LA-UR-14-20932

NMSBA Project: Canadian Watershed River Flow and Precipitation Analysis

Prepared For
Phillip Box, POC;
Franklin McCasland, Manager
Arch Hurley Conservation District AHCD

Prepared by:
CCS-3, DSA-4

December 16, 2013

Table of Contents

1. OBSERVATIONS AND FINDINGS

Key observations and findings discussed in this site down select report are listed below:

- ***The primary goal of this project*** is to determine, within an acceptable range of confidence, what has changed in the Canadian River watershed in the past 10 years resulting in lower water availability for irrigation?
- There are clearly insufficient numbers of weather monitoring stations with long term precipitation series available, especially in the middle to lower watershed; for example Cimarron watershed has the largest number of weather stations but it contributes the smallest river flow volume into Conchas reservoir.
- Predictive linear regression models were constructed that were used to analyze all of the available predictors at the same time rather than perform simple pair wise correlations. This approach provided more insight into the causal factors influencing river flow
- "R-square or R^2" value is a measure of the variability explained by a regression model. In the first model, R^2 was estimated to equal 91%. In the second model, R^2 was estimated to equal 98%.
- ***A major conclusion from this analysis is:*** using R^2 as a measure of fit, both regression models explain the majority of observed variation in the river flow data. This does not preclude changes in other factors over the last 10 years from a minor influence.
- Based on these results, no major changes in the relationships among precipitation and river flows have occurred in the past 10 years as compared to the previous 30 years for the areas bounded by the rectangles shown in Figure 3.
- Approaches that could improve the regression models are:
  - Exploration and down-weighting of outlier data points.
  - Combining multiple regression models into a single graphical model, which represents dependencies among the individual models.
  - Adding hydrologic models such as Voronoi to better distribute rainfall over the watershed.
- It is clear that future analysis should be focused instead on ***adaptation*** measures. This could include method improvements in the predictive model, consideration of management practices in Conchas Reservoir as well as inclusion of climate forecasts derived from Global Climate Models or GCMs.

2. BACKGROUND

Los Alamos National Laboratory (LANL) was tasked to support Tucumcari, NM area small businesses, based on the following scope of work:

> "Using historical data, establish a baseline for the Canadian River watershed inflow of runoff and stream infiltration. Outline baseline statistics, input datasets formatted for computer models, identify causal factors which result in changes to irrigation water levels, and delineate factors that are not associated with drought conditions. A predictive regression model will be created to establish the statistical relationship between monthly rainfall and river flow. A final deliverable consisting of a demonstration of the forecast model, discussion of methods used and an issues summary with recommendations will be provided to the Requester at the completion of the project."

The following Appendices are attached to this report:

- Appendix A- Plotted Factor Data Series
- Appendix B- River Flow Correlation Tables
- Appendix C- Impoundment Examples
- Appendix D- Regression Analysis- Overview
- Appendix E- Glossary of terms and acronyms

3. PROJECT GOALS

In March, LANL staff attended a kickoff meeting with project sponsors to review preliminary results from a preliminary statistical correlation of factors such as rainfall, river flow and allocations to Arch Hurley Irrigation District (AHCD). Since the initial analysis was focused on Conchas Reservoir and outflow to AHCD, it was viewed as being less valuable in terms of broader watershed issues. The project scope of work was reviewed, and all present agreed to enlarge the area of analysis northward beyond the inlet of Conchas. It was proposed that the entire Canadian River watershed be analyzed, since it feeds Conchas. Unfortunately only two USGS river gages are available to provide flow data, Sanchez and Taylor Springs. However, the process used to derive initial results was ultimately applied to a broader watershed analysis, extending into the Upper Canadian watershed.

The primary goal of this project is to determine, within an acceptable range of confidence, what has changed in the Canadian River watershed in the past 10 years resulting in lower water availability for irrigation? This analysis would succeed if the following pattern could be observed: river flow and precipitation are strongly correlated in prior to the year 2000 but weakly correlated later- this indicates a possible change of causation due to drought and non-drought factors. Identification of causal factors such as stockyard tanks (also including dirt or earth ponds) and growth of salt cedar or related brush growth that cannot be easily controlled are considered to be possible issues other than drought. AHCD water rights clearly do not apply to this study, so that causal factor was eliminated.

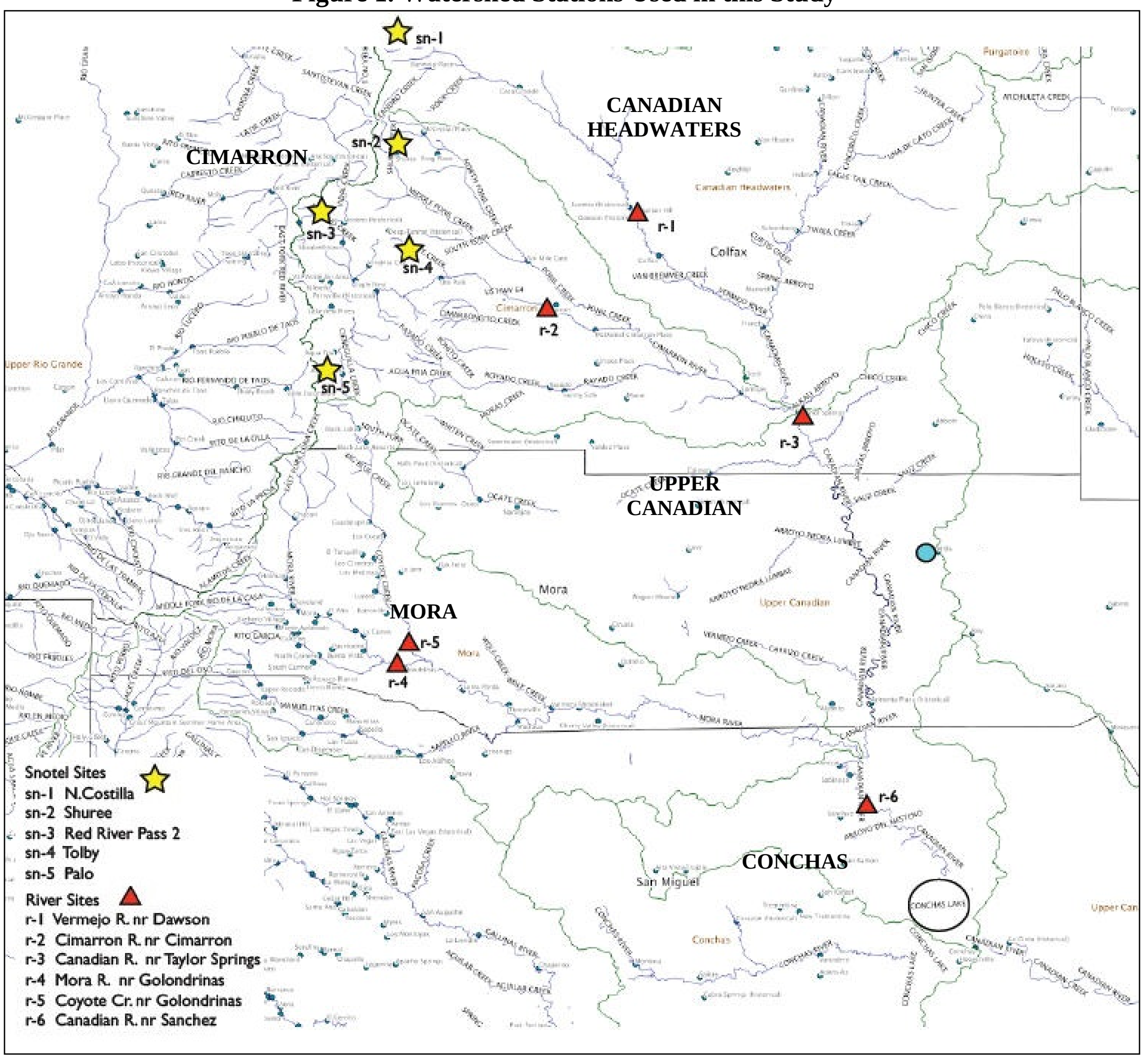
4. WATERSHED DATA SOURCES

USGS maintains six river stations which provide flow data at varying degrees of detail and historic accuracy:

- Vermejo River near Dawson
- Cimarron River near Cimarron
- Canadian River near Taylor Springs
- Mora River near Golondrinas
- Coyote Creek
- Canadian River near Sanchez

Five NCRS weather stations served as the primary source of rainfall data in the upper watershed. Figure 1 displays the location of all stations incorporated into the initial phase of this study.

Figure 1. Watershed Stations Used in this Study



There are clearly insufficient numbers of weather monitoring stations with long term precipitation series available, especially in the middle to lower watershed; for example Cimarron watershed has the largest number of weather stations but it contributes the smallest river flow volume into Conchas reservoir. Figure 1 displays an additional weather station (blue icon) located at Mills Canyon which provided precipitation data only for 2010 to 2013. The remaining stations provided data series extending from 1991 to 2013, or longer periods.

5. HYDROLOGIC PARAMETERS

There are many hydrologic factors that may affect the validity of a statistical correlation analysis based on a limited set of data points. These include, but are not limited to, human-induced changes in runoff (e.g., diversions, storage facilities) and soil conditions during precipitation periods affecting runoff. In addition, hydrology is defined by topographic features, such as watersheds. Understanding how a limited set of precipitation gages represents a wide area potentially crossing multiple watersheds is important in developing accurate statistical correlations.

Precipitation gages are generally part of a sparse network where measurements are taken at a single point in space. These point measurements are often spatially interpolated using geometric techniques to create a two-dimensional representation of precipitation. A challenge in making statistical correlations of point precipitation data e.g., rain gages with stream gages is determining the proper weighting or distribution of precipitation points. This spatial interpolation[1] is used to determine the relative contribution of precipitation measured at a point to the stream gage. A method of spatial interpolation widely used in hydrologic modeling and simulation is the Voronoi Method.

The Voronoi method is a geometric algorithm used to divide watersheds into regions. This method uses a set of points (e.g., rain gages) as a starting point. From the points, a region is defined which consists of space closer to the initial point than any other point. Using the two-dimensional regions derived from the Voronoi method, the contribution of precipitation of measured precipitation at gages can be more accurately divided up by watershed boundaries. This approach could potentially strengthen the outcome of future statistical correlations which are refinements beyond the results described in this report.

6. SANCHEZ -CONCHAS RESERVIOR AND AHCD CANALS

The area extending from Sanchez station to Conchas reservoir to laterals in AHCD's canal distribution system served as LANL's initial focus for analysis. A variety of data correlations were attempted in order to potentially identify causal factors of interest. 11 data variables were chosen for this analysis, 6 variables relate to some aspect of Conchas operation such as reservoir storage, releases to AHCD and elevation changes. The other 5 variables relate to flows at Sanchez station or Tucumcari's rainfall.

In order to avoid inconsistency among data types during this comparison, all factor values were "normalized"[2]. Initially, a variety of graphs were created to plot year-to-year changes among

[1] Interpolation is a method of constructing new data points within the range of a discrete set of known data points.
[2] Normalized values were calculated by setting the minimum value to 0 and the maximum value to 100 (with values in between set proportionately) for each data series.

variables that might indicate correlations or relationships between two or more variables. The figures created were (see Appendix A):

- Figure A-1 plots compares Conchas Reservoir levels with Tucumcari rain and Sanchez River flows between 1971 and 2012 (showing the lowest values in the mid-70's and from 2010 onwards).

- Figure A-2 compares Conchas Reservoir levels with release flows and AHCD allocations between 1991 and 2012. ***Note:*** Army Corps of Engineers has stated that data before 1991 was less reliable.

Legends in the plots include variable description, short variable name, and range of values for the variable in the plot i.e. minimum and maximum. After assembling these data series, correlation coefficients were calculated. Correlation coefficients are pair-wise comparisons of variables and range between 0 and 1. Any correlation coefficient greater than or equal to .05 has a 95% chance of being significant. However, the higher the number, the greater the significance of the correlation. For this study coefficient values less than 0.50 are ignored since the degree of correlation is difficult to assess, in terms of significance.

