

Predicting Agricultural Water Usage in California*

An Examination of the Use of Urban Water usage to predict Agricultural Water Usage

Catharina Castillo

September 24, 2025

Agricultural water use is harder to measure than urban water use due to limitations of the measuring techniques. This paper seeks to explore if measurements of urban water use can be used to predict measurements of agricultural water use with simple linear regression to see if recorded urban water use can be used to assess confidence in future agricultural water use measurements. Unfortunately, it was discovered that measured urban water use cannot accurately predict measured agricultural water use through a simple linear regression, though we still found evidence of a relation between urban water use and agricultural water use. We hope future research may be able to add more variables to create a model that can be used to assess measured agricultural water use.

1 Introduction

Historically, it has been harder to measure agricultural water usage than urban water usage. The infrastructure used in urban settings has allowed water to be measured directly using metering. The agricultural water use cannot always be measured directly. Instead, land use surveys are used to calculate water used based on factors such as amount of land and crops grown (Cooley 2020). Irrigation alone has been noted as one of the more unreliable measurements (“California Water Use” 2018).

With recent technological advancements, such as micro sprinklers, more accurate measurements have been able to be taken. These have all arisen as a result of pushes for efficiency caused by drought concerns in California (PBSNewsHour 2025). The use of satellite and aerial

*Project repository available at: https://github.com/NettleHook/Groundwater_Project.git.

imagery has also added another method of tracking water use across all sectors (ABC10 and Knight 2023).

It is important to have accurate measurements of water usage across sectors, as this information helps to predict future water usage and allows us to track progress towards sustainability. (“Groundwater Sustainability Plans” 2018)

This paper seeks to explore if we can use the measured urban water usage to estimate the agricultural water usage using linear regression. While this can’t replace current methods of measuring agricultural water usage, an estimate can help assess the confidence in the measurement taken.

In Section 2, we will explain the data used. In Section 3, we will introduce the linear regression model. Finally, in Section 4 we will lay out our results and in Section 5 we will discuss the implications and future steps.

2 Data

We will begin by using the Total Water Use dataset provided by the Groundwater Sustainability Plan Annual Report datasets submitted to data.ca.gov

The data provided is sourced through Groundwater Sustainability Agencies and Alternative Agencies as part of the Groundwater Sustainability Plan. Every year, each basin agency submits a report to the Department of Water Resources through the [Sustainable Groundwater Management Act’s Portal](#). While the organizations are concerned with increasing sustainability of groundwater sources, the Total Water Use dataset they provide takes account of all water sources and water usage sectors, which makes it a good fit for our research.

Since the reported values are measured rather than the true values, so we are limited by the accuracy of the measurement tools, but hopefully the large number of samples should reduce the effect on the final result.

Each record represents one annual report from one basin. All water usage and water sources are measured in acre-feet. The variables of interest for this paper are labelled “WUS_URBAN” and “WUS_AGRICULTURAL”.

“WUS_URBAN” stands for “Water Use Sector Urban”, and represents the total volume of water in acre-feet that was applied to the urban sector.

“WUS_AGRICULTURAL” stands for “Water Use Sector Agriculture”, and represents the total volume of water in acre-feet that was applied to the agricultural sector.

There are other variables in this dataset tracking the volume of water applied to other sectors. These were used in aid of cleaning. 24 annual reports were excluded as reported volume of water was 0 for all sectors. 23 of these observations were cross-referenced with groundwater source and use measurements from the [Groundwater Extractions dataset](#) provided as part

of the Groundwater Sustainability Plan. Entries with measurements for water use in the Groundwater Extractions dataset but not in the Total Water Use dataset were dropped. This included:

- * 2021-2024 reports for subbasin 4-002
- * 2022-2024 reports for subbasin 4-013
- * 2022-2024 reports for subbasin 5-022.16
- * 2021 and 2024 reports for subbasin 6-005.01
- * 2022-2024 reports for subbasin 7-021.04
- * 2017-2024 reports for 7-024.01

Finally, the 2024 report for subbasin 5-004 was excluded because it had no entries for water use in either dataset, which didn't match the reports of previous years.

Some outliers with high values for urban water usage and extremely low values for agricultural water usage were excluded, as it was discovered that all of these records were from subbasin 8-001, which is a coastal, urbanized area. This is not the type of area relevant to our research.

The subbasin numbering system referenced is standard in California.

After removing these entries, we still have 450 reports from 2017-2024.

