SPECIAL PUBLICATION SJ2007-SP9

STATISTICAL MODELING OF SPRING DISCHARGE AT PONCE DE LEON, GREEN, AND GEMINI SPRINGS IN VOLUSIA COUNTY, FLORIDA

Statistical Modeling of Spring Discharge at Ponce de Leon, Green, and Gemini Springs in Volusia 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 20, 2007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Currently, the St. Johns River Water Management District (District) is engaged in hydrologic modeling and data analysis in support of the 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 streams in the District. This report describes the development of statistical models for predicting daily spring discharge time series for Ponce de Leon (PDL), Gemini, and Green springs from an assortment of auxiliary data such as: (a) previously recorded spring discharge rates at the springs 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. Furthermore, the correlation is improved using the average water level values rather than the individual measurements which are known to display higher variances.

Data screening indicates that most measurements of spring discharge and groundwater level are at a frequency of ~30 days or greater – necessitating the generation of moving averages with commensurate lags to be used as independent variables for predicting spring daily discharge. Also, the Blue Spring daily discharge values show significant correlation with the daily discharge values at Gemini Springs and some correlation with PDL. Hence, discharge from Blue Spring has been utilized to help estimate discharge at Gemini and PDL springs when groundwater level measurements are scarce. Analysis of data overlap is helpful in determining how to partition the period of record into sub-periods where a common set of variables can be defined.

Forward stepwise regression analysis is used to build multivariate linear input-output models between the response variable (spring discharge) and the independent variables (moving averages of water level measurements and precipitation) at the springs of interest. Typically, two regression models of spring discharge are needed: (a) one for the period when groundwater levels and rainfall data are available, and (b) one for the period when rainfall data are

supplemented with discharge from adjacent springs and perhaps groundwater levels from one or two long-term monitoring wells.

The following regression model is developed for PDL Springs:

- PDL discharge as a function of water level measurements from Floridian aquifer well (FAW) L-0045 (8-, 48- and 52-week moving average), 12-, 48-, and 52-week moving average of PDL discharge and 52-week moving average of the Blue Spring discharge (R^2=0.57).

For Gemini Springs, the regression models developed are as follows:

- Gemini 1995-2000 discharge as a function of 12-, 48-, and 52-week moving average of the Blue Spring discharge, and water levels at FAW S-0257 (12- and 24-week moving average) – in order to predict the daily Gemini Springs discharge prior to 1995 when no measurements are available at the springs (R^2=0.47).
- Gemini 1995-2004 discharge as a function of daily and 4-, 8-, 12-, and 52-week moving average of rainfall at Sanford, Florida, and water levels at S-1230 (4-, 6-, 8-, and 24-week moving average) – in order to predict the daily Gemini Springs discharge during the 1995-2004 period (R^2=0.79).

Finally, the following regression models are developed for Green Springs:

- Green 2000-2004 discharge as a function of water levels at FAW V-0166 (24-, 48-, and 52-week moving average) and 1-, 3-, and 48-week moving average of rainfall at Sanford – in order to predict discharge at Green Springs in the 1987-1996 period (R2=0.89).
- Green 1999-2004 discharge as a function of water levels at FAW V-0810 and FAW V-0772, and 4-, 8-, and 48-week moving average of rainfall at Stanford, Florida – in order to predict the daily Green Springs discharge during this period and for 1996-1999 (R^2=0.97).

Using these models, daily discharge predictions are made for PDL Springs as far back in time as 1966 with reasonable accuracy. However, comparable predictions can only be made until 1996 for Gemini Springs and until 1988 for Green Springs. Flow duration curves are also generated for all three springs along with high- and low-frequency analyses for set durations (1-, 2-, 3-, 4-, 6-, and 12-months) from the simulated daily spring discharge.

This report incorporates comments provided by peer review of an earlier version. The modifications included some comments on potential use of the presented models as well as a clear statement of the models objectives. Further, additional analysis was added to the report to highlight the differences between the variances of the observed and regression-model-generated spring discharge values. 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 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 predictions at various streams in the District. MFLs for three springs in Volusia County, Florida, namely, Ponce de Leon (PDL), Gemini, and Green springs are currently needed. As discussed in the following sections, while the PDL Springs has more data than either Green Springs or Gemini Springs, each of these springs has limited spring flow measurements (Osburn et al., 2002). This study evaluates the application of statistical models to generate long-term daily discharge simulations for each of these three springs.

2.0 OBJECTIVE

The objective of this study is the development of daily spring discharge time series for PDL, Gemini, and Green springs from an assortment of auxiliary data such as: (a) previously recorded springs discharge rates at the springs of interest and at adjacent springs, (b) groundwater level measurements from adjacent monitoring wells, 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 exploratory data analysis (EDA) for rainfall, water level and spring discharge data for PDL, Gemini, and Green 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. Exploratory data analysis is described in Section 3. Section 4 contains the regression modeling methodology and the regression models developed for PDL, Gemini, and Green 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 EXPLORATORY DATA ANALYSIS

This section summarizes the available data and shows the results of the statistical exploratory analyses (EDA) conducted for the available time series. The objective of the EDA is to identify the correlation structure between the spring discharge at the three springs of interest and the other time series. Results from the EDA 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:

- Spring discharge measurements at PDL, Gemini, and Green springs; as well as measurements at Blue Spring.
- Groundwater level measurements at monitoring wells:
  - V-1030. V-0156, V-0742, L-0045, V-0095, and V-0096 for PDL
  - S-1230 and S-0257 for Gemini Springs
  - V-0810, V-0772, and V-0166 for Green Springs
  - V-0083, V-1091, V-0196, L-0059, V-0101, and M-0024 for Blue Spring
- Precipitation measurements at rain gages:
  - SR-40 and SR-11 for PDL
  - Sanford for Gemini and Green Springs
  - Deland for Blue Spring.

Blue Spring is located in close proximity to the three primary springs of interest, and is a potential source of ancillary data because of its extensive period of record. The other wells are chosen because of their locations vis-à-vis the springs of interest, and also if they provide a long-term record of groundwater level measurements.

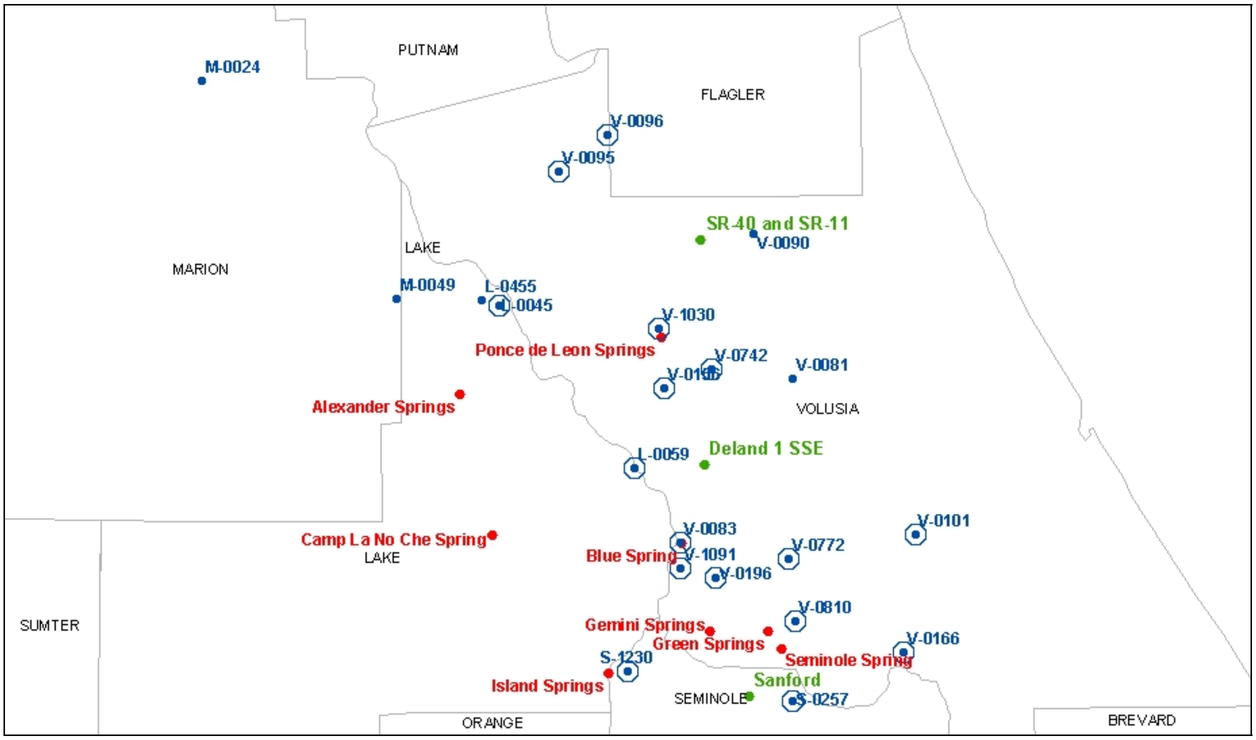


Figure 1 Location of springs and groundwater monitoring wells in region of interest.

In order to conduct exploratory data analysis, a database was compiled of spring discharge (response variable), groundwater levels (explanatory variable), and precipitation (explanatory variable) with a common time basis. For each spring of interest, several groundwater monitor wells in its vicinity and the nearest rain gage were selected. Table 1 shows basic statistics (i.e., minimum, maximum, average and standard deviation) for these various data types at Blue, Green, PDL, and Gemini springs.

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 the springs of interest as well as at adjacent springs at selected lag times such as 1, 2, 3, 4, 6, 8, 12, 24, 48, and 52 weeks. These moving averages act as surrogate predictor variables and carry useful information regarding the physical state of the system prior to the time of interest.

Table 1 Basic statistics for various data types at Blue, Green, PDL, and Gemini springs.

Data Type	Range	Min	Max	Average	Std Dev	Variable type
Blue Spring	3/7/32-12/18/01	63.00	217.73	156.25	19.30	Discharge(cfs)
V-0083	4/20/95-12/1/04	4.42	11.82	7.11	1.37	Water-level(ft)
V-1091	9/25/81-12/01/04	9.44	22.60	19.10	1.67	Water-level(ft)
V-0196	1/7/87-12/22/03	11.75	22.31	16.63	2.87	Water-level(ft)
L-0059	1/31/84-11/18/04	11.80	20.64	16.36	1.47	Water-level(ft)
V-0101	5/28/36-11/22/04	24.49	32.10	29.58	1.36	Water-level(ft)
M-0024	11/21/85-11/16/04	21.56	29.28	24.28	1.54	Water-level(ft)
Deland	2/17/29-6/30/04	0.00	7.77	0.16	0.43	Rainfall(in)
Green Springs	1/20/00-11/11/04	0.00	2.92	1.31	0.94	Discharge(cfs)
V-0810	12/26/96-12/15/04	9.76	18.85	14.31	2.03	Water-level(ft)
V-0772	8/3/95-12/15/04	6.88	17.10	11.66	2.11	Water-level(ft)
V-0166	1/7/87-11/24/03	11.00	18.03	14.22	1.32	Water-level(ft)
Sanford	1/1/48-6/30/04	0.00	6.88	0.14	0.41	Rainfall(in)
Ponce De Leon	1/14/65-1/13/97	16.67	40.90	27.57	4.52	Discharge(cfs)
V-1030	10/6/94-10/19/04	9.78	23.78	18.72	1.64	Water-level(ft)
V-0156	8/29/84-8/25/04	1.78	19.51	14.99	2.13	Water-level(ft)
V-0742	11/1/93-10/19/04	21.81	39.06	32.52	2.23	Water-level(ft)
L-0045	1/24/50-11/18/04	10.69	18.06	14.31	1.60	Water-level(ft)
V-0095	3/20/36-12/1/04	8.72	27.90	23.16	2.03	Water-level(ft)
V-0096	2/18/36-11/18/04	14.51	22.90	20.26	1.42	Water-level(ft)
SR-40 & SR-11	10/1/93-11/17/04	0.00	8.91	0.14	0.44	Rainfall(in)
Gemini Springs	9/22/95-12/1/04	6.20	13.00	9.96	1.40	Discharge(cfs)
S-1230	2/26/96-11/15/04	16.44	22.67	20.22	1.64	Water-level(ft)
S-0257	11/21/52-9/20/99	16.66	26.45	22.61	1.32	Water-level(ft)
Sanford	01/01/48-6/30/04	0.00	6.88	0.14	0.41	Rainfall(in)

3.2 Frequency of Observation

Table 2 shows the mean and standard deviation of frequency of observation for each data type for Blue, Green, PDL, and Gemini springs. For Green Springs, the springs discharge has a period of record dating back to January 2000 at an average frequency of 57 days – although a few isolated observations extend back to February 1972. At well V-0810, groundwater levels are available daily from December 1996. At well V-0772, groundwater levels are available daily from August 1995. At well V-0166, groundwater level measurements are available from January of 1987 at a frequency of 30 days. Finally, for the Sanford rain gage, daily precipitation observations are available from January 1948.

Table 2 Frequency of observation of various data types at Blue, Green, PDL, and Gemini springs.

Data Type	Range	Min	Max	Average	Std Dev	Variable type
Blue Spring	3/7/32-12/18/01	63.00	217.73	156.25	19.30	Discharge(cfs)
V-0083	4/20/95-12/1/04	4.42	11.82	7.11	1.37	Water-level(ft)
V-1091	9/25/81-12/01/04	9.44	22.60	19.10	1.67	Water-level(ft)
V-0196	1/7/87-12/22/03	11.75	22.31	16.63	2.87	Water-level(ft)
L-0059	1/31/84-11/18/04	11.80	20.64	16.36	1.47	Water-level(ft)
V-0101	5/28/36-11/22/04	24.49	32.10	29.58	1.36	Water-level(ft)
M-0024	11/21/85-11/16/04	21.56	29.28	24.28	1.54	Water-level(ft)
Deland	2/17/29-6/30/04	0.00	7.77	0.16	0.43	Rainfall(in)
Green Springs	1/20/00-11/11/04	0.00	2.92	1.31	0.94	Discharge(cfs)
V-0810	12/26/96-12/15/04	9.76	18.85	14.31	2.03	Water-level(ft)
V-0772	8/3/95-12/15/04	6.88	17.10	11.66	2.11	Water-level(ft)
V-0166	1/7/87-11/24/03	11.00	18.03	14.22	1.32	Water-level(ft)
Sanford	1/1/48-6/30/04	0.00	6.88	0.14	0.41	Rainfall(in)
Ponce De Leon	1/14/65-1/13/97	16.67	40.90	27.57	4.52	Discharge(cfs)
V-1030	10/6/94-10/19/04	9.78	23.78	18.72	1.64	Water-level(ft)
V-0156	8/29/84-8/25/04	1.78	19.51	14.99	2.13	Water-level(ft)
V-0742	11/1/93-10/19/04	21.81	39.06	32.52	2.23	Water-level(ft)
L-0045	1/24/50-11/18/04	10.69	18.06	14.31	1.60	Water-level(ft)
V-0095	3/20/36-12/1/04	8.72	27.90	23.16	2.03	Water-level(ft)
V-0096	2/18/36-11/18/04	14.51	22.90	20.26	1.42	Water-level(ft)
SR-40 & SR-11	10/1/93-11/17/04	0.00	8.91	0.14	0.44	Rainfall(in)
Gemini Springs	9/22/95-12/1/04	6.20	13.00	9.96	1.40	Discharge(cfs)
S-1230	2/26/96-11/15/04	16.44	22.67	20.22	1.64	Water-level(ft)
S-0257	11/21/52-9/20/99	16.66	26.45	22.61	1.32	Water-level(ft)
Sanford	01/01/48-6/30/04	0.00	6.88	0.14	0.41	Rainfall(in)

For Blue Spring, the spring's discharge has an extended period of record dating back to March 1932 at an average frequency of 43 days. At well V-0083, groundwater levels are available daily from April 1995 – although a few isolated observations extend back to December 1984. At well V-1091, groundwater levels are available at an average frequency of 52 days between September 1981 and March 2002 and daily thereafter. At well V-0196, groundwater levels are available at an average frequency of 24 days from January 1987. At well L-0059, groundwater levels are available at an average frequency of 26 days from January 1984 – although a few isolated observations extend back to may 1976. At well V-0101, groundwater levels are available at an average frequency of 31 days from May 1936. At well M-0024, groundwater levels are available at an average frequency of 30 days from November 1985 –

although a few isolated observations extend back to September 1980. Finally, for the Deland rain gage, daily precipitation observations are available from February 1929.

For PDL Springs, the springs' discharge has a period of record dating back to January 1965 at an average frequency of 55 days – although a few isolated observations extend back to February 1929. The PDL Springs discharge data, recorded after April 7, 1997, are of questionable quality because of biased measurements introduced by construction of the pool weir. Hence, PDL discharge records measured only before April 7, 1997, are used for the regression model. At well V-1030, groundwater levels are available daily from October 1994. At well V-0156, groundwater levels are available at an average frequency of 24 days between from August 1984. At well V-0742, groundwater levels are available daily from November 1993. At well L-0045, groundwater levels are available at an average frequency of 47 days from January 1950. At wells V-0095 and V-0096, groundwater levels are available from early 1936 at a frequency of 6 and 36 days, respectively. Finally, for the SR-40/SR-11 rain gage, daily precipitation observations are available from October 1993.

For Gemini Springs, the springs' discharge has a period of record dating back to September 1995 at an average frequency of 56 days – although a few isolated observations extend back to June 1966. At well S-1230, groundwater levels are available at an average frequency of 29 days between February 1996 and October 2004 and daily thereafter. At well S-0257, groundwater levels are available at an average frequency of 6.5 days beginning in November 1952. Finally, for the Sanford rain gage, daily precipitation observations are available from January 1966.

3.3 Analysis of Overlap

Periods of overlap between different data types were analyzed for each of the springs of interest. 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. This is useful for determining appropriate lag and moving average windows. Moving averages were calculated for recorded water levels, precipitation and springs discharge at adjacent springs at selected lag times such as 1, 2, 3, 4, 6, 8, 12, 24, 48, and

52 weeks. The moving averages act as surrogate predictor variables and carry useful information regarding the physical state of the system prior to the time of interest.

Figure 2 shows the overlap between various data types for the PDL Springs. Shown here are the periods of record for (a) springs discharge, (b) groundwater levels at monitoring wells V-1030, V-0156, V-0742, L-0045, V-0095, and V-0096, and (c) precipitation measurement at SR-40 and SR-11. Also indicated therein is the average frequency of observation for each data type (as was discussed in detail in the previous section). As mentioned before, PDL data starts from 1965 and is only available through the spring of 1997. Prior to that, groundwater level data at wells L-0045, V-0095 and V-0096 are available starting in the late 1930s. However, the average data frequency for these wells is about 7 weeks. It is likely that a moving average window of 8 weeks or greater will be used to take advantage of this water level measurement. From 1995, several time series are available, but this information cannot be used in the regression model due to short period of overlap with PDL. Information from another spring (Blue) could also be added to the explanatory variables set to help provide more accurate estimates for daily discharge at PDL Springs.

Figure 3 shows the overlap between various data types for the Gemini Springs. Shown here are the periods of record for (a) springs discharge, (b) groundwater levels at monitoring wells S-1230 and S-0257, and (c) precipitation measurement at Stanford. Also indicated therein is the average frequency of observation for each data type (as was discussed in detail in the previous section). The daily groundwater level at well S-1230 is available only since mid-2004. Prior to this timeframe, the well has an average recording frequency of about once a month. The Sanford daily rainfall record has a much longer record starting in the 1940s. It is likely that two different model sets will be used to estimate daily discharge at Gemini Springs. The first set of models will cover recent times since groundwater measurements became available. The early models will use rainfall and water levels at wells S-0257 as supporting explanatory variables. However, as the report will detail later, information from Blue Spring could also be added to the explanatory variables set to help provide more accurate estimates for daily discharge at Gemini Springs.

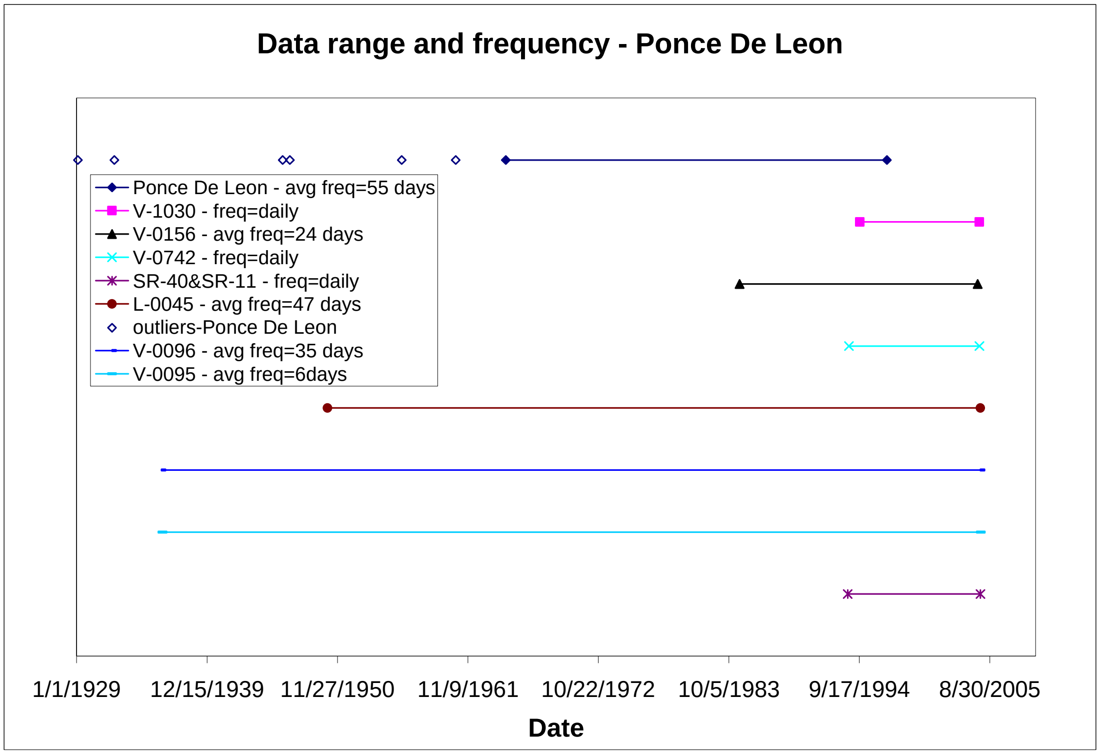


Figure 2 Overlap between various data types, PDL Springs.

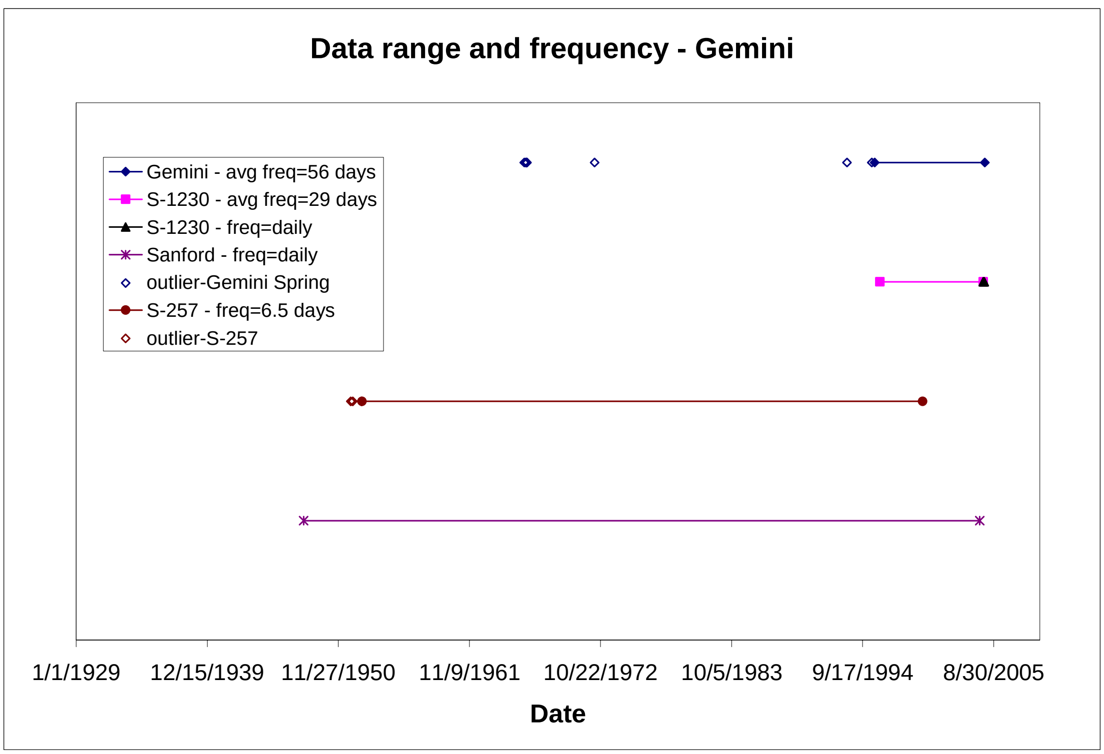


Figure 3 Overlap between various data types, Gemini Springs.

