regression modeling process, the variable that explains the largest fraction of unexplained variance from the previous step is included. This is the variable with the largest absolute value of the partial correlation coefficient. The model generated at every step is tested to ensure that the each of the regression coefficients is significantly different from zero. The test is implemented in two stages. First, a variable selected for entry via the PCC criterion is tested for its significance before it is admitted into the model. Second, after the model is built at that step, each of the variables in the model is tested for significance. If some variables are found to be insignificant, then the “most insignificant” variable is dropped and the model is built again. The sequential dropping of the variables judged to be not significant and rebuilding the model continues until all the variables in the model become significant at the prescribed levels. The significance levels are prescribed separately for the entering and departing variables to avoid possible looping where the same variable can enter and depart from the model with the significance level for the departing variables generally set larger than that for the entering variable. Note that the need for dropping a variable generally arises only in the cases when the input variables are strongly correlated (strong multicolinearity). This step ensures that no significant multicolinearity will be present in the final multiple regression model. The stepwise regression process continues until the input-output model contains all of the input variables that explain statistically significant amounts of variance in the output (i.e., no more variables are found with a statistically significant regression coefficient).

3. The results of predicted versus observed time series are visually satisfactory (e.g. Figure 18 in Intera, 2006); however, their corresponding observed versus predicted plots (e.g. Figure 12 in Intera 2006) display poor fits. An explanation of this visual discrepancy would be helpful. I also noticed that the updated frequency curves for Apopka and Bugg springs are much closer to the pattern exhibited by the observed data. However, the addendum dated July 11, 2006 does not describe the reason for this improvement.

Figure 18 shows that the general pattern displayed by the observed discharge hydrograph for Bugg Spring. While there is significant visual scatter shown in Figure 12, this figure also shows that the vast majority of the predicted discharge values are in agreement with the observed values. Figure 12 also shows that there in no general bias in any direction for the entire range of observed discharge values, a further affirmation for the validity of predictive model. The explanations missing from the July 11, 2006 addendum have been added to the final report.

4. To compare observed versus predicted discharges, the authors should also consider the comparison of their variances. Results like Figure 12 (Intera, 2006) imply that the predicted values are much less variable that measured discharges. Although, such results are not unexpected (estimated values are generally smoother than actual data), the impacts of such smoothings on the frequency curves must be discussed. Specifically, are extreme discharges adequately estimated?

Consider the updated frequency curve for Bugg Spring (Intera addendum dated 7/11/06). While observed discharges in the central portion of the curve match their estimated values, extreme values deviate systematically, i.e. biased results. Similar patterns are present in almost all the generated frequency curves. The authors should address this issue, and if deemed significant, appropriate remedies should be considered.

While it is not anticipated that extreme discharge values will be predicted accurately, it is important that no consistent bias is displayed by the predictive models. Figure 12 clearly shows that predicted values are not biased at either end of the observed discharge values because high and low values are equally spread around the regression line. Further, additional analyses are added to the report to examine the differences between the variances of the observed and regression-model-generated spring discharge values.