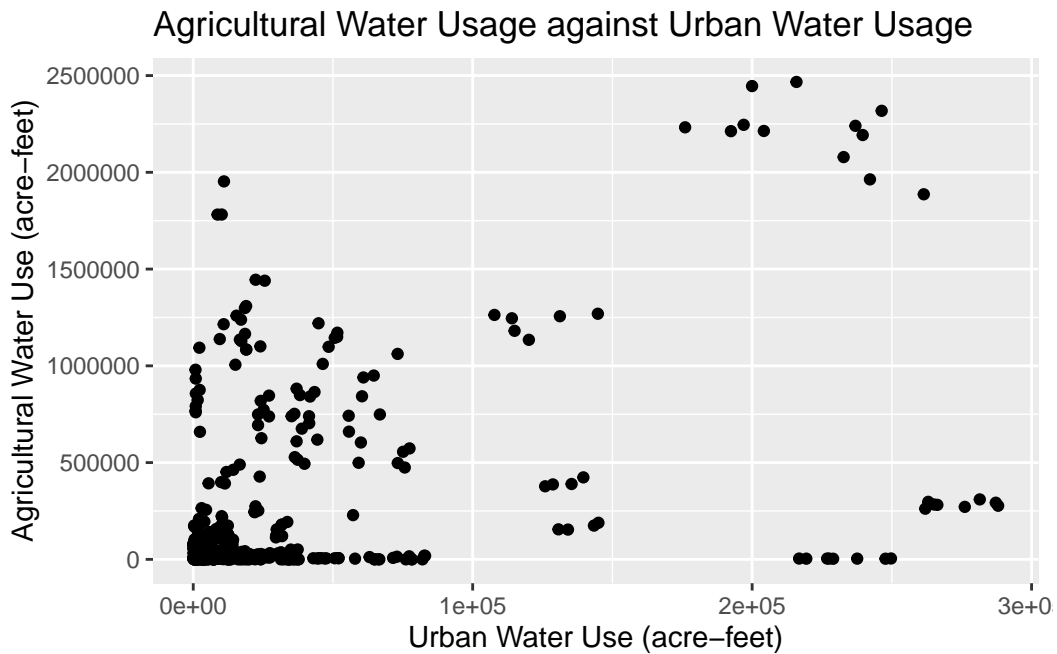


Figure 1: Scatter plot of urban water usage (x-axis) and agricultural water usage (y-axis). Both are measures of volume in units of acre-feet. The plot shows that there are many points in a vaguely linear fashion, though there are also many points at higher urban water use that may give our fit some trouble.

3 Model

We will be using a linear regression model with estimator equation:

$$Y_i = b_0 + b_1 * X_i$$

Where Y_i represents the estimated value for the agricultural water usage in acre-feet, X_i represents the measured value for the urban water usage in acre-feet, b_0 is the estimator of the value of the agricultural water usage in acre-feet when we measure no urban water usage, and b_1 is the estimator of the change in agricultural water usage when we see a unit change in urban water usage(X_i). This analysis has been implemented using the R-language (R Core Team 2025).

4 Results

After building our linear regression model, we end up with the following results:

$$b_0 = 145890.6 \text{ and } b_1 = 3.518$$

From our model, we would expect to measure around 145890.6 acre-feet of agricultural water use when no urban water use is measured. For every acre-foot of urban water use measured, we should expect to see an additional 3.518 acre-feet of agricultural water use measured.

Because the slope is so small relative to the scale of our variables and the equation obtained from the linear regression is a poor fit for the points (as seen in Figure 2), we performed a hypothesis test using the null hypothesis $\beta_1 = 0$ and alternate hypothesis $\beta_1 \neq 0$.

We used the test statistic $t^* = \frac{b_1}{s\{b_1\}} = \frac{3.518}{0.346} = 10.175$. We used a 95% confidence interval, so our test statistic will need to be compared with $t(1 - \frac{\alpha}{2}; n - 2) = t(0.975; 448) = 0.835$. We can confirm that $|t^*| = 10.175 > 0.835$, so we reject the null hypothesis $\beta_1 = 0$.