Some factor coefficients need further explanation; for example C_Inflow (Conchas inflow) and C_AHCD_Rel (release to Arch Hurley) are only moderately correlated. This probably indicates other factors such as pool storage prevent this correlation from reporting higher values. There are correlations that indicate strong relationships between important factors; for example C_Storage (pool storage) and C_AHCD_Rel (release to Arch Hurley) report a highly significant coefficient of 0.91. Finally, some factors indicate a significant non-correlation, in this case Tucumcari's rainfall is shown to not be significantly correlated with any factors.

This correlation indicated a high correlation coefficient[3] between AHCD allocations (or Conchas average release to AHCD) versus Conchas storage level, Conchas average elevation or Conchas January elevation. The correlation coefficients between AHCD annual allocation versus the average river flow at Sanchez, the January river flow at Sanchez, and inflow to Conchas Reservoir were lower at 0.55, 0.86, 0.60 respectively but still significant[4]. AHCD allocations are clearly highly correlated with Conchas Reservoir inflows and storage levels as well as to river flow at Sanchez.

7. CANADIAN WATERSHED AND PRECIPITATION

12 data variables were initially chosen for analysis of watershed and precipitation factors, six variables relate to river flow and six variables relate to precipitation within the interconnected watersheds. In general, this analysis did not reveal any significant, consistent correlations between river flow and precipitation. This result may be due in part to the relatively sparse location of weather sampling stations; these stations do not provide adequate coverage for the

[3] Greater than 0.9; this value for a sample size of 42 (the number of years analyzed) is highly significant i.e. the probability that there is no correlation is virtually zero.
[4] A correlation coefficient of 0.55 for this sample size indicates there is a probability of 0.0001 that there is no correlation.

watersheds involved. Also timing lags between rainfall or snow and infiltration into river flows is probable, measured in months. In order to account for this factor, substantial re-analysis of the data would be required. Finally, the data series represented a limited duration (October 2010 to September 2012) which was the only period reporting a complete data series for analysis of both flow and precipitation data.

In order to correlate watershed factors over the period 1991 to 2010, LANL then chose to only analyze river flows. A correlation table was developed for each five year period, for example 1991 to 1995, 1996 to 2000 and so forth. This analysis approach was intended to reveal any major changes in flows over time. Table 1 summarizes seven variables chosen for the second stage of this analysis.

Table 1. Data Variables of Interest

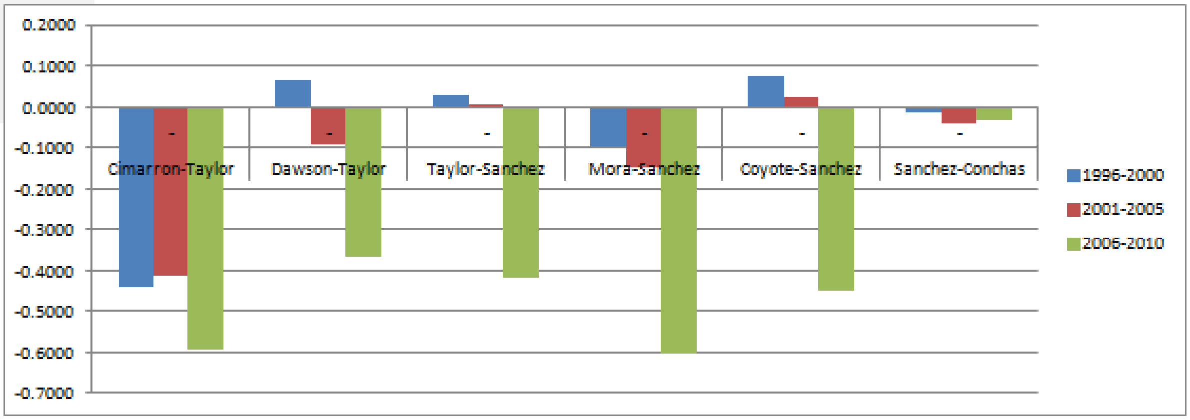
Factor	Description	Units	Source
dawson_cfs	Average flow of Canadian River near Dawson NM	Cubic feet per second	http://nwis.waterdata.usqs.gov/nwis
taylor_cfs	Average flow of Canadian River near Taylor Springs NM	"	http://nwis.waterdata.usqs.gov/nwis
cimarron_cfs	Average flow of Canadian River near Cimarron NM	"	http://nwis.waterdata.usqs.gov/nwis
mora_cfs	Average flow of Canadian River near Golondrinas NM	"	http://nwis.waterdata.usqs.gov/nwis
coyote_cfs	Average flow of Canadian River near Golondrinas NM	"	http://nwis.waterdata.usqs.gov/nwis
sanchez_cfs	Average flow of Canadian River near Sanchez NM	"	http://nwis.waterdata.usqs.gov/nwis
conchas_inflow_cfs	Average calculated inflow at Conchas Reservoir (	"	US Army Corps of Engineers operational records
Note: Data was either averaged over a month as appropriate			

Figure A-3 (see Appendix A) plots the entire data series used in this analysis. Five year periods are also plotted in Figures A-4 to A-7; note that vertical scale varies on these plots which allows month-to-month variations to be more easily observed. A total of 60 months' data is included, however not all data was considered usable. Only those months reporting flow data for all seven stations simultaneously were actually included in the analysis[5]. As a result, no five year period reported more than 51 months of data, and the last period reported only 36 months of data.

Tables B-1 (see Appendix B) summarizes all correlation coefficients estimated for this set of factors, for each five year period. Table B-2 summarizes cumulative changes in correlation between flow stations of interest. Data listed in Table B-2 is also plotted in Figure 2's bar chart.

[5] The number of observations with values for all flow rates for the 5 year periods were: 1991-1995: 47; 1996-2000: 51; 2001-2005: 51; 2006-2010: 36

Figure 2. Changes in Correlation: Watershed River Flow



In this analysis, 1991-1995 is treated as the baseline period, and all changes are measured relative to that time. The findings of interest and observations related to Figure 2 are:

- Sanchez-Conchas flow indicated no significant change from baseline for any period.
- Taylor-Sanchez flow indicated no significant change from baseline until the 2006-2010 period; the observed change (green bar) is statistically significant.
- Dawson-Taylor flow indicated no significant change from baseline until the 2006-2010 period; the observed change (green bar) is statistically significant.
- Mora-Sanchez and Coyote-Sanchez flow indicated no significant change from baseline until the 2006-2010 period; the observed changes (green bar) are statistically significant.
- Cimarron-Taylor flow indicated a large change from baseline in all periods; since the plotted changes are cumulative values, a change in flow occurred in 1996-2000 which has been maintained consistently since that period.
- Since Cimarron's flow is generally only a relatively small portion of total flow into Conchas, changes in Cimarron's watershed cannot account completely for the relatively large, sudden change in Taylor-Sanchez flow during 2006-2010.

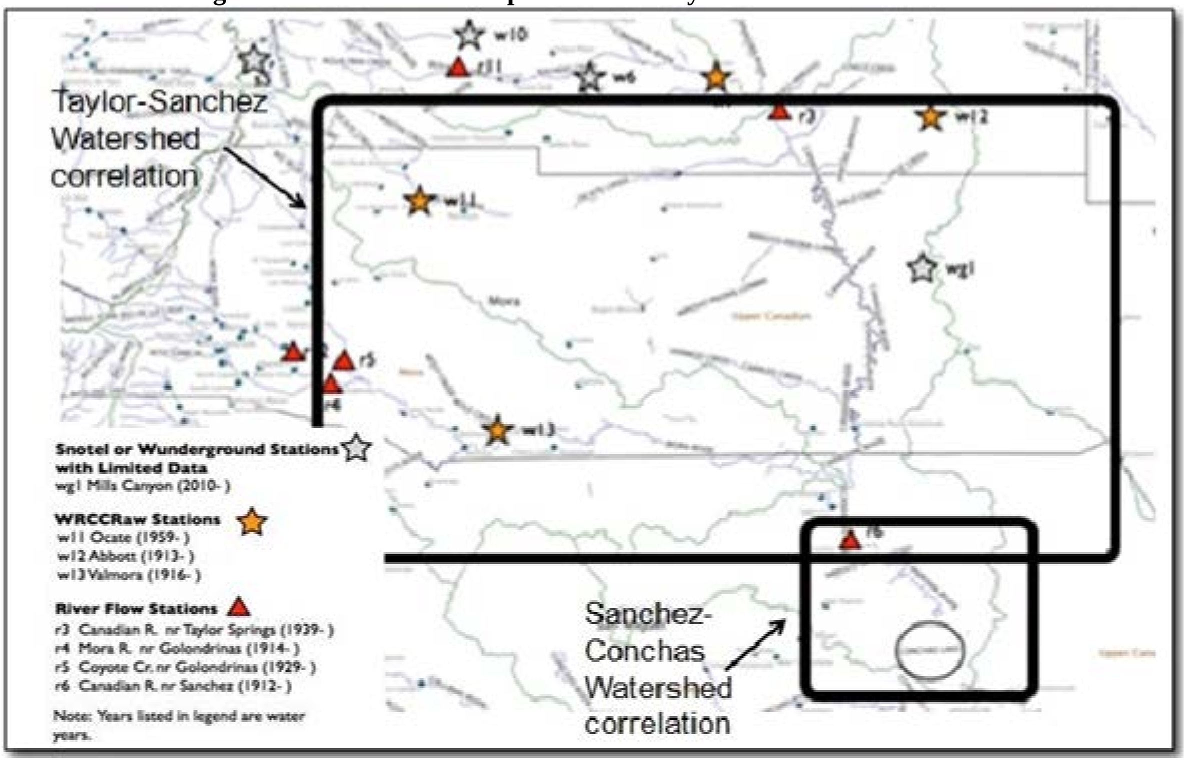
The effect of unequal sampling during each five year period is a potential issue. For example the 2006-2010 period only reported 36 of 60 possible sample months. Does this impact the finding stated above? To address this issue, a measure of "signal-to-noise" or S/N in the flow data was estimated by calculating average flow divided by standard flow deviation for each month in the reported series; this result indicated large, inconsistent changes appear to occur throughout the entire data series. However, when evaluated for each five-year sampling period, a fairly stable pattern was observed. A general trend to lower S/N ratios occurred from 1991 to 2005, then ratios increased. This suggests fewer sampling months in the last interval 2006-2010 cannot completely account for the sudden drop in correlation coefficient values plotted in Figure 2.

It is important to recognize that analysis of river flow data alone, as presented in this section, cannot provide sufficiently detailed results to exclude drought as being the primary cause of observed changes. Therefore, effort was undertaken to locate additional rainfall data which fills previous data gaps. The next section describes how this process was conducted.

8. PREDICTIVE FLOW/PRECIPITATION MODEL

LANL's watershed analysis was expanded beyond Sanchez into the upper Canadian and Headwaters areas and granularity was increased to utilize monthly data; Additional data from river flow and precipitation stations was located as shown in Figure 3[6].

Figure 3. Water Flow/Precipitation Monthly Data Used in Correlation



This approach could uncover changes in the relationships among flows and precipitation in recent years. Unfortunately, river flow and precipitation data is not available at the density and distribution needed to conduct this analysis with high confidence.

Data sources include river flows from the US Geological Survey (USGS) and precipitation from the WRCC, and Snotel (US Department of Agriculture). Some reporting stations archive very limited data series. For example, a subset of Snotel stations started recording data in 2010. However, LANL processed complete monthly data series from 1970-2012 for ten USGS river flow stations and ten WRCC precipitation stations upstream of Sanchez. To be conservative,

[6] Obtained from the WRCC Western Regional Climate Center database, http://www.wrcc.dri.edu/.

data from months with missing data values was excluded. In most cases, the excluded months were missing all or most days suggesting a failure in the measurement equipment. Monthly river flow values (unit: cubic feet per second cfs) were estimated by calculating the average of hourly flow rates. Monthly precipitation values (unit: inches) was estimated by calculating the sum of daily precipitation values.

With this expanded dataset, predictive linear regression[7] models were constructed that was used to analyze all of the available predictors at the same time rather than perform simple pair wise correlations as before. This approach provided more insight into the causal factors influencing river flow. In particular, it was used to estimate percent variability explained by the model and to predict the value of the response variable (river flow) for arbitrary values of the explanatory variables (precipitation) within the range of values covered by the original dataset.
Two linear regression models were created, as described below.