Figure 4 shows the overlap between various data types for the Green Springs. Shown here are the periods of record for: (a) springs discharge, (b) groundwater levels at monitoring wells V-0810, V-0772, and V-0166, and (c) precipitation measurement at Sanford. Also indicated therein is the average frequency of observation for each data type (as was discussed in detail in the previous section). Data availability issues for Green Springs are similar to Gemini Springs. The models developed for both these springs are expected to be similar in structure, i.e., the first set of models will cover recent times since groundwater measurements became available. The early models will use rainfall and perhaps discharge at Gemini Springs as supporting explanatory variables. However, as the report will detail later, information from another spring (Blue) could also be added to the explanatory variables set to help provide more accurate estimates for daily discharge at Green Springs.

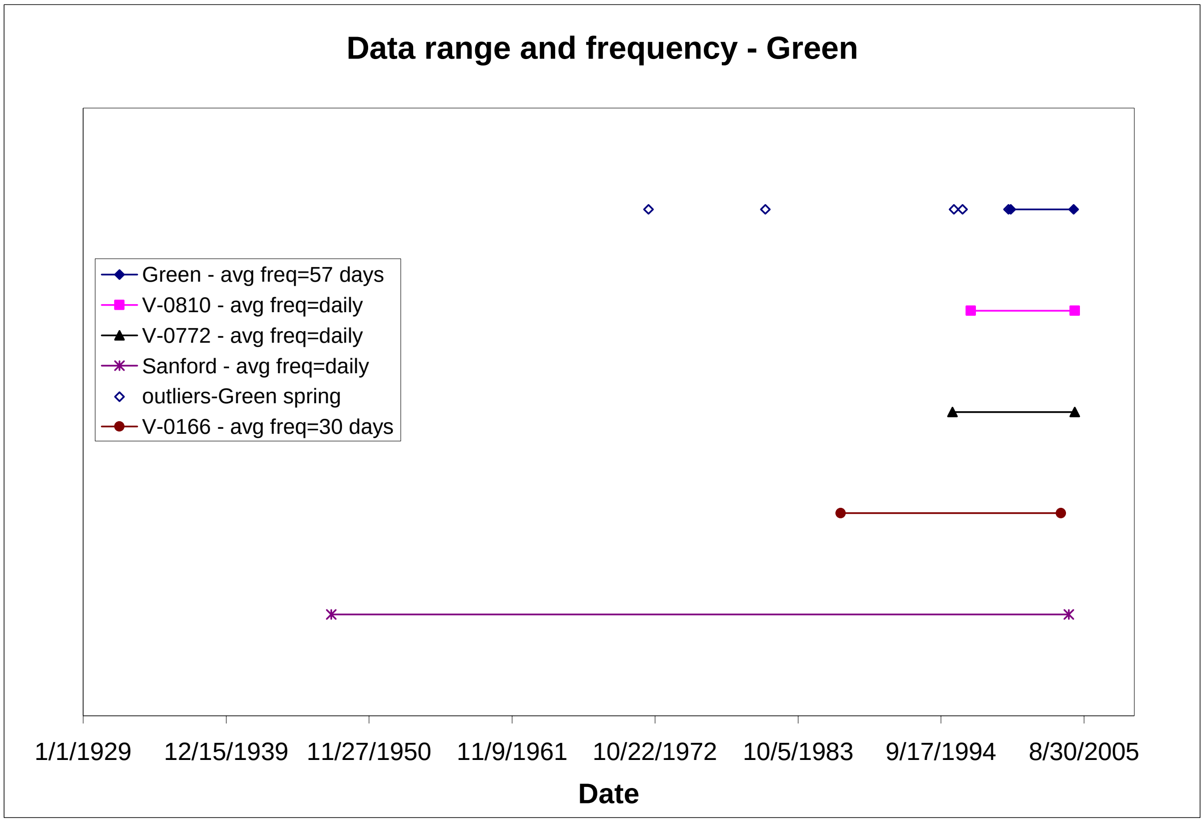


Figure 4 Overlap between various data types, Green Springs.

Figure 5 shows the overlap between various data types for Blue Spring. Shown here are the periods of record for: (a) springs discharge, (b) groundwater levels at monitoring wells

V-0083, V-1091, V-0156, L-0059, V-0101, and M-0024, and (c) precipitation measurement at Deland 1 SSE. Also indicated therein is the average frequency of observation for each data type (as was discussed in detail in the previous section). In general, the data records for Blue Spring could be broken into pre- and post-1984. Late time models can take advantage of all available data at the majority of the monitoring wells as well as precipitation data from the Deland gauge. Early time models will only have the water level record at V-0101 and the Deland rainfall record as supporting explanatory variables.

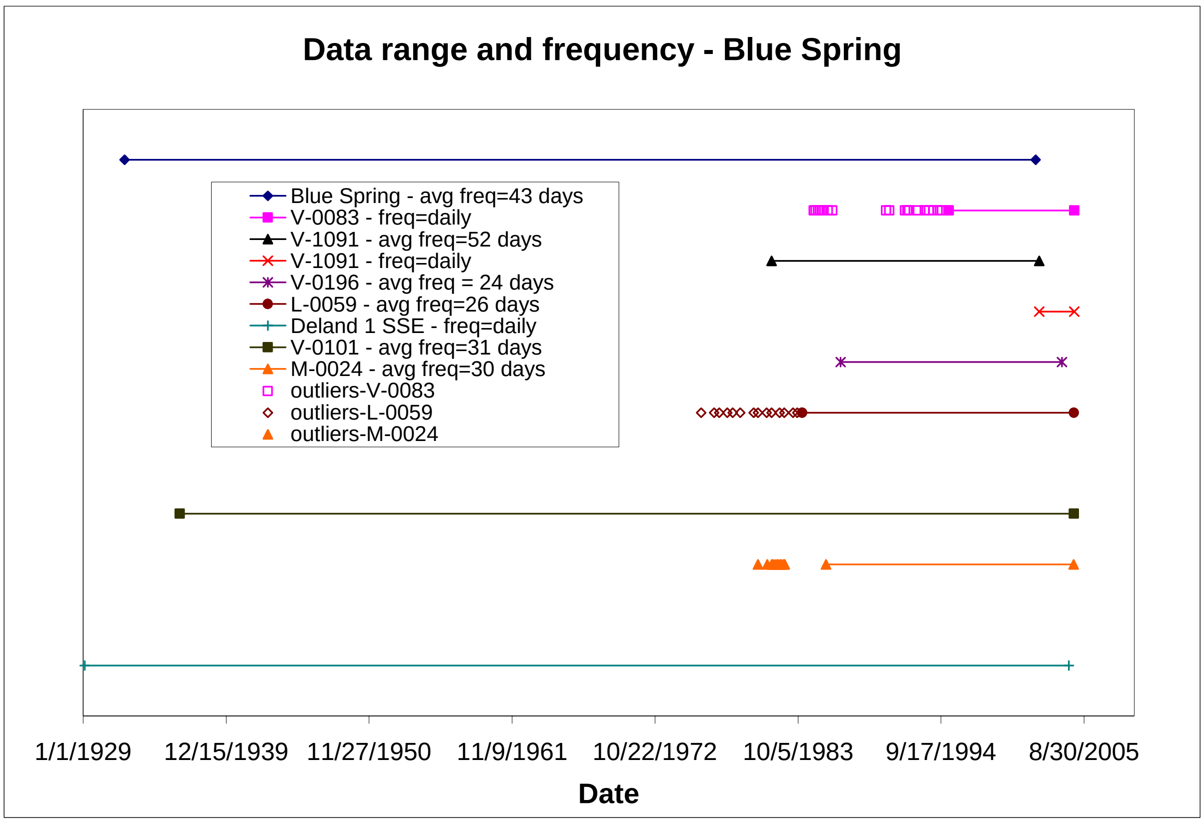


Figure 5 Overlap between various data types, Blue Spring.

3.4 Correlation Analysis – Spring to Spring

The correlation between spring discharge at a given spring, and the 6, 8, 12, 24, 48, and 52-week moving averages of spring discharge at Blue Spring is presented in this section. The motivation here is to determine if any of the moving-average spring discharge at Blue Spring can be used as a predictor variable for discharge at PDL, Gemini, or Green springs. The rationale for selecting Blue Spring as a potential "global" predictor is twofold: (a) Blue Spring has an extensive period of record dating back to March 1932, and (b) its location is between PDL and Gemini/Green springs. This implies that Blue Spring can be of potential value as an auxiliary source of data for PDL as well as for Gemini and Green springs.

Table 3 suggests that discharge at PDL is maximally correlated to the 48-week (or 52-week) moving average discharge at Blue Spring. Similarly, as shown in Table 4, the maximum correlation for Gemini Springs occurs at a lag of 12 weeks. Finally, Table 5 indicates that the maximum correlation for Green Springs occurs at a lag of 24 weeks – although the low value of the correlation coefficient and the small sample size make this result of questionable value.

Thus, on the basis of these spring-spring correlation analyses, it seems plausible that daily Blue Spring discharge may be used as one of the explanatory variables for PDL Springs and Gemini Springs.

Table 3 Correlation coefficients between discharge at PDL Springs and moving averages of discharge at Blue Spring.

	Ponce de Leon	Count
Blue Spring	0.68	13
Blue Spring- 6 week	0.40	150
Blue Spring- 8 weeks	0.43	220
Blue Spring- 12 week	0.47	244
Blue Spring- 24 week	0.49	250
Blue Spring- 48 week	0.55	252
Blue Spring-52week	0.55	253

Table 4 Correlation coefficients between discharge at Gemini Springs and moving averages of discharge at Blue Spring.

	Gemini	Count
Blue Spring	0.99	3
Blue Spring- 6 week	0.45	39
Blue Spring- 8 weeks	0.42	52
Blue Spring- 12 week	0.49	54
Blue Spring- 24 week	0.48	58
Blue Spring- 48 week	0.41	59
Blue Spring-52week	0.39	60

Table 5 Correlation coefficients between discharge at Green Springs and moving averages of discharge at Blue Spring.

	Green	Count
Blue Spring	N/A	2
Blue Spring- 6 week	-0.01	10
Blue Spring- 8 weeks	0.11	11
Blue Spring- 12 week	0.14	15
Blue Spring- 24 week	0.23	18
Blue Spring- 48 week	0.18	20
Blue Spring-52week	0.16	21

It should be pointed out that the correlation between Gemini and Green springs was also evaluated, but not considered for additional analysis because of the limited number of data points (<10) or the common days at which measurements for both springs were recorded.

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 (moving averages of water level measurements and precipitation) at the springs of interest. Such a relationship can be expressed by:

$$q_t = b_0 + b_1\, q_{t-i} + b_2\, q_{t-j} + \ldots + b_3\, h_{t-k} + b_4\, r_{t-l} + \varepsilon \qquad (1)$$

where q is spring discharge; q^* is discharge at an adjacent spring, h is groundwater level; r is precipitation; ε is an error term; b_0, b_1, b_2, b_3 and b_4 are regression coefficients; t is time, and i, j, k and l denote lags that maximize the correlation between the response and predictor variable pair of interest. Here, the use of surrogate predictors is necessitated by the fact that most predictor variables are not measured on a daily basis. Generation of daily discharge thus requires the use of predictor variables for which daily values can be generated, e.g., on the basis of averaging over some moving time window.

Eq. (1) can be symbolically re-stated as follows, where MA denotes moving average:

$$\begin{aligned}\text{[Spring discharge]} = \boldsymbol{f}\,\{\, &\text{[same spring MA]} + \text{[water level MA]} \\ &+ \text{[precipitation MA]} + \text{[adjacent spring MA]}\,\}\end{aligned} \qquad (2)$$

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 the recent period since mid-1990s, when detailed measurements of groundwater levels are available. On the other hand, for springs such as Gemini and Green, discharge measurements are not available prior to this time. Thus, early-time regression models for these springs will have to rely only on rainfall, discharge at adjacent springs and, when possible, water levels from long-term monitoring wells.

The model building process can be carried out using forward stepwise regression, where variables are added 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. This is the variable with the largest absolute value of the partial correlation coefficient (*PCC*), which measures the correlation between the output and the selected input variable after the linear influence of the other variables have been eliminated.

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 - α) % confidence level. 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.

If necessary, piecewise regression or non-parametric regression (e.g., Alternating Conditional Expectation or ACE) can be used as an alternative to stepwise regression to improve the linear model goodness of fit.

- In piecewise regression, the algorithm automatically splits the data into two or more subsets such that model predictions have the highest possible correlation with observed values of the response variable (daily spring discharge).
- In ACE, the algorithm automatically selects optimal non-parametric transformations for each of the variables such that the transformed response variable can be expressed as the sum of all the transformed explanatory variables and the input-output correlation coefficient is maximized.

Note that the number of potential explanatory variables can be quite high, given that moving averages from multiple lags are considered for each of the terms in Eq. (1). It is therefore necessary to ensure that the regression model includes only those predictor variables that have the highest correlation with the response variable, while taking into account any predictor-predictor correlations. However, the selection of the most relevant predictors is carried out automatically as part of the stepwise regression process – thus, eliminating this onerous pre-processing step. On the other hand, both piecewise regression and ACE require the variables to be included in the model be specified a priori. A careful examination of correlation and partial correlation coefficients is warranted in such cases to assist in the parsimonious selection of predictor variables and to avoid over-parameterization of the model. An alternative would be to use a data reduction technique such as principal component analysis (PCA) to combine the predictors into surrogate variables and apply principal component regression. However, exploratory analysis with such an approach using data from Blue Spring did not yield regression models superior to those generated using stepwise regression.

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

- Split the period of record into a late-time period, where detailed groundwater level measurements are available, and an early time period where only limited or no groundwater level measurements are available.
- For each period, organize the spring discharge data (response variable) and the corresponding 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 forward 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 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 PDL Springs

One modeling period can be identified for PDL:

- 1965-1997 period, when groundwater level measurements are available from L-0045, V-095, and V-0096; along with discharge data from Blue Spring.

Stepwise regression analyses were performed for the above mentioned modeling period and the results are presented below. The stepwise regression analysis for the PDL data produced the following model:

$$\begin{aligned} \text{PDL} = {} & \text{PDL..12.week} + \text{PDL..48.week} + \text{PDL..52.week} + \text{Blue-Spring..52.week} \\ & + \text{L.0045..8.week} + \text{L.0045..48.week} + \text{L.0045..52.week} \end{aligned} \tag{3}$$

The multiple R^2 for this regression model was 0.57. The standard error of estimate was 2.89. The F-statistic was 29.605, and the p-value was 0. Estimated regression coefficients and their statistics are given below in Table 6.

In Table 6, the "B" column contains the regression coefficients in actual units. The "beta" column denotes the standardized regression coefficients (*SRC*) that would have resulted if the predictor variables had been normalized to zero mean and unit standard deviation. The absolute value of the *SRC*s can be used as an indicator of variable importance (Draper and Smith, 1981). Thus, the most important predictor variables can be identified as [L-0045 48-week], [L-0045 52-week] and [PDL 48-week].

Table 6 PDL– regression coefficient statistics.

Regression Summary for Dependent Variable: Ponce de Leon (PDL_Data_until_1997_Regression.sta)
R= .77142811 R²= .59510133 Adjusted R²= .57499998
F(7,141)=29.605 p<0.0000 Std.Error of estimate: 2.8923

N = 149	Beta	Std.Err.	B	Std.Err.	t(135)	p-level
Intercept			-1.67405	3.717885	-0.45027	0.653207
Ponce De Leon- 52 week	-0.43221	0.519382	-0.55553	0.667581	-0.83216	0.406727
L-0045-8week	0.53170	0.096945	1.93914	0.353564	5.48456	0.000000
L-0045-48week	-1.20757	0.703044	-5.14812	2.997218	-1.71763	0.088059
Ponce De Leon- 12 week	0.17028	0.090858	0.17495	0.093350	1.87414	0.062980
Ponce De Leon- 48 week	0.87540	0.535413	1.12340	0.687096	1.63499	0.104281
Blue Spring- 52 week	0.10275	0.085744	0.03471	0.028963	1.19828	0.232819
L-0045-52week	0.80951	0.695952	3.45109	2.966979	1.16317	0.246726

Figure 6 shows a comparison between the observed and fitted values of PDL Springs discharge. The scatter in the data is consistent with a final R^2 of 0.57. Note also the resulting under prediction of some high discharge values and over prediction of some low discharge values (i.e., the outliers in Figure 6).

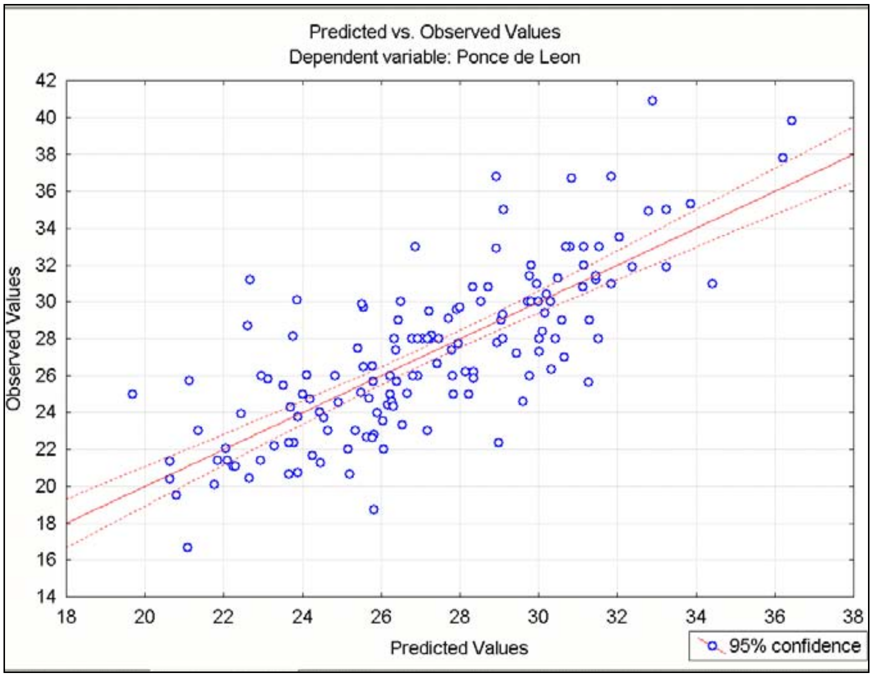


Figure 6 PDL – comparison of observed and predicted values.

Figure 7 shows a normal probability plot of the residuals for the PDL regression. The linearity of the data suggests that standard assumptions for normally distributed errors in a multivariate linear regression model have been satisfied and the model is properly parameterized.

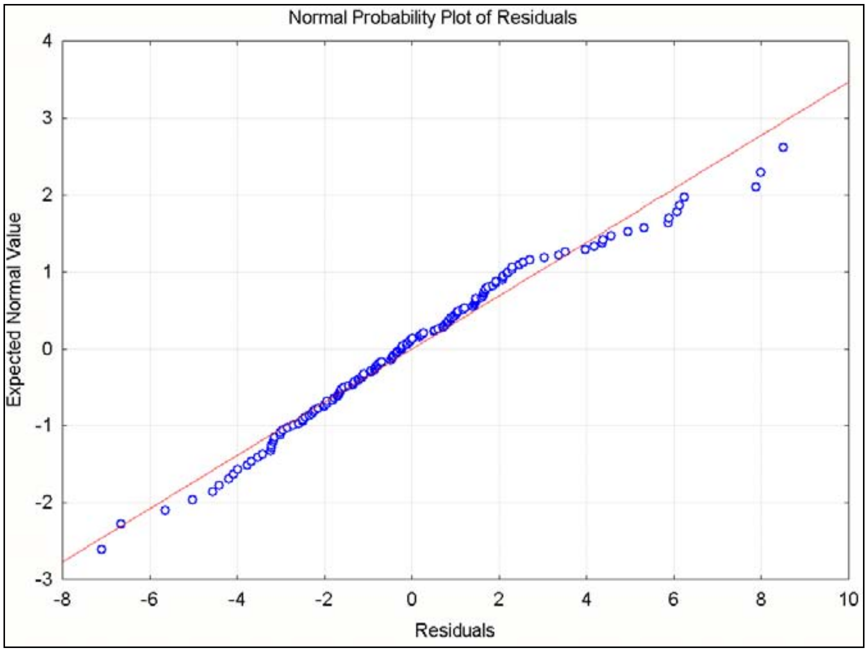


Figure 7 PDL – normal probability plot of residuals.

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 7 shows the F-test and K-S test between observed PDL Springs time-series and predicted PDL Springs time-series on days corresponding to observed data. Results for the F-test indicate that there is a significant difference between the two variances. However, the K-S D statistic does not show a significant difference between the two empirical CDFs.

Table 7 F-test and K-S test between observed and predicted PDL timeseries.

	PDL (observed)	*PDL (predicted)*
Mean	26.64	26.67
Variance	24.44	13.47
Observations	247	247
df	246	246
F	1.81	
P(F<=f) one-tail	0.00	
F Critical one-tail	1.23	
K-S D statistic	0.08	
p-value for K-S test	0.35	

Figure 8 shows the box-whisker plots for three data sets:

(1) observed discharge values at PDL Springs for the time period 1965-1997,

(2) regression-predicted values for the same dates at which observed discharge value are available, and

(3) regression-predicted values from the regression model for each day in the time period 1965-1997.

The plots show that the observed discharge values at PDL 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 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.

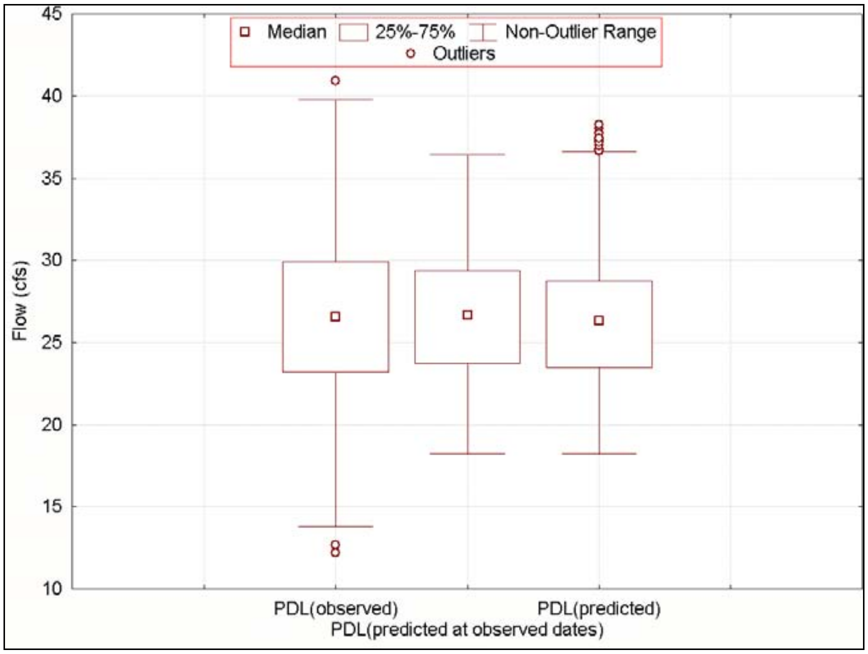


Figure 8 PDL – Box and Whisker plot.

4.3 Regression Models for Gemini Springs

The two modeling cases for Gemini Springs are as follows:

- Gemini 1995-2000, where a relationship is sought between Gemini Springs discharge, Blue Spring discharge, and water levels at S-0257 in order to predict the daily Gemini Springs discharge prior to 1995 (when no measurements are available at the spring of interest itself, and

- Gemini 1995-2004, where a relationship is sought between Gemini Springs discharge, Blue Spring discharge, rainfall at Sanford and water levels at S-0257 and S-1230 in order to predict the daily Green Springs discharge during this period (1995-2004).

Stepwise regression analyses were performed for both of these modeling periods and the results are presented below.

The stepwise regression analysis of the 1995-2000 Gemini Springs discharge data produced the following model:

$$\text{Gemini} = \text{Blue.Spring..12.week} + \text{Blue.Spring..48.week} + \text{Blue.Spring..52.week} + \text{S.0257..12.week} + \text{S.0257..24.week} \quad (4)$$

The multiple R^2 for this regression model was 0.47. The residual standard error was 0.87. The F-statistic was 6.9 and the p-value was 0. Estimated regression coefficients and their statistics are given below in Table 8. The most important variables in the regression model, identified on the basis of the absolute value of the *SRC*, are [Blue Spring 52-week], [Blue Spring 48-week], [S-0257 24-week] and [S-0257 12-week].

An alternative model that included moving averages of rainfall recorded at Stanford produced a better fit for this data set. However, its application for daily discharge predictions produced too much noise – indicating that the fluctuations in the rainfall time series were not being damped by other explanatory variables. It was therefore decided to exclude Sanford as an explanatory variable in order to retain a reasonable degree of variability in the daily springflow predictions.

Table 8 Gemini – 1995-2000 – regression coefficient statistics.

Regression Summary for Dependent Variable: Gemini (Rerun2_Gemini_woutGemlangs_1995-2000.sta)
R= .74189587 R²= .55040949 Adjusted R²= .47012547
F(5,28)=6.8558 p<.00027 Std.Error of estimate: .86705

N = 34	Beta	Std.Err.	B	Std.Err.	t(28)	p-level
Intercept			6.313614	7.493850	0.84251	0.406644
S-0257-24 week	0.96838	0.403572	1.433453	0.597390	2.39953	0.023314
Blue Spring-52week	-1.38467	0.596370	-0.321406	0.138428	-2.32182	0.027739
Blue Spring-12week	0.12028	0.213227	0.012948	0.022954	0.56410	0.577174
Blue Spring-48week	1.06485	0.611768	0.246067	0.141369	1.74060	0.092734
S-0257-12 week	-0.60016	0.350785	-0.809407	0.473087	-1.71090	0.098158

Figure 9 shows a comparison between the observed and fitted values of the 1995-2000 Gemini Springs discharge indicating agreement that is consistent with the moderate R^2 value of 0.47.

