Factors Effecting Water Level in the Edwards Aquifer

Analysis Summary

Paul A. Wilson

Problem and Hypothesis

Sustained water availability in the Edwards Aquifer is often questioned due to population growth. This analysis attempts to answer the question: Are any of the major factors believed to effect water availability statistically significant at an aggregated annualized level?

The Edwards Aquifer is in south central Texas. It covers approximately 3,600 square miles and is about 180 miles long. Its honeycombed carbonate limestone structure supports water flow and storage, caves, sinking streams, sinkholes, artesian springs (Eckhardt, 2018a). Figure 1 shows the aquifer's three regions. This analysis concentrates on water availability for more than 2 million people that live in the six counties over the Recharge Zone and Artesian Zone. These are Kinney, Uvalde, Medina, Bexar, Comal, & Hays counties. The most important measure of water level is the J17 Bexar Index Well measured in mean feet above sea level (msl).

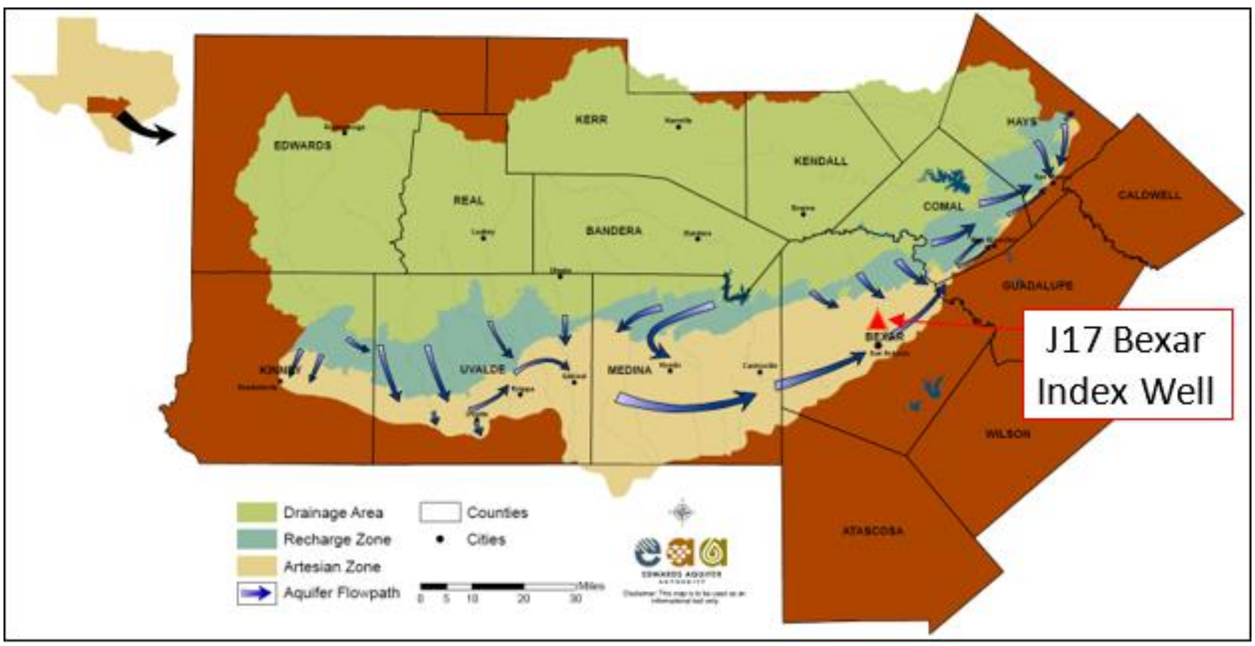

Figure 1. The Edwards Aquifer and its zones. Water drains from the north on the surface (Drainage Area), then plunges below the surface in the Recharge Zone. Water can surface in the Artesian Zone through natural springs or be pumped out (Edwards Aquifer Authority [EAA], 2018d).

This analysis will attempt to answer the question with a multiple linear regression equation using more than six decades of data. Response variable: J17 well level ft. msl. Predictor variables: Total population, San Antonio rainfall in inches (in.), Total Recharge, Spring flow, and various types of pumping all in 1,000 acre feet (af).

The null hypothesis is that parameter estimates for all regression predictor variables are equal to zero. The alternative hypothesis is that at least one regression predictor variable's parameter estimate is not equal to zero.