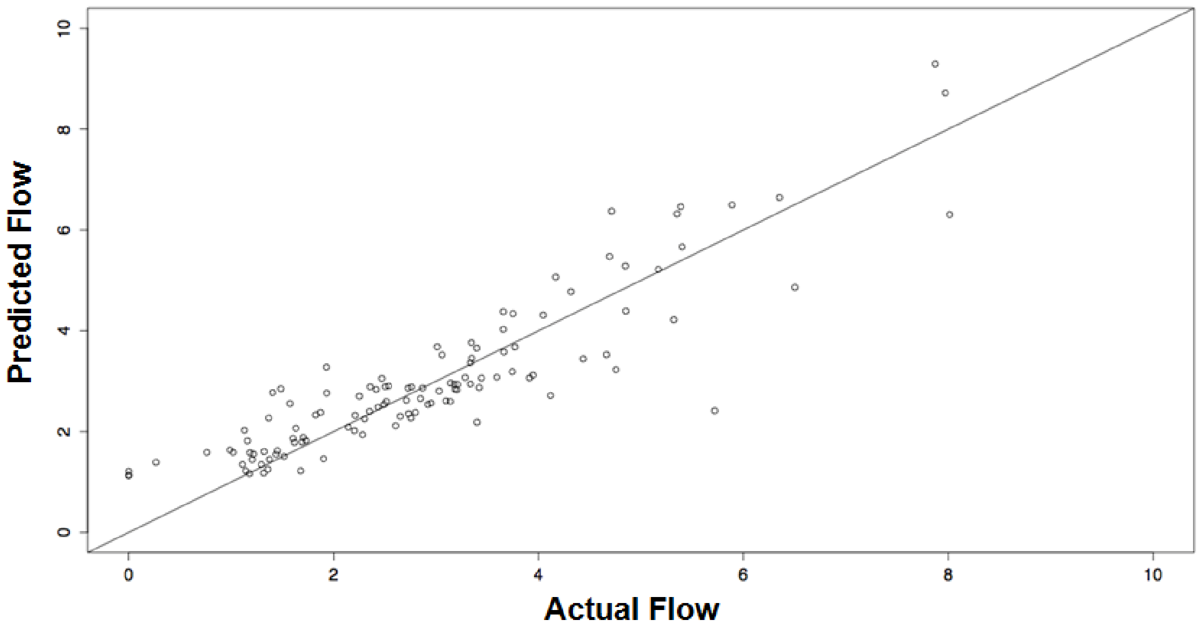
- The first model is represented in Figure 3 by the lower rectangle
  - Response variable: calculated inflow of water to Conchas Reservoir
  - Explanatory variable: Canadian river flow near Sanchez.
- The second model incorporates more data from upstream areas and is more complex, represented in Figure 3 by the upper rectangle
  - Response variable: Canadian river flow near Sanchez
  - Explanatory variables: three river flow stations (Mora River near Golondrinas, Coyote Creek near Golondrinas, Canadian River near Taylor Springs) and three WRCC precipitation stations (Abbott, Ocate, and Valmora) upriver from Sanchez..

Both regression models were derived from monthly values from 1970-2002. LANL used this data to predict the river flow using the historical values of precipitation from 2003-2012, then compared the predicted versus actual values. This method measures the accuracy of “fit” i.e. how well data from the past ten years fits a model based on the previous thirty years. This approach can potentially be used to determine whether or not it is likely that a major change in relationship between upstream precipitation and Canadian River flows has occurred. For the second model, exploratory data analysis and an examination of the regression residuals[8] led to modifications of the model in two ways, using a hydrologic perspective[9]. This model is not based on a complete hydrologic representation describing how rainfall is distributed over the geographic area or how rainfall is transported to a particular body of water based on distance and other factors (vegetation, soil type, etc).

[7] A regression model mathematically captures the relationship between a response variable e.g., flow rate at location x and multiple explanatory variables such as river flow rates and precipitation directly upstream of location x.
[8] The residual of an observed value is the difference between the observed value and the estimated value.
[9] First, river flow values were scaled by the cube root based on the observation that river flows are measured in cubic feet per second while precipitation is measured linearly in inches. Second, a cutoff value for precipitation was implemented thus setting very low precipitation values to zero; this is based on the observation that light precipitation results in moisture which either evaporates or not transport to the river.

When the second model is used to predict Canadian river flow rates at Sanchez from 2003-2012, a very accurate match to actual flow is obtained (see Figure 4).

Figure 4. Predicting Sanchez Flow (2003-2012) Using Regression Model



"R-square or R^2" value is a measure of the variability explained by a regression model. In the first model, R^2 was estimated to equal 91%. In the second model, R^2 was estimated to equal 98%. The intercept[10] and Ocate/Abbott precipitation variables were statistically insignificant. ***A major conclusion from this analysis is:*** using R^2 as a measure of fit, both regression models explain the majority of observed variation in the river flow data .

See Appendix D for additional discussion of these findings, including a description of the regression analysis method, results from the second model and a baseball statistic example.

Based on these results, no major changes in the relationships among precipitation and river flows have occurred in the past 10 years as compared to the previous 30 years for the areas bounded by the rectangles shown in Figure 3. At the upstream end, this area includes Taylor Springs in the Upper Canadian watershed and the Mora River to Golondrinas in the Mora watershed. These results do not include the Cimarron River watershed or the Canadian Headwaters watershed, except by proxy of the river flow at Taylor. Adding the Cimarron River and Canadian Headwater watersheds would include areas where precipitation in the form of snow is more prevalent; this complicates the analysis due to time lags in the release of snow and ice to water flows[11].

[10] The intercept is a point where the graph intersects with the *y*-axis; as such, these points satisfy *x*=0

[11] Time lag analysis might also improve the fit of the current regression models but this feature was not been incorporated into this study.

At the downstream end, the regression models include inflow to Conchas Reservoir. For AHCD, annual water allocations from Conchas are most important. The correlation between annual AHCD allocations and average annual inflows, calendar year average elevation, water year average elevation, and calendar year average reservoir storage at Conchas Reservoir from 1971 to 2012 was re-estimated. Storage and elevation are highly correlated with AHCD allocations, indicating a high correlation of 0.95. However, average inflow to Conchas Reservoir has lower correlation of 0.69. These numbers were obtained by using only those data points where all values are available.

The lower correlation of inflow to allocation is probably credible since below a certain level of water in Conchas, no water can be released- the allocation would be zero. Storage measurements indicate how much water is above this minimum level and available for release. Therefore, successive years of water levels below the minimum may require a longer period of increased inflow to have a significant impact on storage. The correlation between allocations and the average inflow over various years was then evaluated. For this calculation, average release flow rate (unit: cfs) weighted by the number of months when water was released to AHCD was substituted for the allocation (unit: inches).

Correlation coefficients for release versus inflow were averaged over 1, 2, 3, 4, or 5 years for the period between 1991 and 2012 when release flow rates were available. The best correlation (0.93) was obtained when release rate was correlated with inflow for three years. This result suggests unexplained variability in the regression model relates directly, with high confidence, to changes in watershed flow, inflow and release from Conchas. Further, unexplained variations are roughly proportional: that is, a given percentage variation in watershed flow will result in approximately the same percentage variation in releases.

9. POSSIBLE CAUSAL FACTORS

Causal factors for unexplained variations in river flow could include several possible effects, including stockyard tanks also called dirt or earth ponds. Landowners have been encouraged to impound water up to 10 acre feet in volume or a 200x200 foot footprint. The federal “Equip” program provided funding in some cases, the main driver was to capture surface water runoff. If not in a designated watershed, the acre-foot limit cannot be enforced, there is also no formal permitting for these tanks. Other possible causal factors are growth of salt cedar or related brush growth that cannot be easily controlled.

Any of the dirt tanks that are built with cost share money from the USDA through the NRCS would utilize between 1 acre-foot to 10 acre-foot storage capacity. These are built to water livestock and wildlife, not accomplish flood control. Anyone who builds a water storage impoundment and without a permit would be limited only by cost, an ongoing concern of some Arch Hurley district constituents. The impoundments that include dams built across a stream bed or a diversion of a stream are of particular concern. Appendix C discusses specific examples of impoundments that were identified during this study. Using imagery, only playa lakes can be easily identified. However a ground level survey did identify candidate impoundments of interest.

A simple pond model can be constructed that accounts for rainfall, evaporation and infiltration. However, infiltration and evaporation are difficult to estimate due to varying hydrologic and meteorological factors. If these factors are ignored, the Taylor to Sanchez drainage area (assuming one mile infiltration buffer on each side of the Canadian River) could impound approximately 1,000 acre-feet or more of water annually. This estimate assumes:

- Average pond capacity of ten acre-feet, ten ponds per square mile
- A total buffer size of 120 square miles[12].
- An improbable count of 1,200 ponds within the buffer area.
- The average river flow deficit is estimated to equal 7.6 cfs in this case.
- In general, fewer ponds or lower pond capacity causes proportionately lower river flow impacts.

Several additional steps need to be completed to obtain sufficient data to eliminate dirt ponds as possible causal factors. The first step would be to contact the NM State Engineers Office to obtain a list of all types of water diversion and storage structures that have been permitted within the last twenty years. Then, determine how much additional water impoundment has been permitted. This data could answer a variety of questions such as:

- Are there any impoundments not permitted?
- How many irrigation and domestic wells have also been permitted in this time frame?
- How many circle sprinklers have been constructed with in this same time period? Are they replacing other irrigation methods or is new land being brought into production with irrigation water?

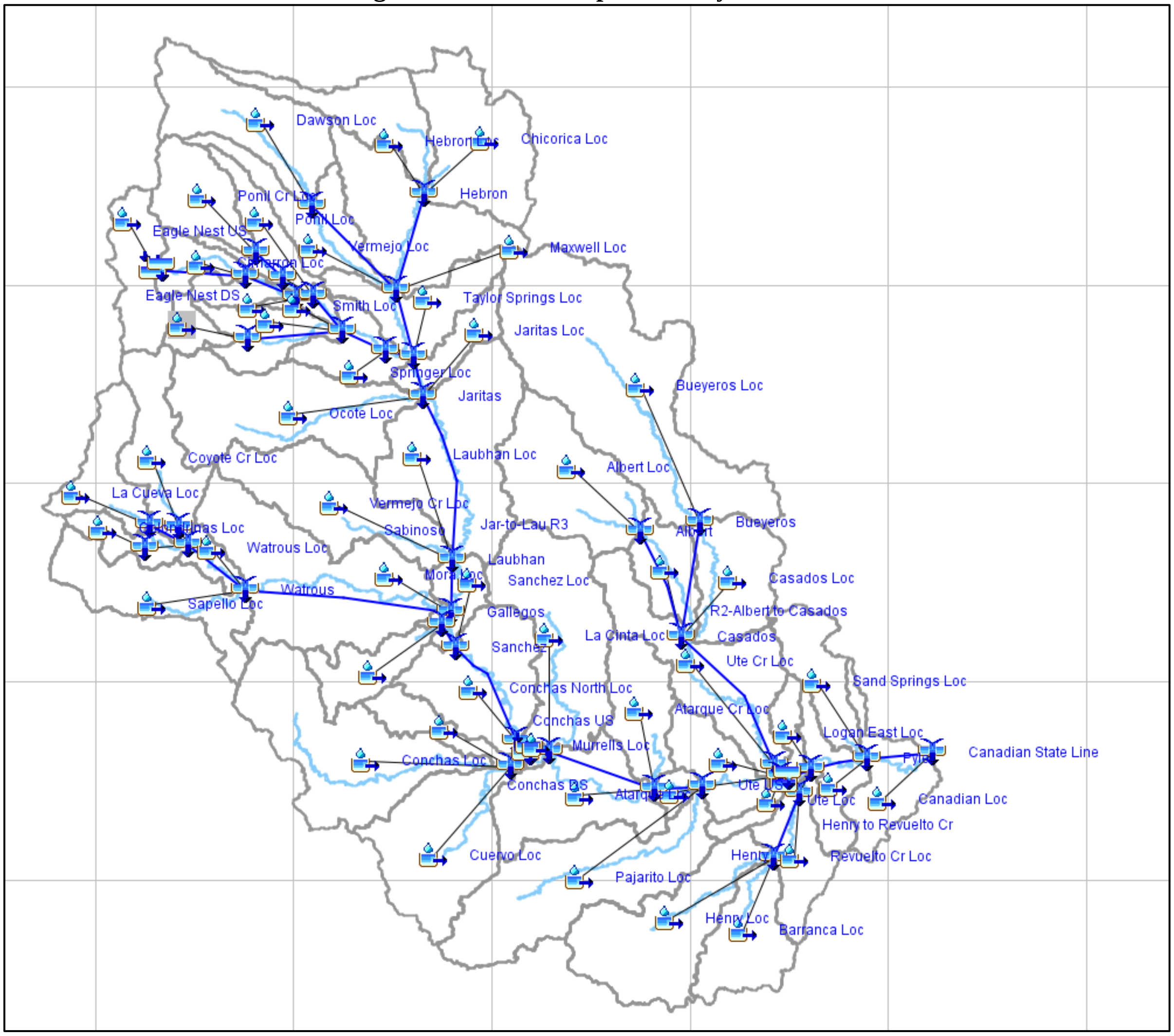
10. USACE WATERSHED MODEL (HMS)