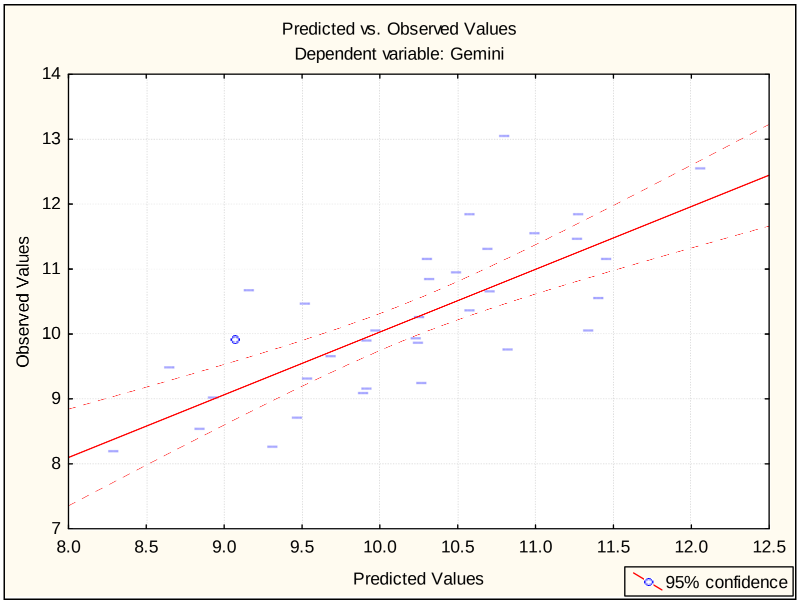


Figure 9 Gemini – 1995-2000 – comparison of observed and predicted values.

Figure 10 shows a normal probability plot of the residuals for the 1995-2000 period of Gemini Springs discharge, with minor deviations from linearity at low values of residuals.

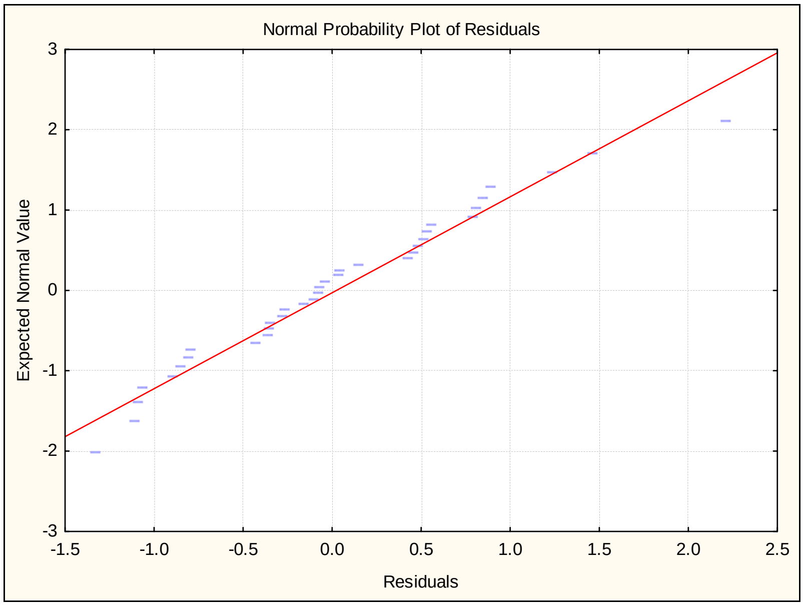


Figure 10 Gemini – 1995-2000 – normal probability plot of residuals.

A stepwise regression analysis of the 1995-2004 Gemini Springs discharge data produced the following model:

$$\text{Gemini} = \text{S.1230..4.week} + \text{S.1230..6.week} + \text{S.1230..8.week} + \text{S.1230..24.week} + \text{Sanford} + \text{Sanford..4.week} + \text{Sanford..8.week} + \text{Sanford..12.week} + \text{Sanford..52.week} \quad (5)$$

The multiple R^2 for this regression model was 0.79. The standard error of estimate was 0.77. The F-statistic was 16.58, and the p-value was 0. Estimated regression coefficients and their statistics are given below in Table 9. The most important variables in the regression model, identified on the basis of the absolute value of the *SRC*, are [S-1230 4-week], [S-1230 8-week] and [S-1230 24-week].

Table 9 Gemini – 1995-2004 – regression coefficient statistics.

Regression Summary for Dependent Variable: Gemini (Rerun_Gemini_PCC_STAT_Input_1995-2004)
R= .91500908 R²= .83724162 Adjusted R²= .78673040
F(9,29)=16.575 p<.00000 Std.Error of estimate: .60752

N = 39	Beta	Std.Err.	B	Std.Err.	t(29)	p-level
Intercept			-2.67901	3.170252	-0.84505	0.405003
S-1230-24 week	1.14099	0.221150	1.30807	0.253533	5.15936	0.000016
Sanford- 12 week	0.69916	0.176118	11.43520	2.880518	3.96984	0.000434
S-1230-8 week	-1.58770	0.523912	-1.51468	0.499817	-3.03047	0.005096
Sanford- 8 week	-0.34810	0.190757	-4.94203	2.708199	-1.82484	0.078343
S-1230-4 week	1.65641	0.665143	1.52284	0.611504	2.49031	0.018740
Sanford	-0.16438	0.078362	-0.55164	0.262974	-2.09772	0.044758
Sanford-52 week	0.33889	0.179265	12.16646	6.435694	1.89047	0.068726
Sanford- 4 week	-0.23539	0.128170	-2.68144	1.460054	-1.83653	0.076549
S-1230-6 week	-0.86939	0.756274	-0.80867	0.703455	-1.14957	0.259713

Figure 11 shows a comparison between the observed and fitted values of the 1995-2004 Gemini Springs discharge indicating good agreement.

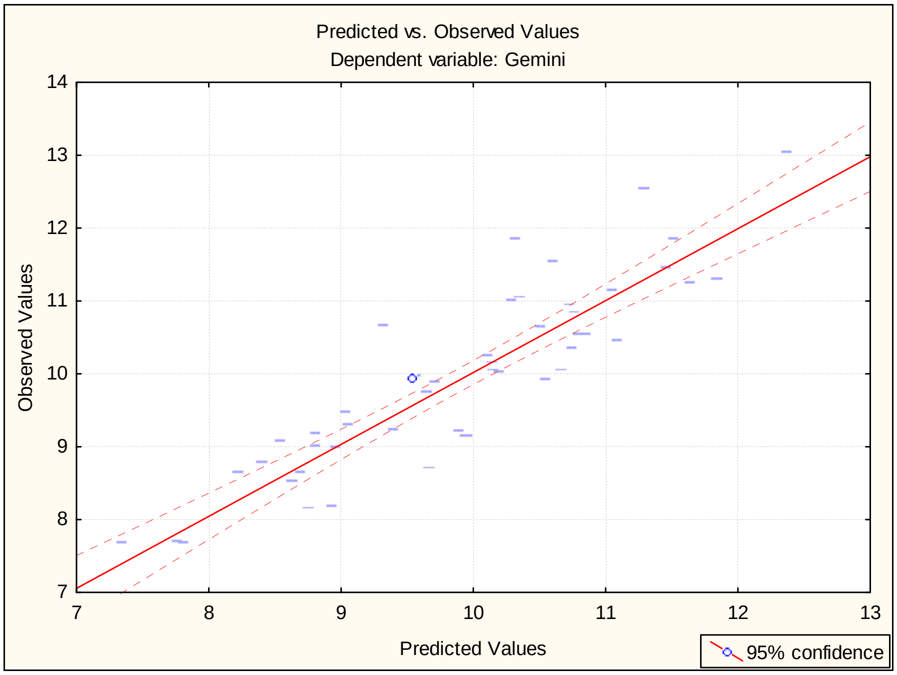


Figure 11 Gemini – 1995-2004 – comparison of observed and predicted values.

Figure 12 shows a normal probability plot of the residuals for the 1995-2000 period of Gemini Springs discharge, with minor deviations from linearity at low and high values of residuals.

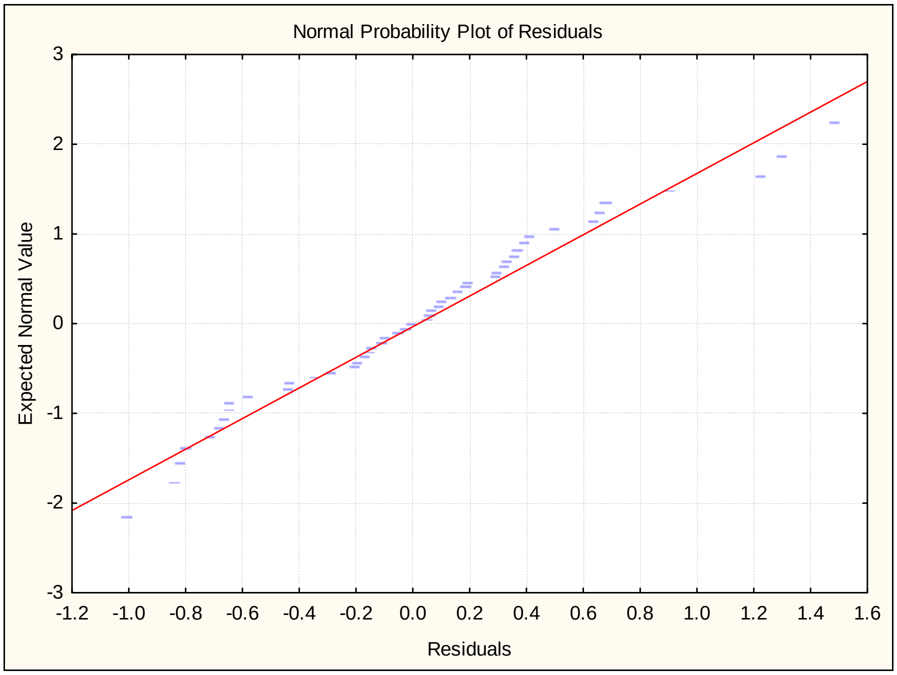


Figure 12 Gemini – 1995-2004 – normal probability plot of residuals.

To compare observed versus predicted discharges, the same methods described before for PDL Springs are used for Gemini 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 no statistically significant difference between the two variances; with values of 1.88 and 1.50 for the observed and regression-predicted values, respectively. Similarly, the K-S D statistic shows no significant difference between the two empirical CDFs.

As mentioned before for PDL 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 13 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 one low-value outlier in the observed time series. The non-outlier range is almost identical for the two time series, with a slight difference at the upper end. 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 higher than 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 Gemini Springs - 1995-2004 Observed and Regression-Predicted Variance Statistics

	Gemini (observed)	*Gemini (predicted)*
Mean	9.93	10.05
Variance	1.88	1.50
Observations	68	68
df	67	67
F	1.25	
P(F<=f) one-tail	0.18	
F Critical one-tail	1.50	
K-S D statistic	0.09	
p-value for K-S test	0.96	

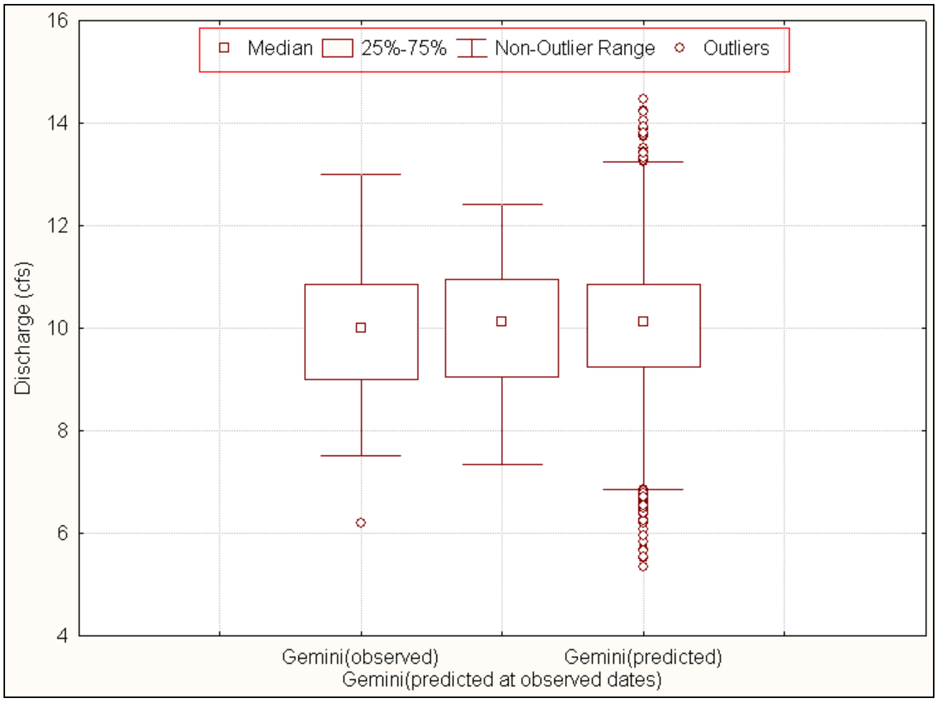


Figure 13 Gemini – Box and Whisker plot.

4.4 Regression Models for Green Springs

The two modeling cases for Green Springs are as follows,

- Green pre-1996, where a relationship is sought between Green Springs discharge, water levels at V-0166 and rainfall at Sanford during the 2000-2004 period in order to predict discharge at Green Springs in the 1987-1996 period, and
- Green 1999-2004, where a relationship is sought between Green Springs discharge, Blue Spring discharge, rainfall at Sanford and water levels at V-0810, V-0772, and V-0166 in order to predict daily Green Springs discharge during this period and also for 1996-1999.

Stepwise regression analyses were performed for both of these modeling periods and the results are presented below. The stepwise regression analysis of the first dataset (referred to as Green Springs pre-1996) produced the following model:

$$\text{Green} = \text{V.0166.24.week} + \text{V.0166..48.week} + \text{V.0166..52} + \text{Sanford..1.week} + \\ + \text{Sanford..3.week} + \text{Sanford..48.week} \quad (6)$$

The multiple R^2 for this regression model was 0.89. The standard error of estimate was 0.28. The F-statistic was 35.1, and the p-value was 0. Estimated regression coefficients and their statistics are given below in Table 11. The most important variables were [V-0166 52-week] and [V-0166 48-week].

Table 11 Green - pre-1996 – regression coefficient statistics.

Regression Summary for Dependent Variable: Green (Green-pre2000_V0166) R= .95981092 R²= .92123700 Adjusted R²= .89498266 F(7,21)=35.089 p<.00000 Std.Error of estimate: .28398						
N = 29	Beta	Std.Err.	B	Std.Err.	t(21)	p-level
Intercept			-10.3182	1.912913	-5.39399	0.000024
Sanford_48week	0.371090	0.155638	9.9353	4.166933	2.38432	0.026621
V-0166_24week	0.450713	0.111107	0.5010	0.123506	4.05658	0.000568
Sanford_12week	0.372201	0.123698	3.8407	1.276435	3.00895	0.006684
Sanford-3week	-0.345206	0.114099	-2.0444	0.675725	-3.02550	0.006435
V-0166_52week	1.017014	0.415291	1.3844	0.565316	2.44892	0.023193
V-0166_48week	-0.926160	0.435435	-1.2006	0.564475	-2.12698	0.045439
Sanford-1week	0.193179	0.094735	1.0750	0.527168	2.03915	0.054219

Figure 14 shows a comparison between the observed and fitted values for the Green Springs pre-1996 data set indicating good agreement. Figure 15 shows a normal probability plot of the residuals for the Green pre-1996 model, indicating proper linear-type diagnostics.

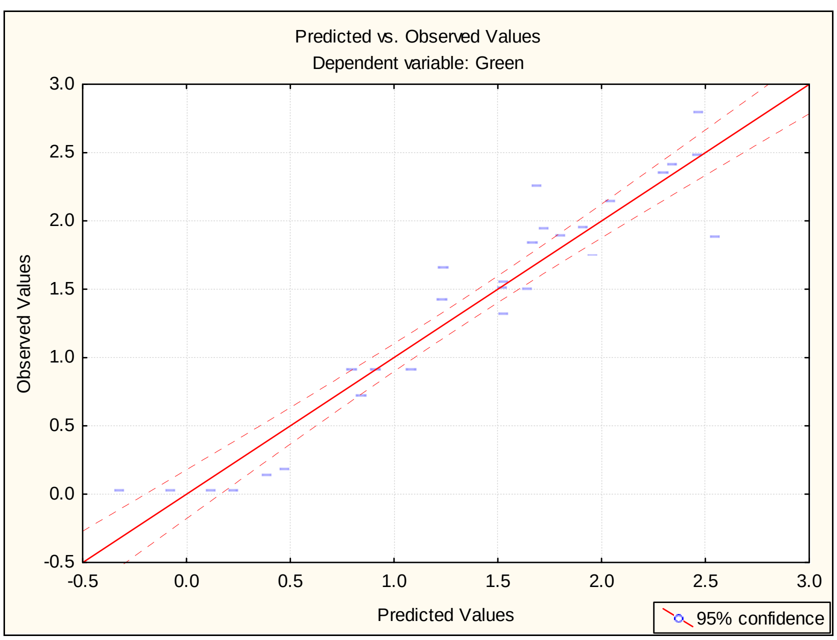


Figure 14 Green – pre-1996 – comparison of observed and predicted values.

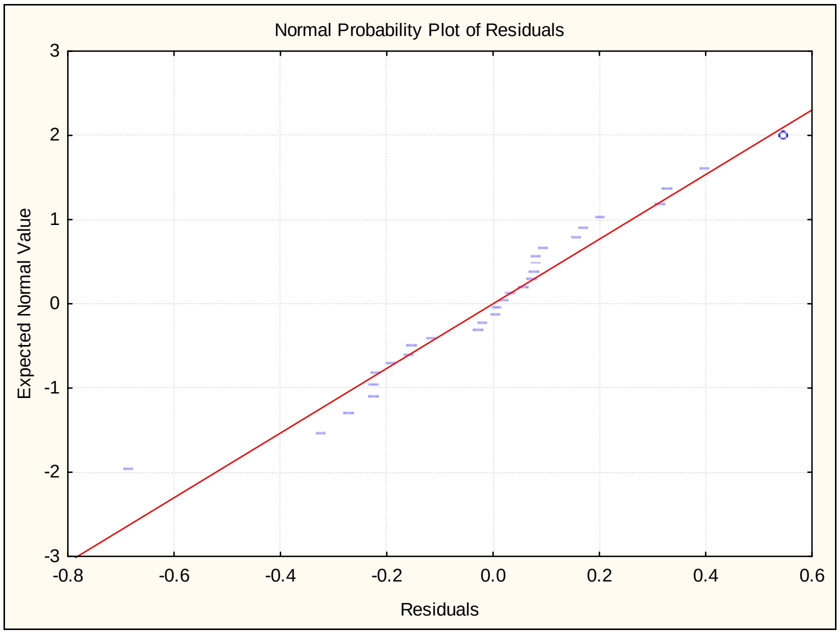


Figure 15 Green – pre-1996 – normal probability plot of residuals.

The stepwise regression analysis of the second dataset (referred to as Green Springs 1999-2004) produced the following model:

$$\text{Green} = \text{V.0810} + \text{V.0772} + \text{Sanford..4.week} + \text{Sanford..8.week} + \text{Sanford..48.week} \tag{7}$$

The multiple R^2 for this regression model was 0.97. The standard error of estimate was 0.14. The F-statistic was 189.3, and the p-value was 0. Estimated regression coefficients and their statistics are given below in Table 12. The most important variable was [V-0810].

Table 12 Green – 1999-2004 – regression coefficient statistics.

Regression Summary for Dependent Variable: Green (Green-post2000) R= .98858006 R²= .97729053 Adjusted R²= .97212929 F(5,22)=189.35 p<.00000 Std.Error of estimate: .14838						
N = 28	Beta	Std.Err.	B	Std.Err.	t(22)	p-level
Intercept			-3.79682	0.435729	-8.71371	0.000000
V-0772	-0.056706	0.324536	-0.02285	0.130799	-0.17473	0.862889
Sanford-8week	0.378254	0.072825	3.67987	0.708480	5.19404	0.000033
Sanford-4week	-0.384526	0.074412	-2.25024	0.435461	-5.16749	0.000035
Sanford-48week	0.271348	0.093230	7.73961	2.659183	2.91052	0.008107
V-0810	0.702443	0.273103	0.28869	0.112240	2.57208	0.017386

Figure 16 shows a comparison between the observed and fitted values for the Green Springs 1999-2004 data set indicating good agreement.

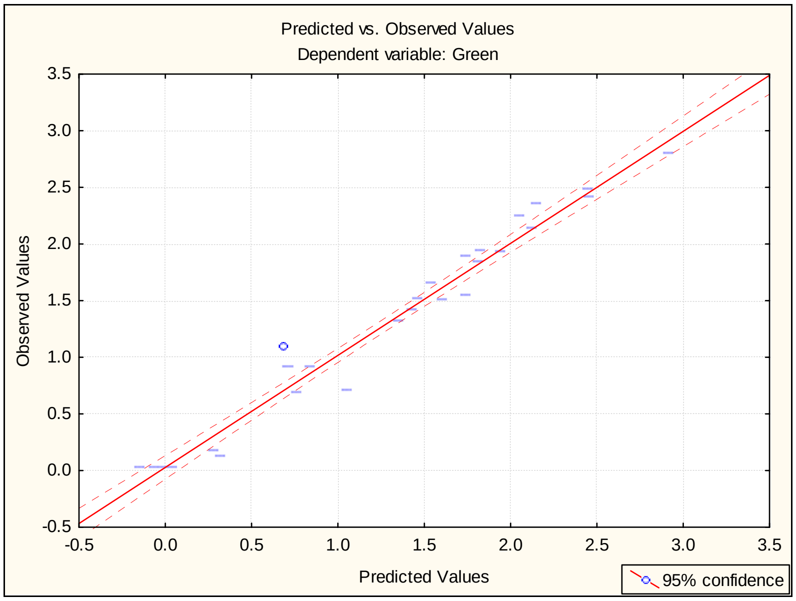


Figure 16 Green – 1999-2004 – comparison of observed and predicted values.

Figure 17 shows a normal probability plot of the residuals for the Green 1999-2004 model, indicating proper linear type diagnostics excepting in the very high residual range.

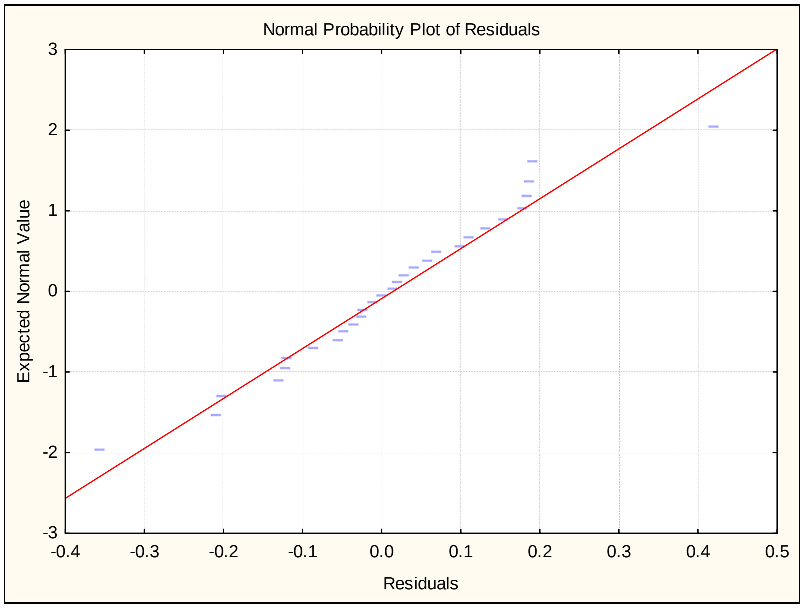


Figure 17 Green – 1999-2004 – normal probability plot of residuals.

To compare observed versus predicted discharges, the same methods described before for PDL Springs are used for Green Springs. Results for the F-test and K-S D statistic are shown in Table 13. Results for the F-test indicate that there is no statistically significant difference between the two variances; with values of 0.73 and 0.75 for the observed and regression-predicted values, respectively. Similarly, the K-S D statistic shows no significant difference between the two empirical CDFs.

As mentioned before for PDL 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 18 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 there are slight differences between the observed and predicted values at the upper end. The lower end is naturally bounded by the zero discharge observations and the physical constraint used to ensure that regression-generated discharge values are not negative. The non-outlier range is almost identical for the two time series. 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 higher than 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 13 Green – 1999-2004 Observed and Regression-Predicted Variance Statistics.