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ $\qquad H_a: at\ least\ one\ \beta_1\ ,\ \ \beta_2\ , \cdots \beta_k \neq 0$

Data Analysis Process

Data acquisition and preparation uses Microsoft Excel's Power Query. Aggregated annual aquifer data for both recharge and discharge variables is accessed from two html tables (Eckhardt, 2018b). Daily J17 well data is accessed from a csv table (EAA, 2018b). Annual county population data is accessed from two csv files and hand transcribed from two pdf documents (Roth, 2016; U.S. Census Bureau, 2015, 2016a, 2016b). Power Query performs the following operations on the data: connect to each data source, row and column filtering, data type transformations, grouping and aggregations, combine to a single csv file with 13 columns and 62 rows (1954 to 2015).

Remaining data cleaning and analysis uses SAS University Edition. Data cleaning uses a SAS DATA step to create custom variables to compare variances between summary and detail variables. Several 1954 detail observations are found to be erroneous. A new dataset is created without 1954 and the three total variables, which are no longer needed. The remaining variables are: Year (observation), AvgJ17Year (response), TtlRecharge, SAPrecip, Muni_Mil, Ind, Dom_Livestk, Irrig, Spring, PopTtl (predictors).

Preliminary data exploration uses PROC SGSCATTER and PROC REG. SGSCATTER reveals interesting correlations between some predictor variables and J17AvgYear. Figure 2 highlights a cluster of low Dom_Livestk values from the mid-1990s to 2015 and an extreme Ind observation for 1991.

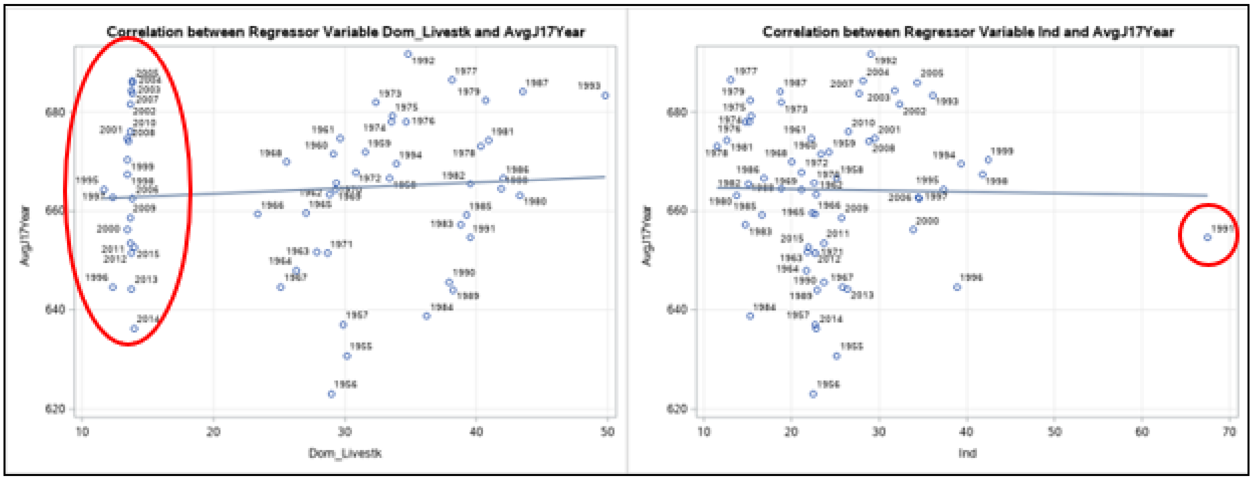

Figure 2. Correlations between AvgJ17Year and regressor variables Dom_Livestk and Ind.

The cluster of lower values in Dom_Livestk makes sense because of the regulatory limits on pumping enforced by the Edwards Aquifer Authority (EAA) since late 1993 (EAA, 2018c). According to the 1991 Hydrologic report from the EAA (1992), the observation for Ind is transcribed correctly from the original data. The initial PROC REG exploration includes Variance Inflation statistics that verify no collinearity problems. The ANOVA table in figure 3 shows that the model is statistically significant at the 0.05 alpha level and an Adjusted R-square of 0.9112.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	13297	1662.11942	77.95	<.0001
Error	52	1108.76662	21.32243		
Corrected Total	60	14406			

Root MSE	4.61762	R-Square	0.9230
Dependent Mean	664.28852	Adj R-Sq	0.9112
Coeff Var	0.69512		

Figure 3. ANOVA table and Adjusted R-Square from initial PROC REG model with all variables.