Computer models are representations of real-world systems. These models can be used to simulate historical and future events based on model inputs, such as precipitation and temperature. The United States Army Corp of Engineers (USACE) provided LANL with a hydrologic model of the Canadian River Watershed. This model was developed using their Hydrologic Modeling System (HMS), a standard modeling approach used nationally and internationally. The challenge in using hydrologic models is ensuring that the model is adequately calibrated to represent actual conditions throughout the watershed. These models use many parameters to represent infiltration and runoff processes and must be calibrated against a range of hydrologic conditions.

The HMS model extents shown in Figure 5.

[12] Valmora's ten-year average rainfall from 2002-2012 was assumed to provide pond inflow, or 10.3 inches per year. Pond collection areas could be significantly larger than assumed which would increase inflow of water actually stored.

Figure 5. HMS model provided by USACE



The model shown in Figure 5 represents fifty sub-watersheds throughout the Canadian River Basin. In this case, USACE indicated the model was primarily developed for high-flow discharge events e.g., spring snowmelt or torrential rainfall. The model is not calibrated to accommodate low-flow events such as those common during period in which water allocations are likely to occur. In its current state, this model cannot be used as a forecasting model for AHCD.

To expand the HMS model to include low-flow events, loss functions such as infiltration and surface roughness must be estimated for each of the sub-watersheds. This process requires gathering additional data mainly soil moisture, land cover, and infiltration parameters. It would represent a substantial data gathering effort and significant data requirement by volume. Ideally, the process would be accomplished in collaboration with the USACE. While it would not solve water availability concerns, an HMS model could be used to forecast AHCD water availability and assess what-if scenarios for future conditions.

11. CONCLUSIONS/FUTURE TASKS

LANL was tasked to evaluate within an acceptable range of confidence, what has changed in the Canadian River watershed in the past 10 years resulting in lower water availability for irrigation? Identifying any non-drought causal factors that could be favorably influenced as a ***preventative*** measure in the near-future could lead to successful outcomes. However, given the conclusions stated below, it is clear that future analysis should be focused instead on ***adaptation*** measures. This could include method improvements in the predictive model, consideration of management practices in Conchas Reservoir as well as inclusion of climate forecasts derived from Global Climate Models or GCMs. The latter approach would extend the time frame of analysis forward, as opposed to analyzing historic data trends.

Important conclusions were identified from the development and application of predictive regression models during this study, namely:

- The vast majority of variation in the inflow to Conchas Reservoir and the flow rates on the Canadian River near Sanchez between 1970-2002 can be explained by the regression models.
- The prediction of river flow near Sanchez for 2003-2012 using the 1970-2002 model fits the actual data very closely.
- The relationship among water flows and precipitation for the analyzed area of the watershed has not changed over the past 10 years with respect to the previous 30 years. ***This does not preclude changes*** in other factors over the last 10 years from a minor influence.
- Approaches that could improve the regression models are:
  - Exploration and down-weighting of outlier[13] data points.
  - Combining multiple regression models into a single graphical model, which represents dependencies[14] among the individual models.
  - Adding hydrologic models such as Voronoi to better distribute rainfall over the watershed.

Important questions remaining to be answered include development of more statistical analysis of the following relationships:

- Has the relationship between upstream precipitation, Canadian River flows and inflow to Conchas Reservoir changed in a statistically significant manner over ten years or more?
- Is inflow to Conchas determined in part by other causal factors such as maintenance practices, leakage etc?

---

[13] Outliers are “extreme” observations in a regression analysis, which can bias the model.

[14] Variables used in modeling can be divided into three types: "dependent variable", "independent variable", or other. The "dependent variable" represents the output or effect, or is tested to see if it is the effect.

APPENDIX A. PLOTTED FACTOR DATA SERIES

Figure A-1. Rainfall-Sanchez flow- Conchas elevation

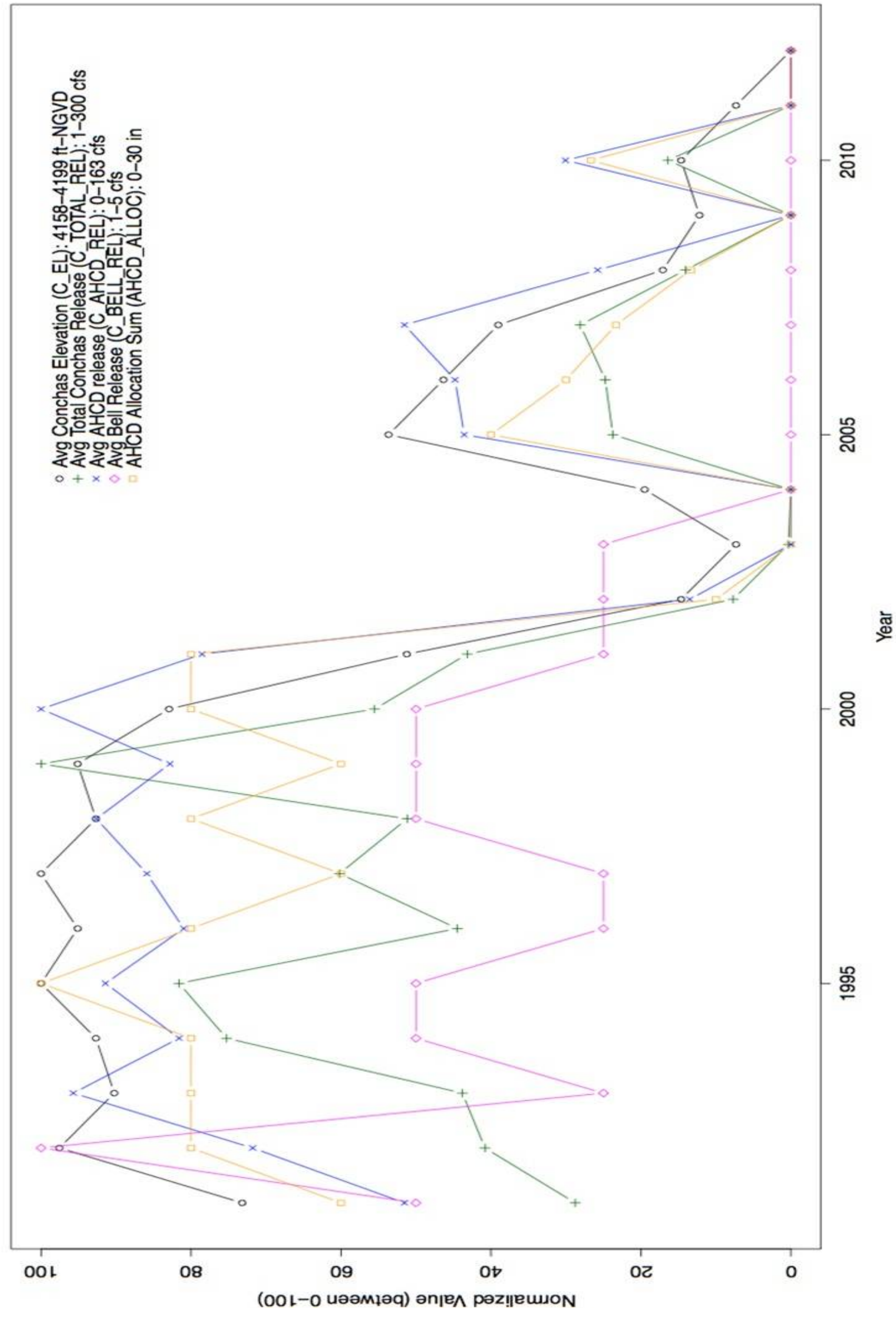


Figure A-2. Conchas elevation-Total release-AHCD release-Bell release-AHCD allocation