	Green (observed)	*Green (predicted)*
Mean	1.28	1.30
Variance	0.73	0.75
Observations	31	31
Df	30	30
F	0.97	
P(F<=f) one-tail	0.47	
F Critical one-tail	0.54	
K-S D statistic	0.10	
p-value for K-S test	1.00	

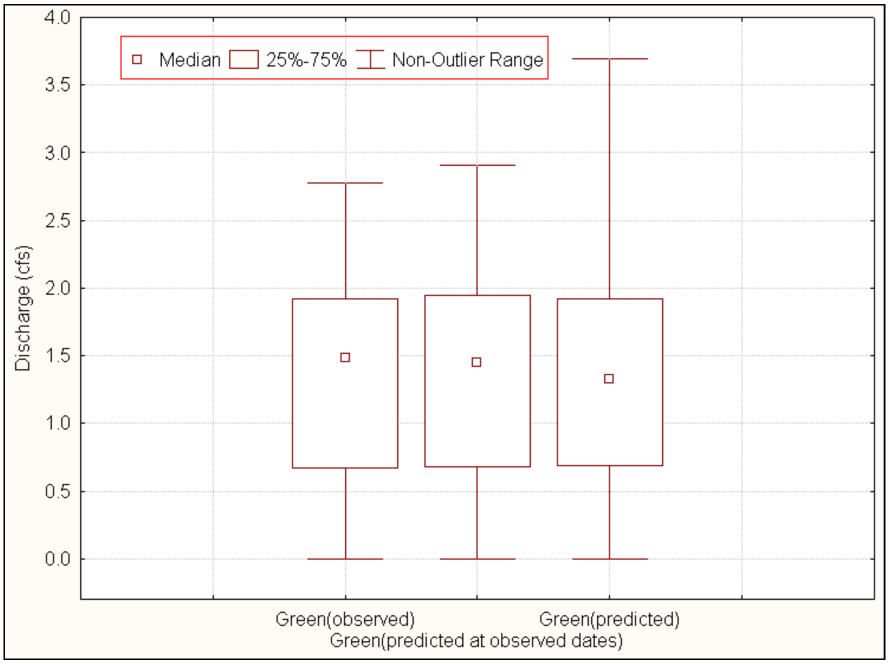


Figure 18 Green – Box and Whisker plot.

5.0 PREDICTION OF DAILY DISCHARGE AND FLOW DURATION

5.1 Daily Discharge Predictions and Flow Duration Curves for PDL

Predictions of daily discharge and flow duration curves for PDL are carried out with the help of Eq. (3). Figure 19 shows these daily predictions juxtaposed with actual measurements of PDL discharge (at an average frequency of 55 days). The agreement between both the two time series is quite good and the absence of any significant divergent trends indicates that the linear model is able to capture the general trend of the spring discharge.

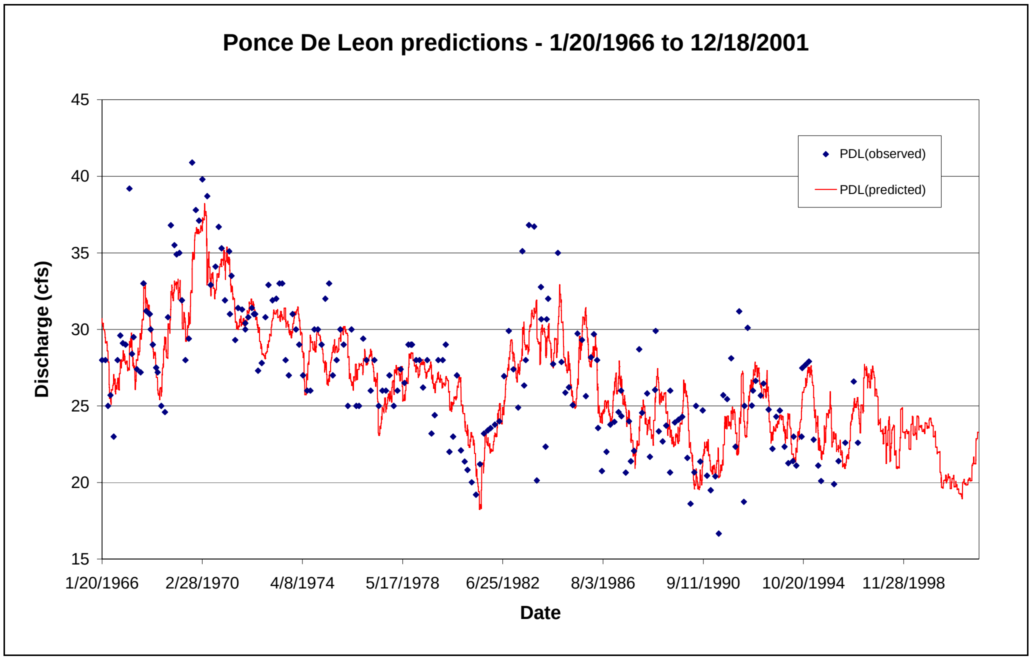


Figure 19 Daily discharge predictions for PDL, 1966-2001.

Figure 20 shows the flow duration curve, i.e., discharge versus percent exceedance for the long-term simulation generated from the results of the statistical modeling, indicating good agreement between the statistical characteristics of the observed and predicted spring discharge.

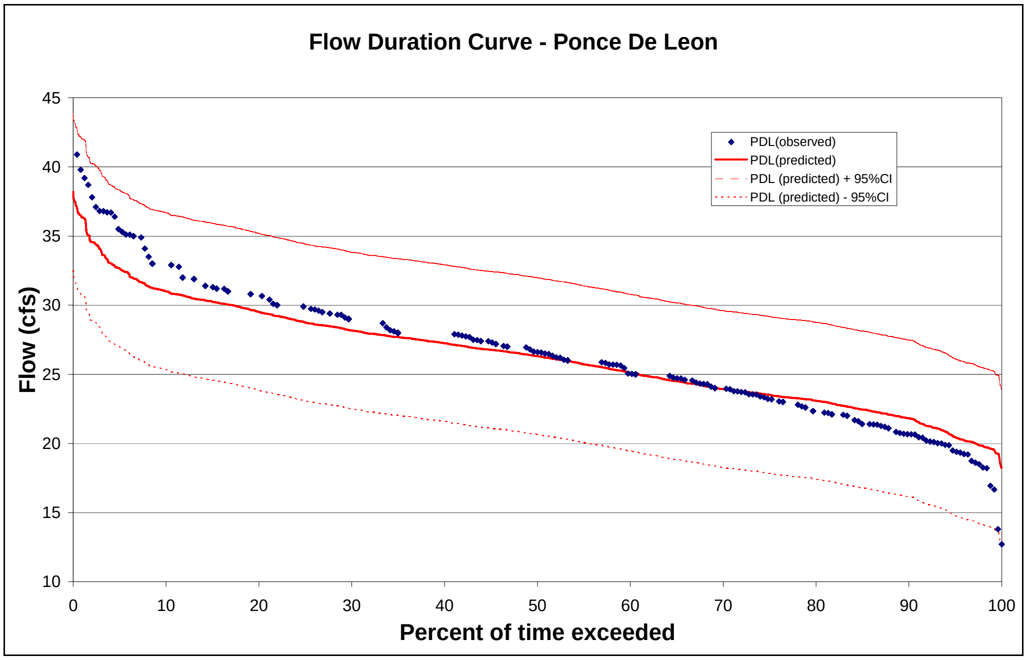
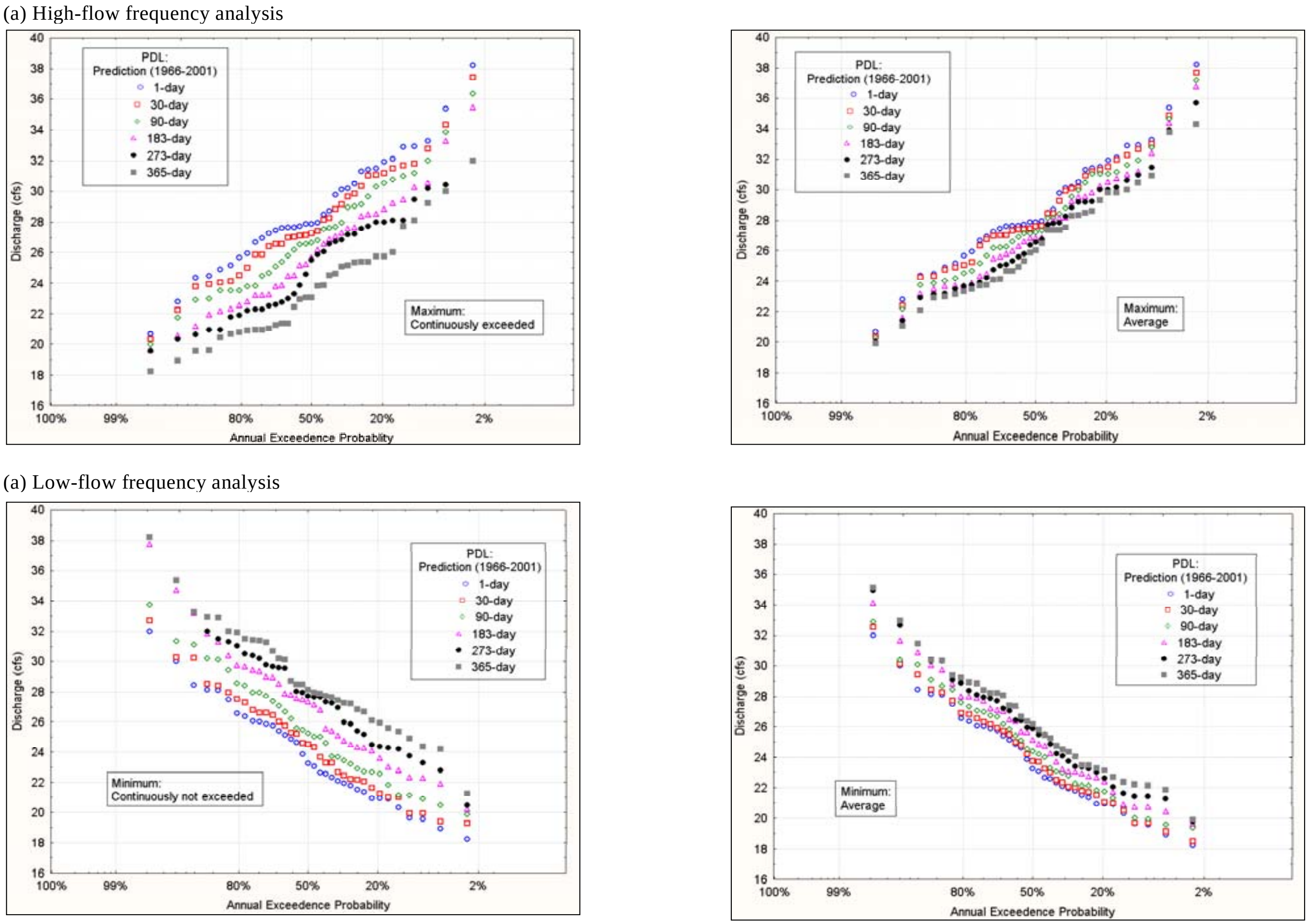


Figure 20 Flow duration curve for PDL Springs.

The corresponding high- and low-flow frequency analyses for the system (frequency of spring discharge for durations of 1 month, 2 months, 3 months, 4 months, 6 months, and 1 year) are shown in Figure 21.

5.2 Daily Discharge Predictions and Flow Duration Curves for Gemini Springs

Predictions of daily discharge and flow duration curves for Gemini Springs are carried out with the help of Eq. (4) for the pre-1995 period and Eq. (5) for the post-1995 period. Figures 22 and 23 show these daily predictions juxtaposed with actual measurements of Gemini Springs discharge (at an average frequency of 56 days). The agreement between both the two time series is quite good and the absence of any significant divergent trends indicates that the linear model is able to capture the general trend of the spring discharge.



Date: July 5, 2007

File: Fig 21.pdf

High- and low-frequency analysis of discharge for PDL Springs

INTERA

St. Johns River Water Management District
Palatka, Florida

Figure 21

Figure 21 High- and low-frequency analysis of discharge for PDL Springs.

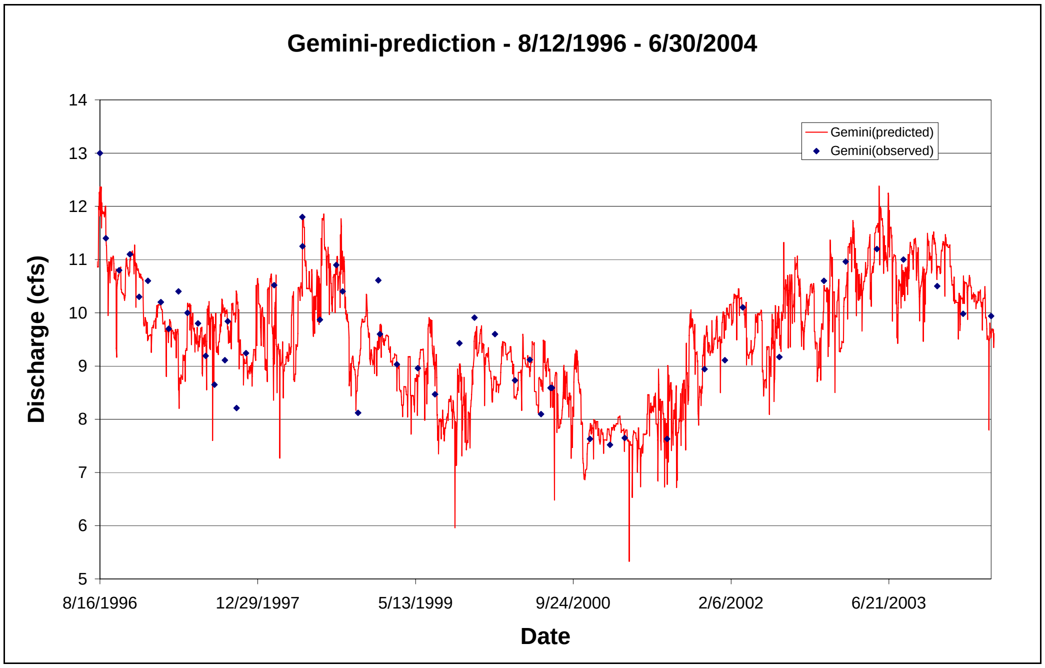


Figure 22 Daily discharge predictions, Gemini Springs, 1996-2004.

The sparsity of actual observations of Gemini Springs discharge during the 1953-1996 period preclude a meaningful evaluation of the reliability of the daily predictions shown in Figure 23, generated using Eq. (5). However, it should be noted that the few measurements that are available are generally consistent with the predictions – excepting for the outliers in the 1993 time frame.

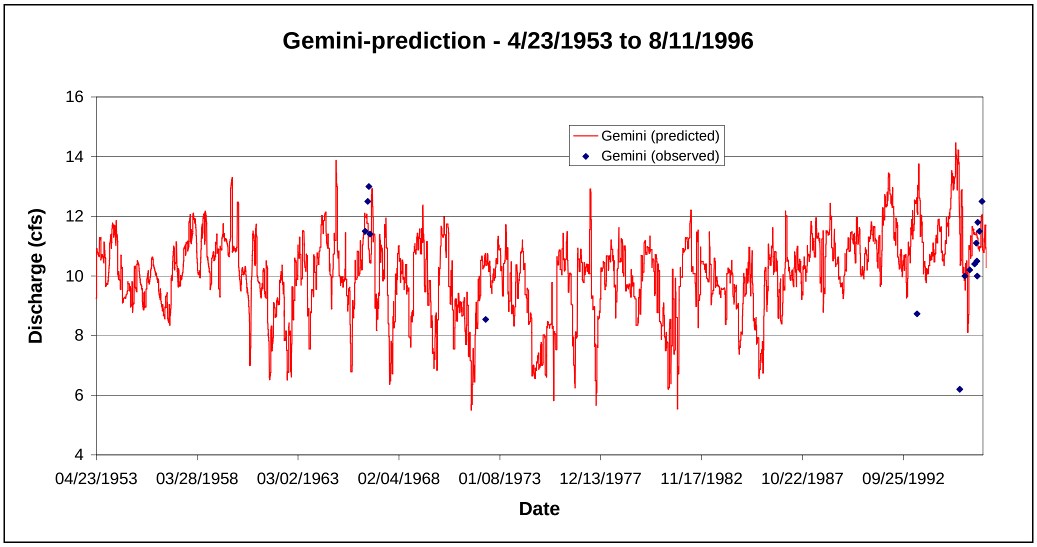


Figure 23 Daily discharge predictions, Gemini Springs, 1953-1996.

Figure 24 shows the flow duration curve for Gemini Springs, i.e., discharge versus percent exceedance for the long-term simulation generated from the results of the statistical modeling, indicating good agreement between the statistical characteristics of the observed and predicted spring discharge.

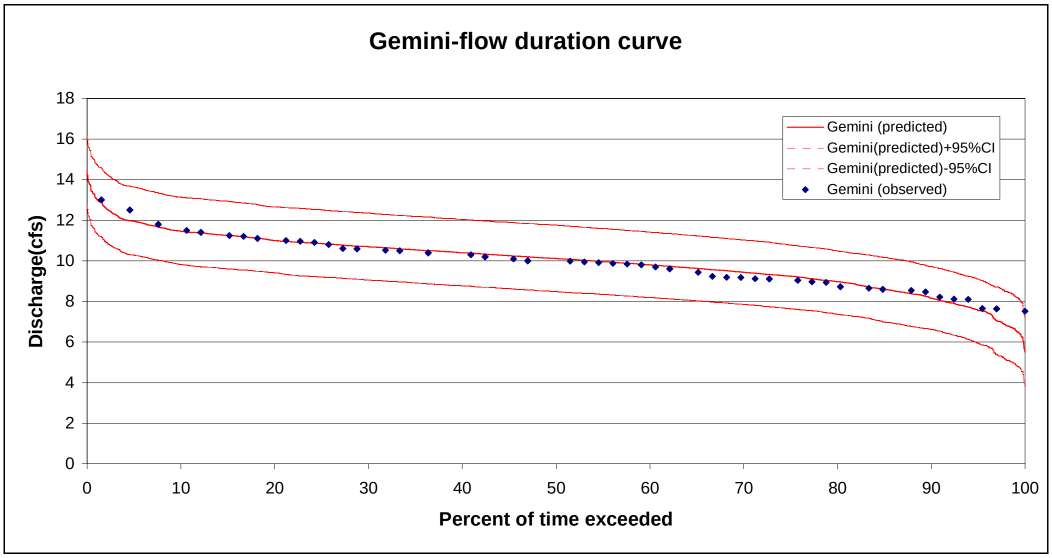
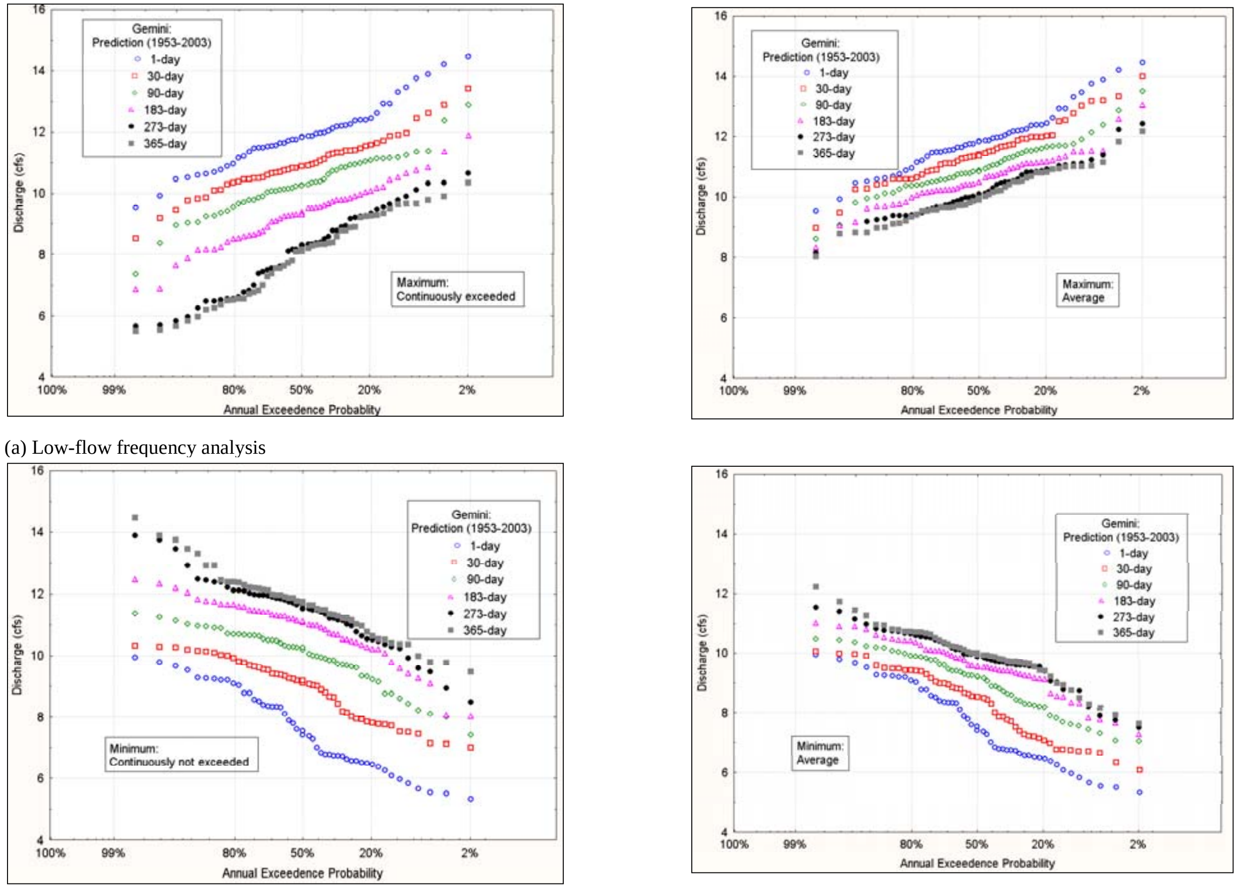


Figure 24 Flow duration curve for Gemini Springs.

The corresponding high- and low-flow frequency analyses for the system (frequency of spring discharge for durations of 1 month, 2 months, 3 months, 4 months, 6 months and I year) are shown in Figure 25.

(a) High-flow frequency analysis

(a) Low-flow frequency analysis



Date: July 5, 2007	High- and low-frequency analysis of discharge for Gemini Springs	
File: Fig 25.pdf		
INTERA	St. Johns River Water Management District Palatka, Florida	Figure 25

Figure 25 High- and low-frequency analysis of discharge for Gemini Springs.

5.3 Daily Discharge Predictions and Flow Duration Curves for Green Springs

Predictions of daily discharge and flow duration curves for Green Springs discharge are carried out with the help of Eq. (6) for the pre-1996 period and Eq. (7) for the post-1996 period. Figures 26 and 27 show these daily predictions juxtaposed with actual measurements of PDL discharge (at an average frequency of 57 days). The agreement between both the two time series for the post-1996 period (Figure 26) is quite good and the absence of any significant divergent trends indicates that the linear model is able to capture the general trend of the spring discharge.

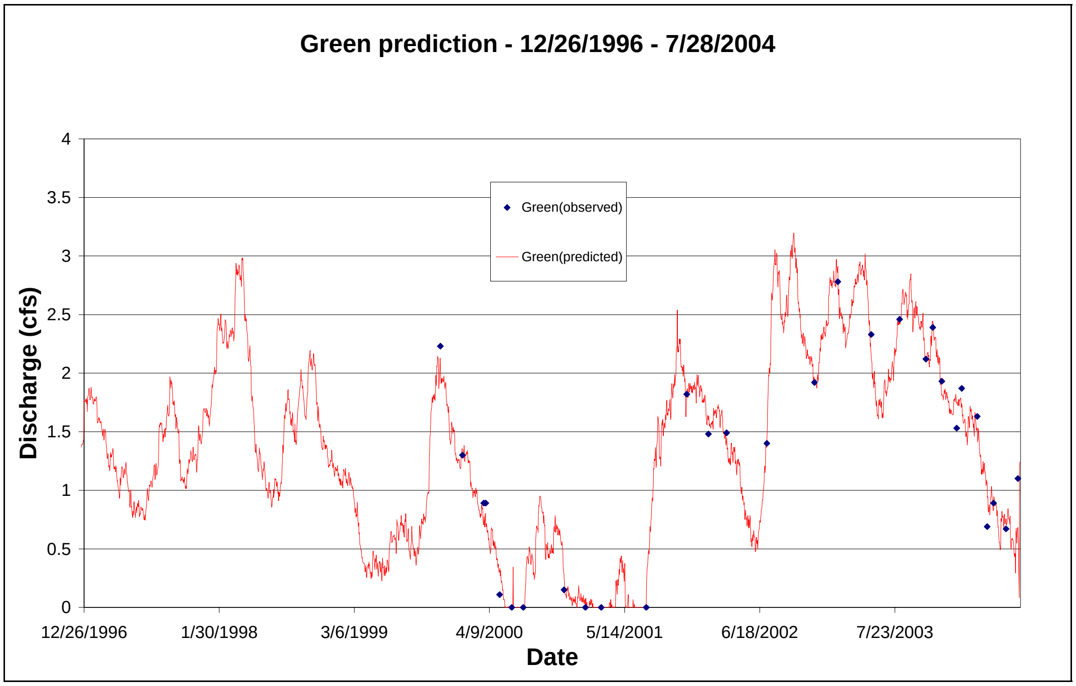


Figure 26 Daily discharge predictions, Green Springs, 1996-2004.