Examining Cook's D, DFBETAS, and Residual by Regressors plots (figure 4) shows 1992's highly influential values. 1991 Ind is an isolated extreme observation. The 1992 values from the original EAA report show that they are all correctly transcribed (EAA, 1993).

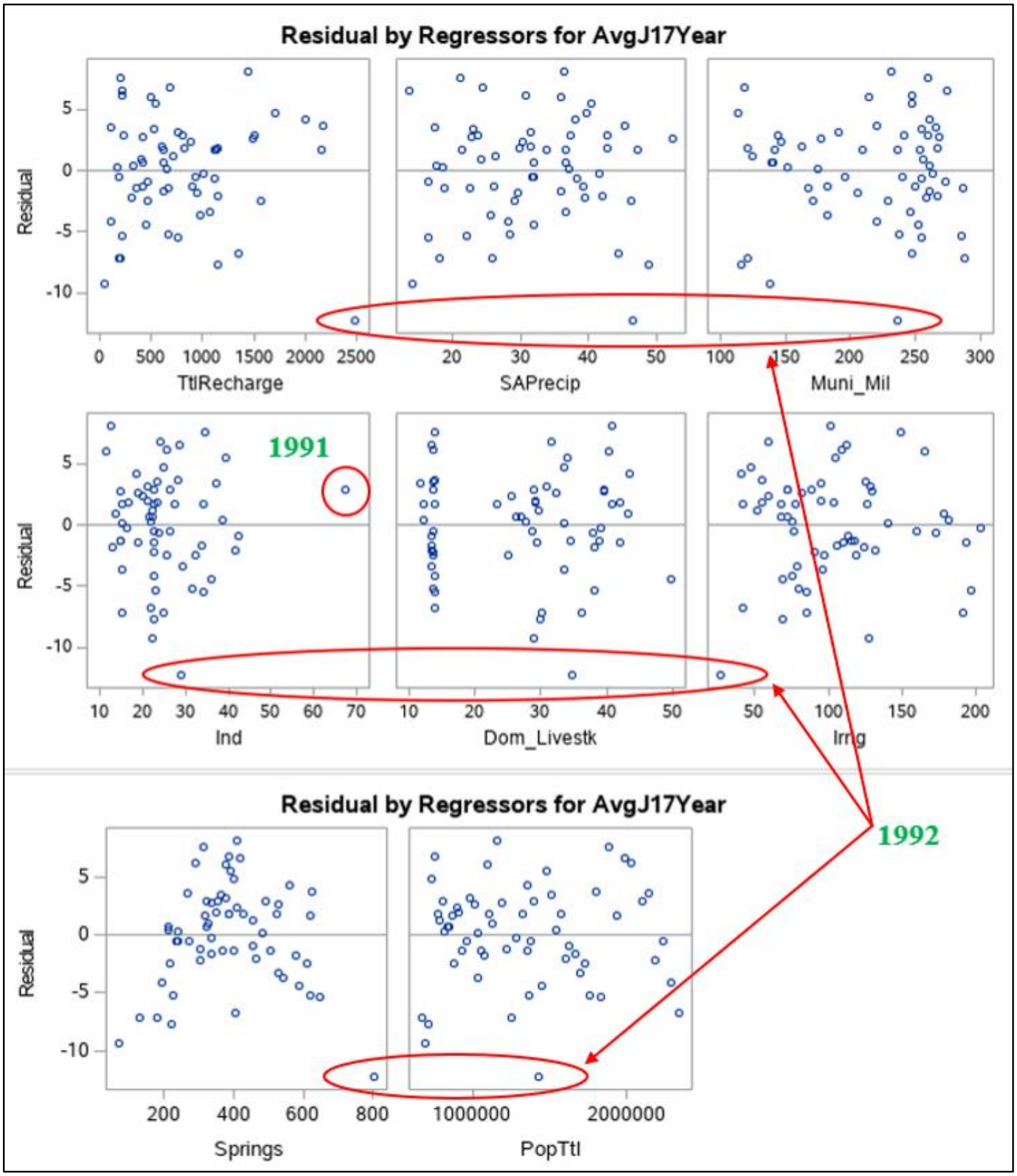

Figure 4. Residuals by Regressors plots clearly show 1992 as the most extreme residual. 1991's residual stands out in Ind because of the extreme value.