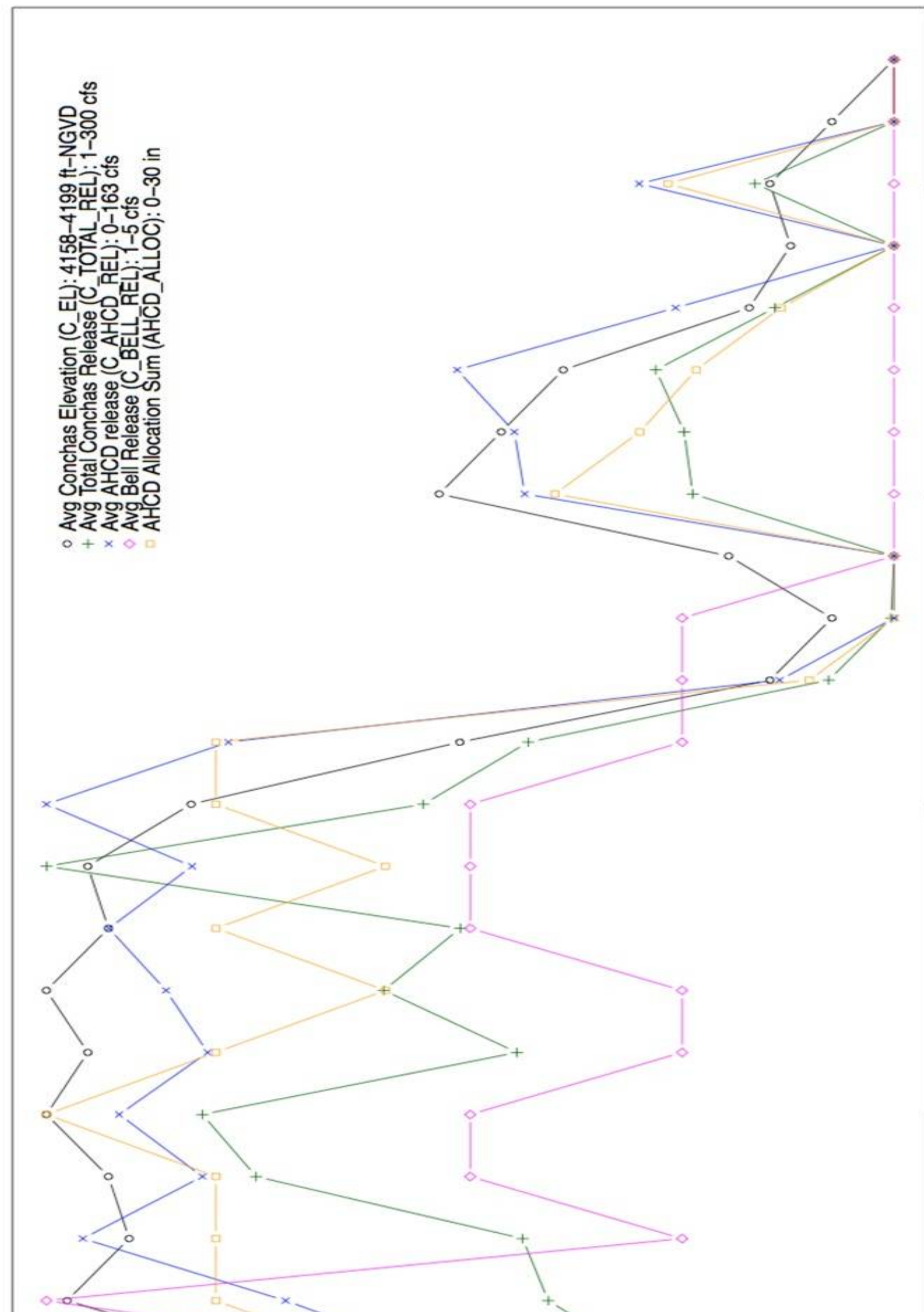


Figure A-3. Monthly Average Flow Rates- 1991 to 2010

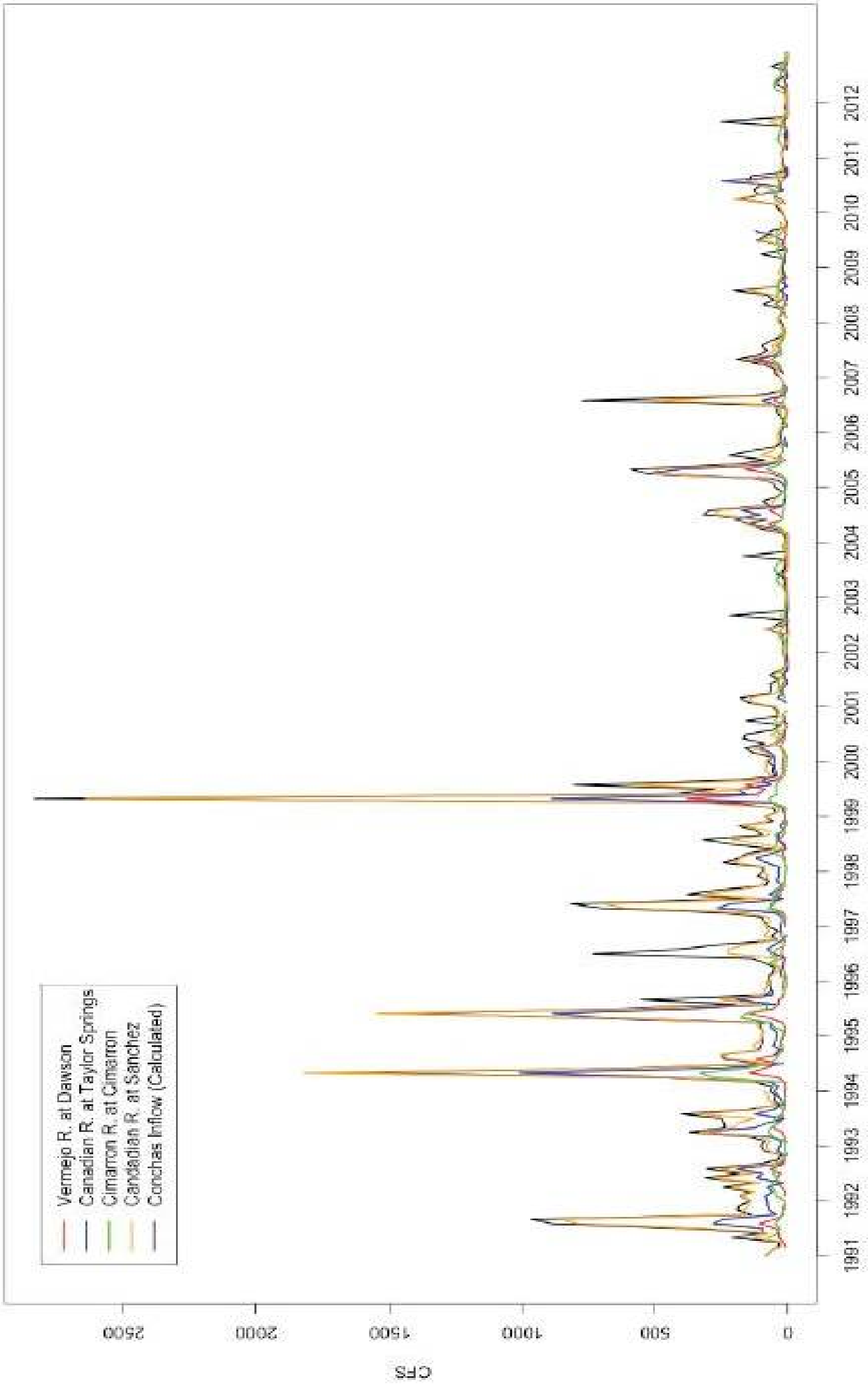


Figure A-4. Monthly Average Flow Rates- 1991 to 1995

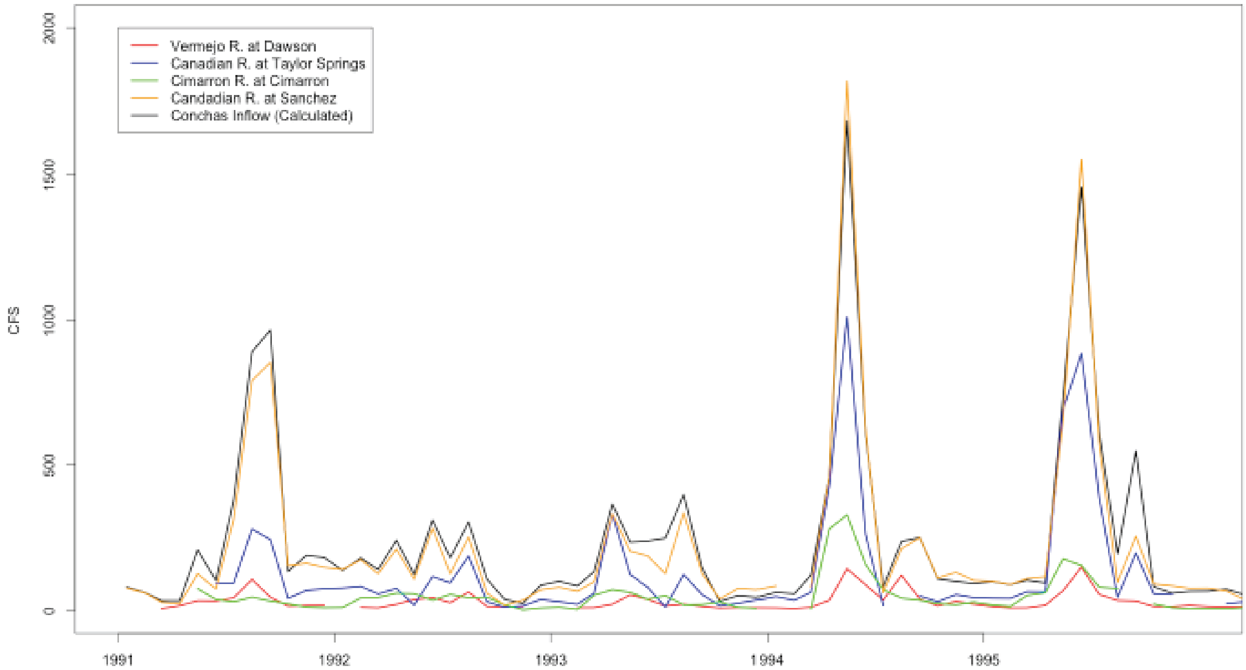


Figure A-5. Monthly Average Flow Rates- 1996 to 2000

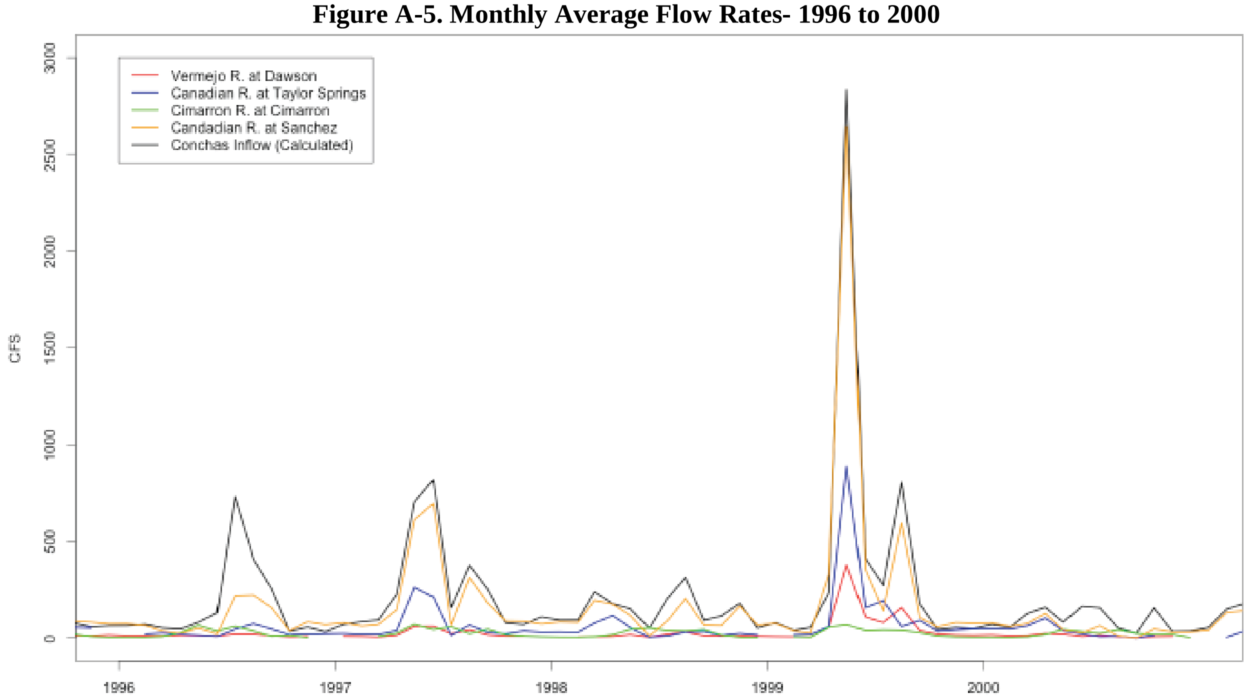


Figure A-6. Monthly Average Flow Rates- 2001 to 2005

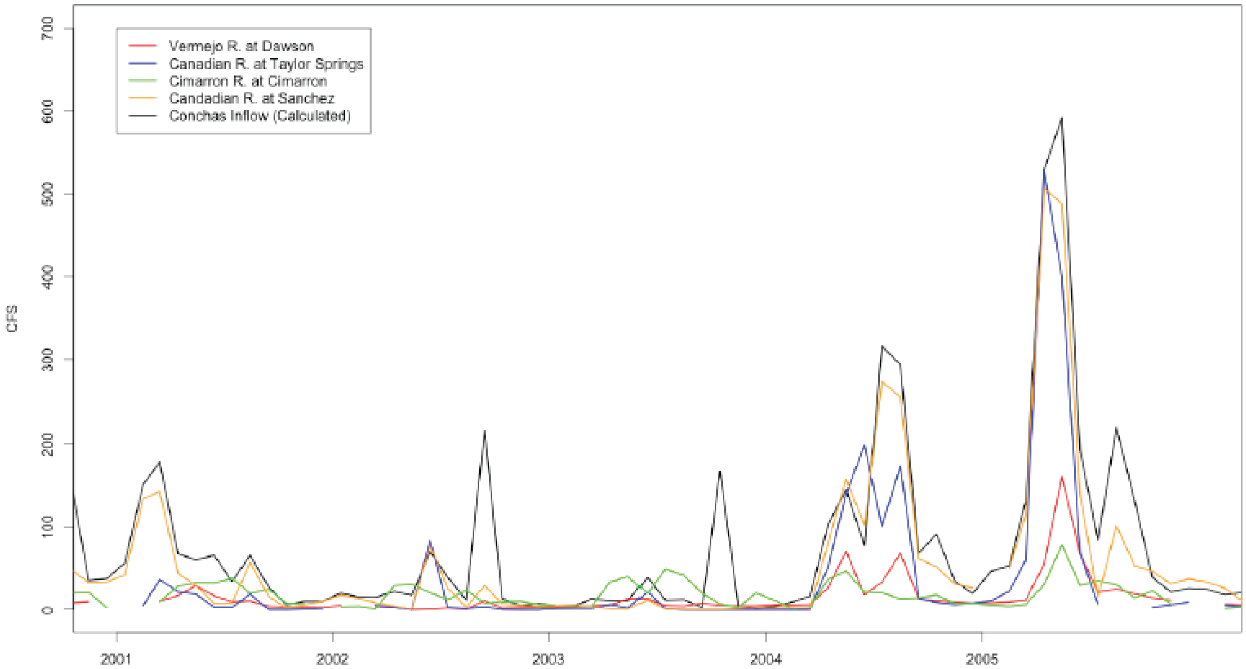
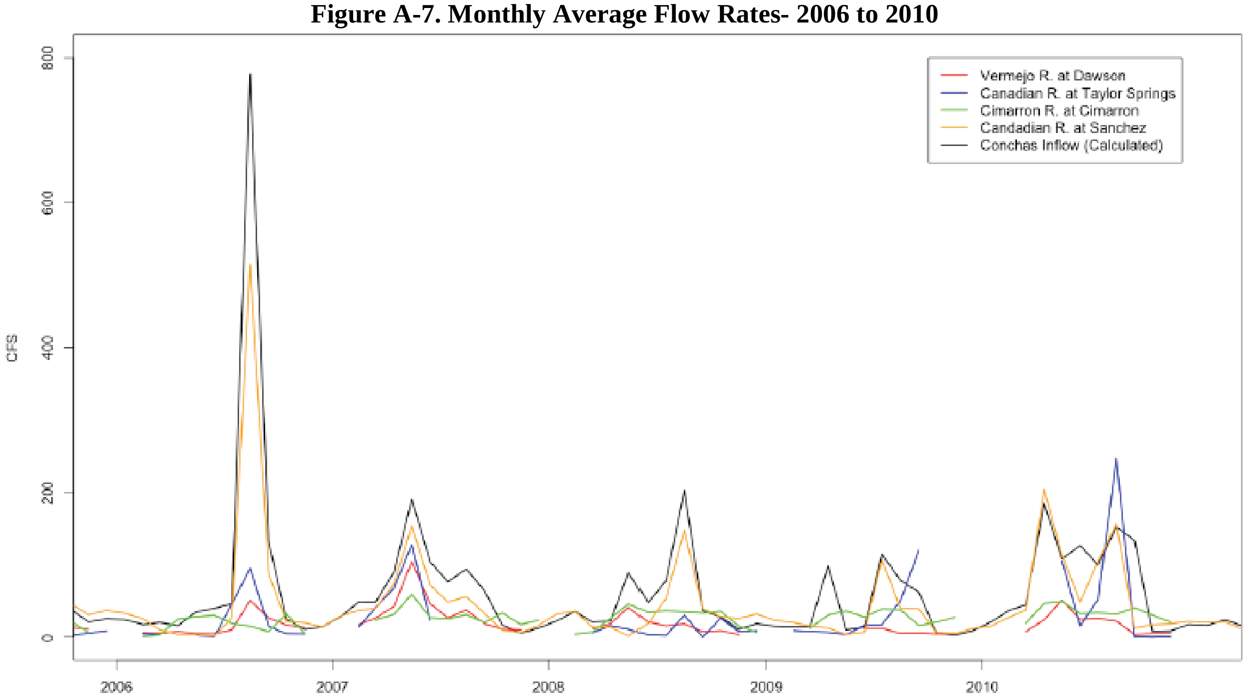


Figure A-7. Monthly Average Flow Rates- 2006 to 2010



APPENDIX B. RIVER FLOW CORRELATION TABLES

Table B-1. Correlation Coefficients: Watershed River Flow

1991-1995

Factor	dawson_cfs	taylor_cfs	cimarron_cfs	mora_cfs	coyote_cfs	sanchez_cfs	conchas_inflow_cfs
dawson_cfs	1	0.842	0.675	0.930	0.674	0.906	0.904
taylor_cfs	0.842	1	0.827	0.814	0.653	0.931	0.917
cimarron_cfs	0.675	0.827	1	0.675	0.527	0.731	0.712
mora_cfs	0.930	0.814	0.675	1	0.794	0.933	0.927
coyote_cfs	0.674	0.653	0.527	0.794	1	0.802	0.813
sanchez_cfs	0.906	0.931	0.731	0.933	0.802	1	0.992
conchas_inflow_cfs	0.904	0.917	0.712	0.927	0.813	0.992	1

1996-2000

Factor	dawson_cfs	taylor_cfs	cimarron_cfs	mora_cfs	coyote_cfs	sanchez_cfs	conchas_inflow_cfs
dawson_cfs	1	0.909	0.438	0.798	0.876	0.947	0.933
taylor_cfs	0.909	1	0.388	0.800	0.817	0.958	0.929
cimarron_cfs	0.438	0.388	1	0.514	0.344	0.422	0.488
mora_cfs	0.798	0.800	0.514	1	0.873	0.834	0.833
coyote_cfs	0.876	0.817	0.344	0.873	1	0.876	0.856
sanchez_cfs	0.947	0.958	0.422	0.834	0.876	1	0.980
conchas_inflow_cfs	0.933	0.929	0.488	0.833	0.856	0.980	1

2001-2005

Factor	dawson_cfs	taylor_cfs	cimarron_cfs	mora_cfs	coyote_cfs	sanchez_cfs	conchas_inflow_cfs
dawson_cfs	1	0.749	0.604	0.934	0.812	0.818	0.821