The sparsity of actual observations of Green Springs discharge during the pre-1996 period preclude a meaningful evaluation of the reliability of the daily predictions shown in Figure 27, generated using Eq. (7).

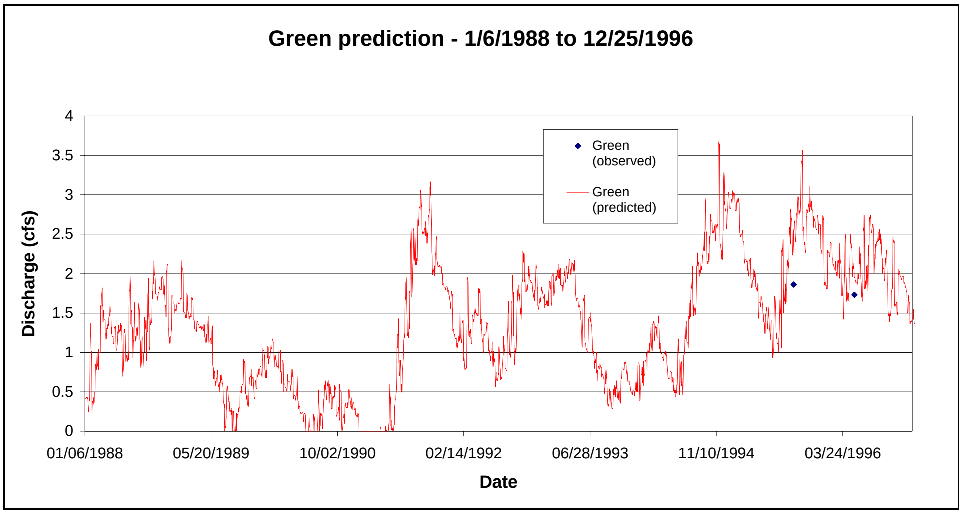


Figure 27 Daily discharge predictions, Green Springs, 1988 – 1996.

Figure 28 shows the flow duration curve for Green Springs, i.e., discharge versus percent exceedance for the long-term simulation generated from the results of the statistical modeling, indicating good agreement between the statistical characteristics of the observed and predicted spring discharge.

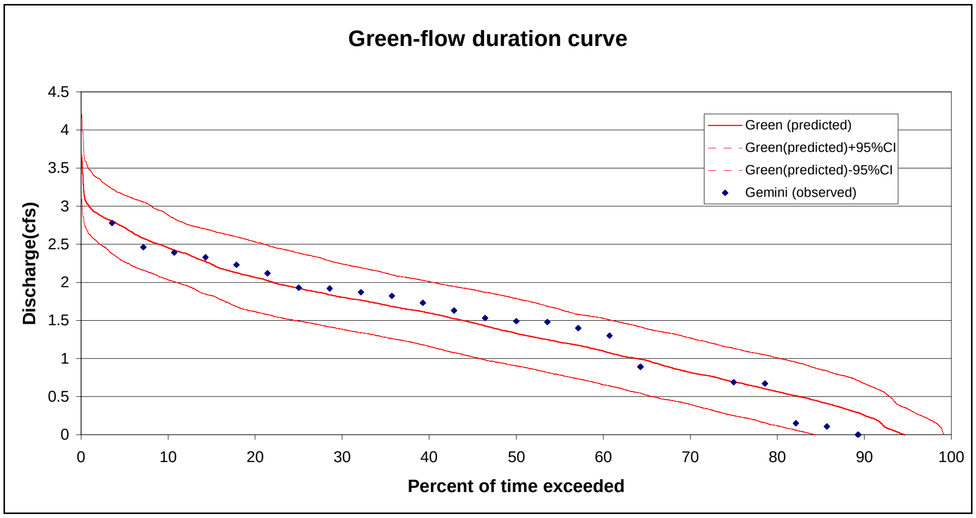
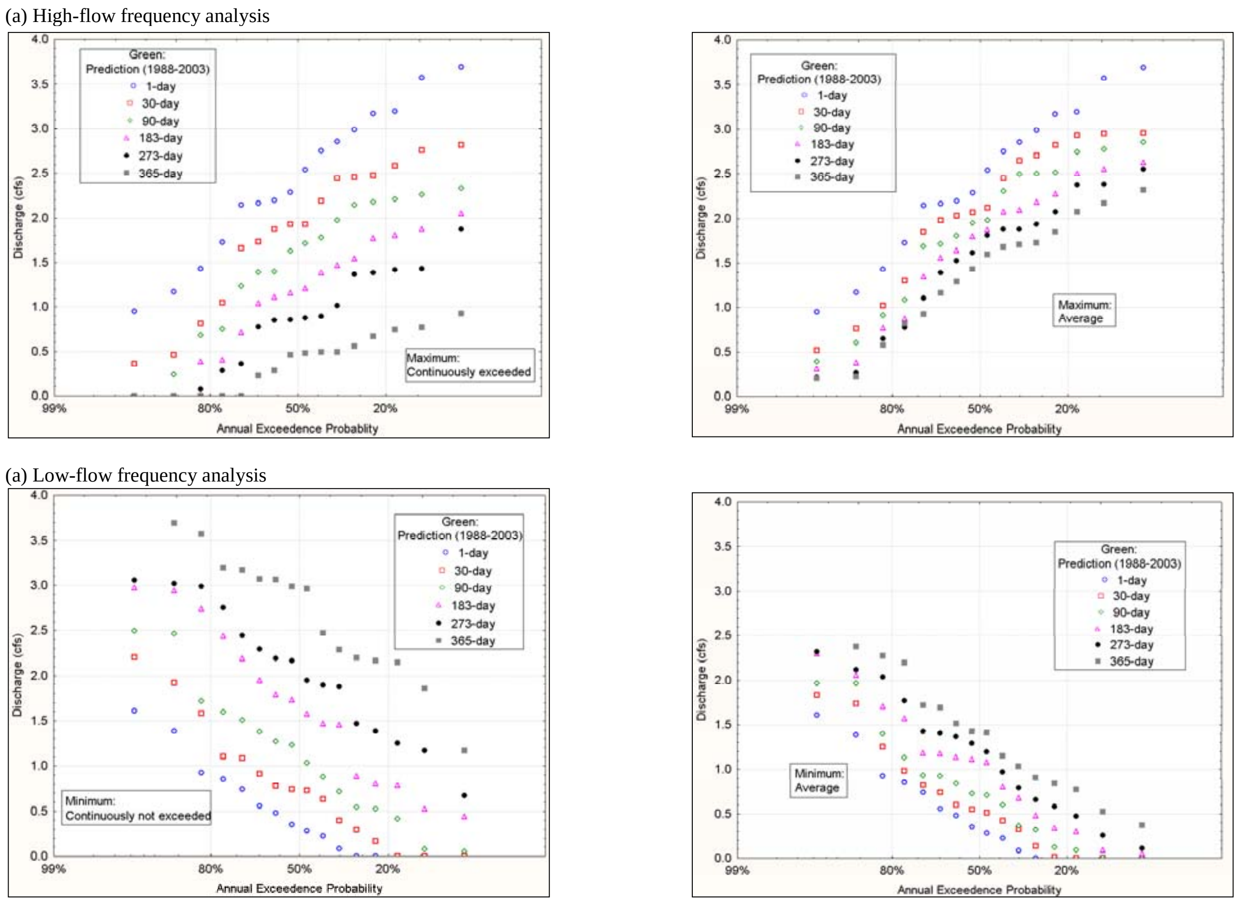


Figure 28 Flow duration curve for Green Springs.

The corresponding high- and low-flow frequency analyses for the system (frequency of spring discharge for durations of 1 month, 2 months, 3 months, 4 months, 6 months and 1 year) are shown in Figure 29.



Date: July 5, 2007	**High- and low-frequency analysis of discharge for Green Springs**
File: Fig 29.pdf	

INTERA	**St. Johns River Water Management District** **Palatka, Florida**	**Figure 29**

Figure 29 High- and low-frequency analysis of discharge for Green Springs.

6.0 CONCLUSIONS AND RECOMMENDATIONS

This document presents an evaluation of the spring discharge data for PDL, Gemini, Green, and Blue 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 daily spring discharge records at PDL, Gemini, and Green springs. Flow duration curves are then 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.

- Most measurements of spring discharge and groundwater level are at a frequency of ~30 days greater – necessitating the generation of moving averages with commensurate lags to be used as surrogate predictor variables.
- The Blue Spring daily discharge values show significant correlation with the daily discharge values at Gemini Springs and some correlation with PDL. This fact, along with the fact that more data are available for estimating early time daily discharge at Blue Spring, has been utilized to help estimate discharge at the PDL and Gemini springs during the early time period when groundwater level measurements are scarce.
- Typically, two regression models of spring discharge are needed: (a) one for the period when groundwater levels and rainfall data are available, and (b) one for the period when rainfall data are supplemented with discharge from adjacent springs and perhaps groundwater levels from one or two long-term monitoring wells.
- Stepwise regression is a good starting point for regression modeling – as indicated by the linearity of the residuals in a probability plot and the reasonable nature of daily discharge predictions compared to actual observations recorded at less frequent intervals.
- Daily discharge predictions can be made for PDL as far back in time as 1966 with reasonable accuracy. However, comparable predictions can only be made until 1996 for Gemini Springs and until 1988 for Green Springs.

Based on the data evaluation, regression model building and discharge prediction exercises undertaken during this study, the following recommendations are offered to improve the modeling process in a subsequent phase.

- Given that some of the regression models produced R^2 values of ~0.50, it might be useful exploring other regression techniques for these datasets. Hastie et al. (2001) describe several alternatives to multivariate linear regression such as (a) generalized additive modeling, (b) tree-based methods, (c) multivariate adaptive regression splines and (d) neural networks. Such models could potentially improve the accuracy of the daily predictions by capturing non-linear trends and/or variable interactions between the response and predictor variables.

- Many of the data sets have an average frequency of 30 days or more. As such, there are data gaps even after the computation of moving averages with lags as long as 8 and 12 weeks. In this study, such gaps were generally filled using simple linear interpolation. It is recommended that a more advanced approach such as spline fitting be employed to fill the data gaps remaining after the computation of moving averages.

- The number of groundwater monitoring wells associated with a spring was limited in this study to those in its geographical vicinity so as to simplify the regression modeling process. However, in stepwise regression, the number of potential predictor variables is not a constraint. As such, it is recommended that groundwater monitoring wells falling within a larger radius than that used in this study be used as candidate predictor variables.

- The generation of daily spring discharge based only on rainfall records and perhaps the discharge at an adjacent spring does not appear to a feasible proposition. It is recommended that daily spring discharge prediction exercises be limited to situations where ancillary groundwater level measurements are available.

In summary, we note that the although reasonable predictions of daily discharge have been made for all three springs of interest using the best available data, the corresponding periods of record are only ~10 years for Gemini Springs and ~20 years for Green Springs.

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 Sttistics*, Vol. 12, No. 4, pp. 412-416.

Draper, N.R. and H. Smith, 1981. *Applied Regression Analysis.* John Wiley, New York.

Hastie, T., R. Tibshirani, and J. Friedman, 2001. *The Elements of Statistical Learning – Data Mining, Inference and Prediction,* Springer-Verlag, New York.

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

Osburn, William, 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

Appendix A: Model Usage Notes

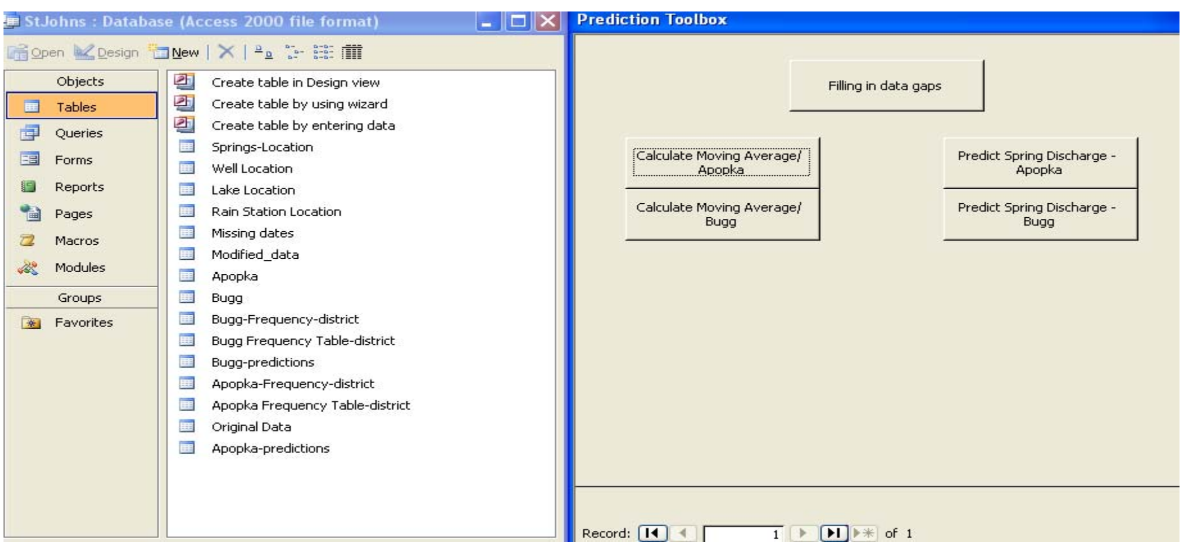
This Appendix describes the structure and operation of an ACCESS database created to facilitate predictive applications of the statistical spring discharge models described earlier in Section 4. An example using Bugg Spring data is also presented.

1. *Folder*: <u>Spring Daily Predictions</u> –

The folder **<u>Spring Daily Predictions</u>** has two files as shown below:

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

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

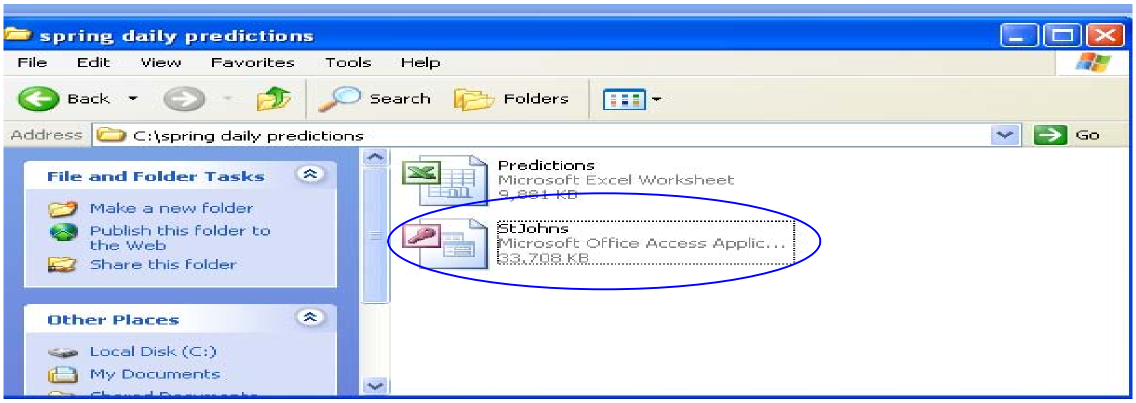


On the left, are the different tables present in the database and on the right is a prediction toolbox. The prediction toolbox executes ACCESS queries and/or VISUAL BASIC APPLICATION Modules, on the click of different buttons. **<u>Predictions.xls – EXCEL file</u>** is used to graphically display the 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.

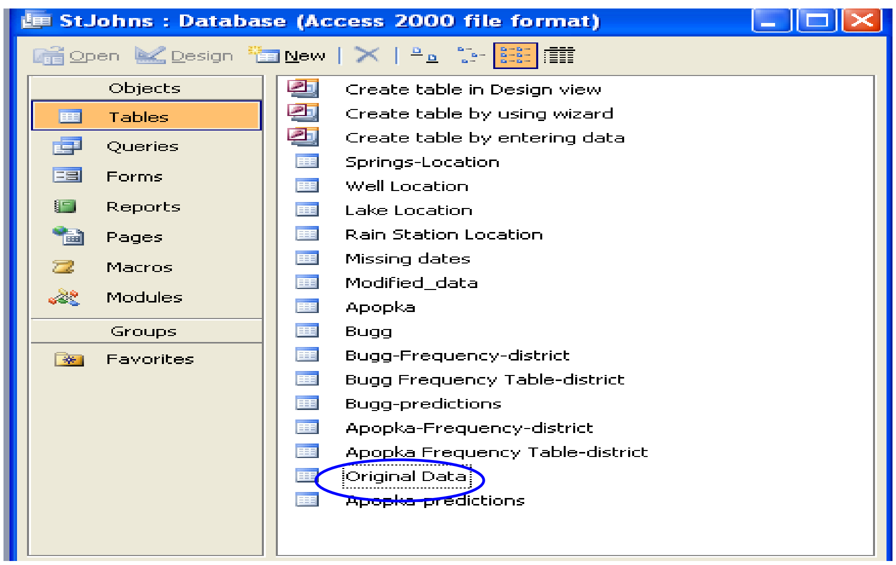
In the example below, our primary task would be to get Bugg Spring daily predictions from 10/27/1973 to 11/28/2005.

2. Open St.Johns.mdb

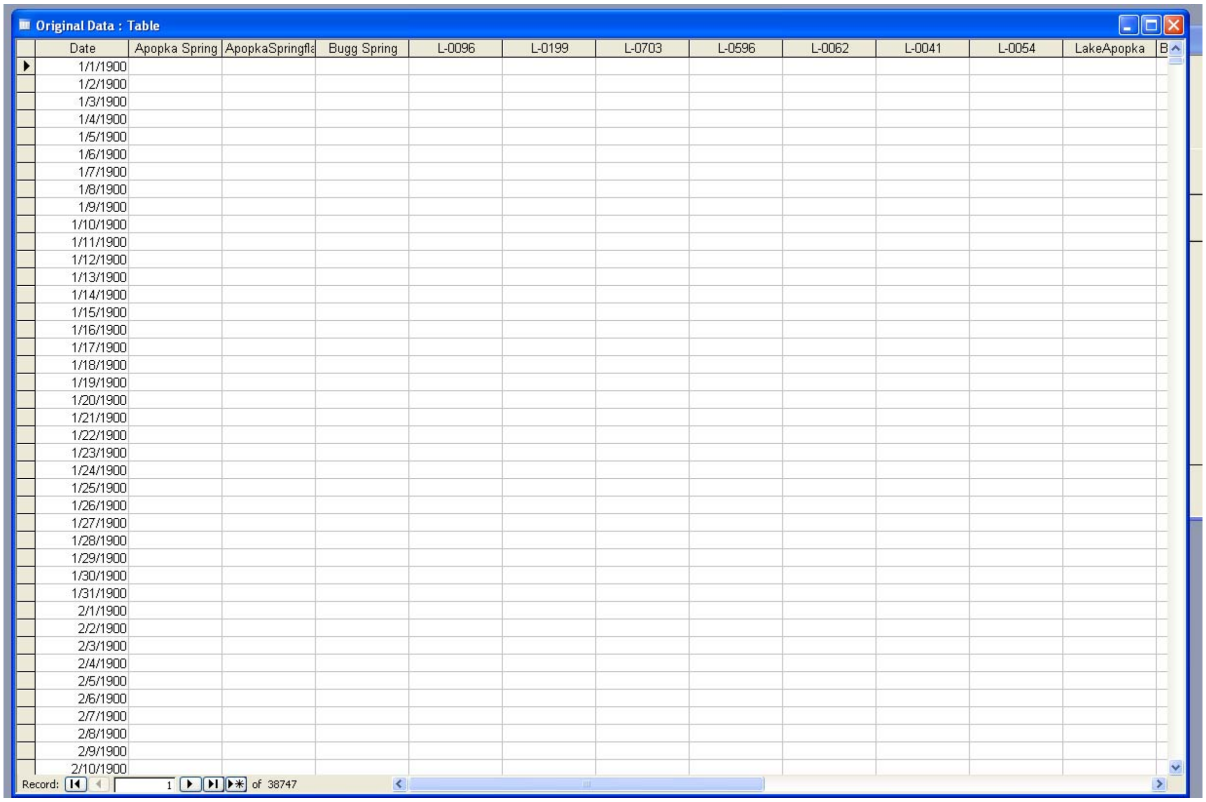
Open **St.Johns.mdb** (highlighted below) by double clicking the file.



The original spring discharge, groundwater elevation, lake level 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.



Date	Apopka Spring	ApopkaSpringfla	Bugg Spring	L-0096	L-0199	L-0703	L-0596	L-0062	L-0041	L-0054	LakeApopka	B
1/1/1900												
1/2/1900												
1/3/1900												
1/4/1900												
1/5/1900												
1/6/1900												
1/7/1900												
1/8/1900												
1/9/1900												
1/10/1900												
1/11/1900												
1/12/1900												
1/13/1900												
1/14/1900												
1/15/1900												
1/16/1900												
1/17/1900												
1/18/1900												
1/19/1900												
1/20/1900												
1/21/1900												
1/22/1900												
1/23/1900												
1/24/1900												
1/25/1900												
1/26/1900												
1/27/1900												
1/28/1900												
1/29/1900												
1/30/1900												
1/31/1900												
2/1/1900												
2/2/1900												
2/3/1900												
2/4/1900												
2/5/1900												
2/6/1900												
2/7/1900												
2/8/1900												
2/9/1900												
2/10/1900												

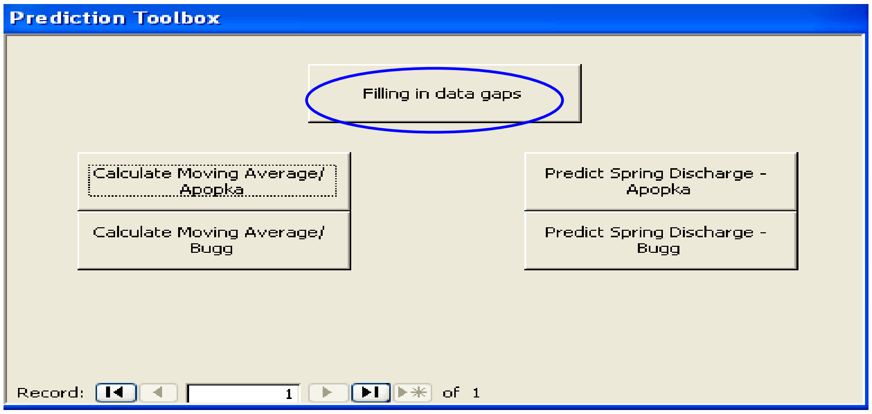
The table has 38747 records for dates ranging from 1/1/1900 to 1/31/2006. If the user wants to change a particular data time series, pasting the new time series (with dates from 1/1/1900 to 1/31/2006) over the old one is one of the ways to do it.

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 L-0096 becomes available, append the new data column as *L-0096(new)* using the *Append Query*. Then delete the old *L-0096* column from **Original Data** table and rename L-0096(new) as L-0096. If data is not available for a particular date, the user can leave it blank as seen in **Original Data** table for different variables.

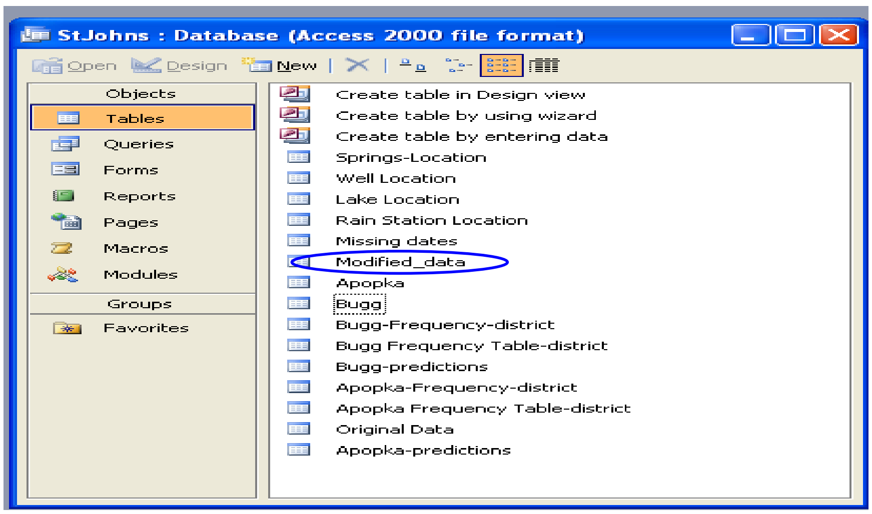
3. Data Gap Filling to create "Modified Data" Table

Gaps in the data (over continuous periods) are filled by regressing against more frequently observed data for a related variable. The need to fill data gaps for some wells arises during the calculation of moving averages. For example, groundwater elevations at L-0703 can be predicted from water levels at L-0096 using a simple linear regression model. Such relationships, developed for well pairs L-0703/L-0096 and L-0054/L-0096 have been pre-programmed, and are invoked to fill in the gaps in the **Original Data** table.