Running PROC REG again without 1992 improves Adjusted R-square to 0.9224 and improves the Residual by Regressor plots.

Automated model selection uses PROC GLMSELECT with Leave-one-out k-fold cross-validation. Figure 5 shows how Stepwise selection uses the predicted residual error sum of squares (PRESS) statistic to select and remove variables. The cvmethod=split(60) option splits the data into 60 folds with each missing a single observation. This model improved Adjusted R-square to 0.9264. Five selected variables are: Springs, Ind, TtlRecharge, PopTtl, and Muni_Mil.

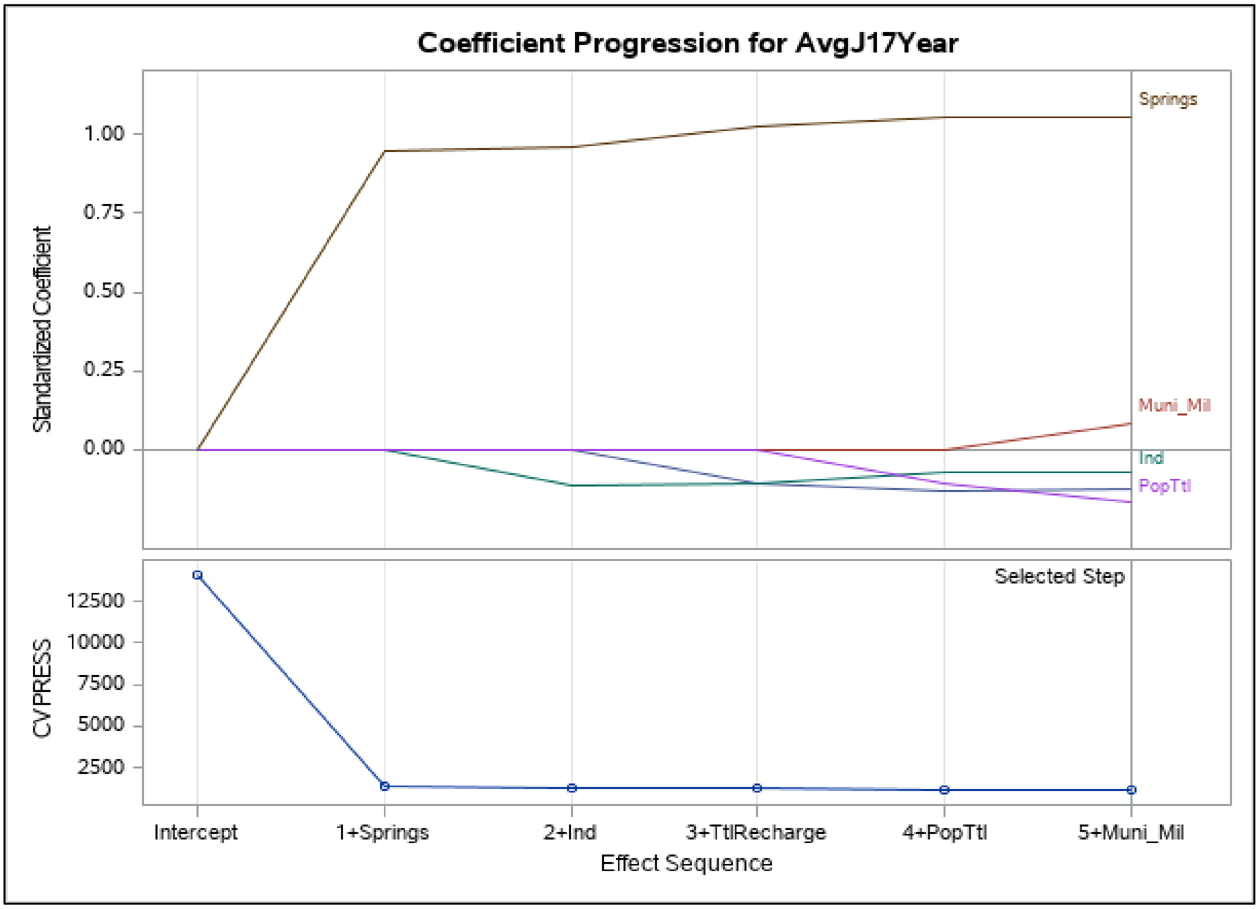

Figure 5. The Coefficient Progression plot shows how Springs is the dominant regressor and the others contribute only small improvements.