taylor_cfs	0.749	1	0.415	0.752	0.815	0.938	0.870
cimarron_cfs	0.604	0.415	1	0.639	0.490	0.394	0.390
mora_cfs	0.934	0.752	0.639	1	0.820	0.785	0.766
coyote_cfs	0.812	0.815	0.490	0.820	1	0.826	0.819
sanchez_cfs	0.818	0.938	0.394	0.785	0.826	1	0.951
conchas_inflow_cfs	0.821	0.870	0.390	0.766	0.819	0.951	1

2006-2010

Factor	dawson_cfs	taylor_cfs	cimarron_cfs	mora_cfs	coyote_cfs	sanchez_cfs	conchas_inflow_cfs
dawson_cfs	1	0.474	0.473	0.556	0.169	0.516	0.473
taylor_cfs	0.474	1	0.233	0.636	0.542	0.515	0.397
cimarron_cfs	0.473	0.233	1	0.467	0.110	0.055	0.065
mora_cfs	0.556	0.636	0.467	1	0.695	0.329	0.214
coyote_cfs	0.169	0.542	0.110	0.695	1	0.353	0.234
sanchez_cfs	0.516	0.515	0.055	0.329	0.353	1	0.962
conchas_inflow_cfs	0.473	0.397	0.065	0.214	0.234	0.962	1

Table B-2. Changes in Correlation: Watershed River Flow

Years	Cimarron-Taylor	Dawson-Taylor	Taylor-Sanchez	Mora-Sanchez	Coyote-Sanchez	Sanchez-Conchas
1991-1995	-	-	-	-	-	-
1996-2000	-0.439	0.068	0.027	-0.099	0.074	-0.012
2001-2005	-0.412	-0.092	0.007	-0.147	0.024	-0.041
2006-2010	-0.595	-0.368	-0.416	-0.604	-0.449	-0.029

APPENDIX C. IMPOUNDMENT EXAMPLES

To illustrate the possible value of locating impoundments using Google Earth imagery, seven examples are displayed in this Appendix.

Some examples clearly display water containment. Surface area ranges from 1,800 square feet to 2.8 acres which is approximately the correct size range for dirt ponds. However all of the examples shown in Figures C-1, C-2 and C-3 are playa lakes which do not include dams or similar structures. These sites range from 36 to 57 miles north of Conchas Reservoir.

Figure C-1. Example 1 (36 miles north of Conchas Reservoir)