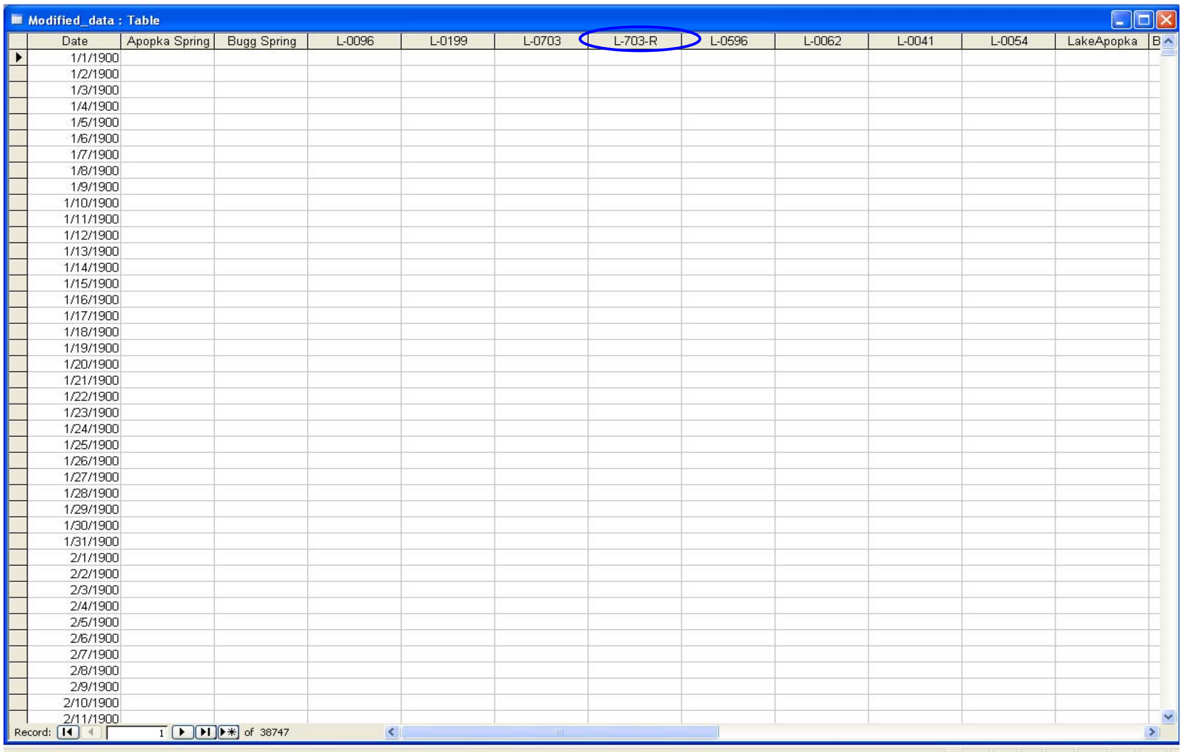
Therefore the next step is clicking the "Filling in data gaps" button on the prediction toolbox.



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:

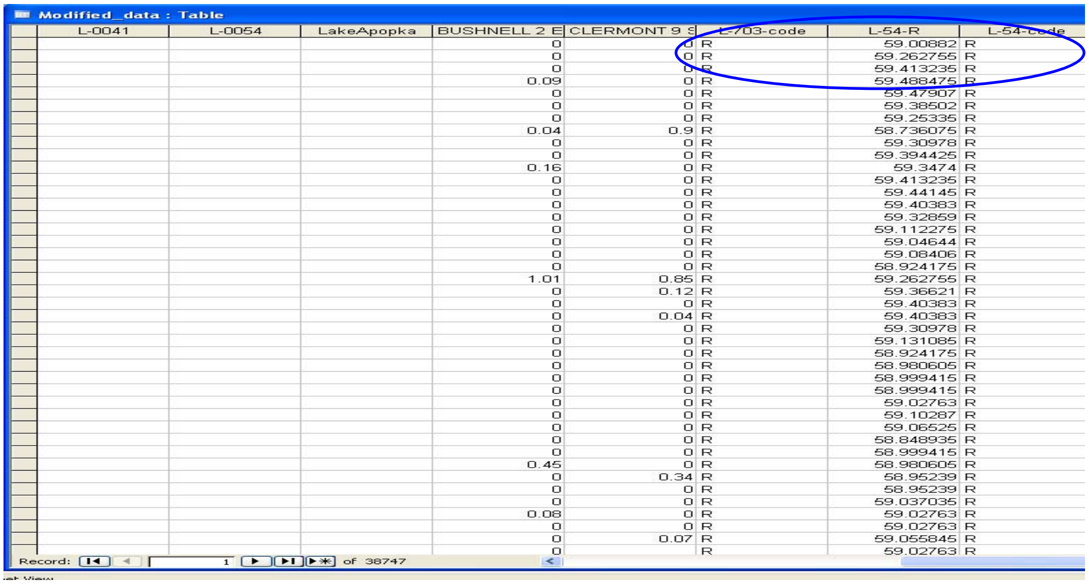


Modified_data : Table

Date	Apopka Spring	Bugg Spring	L-0096	L-0199	L-0703	L-703-R	L-0596	L-0062	L-0041	L-0054	LakeApopka	B
1/1/1900												
1/2/1900												
1/3/1900												
1/4/1900												
1/5/1900												
1/6/1900												
1/7/1900												
1/8/1900												
1/9/1900												
1/10/1900												
1/11/1900												
1/12/1900												
1/13/1900												
1/14/1900												
1/15/1900												
1/16/1900												
1/17/1900												
1/18/1900												
1/19/1900												
1/20/1900												
1/21/1900												
1/22/1900												
1/23/1900												
1/24/1900												
1/25/1900												
1/26/1900												
1/27/1900												
1/28/1900												
1/29/1900												
1/30/1900												
1/31/1900												
2/1/1900												
2/2/1900												
2/3/1900												
2/4/1900												
2/5/1900												
2/6/1900												
2/7/1900												
2/8/1900												
2/9/1900												
2/10/1900												
2/11/1900												

Record: 1 of 38747

The user would notice some new variables present in the **Modified_Table**. For example, we see L-703-R highlighted in the above screenshot. L-703-R has all the original well-data for L-0703 and some regressed data values from L-0096 using a simple linear regression model. Similarly, **Modified_Table** will also have L0054-R as new variable. **Modified_Table** also has additional columns called L-703-code and L-54-code, which flag the water-level data values filled by regression with letter "R". This is highlighted in screenshot below:



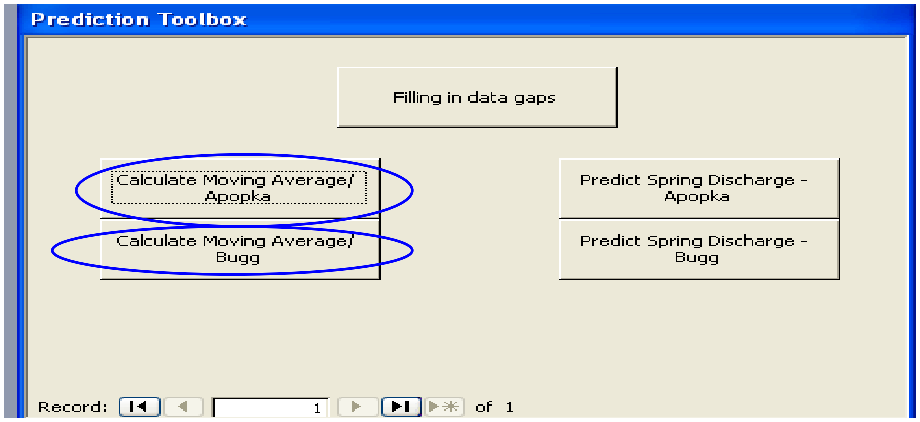
Modified_data : Table

L-0041	L-0054	LakeApopka	BUSHNELL 2 E	CLERMONT 9 S	L-703-code	L-54-R	L-54-code
			0	0	R	59.00882	R
			0	0	R	59.262755	R
			0	0	R	59.413235	R
			0.09	0	R	59.488475	R
			0	0	R	59.47907	R
			0	0	R	59.38502	R
			0	0	R	59.25335	R
			0.04	0.9	R	58.736075	R
			0	0	R	59.30978	R
			0	0	R	59.394425	R
			0.16	0	R	59.3474	R
			0	0	R	59.413235	R
			0	0	R	59.44145	R
			0	0	R	59.40383	R
			0	0	R	59.32859	R
			0	0	R	59.112275	R
			0	0	R	59.04644	R
			0	0	R	59.08406	R
			0	0	R	58.924175	R
			1.01	0.85	R	59.262755	R
			0	0.12	R	59.36621	R
			0	0	R	59.40383	R
			0	0.04	R	59.40383	R
			0	0	R	59.30978	R
			0	0	R	59.131085	R
			0	0	R	58.924175	R
			0	0	R	58.980605	R
			0	0	R	58.999415	R
			0	0	R	58.999415	R
			0	0	R	59.02763	R
			0	0	R	59.10287	R
			0	0	R	59.06525	R
			0	0	R	58.848935	R
			0	0	R	58.999415	R
			0.45	0	R	58.980605	R
			0	0.34	R	58.95239	R
			0	0	R	58.95239	R
			0	0	R	59.037035	R
			0.08	0	R	59.02763	R
			0	0	R	59.02763	R
			0	0.07	R	59.055845	R
			0		R	59.02763	R

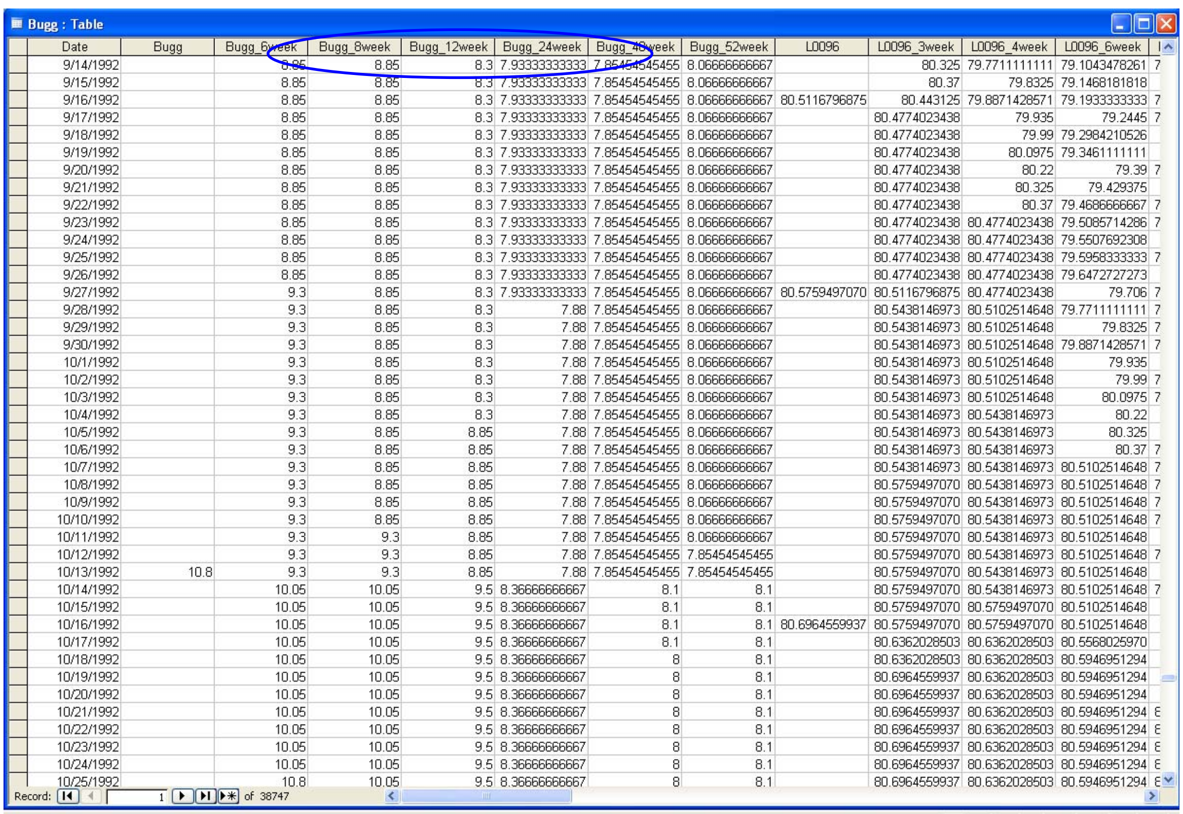
Record: 1 of 38747

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, lake level, and rainfall data) for predicting daily discharge for each spring. Computation of these variables, for each spring, is then performed by clicking the two buttons highlighted below.



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



Bugg : Table

Date	Bugg	Bugg_6week	Bugg_8week	Bugg_12week	Bugg_24week	Bugg_40week	Bugg_52week	L0096	L0096_3week	L0096_4week	L0096_6week
9/14/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.325	79.7711111111	79.1043478261
9/15/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.37	79.8325	79.1468181818
9/16/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667	80.5116796875	80.443125	79.8871428571	79.1933333333
9/17/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	79.935	79.2445
9/18/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	79.99	79.2984210526
9/19/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.0975	79.3461111111
9/20/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.22	79.39
9/21/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.325	79.429375
9/22/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.37	79.4686666667
9/23/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.4774023438	79.5085714286
9/24/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.4774023438	79.5507692308
9/25/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.4774023438	79.5958333333
9/26/1992		8.85	8.85	8.3	7.93333333333	7.85454545455	8.06666666667		80.4774023438	80.4774023438	79.6472727273
9/27/1992		9.3	8.85	8.3	7.93333333333	7.85454545455	8.06666666667	80.5759497070	80.5116796875	80.4774023438	79.706
9/28/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	79.7711111111
9/29/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	79.8325
9/30/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	79.8871428571
10/1/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	79.935
10/2/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	79.99
10/3/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5102514648	80.0975
10/4/1992		9.3	8.85	8.3	7.88	7.85454545455	8.06666666667		80.5438146973	80.5438146973	80.22
10/5/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5438146973	80.5438146973	80.325
10/6/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5438146973	80.5438146973	80.37
10/7/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5438146973	80.5438146973	80.5102514648
10/8/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5759497070	80.5438146973	80.5102514648
10/9/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5759497070	80.5438146973	80.5102514648
10/10/1992		9.3	8.85	8.85	7.88	7.85454545455	8.06666666667		80.5759497070	80.5438146973	80.5102514648
10/11/1992		9.3	9.3	8.85	7.88	7.85454545455	8.06666666667		80.5759497070	80.5438146973	80.5102514648
10/12/1992		9.3	9.3	8.85	7.88	7.85454545455	7.85454545455		80.5759497070	80.5438146973	80.5102514648
10/13/1992	10.8	9.3	9.3	8.85	7.88	7.85454545455	7.85454545455		80.5759497070	80.5438146973	80.5102514648
10/14/1992		10.05	10.05	9.5	8.36666666667	8.1	8.1		80.5759497070	80.5438146973	80.5102514648
10/15/1992		10.05	10.05	9.5	8.36666666667	8.1	8.1		80.5759497070	80.5759497070	80.5102514648
10/16/1992		10.05	10.05	9.5	8.36666666667	8.1	8.1	80.6964559937	80.5759497070	80.5759497070	80.5102514648
10/17/1992		10.05	10.05	9.5	8.36666666667	8.1	8.1		80.6362028503	80.6362028503	80.5568025970
10/18/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6362028503	80.6362028503	80.5946951294
10/19/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/20/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/21/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/22/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/23/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/24/1992		10.05	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294
10/25/1992		10.8	10.05	9.5	8.36666666667	8	8.1		80.6964559937	80.6362028503	80.5946951294

Record: 1 of 38747

The highlighted columns in the **Bugg** table above show some of the calculated moving averages to be used in the Bugg statistical model for daily discharge predictions. One extra piece of information generated on clicking *Calculate Moving Average/Bugg* is in the table **Missing Dates** shown below:

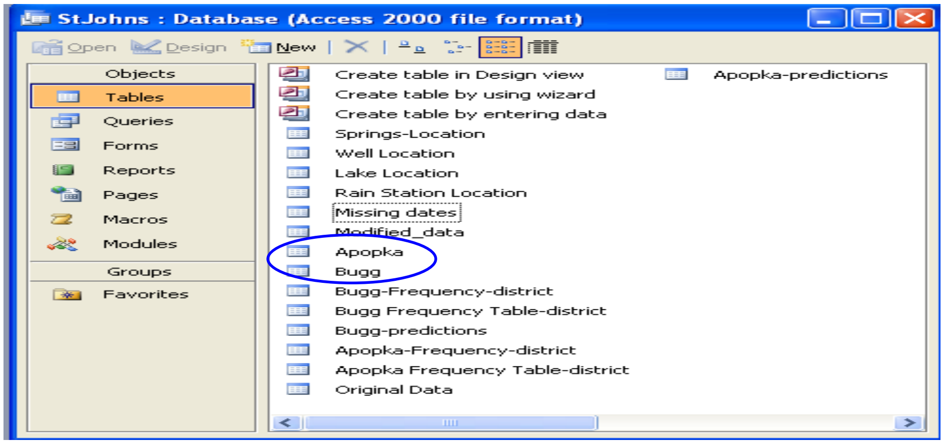
Missing dates : Table

variable	startdate	startvalue	enddate	endvalue	gap	dateint	Interpolated_val	Flag
Bugg	3/2/2002	11.9	4/13/2002	12.2	42	3/23/2002	12.05	Int
Bushnell	12/7/1944	0.66796034828	9/19/1946	2.08398017414	651	10/28/1945	1.37597026121	Int
Bushnell	12/18/1944	0.69008565806	1/28/1945	0.77789548124	41	1/8/1945	0.73399056965	Int
Bushnell	12/18/1944	0.69008565806	3/9/1945	0.86570530443	81	1/28/1945	0.77789548124	Int
Bushnell	12/18/1944	0.69008565806	5/28/1945	1.04132495080	161	3/9/1945	0.86570530443	Int
Bushnell	12/18/1944	0.69008565806	11/6/1945	1.39256424354	323	5/28/1945	1.04132495080	Int
Bushnell	12/18/1944	0.69008565806	9/25/1946	2.09504282903	646	11/6/1945	1.39256424354	Int
Bushnell	12/18/1944	0.69008565806	7/2/1948	3.5	1292	9/25/1946	2.09504282903	Int
Bushnell	12/7/1944	0.66796034828	12/28/1944	0.71221096784	21	12/18/1944	0.69008565806	Int
Bushnell	12/7/1944	0.66796034828	1/17/1945	0.75646158739	41	12/28/1944	0.71221096784	Int
Bushnell	12/7/1944	0.66796034828	2/26/1945	0.84496282651	81	1/17/1945	0.75646158739	Int
Bushnell	2/16/1945	0.81958688313	12/21/1945	1.48969016235	308	7/20/1945	1.15463852274	Int
Bushnell	12/7/1944	0.66796034828	10/28/1945	1.37597026121	325	5/19/1945	1.02196530474	Int
Bushnell	12/29/1944	0.71203811385	10/1/1946	2.10601905693	641	11/15/1945	1.40902858539	Int
Bushnell	12/7/1944	0.66796034828	7/2/1948	3.5	1303	9/19/1946	2.08398017414	Int
Bushnell	11/26/1944	0.64566082346	12/17/1944	0.69025987309	21	12/7/1944	0.66796034828	Int
Bushnell	11/26/1944	0.64566082346	1/6/1945	0.73485892273	41	12/17/1944	0.69025987309	Int
Bushnell	11/26/1944	0.64566082346	2/17/1945	0.824057022	83	1/6/1945	0.73485892273	Int
Bushnell	11/26/1944	0.64566082346	5/10/1945	1.00245322053	165	2/17/1945	0.824057022	Int
Bushnell	11/26/1944	0.64566082346	10/21/1945	1.3592456176	329	5/10/1945	1.00245322053	Int
Bushnell	11/26/1944	0.64566082346	9/14/1946	2.07283041173	657	10/21/1945	1.3592456176	Int
Bushnell	11/26/1944	0.64566082346	7/2/1948	3.5	1314	9/14/1946	2.07283041173	Int
Bushnell	12/7/1944	0.66796034828	5/19/1945	1.02196530474	163	2/26/1945	0.84496282651	Int
Bushnell	1/9/1945	0.73381906609	6/17/1945	1.07959168283	159	3/29/1945	0.90670537446	Int
Bushnell	11/2/1945	1.38353853005	7/3/1946	1.91265389754	243	3/3/1946	1.64809621379	Int
Bushnell	2/16/1945	0.81958688313	7/2/1948	3.5	1232	10/25/1946	2.15979344157	Int
Bushnell	1/28/1945	0.77704064318	3/8/1945	0.86213312308	39	2/16/1945	0.81958688313	Int
Bushnell	1/28/1945	0.77704064318	4/17/1945	0.94722560298	79	3/8/1945	0.86213312308	Int
Bushnell	1/28/1945	0.77704064318	7/4/1945	1.11741056278	157	4/17/1945	0.94722560298	Int
Bushnell	1/28/1945	0.77704064318	12/7/1945	1.45778048239	313	7/4/1945	1.11741056278	Int
Bushnell	1/28/1945	0.77704064318	10/15/1946	2.13852032159	625	12/7/1945	1.45778048239	Int
Bushnell	1/28/1945	0.77704064318	7/2/1948	3.5	1251	10/15/1946	2.13852032159	Int
Bushnell	12/18/1944	0.69008565806	1/8/1945	0.73399056965	21	12/29/1944	0.71203811385	Int
Bushnell	1/9/1945	0.73381906609	3/29/1945	0.90670537446	79	2/17/1945	0.82026222027	Int
Bushnell	12/29/1944	0.71203811385	7/2/1948	3.5	1281	10/1/1946	2.10601905693	Int
Bushnell	1/9/1945	0.73381906609	11/22/1945	1.42536429957	317	6/17/1945	1.07959168283	Int
Bushnell	1/9/1945	0.73381906609	10/6/1946	2.11690953304	635	11/22/1945	1.42536429957	Int
Bushnell	1/9/1945	0.73381906609	7/2/1948	3.5	1270	10/6/1946	2.11690953304	Int
Bushnell	12/29/1944	0.71203811385	1/19/1945	0.75560001832	21	1/9/1945	0.73381906609	Int
Bushnell	12/29/1944	0.71203811385	2/8/1945	0.7991619228	41	1/19/1945	0.75560001832	Int
Bushnell	12/29/1944	0.71203811385	3/20/1945	0.88628573174	81	2/8/1945	0.7991619228	Int
Bushnell	12/29/1944	0.71203811385	6/8/1945	1.06053334962	161	3/20/1945	0.88628573174	Int

Record: 1 of 883

The table above informs the user about interpolated values added to a particular data time-series to facilitate calculation of certain moving average variables. For example, in the first row, a linear interpolated value (12.05) is added on 3/23/2002 to fill a 42 day gap between 3/2/2002 and 4/13/2002. Values in columns *startvalue* (11.9) and *endvalue* (12.2) are the data associated with 3/2/2002 and 4/13/2002 respectively. This interpolation would then help in calculation of Bugg-6-week moving average variable.

Similarly, clicking *Calculate Moving Average/Apopka*, would fill the table **Apopka** with required moving average variables. Also, the **Missing Dates** table is updated for each spring. The following screenshot indicates the two tables being filled with moving average variables.