Final model selection uses PROC REG and the five variables selected by GLMSELECT. This produces the same Adjusted R-square as GLMSELECT (0.9264). However, p-values show Muni_Mil and Ind are not significant. Removing Muni_Mil and Ind and running PROC REG again lowers Adjusted R-square only slightly to 0.9205. All variable p-values are significant at the 0.05 alpha level and plots verify multiple linear regression assumptions. J17AvgYear is linearly related to a function of the predictor variables, errors have a normal distribution, errors have constant variance, and errors are independent. Figure 6 shows the parameter estimates and that the model explains 92.1% of the variability in J17 well levels.

Root MSE	4.28505	R-Square	0.9246
Dependent Mean	663.82833	Adj R-Sq	0.9205
Coeff Var	0.64551		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation	95% Confidence Limits	
Intercept	1	626.95770	2.11097	297.00	<.0001	0	622.72892	631.18649
Springs	1	0.11953	0.00542	22.04	<.0001	1.71245	0.10867	0.13040
TtlRecharge	1	-0.00396	0.00138	-2.87	0.0057	1.64388	-0.00672	-0.00120
PopTtl	1	-0.00000419	0.00000122	-3.42	0.0012	1.06503	-0.00000664	-0.00000174

Figure 6. Parameter estimates and Adjusted R-square for model with final three variables.

This model is the most parsimonious of all models considered and is the final selection. The final multiple linear regression model is:

AvgJ17Year = 626.95770 +0.11953(Springs) +(-0.00396)(TtlRecharge) +(-0.00000419)(PopTtl)

Findings

Springs, Total recharge, and Population have a statistically significant effect on J17 Bexar index well levels. Therefore, the null hypothesis is rejected in favor of the alternative.

$$H_a: at\ least\ one\ \beta_1\ ,\ \ \beta_2\ , \cdots \beta_k \neq 0$$

Intercept 626.96 is meaningful as a possible level for the J17 well. Springs is the most dominant variable: J17's well level increases an average 0.11953 feet for every 1,000 af increase in Springs output (1 7/16 in.). J17 decreases an average 0.00396 feet for every 1,000 af increase in Total Recharge (3/64 in.). J17's decrease from Total Recharge is counterintuitive because rainfall and runoff are assumed to increase aquifer level. Figure 7 illustrates that this assumption is true. However, comparing the rate of J17 increase for Springs (left plot) to Total Recharge (right plot), shows that Springs' rate increases faster than Total Recharge's rate. This means that Total Recharge has ***less*** of a positive correlation than Springs.

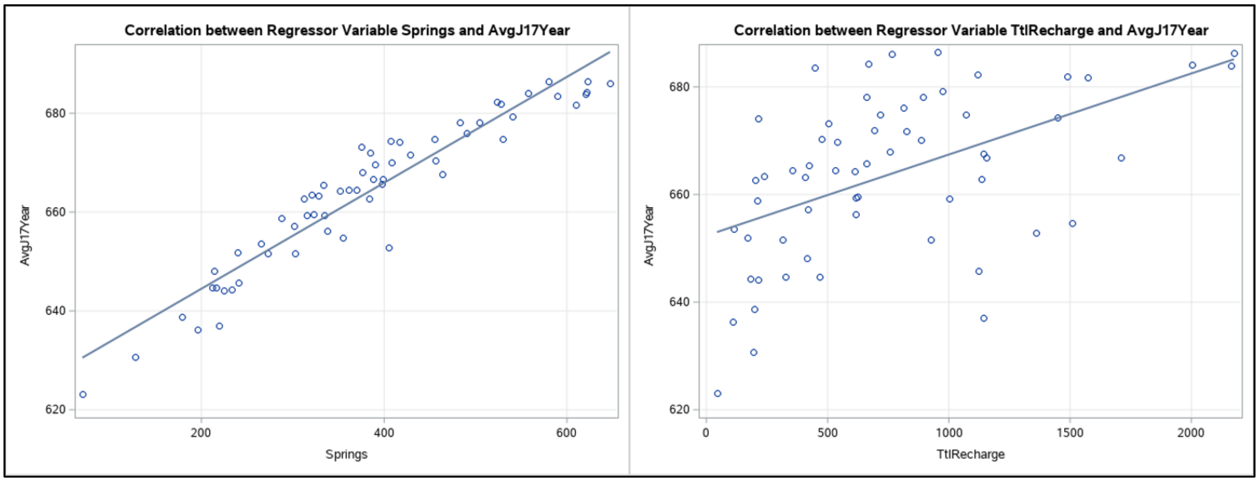

Figure 7. Regression plots for Springs and TtlRecharge versus AvgJ17Year shows that although both have a positive correlation to AvgJ17Year, TtlRecharge's correlation is not as steep (not as strong) and calculates as a *negative* correlation relative to Springs.