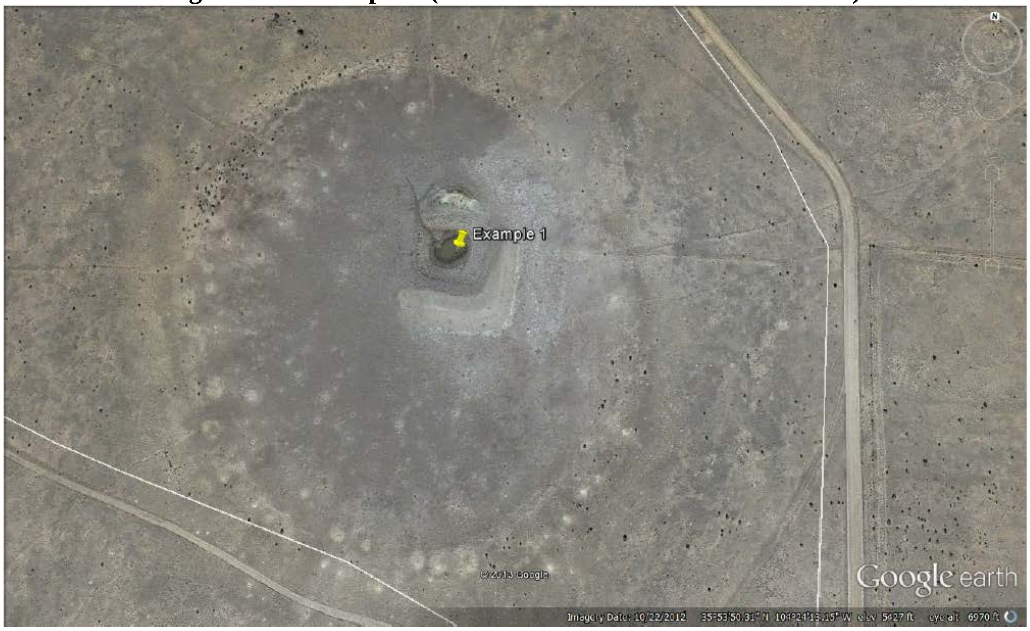


Figure C-2. Example 4, 5 Cluster of lakes

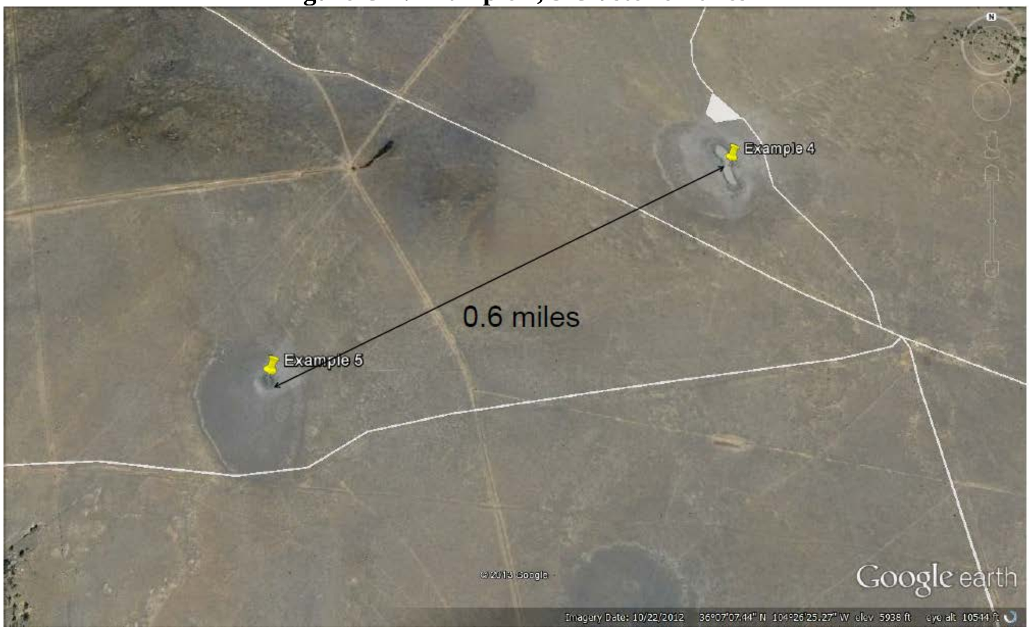
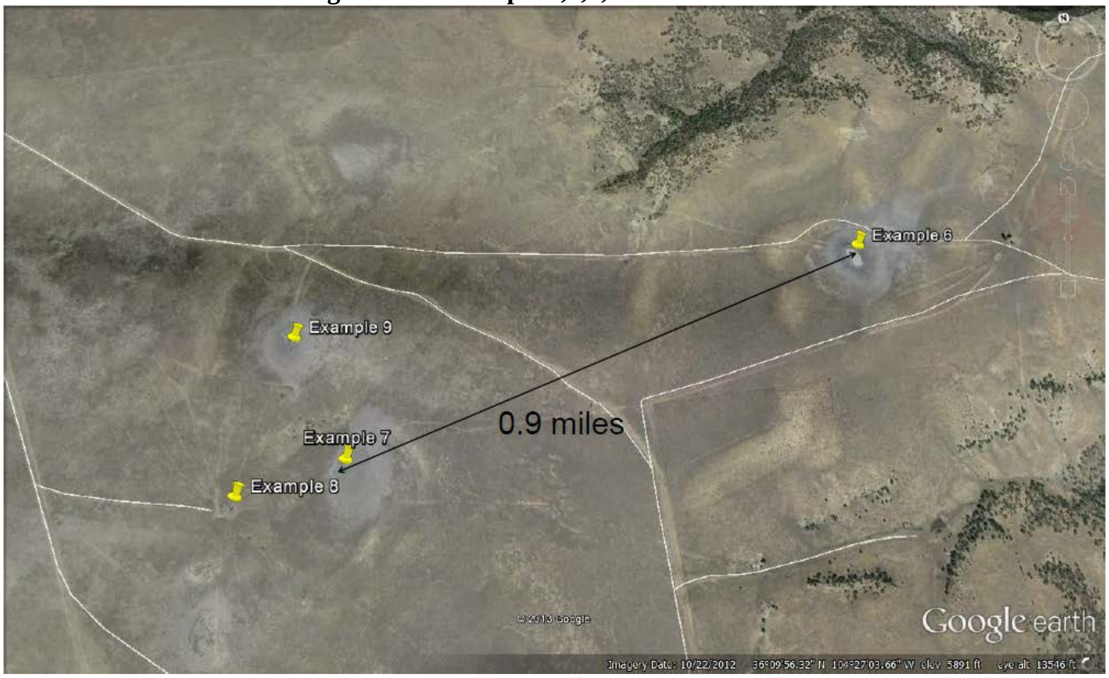


Figure C-3. Example 6,7,8,9 Cluster of lakes



A ground survey was also conducted within the Canadian River watershed and its tributaries. From the roadway a variety of water impoundments (pits, tanks, ponds, etc.) could be identified. In some areas impoundments within the same pathway (ravine, creek, etc.) were also observed. Potentially, a large number of these have been constructed within the last twenty years based upon the vegetation on and around the spoil piles. Table C-1 summarizes the findings from this survey.

Table C-1. Observed Impoundments in Canadian River Watershed

Latitude	Longitude	ID	Comment
N 35– 42.9591'	W 104– 09.0274'	33	Two tanks on south side of 419
N 35– 36.9883'	W 105– 12.8893'	34	N/A
N 35– 36.9883'	W 105– 12.8893'	35	"
N 36– 38.5548'	W 104– 45.8368'	36	"
N 36– 45.3831'	W 104– 29.9937'	37	"
N 36– 45.3840'	W 104– 29.9921'	38	"
N 36– 53.6484'	W 104– 55.0555'	39	Hiway 555 near Vermejo Park Ranch

Travel Routes driven:
Highway 419 from Mosquero to Trementina, then 104 on to Las VegasI-25 from Las Vegas to Watrous, then 161 to Ledous, 518 to La Cueva Highway 442 to Ocate, then 102 to Black Lake and 434 in to Angel Fire Highway 64 to Eagle Nest, Cimarron to a dirt road heading Northwest near the Highway 505 intersection until we reached the Dawson Cemetery About ½ mile North of the 505 intersection on the east side of Highway 64 there was a very large spoils pile with several pieces of heavy equipment. On up 64 to Raton Then Highway 555 over to Vermejo Park

APPENDIX D. REGRESSION ANALYSIS- OVERVIEW

Overview

- Datasets
- Data Exploration
- Regression
- Residual Analysis
- Problems with Historical Data

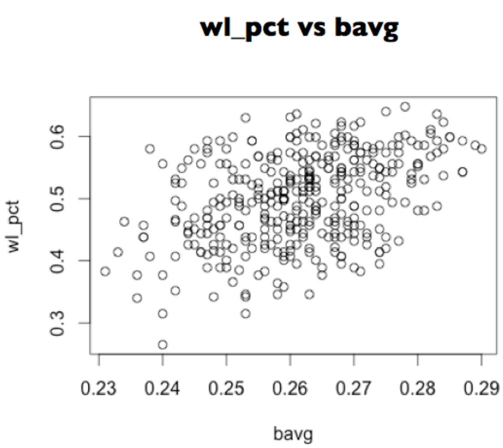
Dataset: Baseball Statistics

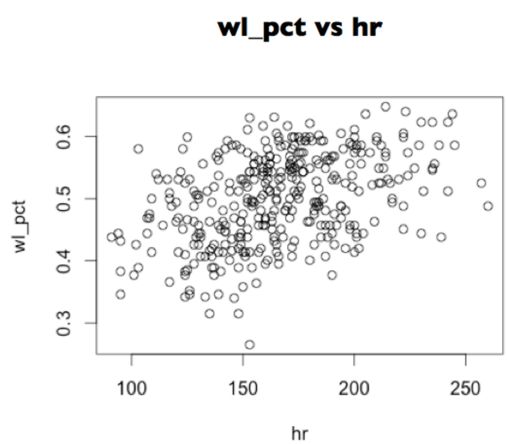
- Annual team data for regular season: 2002-2013
- Data Sources: mlb.com, baseball-reference.com
- Response Variable: Win-Loss Percent (wl_pct)
- Predictor Variables: Batting Avg (bavg), Home Runs (hr), Walks (bb), Strikeouts (so), Stolen Bases (sb), Runs Allowed (ra)

Dataset: Canadian R. at Sanchez

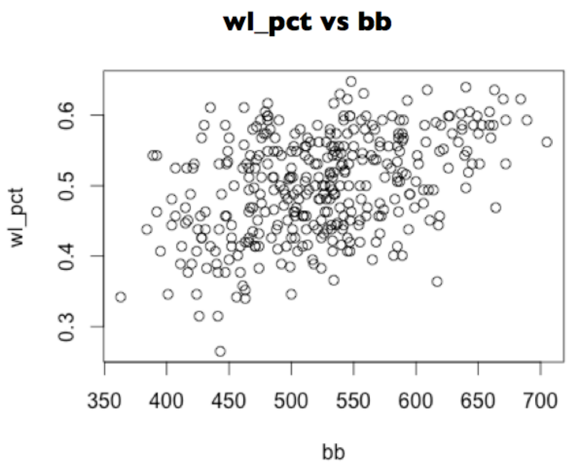
- Monthly water data: 1970-2012
- Data Sources: US Geological Survey, Western Regional Climate Center
- Response Variable: flow on Canadian R. nr Sanchez (sanchez_cfs)
- Predictor Variables: average flow on Canadian R. nr Taylor Springs (taylor_cfs), Mora R. nr Golondrinas (mora_gol_cfs), and Coyote Cr. nr Golondrinas (coyote_cfs); rainfall nr Ocate (ocate_in), Abbott (abbott_in), and Valmora (valmora_in)

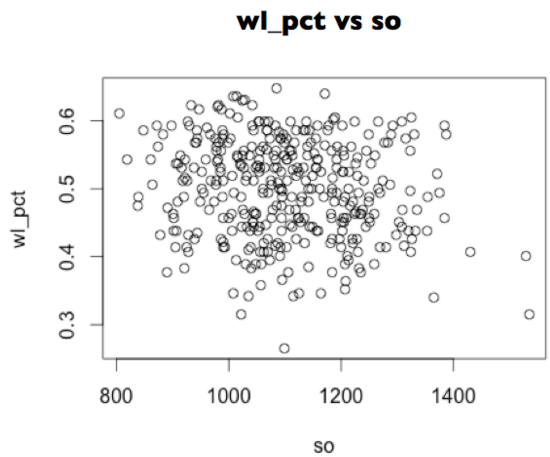
Data Exploration: Baseball



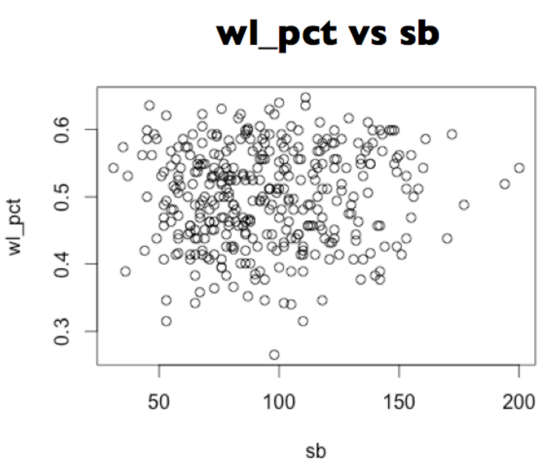


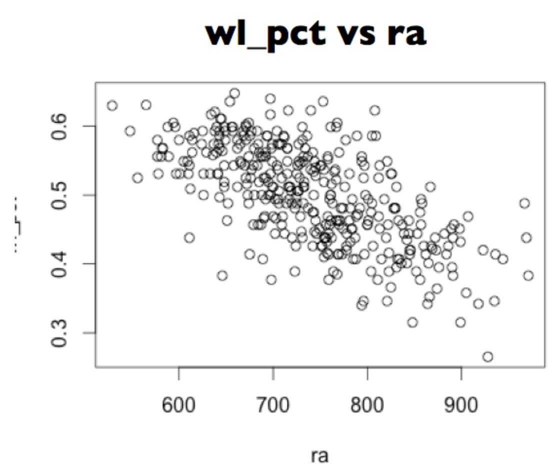
Data Exploration: Baseball





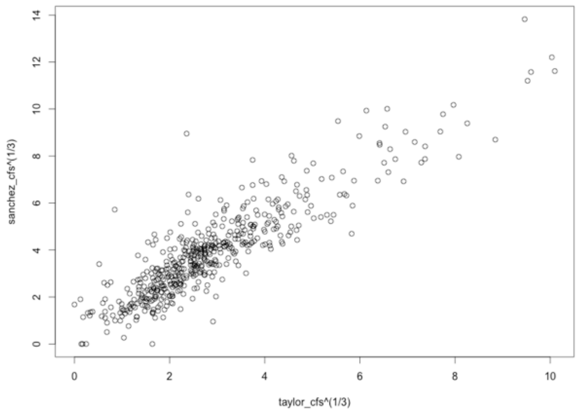
Data Exploration: Baseball



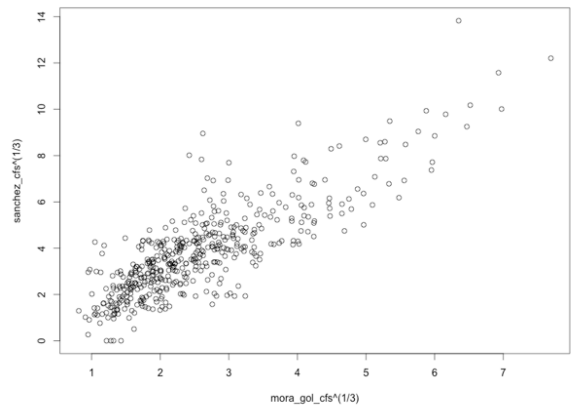


Data Exploration: Sanchez

sanchez_cfs vs taylor_cfs



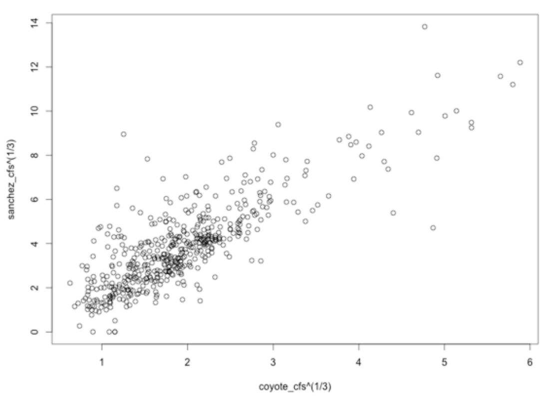
mora_gol_cfs vs taylor_cfs



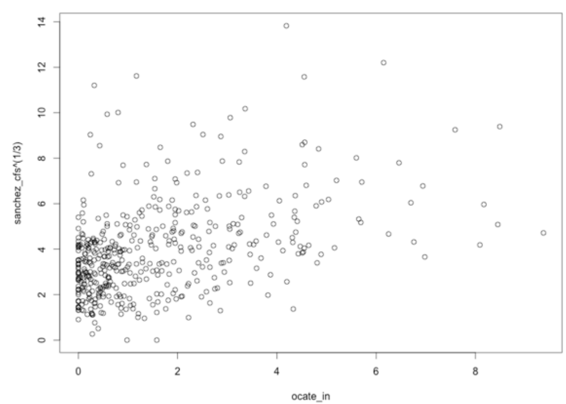
River flows have been scaled by the cube root