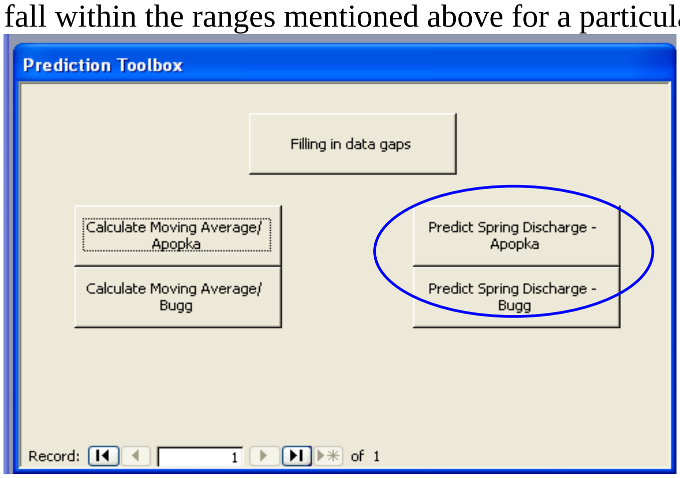


5. Calculate Spring discharge predictions and frequency analysis

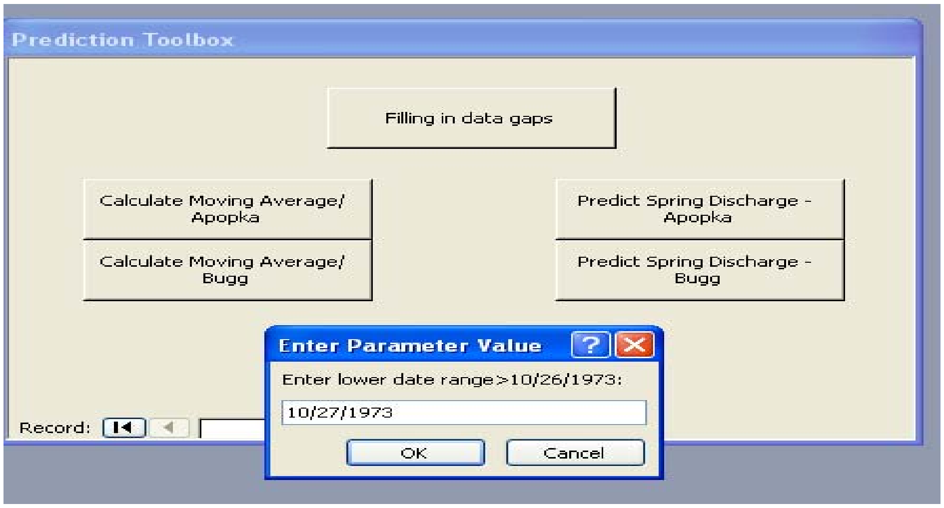
Spring discharge daily predictions are limited by a range of lower and upper date. This is due to limited date range coverage for explanatory variables in the statistical model for a particular spring. The following are the dates for the two springs for which daily discharge predictions can be computed:

Spring	Date Range for discharge predictions
Apopka	6/2/1949 to 12/31/2005
Bugg	10/27/1973 to 11/28/2005

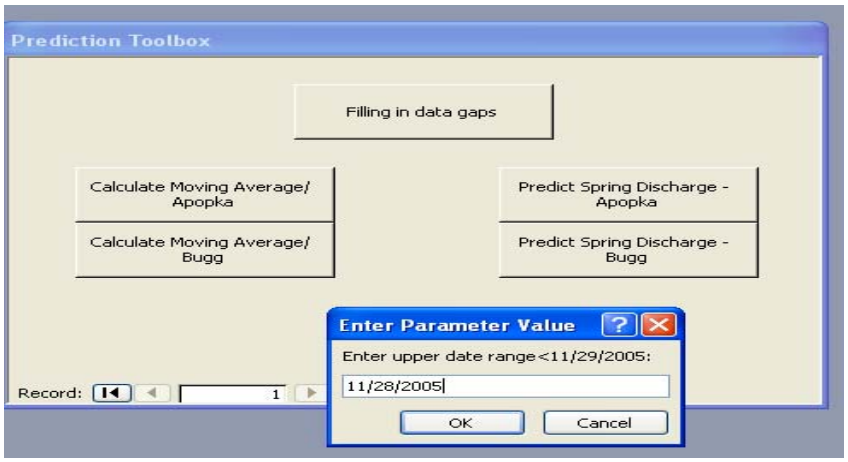
Clicking the buttons highlighted below give daily discharge predictions and maximum and minimum frequencies for date ranges specified by the user. Note that these date ranges have to fall within the ranges mentioned above for a particular spring



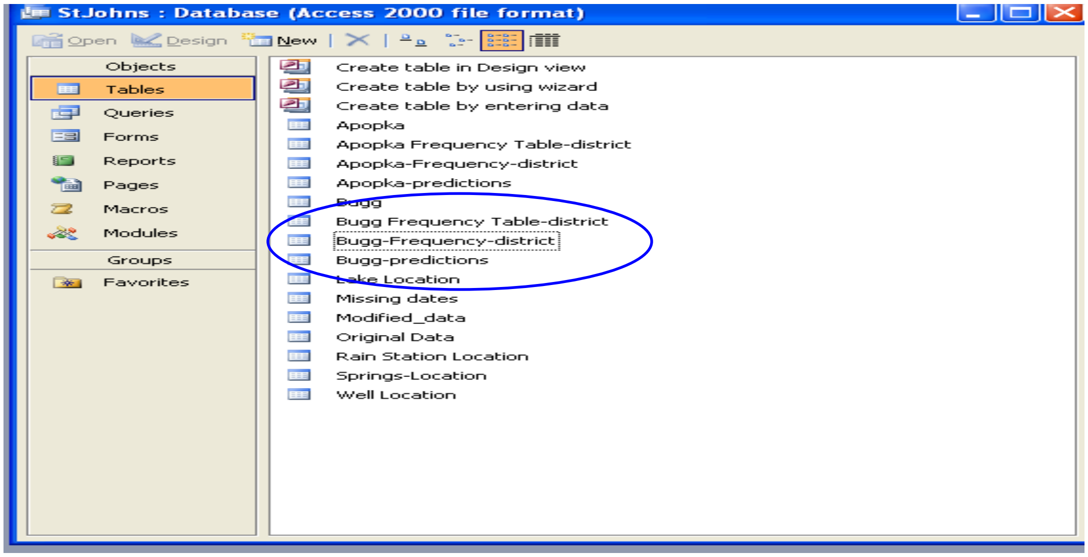
For example, on clicking *Predict Spring Discharge - Bugg*, we see a pop-up window asking for the date from which predictions are needed. For our example enter 10/27/1973. As noted earlier, the date entered should be greater than 10/26/1973, since Bugg Spring predictions are only available since that date.



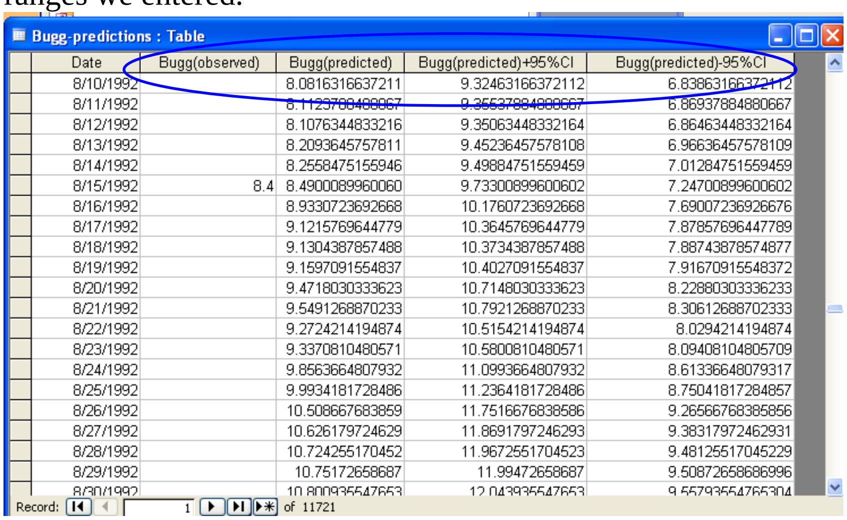
Press OK. Another window asking for the date till which predictions are needed. For our example enter 11/28/2005. Again the date entered should be less than 11/29/2005, since Bugg Spring predictions are only available till 11/28/2005.



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



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



Bugg-predictions : Table

Date	Bugg(observed)	Bugg(predicted)	Bugg(predicted)+95%CI	Bugg(predicted)-95%CI
8/10/1992		8.0816316637211	9.32463166372112	6.83863166372112
8/11/1992		8.11237884880667	9.35537884880667	6.86937884880667
8/12/1992		8.1076344833216	9.35063448332164	6.86463448332164
8/13/1992		8.2093645757811	9.45236457578108	6.96636457578109
8/14/1992		8.2558475155946	9.49884751559459	7.01284751559459
8/15/1992	8.4	8.4900089960060	9.73300899600602	7.24700899600602
8/16/1992		8.9330723692668	10.1760723692668	7.69007236926676
8/17/1992		9.1215769644779	10.3645769644779	7.87857696447789
8/18/1992		9.1304387857488	10.3734387857488	7.88743878574877
8/19/1992		9.1597091554837	10.4027091554837	7.91670915548372
8/20/1992		9.4718030333623	10.7148030333623	8.22880303336233
8/21/1992		9.5491268870233	10.7921268870233	8.30612688702333
8/22/1992		9.2724214194874	10.5154214194874	8.0294214194874
8/23/1992		9.3370810480571	10.5800810480571	8.09408104805709
8/24/1992		9.8563664807932	11.0993664807932	8.61336648079317
8/25/1992		9.9934181728486	11.2364181728486	8.75041817284857
8/26/1992		10.508667683859	11.7516676838586	9.26566768385856
8/27/1992		10.626179724629	11.8691797246293	9.38317972462931
8/28/1992		10.724255170452	11.9672551704523	9.48125517045229
8/29/1992		10.75172658687	11.99472658687	9.50872658686996
8/30/1992		10.800935547653	12.043935547653	9.55793554765304

Record: 1 of 11721

The highlighted columns above show Observed Bugg Discharge data, Bugg discharge predictions, Bugg discharge predictions upper (+) and lower (-) 95% confidence interval.

Double-click table **<u>Bugg-Frequency-district</u>** 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.

Bugg-Frequency-district : Table

Date	Bugg	Cont_exceeded_30days	Average_maximum_30days	Cont_not_exceeded_30days	Average_minimum_30days	Cont_exceeded_90days	Average_maxim
1/28/1974	9.78132326828	9.60577971768708	9.8771075537415	10.1628741573129	9.8771075537415	8.86758917772108	9.85663210341
1/29/1974	9.78132326828	9.60577971768708	9.8585377390873	10.0638001113946	9.8585377390873	8.86758917772108	9.84619218837
1/30/1974	9.77907294685	9.60127907482993	9.84312037120182	10.0630500042517	9.84312037120182	8.86758917772108	9.83564495386
1/31/1974	9.77607251828	9.60127907482993	9.82772800688776	10.0520131960034	9.82772800688776	8.86758917772108	9.82349914288
2/1/1974	9.79739540349	9.60127907482993	9.81306074614512	10.0520131960034	9.81306074614512	8.86758917772108	9.81188535318
2/2/1974	9.80859470961	9.57534926530612	9.80072122101757	10.0520131960034	9.80072122101757	8.86758917772108	9.80040715198
2/3/1974	9.80859470961	9.57359901530612	9.784774081661	10.0520131960034	9.784774081661	8.86758917772108	9.78900870871
2/4/1974	9.81562956675	9.56309751530612	9.76847689230442	10.0128872257653	9.76847689230442	8.86758917772108	9.77563312765
2/5/1974	9.81400895366	9.56034712244898	9.75339222219388	9.99668109481293	9.75339222219388	8.85208395493197	9.76114974591
2/6/1974	9.81150859651	9.56034712244898	9.74778564105726	9.98117888052722	9.74778564105726	8.80581220238094	9.74615223358
2/7/1974	9.80400752509	9.56034712244898	9.74256182777778	9.98117888052722	9.74256182777778	8.80581220238094	9.73416032398
2/8/1974	9.79975691794	9.56034712244898	9.73753854679705	9.88684504379252	9.73753854679705	8.80581220238094	9.72791702245
2/9/1974	9.77926028061	9.56034712244898	9.73624469591837	9.84802951743198	9.73624469591837	8.80581220238094	9.72212389123
2/10/1974	9.77601905442	9.56034712244898	9.73799316969955	9.84802951743198	9.73799316969955	8.80581220238094	9.71675659563
2/11/1974	9.77087864626	9.56034712244898	9.7345211117347	9.84802951743198	9.7345211117347	8.80581220238094	9.71007766724
2/12/1974	9.60577971769	9.56034712244898	9.73078234900794	9.84802951743198	9.73078234900794	8.80581220238094	9.70141634433
2/13/1974	9.60577971769	9.56034712244898	9.72704358628118	9.84802951743198	9.72704358628118	8.80581220238094	9.69052063475
2/14/1974	9.60127907483	9.56034712244898	9.72193072440476	9.84802951743198	9.72193072440476	8.80581220238094	9.67920661709
2/15/1974	9.60127907483	9.5303744659864	9.71374078932823	9.84802951743198	9.71374078932823	8.80581220238094	9.66858951405
2/16/1974	9.60127907483	9.41451464455783	9.70097809736395	9.84802951743198	9.70097809736395	8.80581220238094	9.65806812054
2/17/1974	9.57534926531	9.37905082312926	9.68665996781463	9.84802951743198	9.68665996781463	8.80581220238094	9.64754672703
2/18/1974	9.57359901531	9.37905082312926	9.67234183826531	9.84802951743198	9.67234183826531	8.80581220238094	9.63715860335
2/19/1974	9.56309751531	9.29153771598639	9.65487210990647	9.84802951743198	9.65487210990647	8.80581220238094	9.62712338355
2/20/1974	9.56034712245	9.28505526360545	9.6372403202381	9.84802951743198	9.6372403202381	8.80581220238094	9.61891056098
2/21/1974	9.82848366071	9.12137334098639	9.61423581172052	9.84802951743198	9.61423581172052	8.80581220238094	9.60980153104
2/22/1974	9.82446448214	9.10037171343537	9.59078128466554	9.84802951743198	9.59078128466554	8.80581220238094	9.59953531578
2/23/1974	9.83048045111	9.10037171343537	9.56746844451531	9.84802951743198	9.56746844451531	8.80581220238094	9.58926354417
2/24/1974	9.84802951743	8.94505757057822	9.53966168751417	9.84802951743198	9.53966168751417	8.80581220238094	9.57899177256
2/25/1974	9.83477762457	8.87183978486394	9.50952237852891	9.84802951743198	9.50952237852891	8.80581220238094	9.56831484768
2/26/1974	9.67816167219	8.86758917772108	9.47941272957766	9.84802951743198	9.47941272957766	8.80581220238094	9.55508919338
2/27/1974	9.66916038648	8.86758917772108	9.46214871554705	9.84802951743198	9.46214871554705	8.80581220238094	9.54388140809
2/28/1974	9.66916038648	8.86758917772108	9.44459299318311	9.84802951743198	9.44459299318311	8.80581220238094	9.53639651934
3/1/1974	9.62568709056	8.86758917772108	9.42673028272392	9.84802951743198	9.42673028272392	8.80581220238094	9.53372529460
3/2/1974	9.53037446599	8.86758917772108	9.40933656274093	9.84802951743198	9.40933656274093	8.80581220238094	9.53130614141
3/3/1974	9.41451464456	8.86758917772108	9.39194284275794	9.84802951743198	9.39194284275794	8.80581220238094	9.52893961789
3/4/1974	9.37905082313	8.86758917772108	9.37541344975907	9.84802951743198	9.37541344975907	8.80581220238094	9.52763711523
3/5/1974	9.37905082313	8.86758917772108	9.35894239842687	9.84802951743198	9.35894239842687	8.80581220238094	9.52508484553
3/6/1974	9.29153771599	8.86758917772108	9.33739541070011	9.84802951743198	9.33739541070011	8.80581220238094	9.52224699759
3/7/1974	9.28505526361	8.85208395493197	9.31378663844955	9.84802951743198	9.31378663844955	8.80581220238094	9.51801618198
3/8/1974	9.12137334099	8.80581220238094	9.27969758983843	9.84802951743198	9.27969758983843	8.80581220238094	9.51125244129
3/9/1974	9.10037171344	8.80581220238094	9.25475932200963	9.84802951743198	9.25475932200963	8.80581220238094	9.50466094743

Record: 1 of 11721

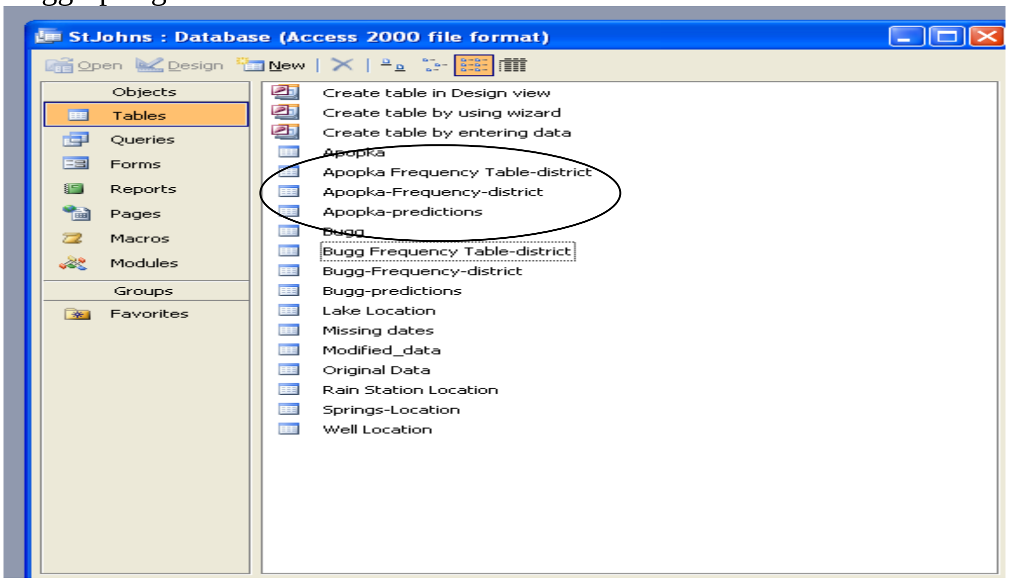
Double-click table **<u>Bugg Frequency Table-district</u>** to view. The table contains the maximums from 1-day, 30-day, 90-day, 183-day, 273-day and 365-day continuously-exceeded and average time-series for each year. The table also contains the minimums from 1-day, 30-day, 90-day, 183-day, 273-day and 365-day continuously-not-exceeded and average time-series for each year. The screenshot below shows a few columns from the table

Bugg Frequency Table-district : Table

Date	1-day(maximum-continuously exceeded)	30-day(maximum-continuously exceeded)	90-day(maximum-continuously exceeded)	183-day(maximum-continuously exceeded)	273-da
1974	12.9873364863019	12.3652778051795	11.8311125152135	9.94466894035592	9.624
1975	10.7120029152749	10.1612351151147	9.94229589398644	9.00918633340136	7.574
1976	10.0136716011905	9.58584862244899	9.22028748062222	9.00818701406851	8.530
1977	11.9180738141203	11.3157542946428	10.9227317491497	10.5872319982993	9.604
1978	11.1105931755952	10.3908891934524	9.90071007780613	9.20334366581636	8.765
1979	12.3007582437474	11.7474281160618	10.6985174306863	10.172534217395	9.323
1980	10.1039171568043	9.62974598854593	8.7724348105578	8.66650139635999	7.645
1981	14.3486929610825	13.6237109440856	11.2364929508616	9.24437599458874	8.824
1982	17.6716803210104	16.0964043531814	15.2071350085633	14.102095820252	12.797
1983	12.1726254076264	11.3983779511963	10.9084294481763	10.690535228025	10.213
1984	12.5956319230442	11.9858310888606	11.0229724466008	10.0915477776709	8.836
1985	12.2847095188492	11.8099947569445	11.525411017432	10.7925601763039	10.056
1986	11.2544365866482	10.304444656379	9.2227712394084	8.61968711143257	8.452
1987	12.9894700100465	12.2414658252008	11.3441930247947	10.635926662433	9.798
1988	12.6829755146687	11.4807375677578	11.0740719604846	10.5026912446445	9.557
1989	11.7341274373908	11.2406368698614	10.8427464963316	10.4493725370055	9.5831
1990	15.7114734327155	10.41829915794	8.47312120334373	7.47213468259883	7.273
1991	17.222739577347	15.3557437312282	11.6217714817397	8.25038339453807	6.349
1992	12.7909850740492	11.4459391257437	10.8746199141869	8.55083067838845	7.420
1993	10.7204862323749	10.0293562734609	8.92299226915057	7.89284465549725	7.414
1994	12.2624149559076	11.4227181230237	10.7130685479875	9.8952567115406	9.666
1995	13.4431485675854	12.5738026674404	11.5364602062692	10.0167220542725	9.728
1996	13.2433622643062	11.8574045416544	10.5222279071597	9.56635064640005	6.8847
1997	15.756156173118	14.3802465375186	12.3180731994123	9.48395906423863	8.689
1998	12.8465324134203	11.4167215790398	10.4790164739841	9.43052147749397	9.070
1999	15.8911193037668	11.4269843089354	9.93691492860102	8.00782841761008	8.007
2000	10.054455420627	9.15003720999122	8.46412095723674	7.62797353962955	7.627
2001	16.2579424298705	14.2278924966622	12.3591055721087	11.4316338719025	10.57
2002	14.8892655187825	14.4220106344657	13.2539204404961	10.8644473201178	10.864
2003	13.8550043070948	13.3321649824161	12.4603928768464	12.3609805281801	12.211
2004	13.7842898567331	12.8182789635149	12.1744373823952	11.9420363266607	8.385

Record: 1 of 31

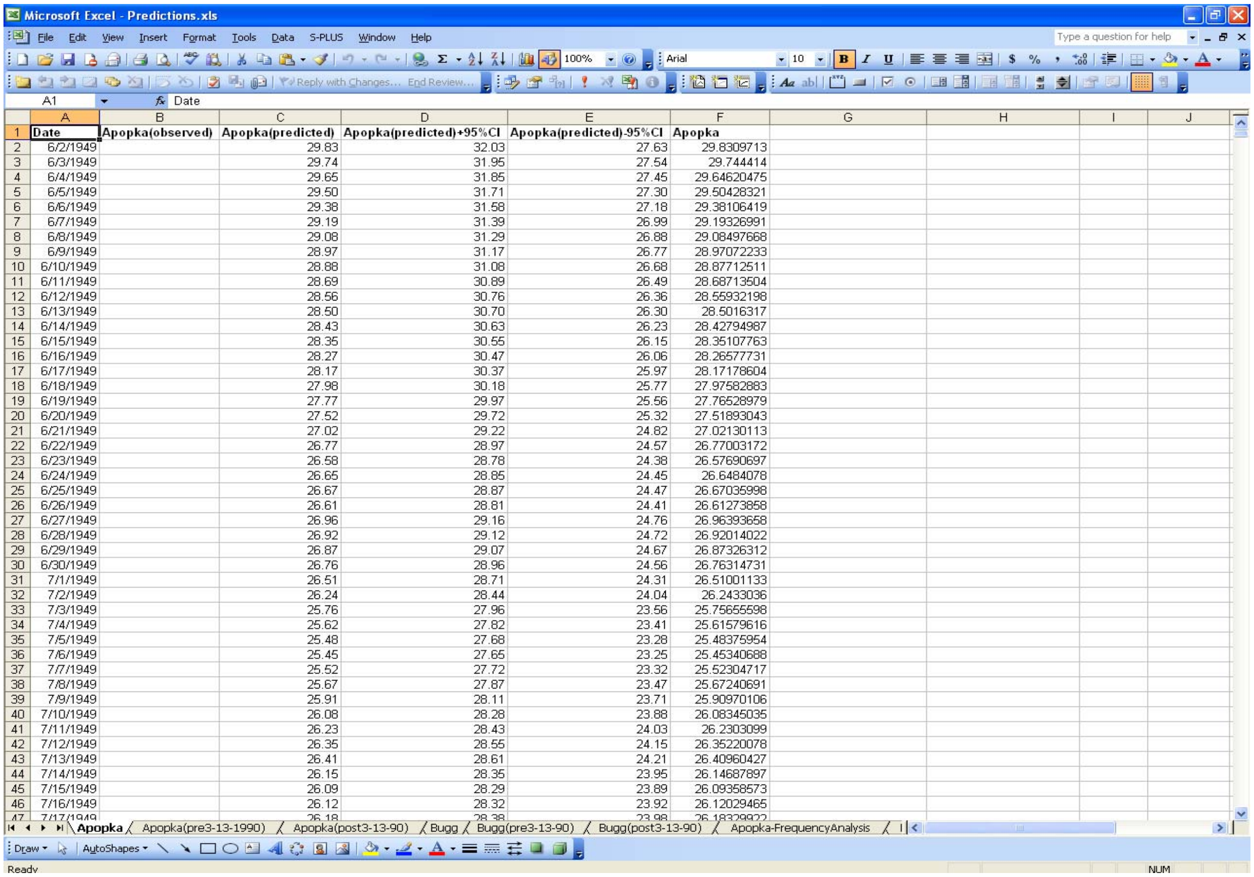
Similarly predictions and, maximum and minimum frequencies, for Apopka Spring can be obtained for any specified upper and lower date ranges. Tables **Apopka-predictions, Apopka-Frequency-District, Apopka Frequency Table-district** (shown below) are added to the database on clicking *Predict Spring Discharge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



6. Viewing prediction plots and maximum and minimum frequencies

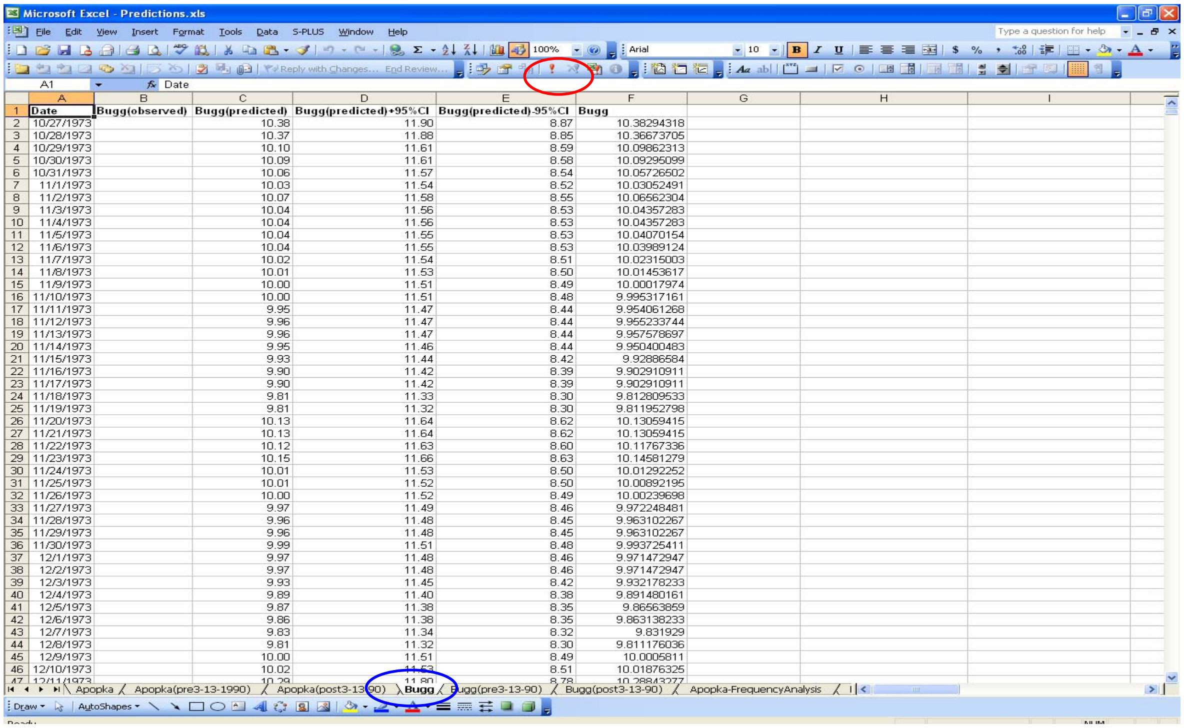
Plots of observed and predicted daily discharge data can be viewed in the EXCEL file **predictions.xls** which is linked to the prediction tables in ACCESS. The file already has been run to include daily predictions and frequencies for Apopka and Bugg springs for the complete date ranges associated with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