J17 decreases an average 0.00000419 feet per one-person increase in population (this is imperceptible). However, we can estimate a total loss from 1955 to 2015 in mean J17 well level as approximately 7 ft. Or 13 3/4 in. per decade. While population influences J17 levels, it is very small compared to spring flow and total recharge. Figure 8 illustrates this dramatically, note how radically J17's level changes each year as compared to the cumulative effect of population.

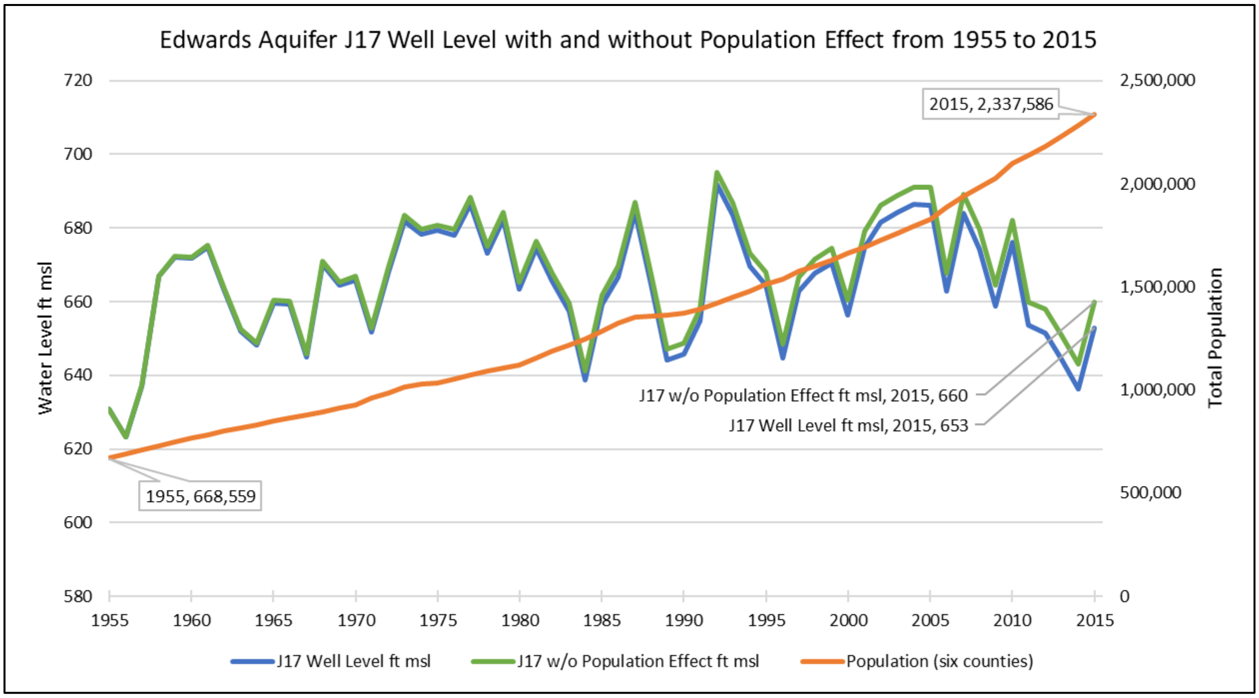

Figure 8. Population from 1955 to 2015 versus J17 Index Well level with and without the estimated effect of population. The cumulative effect over six decades is 7 feet (13 ¾ in. per decade).

Limitations

A limitation of this analysis is that it relies on aggregated annual data making updates to the model slower. Data is available with daily granularity for J17 levels, rainfall, and springs but total recharge is an annual calculation and pumping data is only available annually (EAA, 2018a). Another limitation is the use of SAS University Edition. It is free for non-commercial use, but the license could be revoked at any time (SAS Institute Incorporated, 2015). Another limitation of this analysis is that it only highlights statistically significant *correlations* between J17 levels and the various factors. It cannot explain *causality* or *predict* J17 levels.