Data Exploration: Sanchez

sanchez_cfs vs coyote_cfs



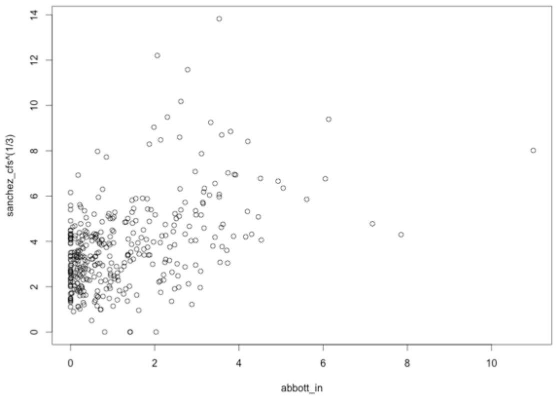
sanchez_cfs vs ocate_in



River flows have been scaled by the cube root

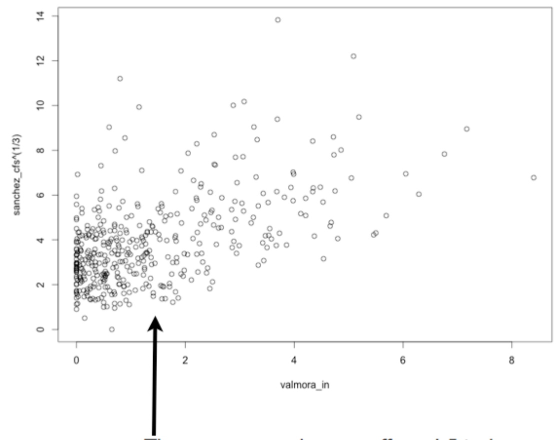
Data Exploration: Sanchez

sanchez_cfs vs abbott_cfs



River flows have been scaled by the cube root

sanchez_cfs vs valmora_in



There appears to be a cutoff at ~1.5 inches

Key Regression Statistics

- Pr(>|t|): probability that the coefficient for a particular variable is zero, i.e. that the variable is not significant (the lower the better)
- Adjusted R-Square: fraction of variability explained (the higher the better)
- F statistic p-value: probability that all coefficients are zero, i.e. that the regression explains none of the variability (the lower the better)

Initial Regression: Baseball

```
Response:  wl_pct
Predictors: bavg + hr + bb + so + sb + ra

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.521e-01  4.961e-02   3.066  0.00233 **
bavg         2.339e+00  1.648e-01  14.196  < 2e-16 ***
hr           7.196e-04  5.050e-05  14.252  < 2e-16 ***
bb           1.954e-04  2.364e-05   8.264 2.88e-15 ***
so          -2.276e-05  1.411e-05  -1.613  0.10764
sb           8.972e-05  4.716e-05   1.902  0.05796 .
ra          -6.341e-04  1.719e-05 -36.890  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02673 on 353 degrees of freedom
Multiple R-squared:  0.8642,   Adjusted R-squared:  0.8619
F-statistic: 374.4 on 6 and 353 DF,  p-value: < 2.2e-16
```

Remove Insignificant Variables

:

```
Response: wl_pct
Predictors:  bavg + hr + bb + ra
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.036e-01  3.328e-02   3.113    0.002 **
bavg         2.511e+00  1.311e-01  19.160  < 2e-16 ***
hr           6.895e-04  4.884e-05  14.117  < 2e-16 ***
bb           1.872e-04  2.337e-05   8.010 1.66e-14 ***
ra          -6.392e-04  1.715e-05 -37.275  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02688 on 355 degrees of freedom
Multiple R-squared:  0.8619,  Adjusted R-squared:  0.8603
F-statistic: 553.9 on 4 and 355 DF,  p-value: < 2.2e-16
```

Regression Notes: Sanchez

- Use cube root of all river flows to scale properly with rainfall units (i.e. river flow is measured as **cubic** feet per second - while rainfall is measured in **linear** inches)
- Rainfall cutoff of 1.5 inches (i.e. if rainfall < 1.5, use 0; if rainfall >1.5, substract 1.5)
- All variables in regressions are transformed as described above
- Data for regressions are restricted to 1970-2002; regression model is then used to predict river flows at Sanchez between 2003 and 2012 (see body of report)

Initial Regression: Sanchez

```
Response Variable:  sanchez_cfs
Predictor Variables: intercept + taylor_cfs + mora_gol_cfs + coyote_cfs +
ocate_in + abbott_in + valmora_in

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.13057    0.09556  -1.366   0.1727
taylor_cfs   0.66877    0.03878  17.247  < 2e-16 ***
mora_gol_cfs 0.40846    0.05629   7.256 2.44e-12 ***
coyote_cfs   0.48470    0.08525   5.686 2.69e-08 ***
ocate_in    -0.08627    0.03780  -2.282   0.0231 *
abbott_in    0.05066    0.04891   1.036   0.3010
valmora_in   0.32021    0.04385   7.302 1.81e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6615 on 362 degrees of freedom
Multiple R-squared: 0.891,  Adjusted R-squared: 0.8892
F-statistic: 493.3 on 6 and 362 DF,  p-value: < 2.2e-16
```

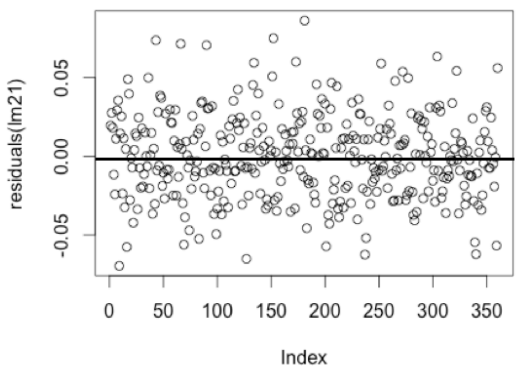
Remove Insignificant Variables + Intercept

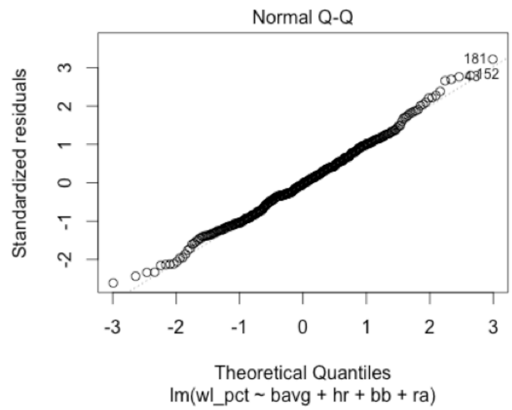
```
Response Variable:  sanchez_cfs*
Predictor Variables: taylor_cfs + mora_gol_cfs + coyote_cfs + valmora_in

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
taylor_cfs    0.66196    0.03818  17.339  < 2e-16 ***
mora_gol_cfs  0.38293    0.05455   7.020 1.09e-11 ***
coyote_cfs    0.46033    0.07869   5.850 1.10e-08 ***
valmora_in    0.29033    0.03635   7.988 1.81e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6647 on 365 degrees of freedom
Multiple R-squared: 0.9791,  Adjusted R-squared: 0.9789
F-statistic:  4278 on 4 and 365 DF,  p-value: < 2.2e-16
```

Residual Analysis: Is the regression model a good fit?

- Residuals = difference between predicted and actual response values
- Common Residual Analysis Plots
  - Residuals vs Sample Order (Is there a pattern of predominantly positive or negative residuals over time?)
  - Normal Probability Plot of Residuals (Are residuals normally distributed?)
  - Residuals vs Fitted Values (What does the fit look like at the extremes?)
  - Leverage Plots (Which sample points have a large influence on model - i.e. outliers?)

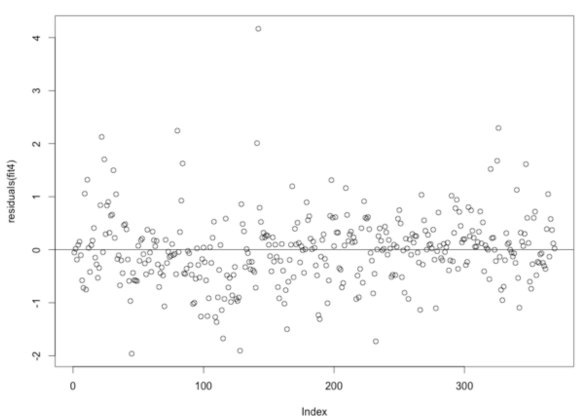
Residuals: Baseball Model

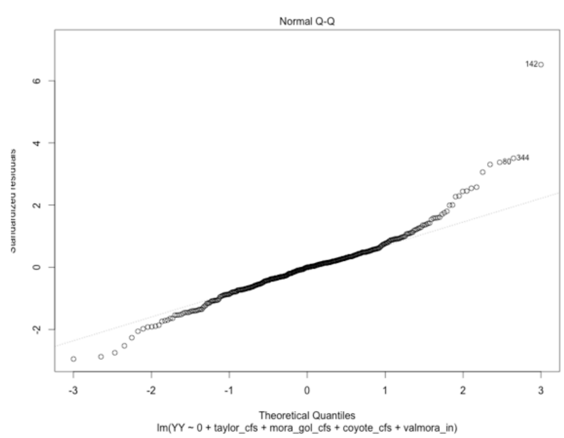





```
      Min        1Q    Median        3Q       Max
-0.069609 -0.019417 -0.000477  0.017382  0.086248
```

Residuals: Sanchez Model





Note: Residuals are in transformed units of flow rate at Sanchez (i.e. cube root of cfs values)

```
    Min      1Q  Median      3Q     Max
-1.9617 -0.3909 -0.0014  0.2926  4.1668
```

Problems with Historical Data

- Inconsistent data measurement
- Range and distribution of variable values
- Missing Data
- Dependencies among variables
- Nonsense correlation due to "lurking" variable

APPENDIX E. GLOSSARY OF TERMS AND ACRONYMS

AHCD	Arch Hurley Conservancy District surrounding the city of Tucumcari, has about 41,000 acres of irrigable land
Correlation	Any of a broad class of statistical relationships involving dependence
HMS	Hydrologic Modeling System; employed by USACE to model flooding events
LANL	Los Alamos National Laboratory, the primary author of this document
NMSBA	New Mexico Small Business Assistance Program funded by Los Alamos and Sandia National Laboratories
NRCS	(Federal) National Resources Conservation Service, a unit of the USDA
Regression	An approach to model the relationship between a scalar dependent variable y and one or more explanatory variables denoted X
R-Square	Goodness of fit of a statistical model, describes how well it fits a set of observations
USACE	(Federal) US Army Corps of Engineers
USDA	(Federal) US Department of Agriculture
USGS	(Federal) Us Geological Survey
Variable	A symbol or factor representing some data, which is commonly a number, but may also be any mathematical object such as a vector, a matrix etc.
WRCC	Western Regional Climate Center, Reno NV