For our example, open **predictions.xls**. The screenshot below shows this file. By default, the *Apopka* worksheet opens up, which contains the predictions for the complete range for which daily discharge values can be computed for Apopka (6/2/1949 to 12/31/2005)



	A	B	C	D	E	F
1	Date	Apopka(observed)	Apopka(predicted)	Apopka(predicted)+95%CI	Apopka(predicted)-95%CI	Apopka
2	6/2/1949		29.83	32.03	27.63	29.8309713
3	6/3/1949		29.74	31.95	27.54	29.744414
4	6/4/1949		29.65	31.85	27.45	29.64620475
5	6/5/1949		29.50	31.71	27.30	29.50428321
6	6/6/1949		29.38	31.58	27.18	29.38106419
7	6/7/1949		29.19	31.39	26.99	29.19326991
8	6/8/1949		29.08	31.29	26.88	29.08497668
9	6/9/1949		28.97	31.17	26.77	28.97072233
10	6/10/1949		28.88	31.08	26.68	28.87712511
11	6/11/1949		28.69	30.89	26.49	28.68713504
12	6/12/1949		28.56	30.76	26.36	28.55932198
13	6/13/1949		28.50	30.70	26.30	28.5016317
14	6/14/1949		28.43	30.63	26.23	28.42794987
15	6/15/1949		28.35	30.55	26.15	28.35107763
16	6/16/1949		28.27	30.47	26.06	28.26577731
17	6/17/1949		28.17	30.37	25.97	28.17178604
18	6/18/1949		27.98	30.18	25.77	27.97582883
19	6/19/1949		27.77	29.97	25.56	27.76528979
20	6/20/1949		27.52	29.72	25.32	27.51893043
21	6/21/1949		27.02	29.22	24.82	27.02130113
22	6/22/1949		26.77	28.97	24.57	26.77003172
23	6/23/1949		26.58	28.78	24.38	26.57690697
24	6/24/1949		26.65	28.85	24.45	26.6484078
25	6/25/1949		26.67	28.87	24.47	26.67035998
26	6/26/1949		26.61	28.81	24.41	26.61273858
27	6/27/1949		26.96	29.16	24.76	26.96393658
28	6/28/1949		26.92	29.12	24.72	26.92014022
29	6/29/1949		26.87	29.07	24.67	26.87326312
30	6/30/1949		26.76	28.96	24.56	26.76314731
31	7/1/1949		26.51	28.71	24.31	26.51001133
32	7/2/1949		26.24	28.44	24.04	26.2433036
33	7/3/1949		25.76	27.96	23.56	25.75655598
34	7/4/1949		25.62	27.82	23.41	25.61579616
35	7/5/1949		25.48	27.68	23.28	25.48375954
36	7/6/1949		25.45	27.65	23.25	25.45340688
37	7/7/1949		25.52	27.72	23.32	25.52304717
38	7/8/1949		25.67	27.87	23.47	25.67240691
39	7/9/1949		25.91	28.11	23.71	25.90970106
40	7/10/1949		26.08	28.28	23.88	26.08345035
41	7/11/1949		26.23	28.43	24.03	26.2303099
42	7/12/1949		26.35	28.55	24.15	26.35220078
43	7/13/1949		26.41	28.61	24.21	26.40960427
44	7/14/1949		26.15	28.35	23.95	26.14687897
45	7/15/1949		26.09	28.29	23.89	26.09358573
46	7/16/1949		26.12	28.32	23.92	26.12029465
47	7/17/1949		26.18	28.38	23.98	26.18329922

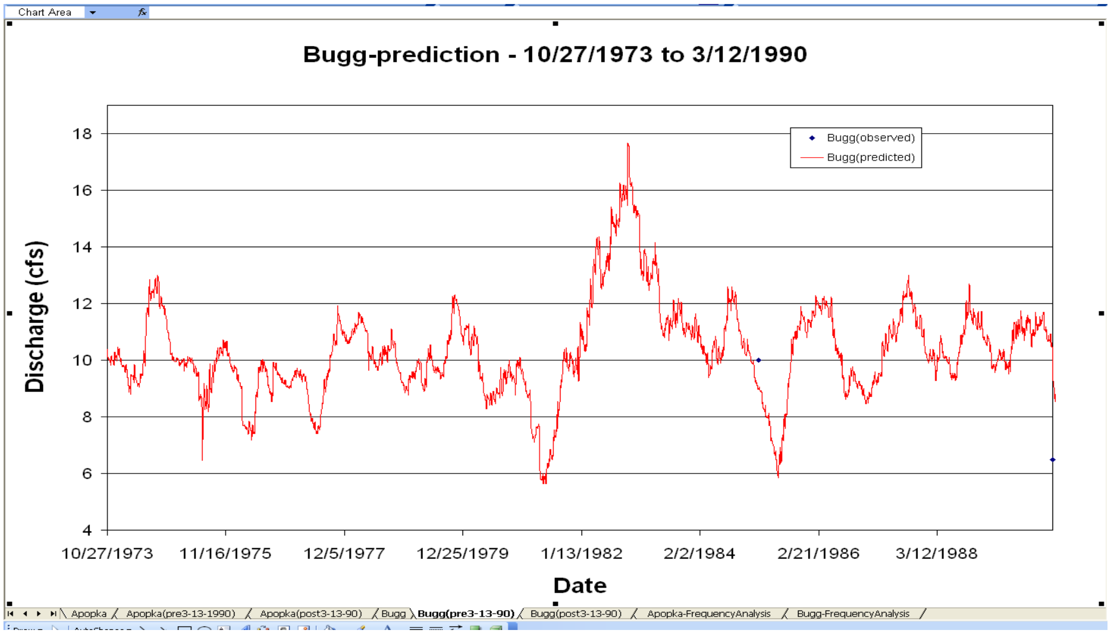
Click worksheet *Bugg* as shown below. We see the daily predictions for Bugg:



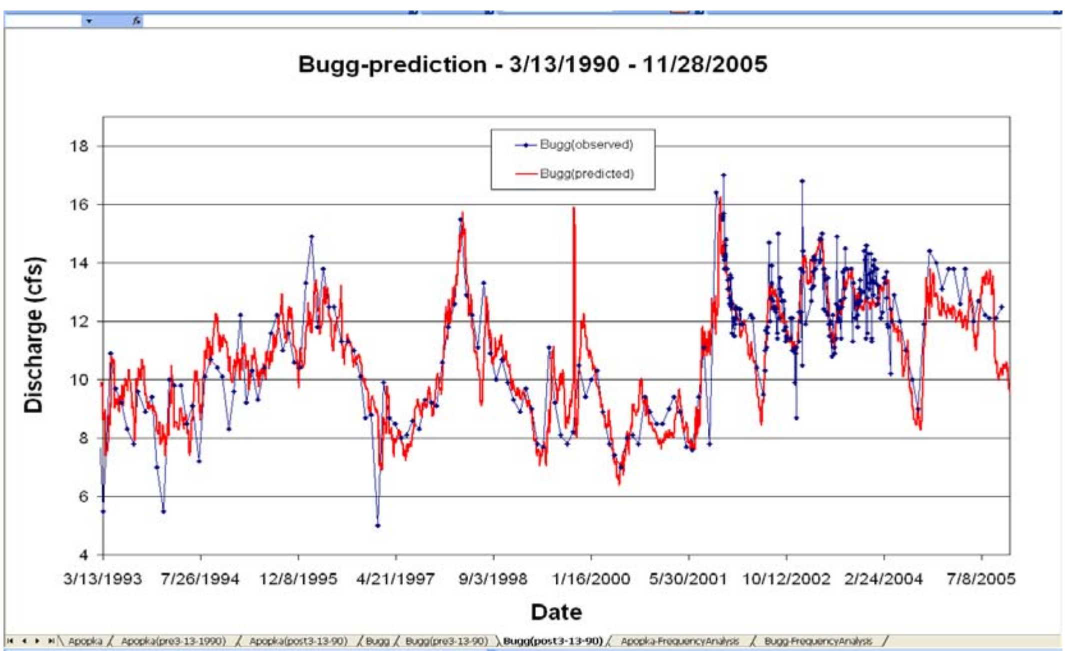
	Date	Bugg(observed)	Bugg(predicted)	Bugg(predicted)+95%CI	Bugg(predicted)-95%CI	Bugg
2	10/27/1973		10.38	11.90	8.87	10.38294318
3	10/28/1973		10.37	11.88	8.85	10.36673705
4	10/29/1973		10.10	11.61	8.59	10.09862313
5	10/30/1973		10.09	11.61	8.58	10.09295099
6	10/31/1973		10.06	11.57	8.54	10.05726502
7	11/1/1973		10.03	11.54	8.52	10.03052491
8	11/2/1973		10.07	11.58	8.55	10.06562304
9	11/3/1973		10.04	11.56	8.53	10.04357283
10	11/4/1973		10.04	11.56	8.53	10.04357283
11	11/5/1973		10.04	11.55	8.53	10.04070154
12	11/6/1973		10.04	11.55	8.53	10.03989124
13	11/7/1973		10.02	11.54	8.51	10.02315003
14	11/8/1973		10.01	11.53	8.50	10.01453617
15	11/9/1973		10.00	11.51	8.49	10.00017974
16	11/10/1973		10.00	11.51	8.48	9.995317161
17	11/11/1973		9.95	11.47	8.44	9.954061268
18	11/12/1973		9.96	11.47	8.44	9.955233744
19	11/13/1973		9.96	11.47	8.44	9.957578697
20	11/14/1973		9.95	11.46	8.44	9.950400483
21	11/15/1973		9.93	11.44	8.42	9.92886584
22	11/16/1973		9.90	11.42	8.39	9.902910911
23	11/17/1973		9.90	11.42	8.39	9.902910911
24	11/18/1973		9.81	11.33	8.30	9.812809533
25	11/19/1973		9.81	11.32	8.30	9.811952798
26	11/20/1973		10.13	11.64	8.62	10.13059415
27	11/21/1973		10.13	11.64	8.62	10.13059415
28	11/22/1973		10.12	11.63	8.60	10.11767336
29	11/23/1973		10.15	11.66	8.63	10.14581279
30	11/24/1973		10.01	11.53	8.50	10.01292252
31	11/25/1973		10.01	11.52	8.50	10.00892195
32	11/26/1973		10.00	11.52	8.49	10.00239698
33	11/27/1973		9.97	11.49	8.46	9.972248481
34	11/28/1973		9.96	11.48	8.45	9.963102267
35	11/29/1973		9.96	11.48	8.45	9.963102267
36	11/30/1973		9.99	11.51	8.48	9.993725411
37	12/1/1973		9.97	11.48	8.46	9.971472947
38	12/2/1973		9.97	11.48	8.46	9.971472947
39	12/3/1973		9.93	11.45	8.42	9.932178233
40	12/4/1973		9.89	11.40	8.38	9.891480161
41	12/5/1973		9.87	11.38	8.35	9.86563859
42	12/6/1973		9.86	11.38	8.35	9.863138233
43	12/7/1973		9.83	11.34	8.32	9.831929
44	12/8/1973		9.81	11.32	8.30	9.811176036
45	12/9/1973		10.00	11.51	8.49	10.0005811
46	12/10/1973		10.02	11.53	8.51	10.01876325

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. 10/27/1973 to 11/28/2005. The exclamation mark is highlighted by a red ellipse in the above figure.

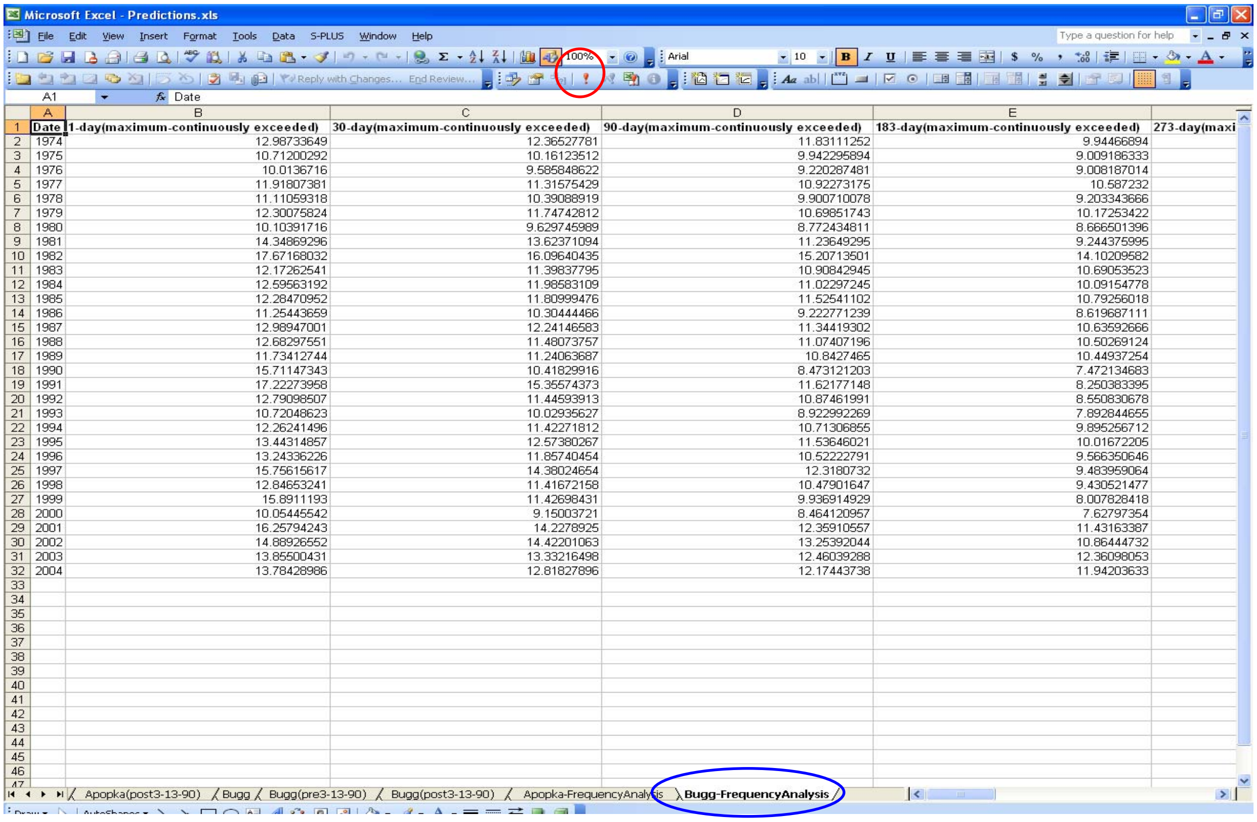
To view the plots for the above data, click on worksheet *Bugg (pre3-13-90)* for predictions before 3/13/1990 and worksheet *Bugg (post3-13-90)* for predictions from 3/13/1990. The worksheets have been highlighted in the figure above. The screenshot below shows worksheet *Bugg (pre3-13-90)*:



Also, the screenshot below shows worksheet *Bugg (post3-13-990)*:



The procedure to view maximum and minimum frequencies is similar to viewing predictions. Click worksheet *Bugg-FrequencyAnalysis* as shown below. We see the maximum and minimum frequencies for Bugg for the year range 1974-2004



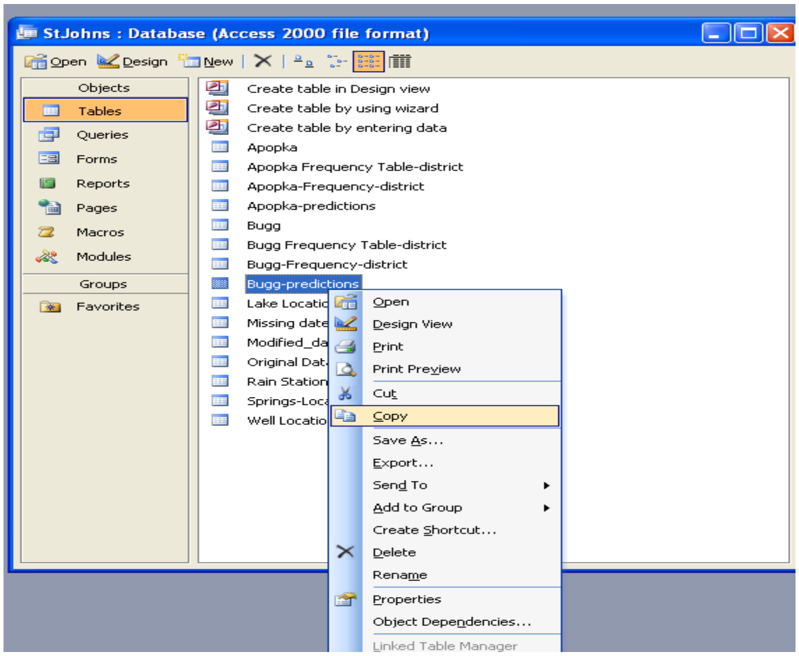
Date	1-day(maximum-continuously exceeded)	30-day(maximum-continuously exceeded)	90-day(maximum-continuously exceeded)	183-day(maximum-continuously exceeded)	273-day(maxi
1974	12.98733649	12.36527781	11.83111252	9.94466894	
1975	10.71200292	10.16123512	9.942295894	9.009186333	
1976	10.0136716	9.585848622	9.220287481	9.008187014	
1977	11.91807381	11.31575429	10.92273175	10.587232	
1978	11.11059318	10.39088919	9.900710078	9.203343666	
1979	12.30075824	11.74742812	10.69851743	10.17253422	
1980	10.10391716	9.629745989	8.772434811	8.666501396	
1981	14.34869296	13.62371094	11.23649295	9.244375995	
1982	17.67168032	16.09640435	15.20713501	14.10209582	
1983	12.17262541	11.39837795	10.90842945	10.69053523	
1984	12.59563192	11.98583109	11.02297245	10.09154778	
1985	12.28470952	11.80999476	11.52541102	10.79256018	
1986	11.25443659	10.30444466	9.222771239	8.619687111	
1987	12.98947001	12.24146583	11.34419302	10.63592666	
1988	12.68297551	11.48073757	11.07407196	10.50269124	
1989	11.73412744	11.24063687	10.8427465	10.44937254	
1990	15.71147343	10.41829916	8.473121203	7.472134683	
1991	17.22273958	15.35574373	11.62177148	8.250383395	
1992	12.79098507	11.44593913	10.87461991	8.550830678	
1993	10.72048623	10.02935627	8.922992269	7.892844655	
1994	12.26241496	11.42271812	10.71306855	9.895256712	
1995	13.44314857	12.57380267	11.53646021	10.01672205	
1996	13.24336226	11.85740454	10.52222791	9.566350646	
1997	15.75615617	14.38024654	12.3180732	9.483959064	
1998	12.84653241	11.41672158	10.47901647	9.430521477	
1999	15.8911193	11.42698431	9.936914929	8.007828418	
2000	10.05445542	9.15003721	8.464120957	7.62797354	
2001	16.25794243	14.2278925	12.35910557	11.43163387	
2002	14.88926552	14.42201063	13.25392044	10.86444732	
2003	13.85500431	13.33216498	12.46039288	12.36098053	
2004	13.78428986	12.81827896	12.17443738	11.94203633	

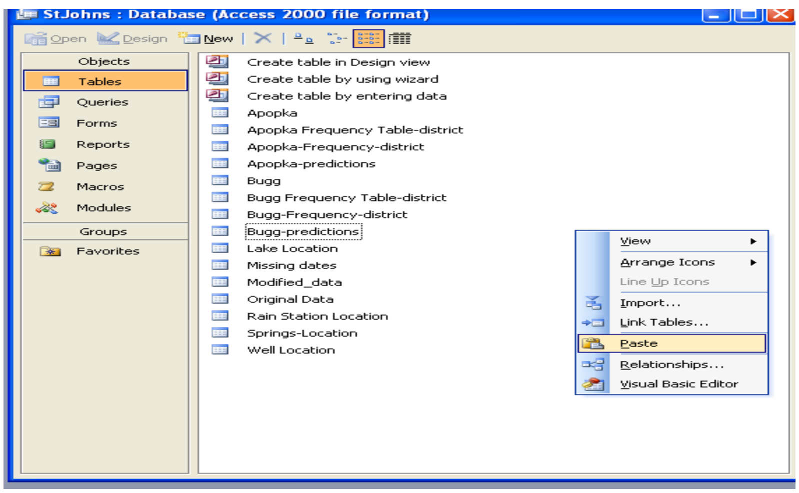
The next step is pressing the red exclamation button to refresh the frequencies for the date range which the user requested for this example, i.e. 10/27/1973 to 11/28/2005. The exclamation mark is highlighted by a red ellipse in the above figure.

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

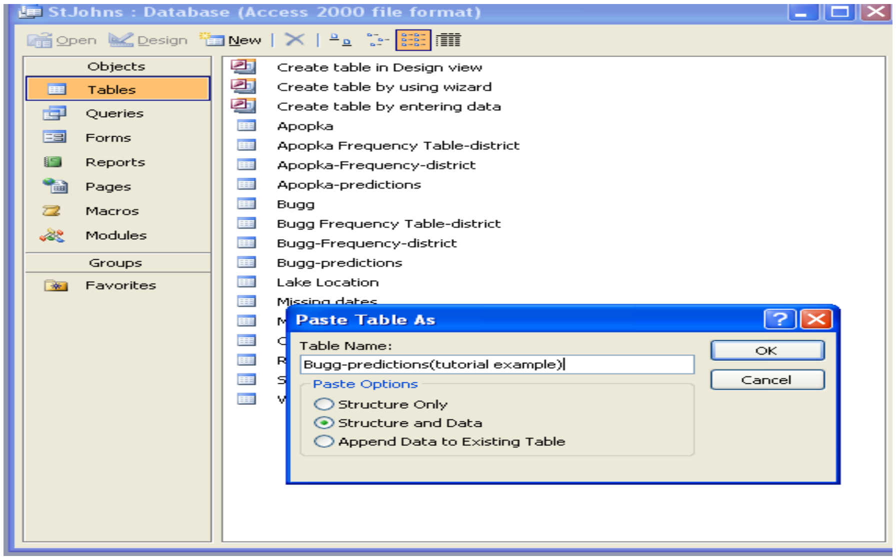
7. <u>Saving results for different cases</u>

To save the daily discharge predictions and frequencies for a particular set of well or spring data in **<u>Original Data</u>** 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 the Bugg-predictions table as shown below:

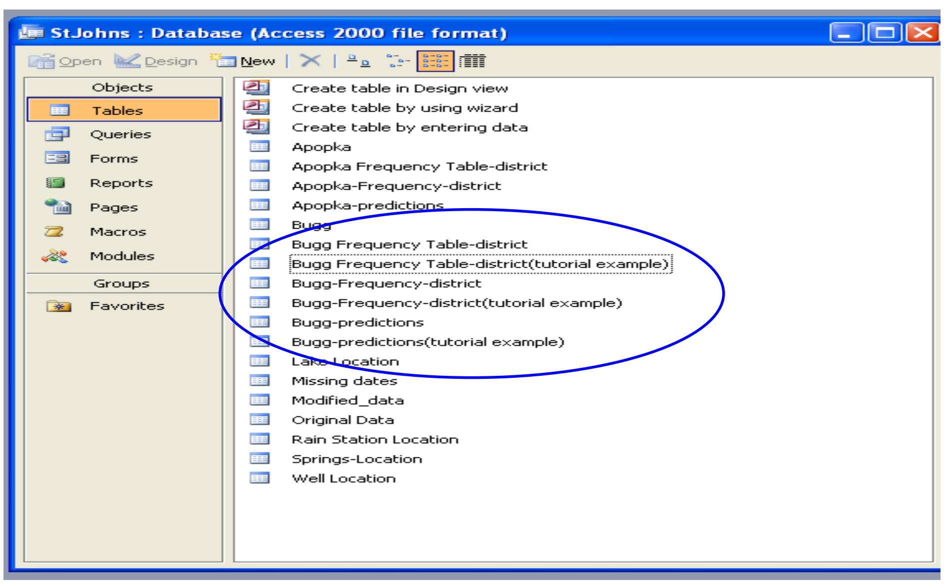




ACCESS prompts for a new name as shown below:



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



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.

NEWFIELDS

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 ith 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.

TECHNICAL MEMORANDUM

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:	July 18, 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.