Proposed Actions

One proposal is to limit the study to daily data to better understand seasonality. J17 well data is available daily along with daily flow from the two major measured springs – San Marcos Springs and Comal Springs (EAA, 2018e). Recharge could be based on rainfall totals from various weather services. That takes care of the two largest variables. For population, the estimated parameter from this model can be applied and adjusted as new population data becomes available. Another proposal is to run the analysis monthly using new automated monthly meter reads from wells. This is possible with a new voluntary program from the EAA (EAA, 2018a). This will allow a correlation between pumping and recharge at a more granular level. A further enhancement would be to add demographic and sociodemographic data about the population served by a well. This would allow the EAA to form usage profiles for each well and better target populations with specific conservation programs.

Expected Benefits

The expected benefits of this analysis are that officials can gain an increased awareness of the statistical significance of spring flow, total recharge, and population. Population growth in Texas is inevitable but armed with an awareness of the true statistical significance, they can understand the extremely small correlation between J17 well level and population as compared to the natural rise and fall correlated with spring flow and recharge. This will free them to concentrate on protecting natural features within the recharge zone from destructive activities that interfere with the aquifer's ability to recharge. They can also increase promotion of water efficient appliances and conservation programs to help counter the effects of population growth. These efforts will help users conserve water so that the Edwards Aquifer remains a viable, thriving source of high-quality water for all future generations of Texans.

Reference

Eckhardt, G. (2018a). *Home.* The Edwards Aquifer Website. Retrieved from http://www.edwardsaquifer.net/index.html

Eckhardt, G. (2018b). *Numeric Data for Charts and Graphs.* The Edwards Aquifer Website. Retrieved from http://www.edwardsaquifer.net/data.html

Edwards Aquifer Authority. (1992). *Compilation of Hydrologic Data for the Edwards Aquifer, San Antonio Area, Texas, 1991, with 1934-91 Summary* (Bulletin 51). Retrieved from https://www.edwardsaquifer.org/files/download/hydro/1992_1991HydrologicData.pdf

Edwards Aquifer Authority. (1993). *Edwards Aquifer Hydrogeologic Status Report For 1992* (Report 93-05) Retrieved from https://www.edwardsaquifer.org/files/download/hydro/1993_1992HydrogeologicReport.pdf

Edwards Aquifer Authority (2018a). *Automated Meter Reading.* Retrieved from https://www.edwardsaquifer.org/groundwater-users/meters/automated-meter-reading

Edwards Aquifer Authority (2018b). *Download J17* [Data file]. Retrieved from https://www.edwardsaquifer.org/science-and-maps/aquifer-data/historical-data

Edwards Aquifer Authority (2018c). *EAA Timeline.* Retrieved from https://www.edwardsaquifer.org/eaa/history/eaa-timeline

Edwards Aquifer Authority (2018d). *General Aquifer Flowpath*. Retrieved from https://www.edwardsaquifer.org/app/main/files-module/local/map_files/generalaquiferflowpath.pdf

Edwards Aquifer Authority (2018e). *Historical Data*. Retrieved from https://www.edwardsaquifer.org/science-and-maps/aquifer-data/historical-data

Roth, J. (2016). *County_population.csv* [Data file]. Census U.S. Intercensal County Population Data, 1970-2014. The National Bureau of Economic Research. Retrieved from https://www.nber.org/data/census-intercensal-county-population.html

SAS Institute Incorporated (2015). *License Agreement for the SAS® University Edition software* (Revised 2015, August 24). Retrieved from http://support.sas.com/legaldocs/univeditionlicense.pdf

U.S. Census Bureau (2015). *Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2015* [Data file]. U.S. Census Bureau. Population Division. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2017_PEPANNRES&src=pt

U.S. Census Bureau (2016a). *1950 Census of Population.* U.S. Department of Commerce. Retrieved from https://www2.census.gov/library/publications/decennial/1950/pc-02/pc-2-43.pdf

U.S. Census Bureau (2016b). *1960 Census of Population.* U.S. Department of Commerce. Retrieved from https://www2.census.gov/library/publications/decennial/1960/population-pc-a1/15611126ch5.pdf