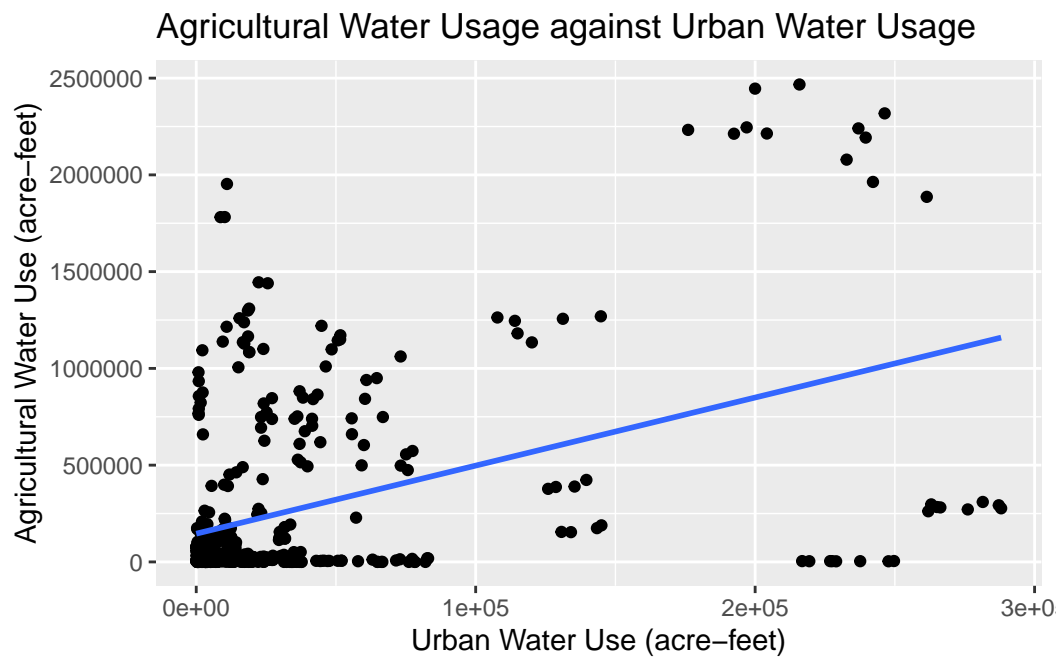


Figure 2: Scatter plot of urban water usage (x-axis) and agricultural water usage (y-axis) once again, with the linear equation found by the regression plotted. Both axes measure volume in units of acre-feet.

By plotting the model residuals against X_i , we receive Figure 3.

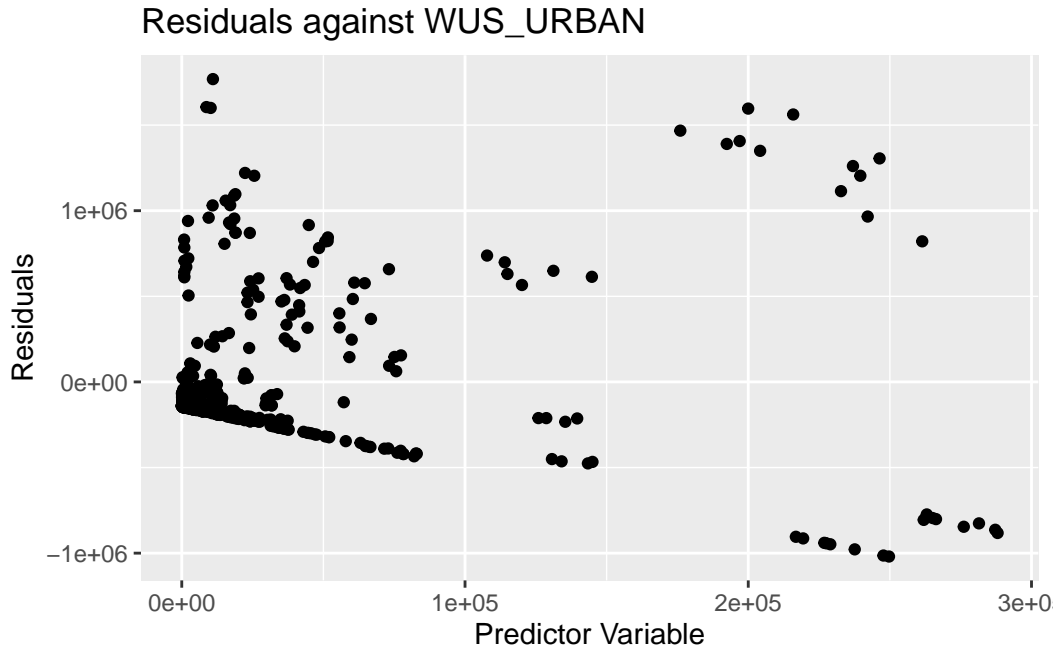


Figure 3: Scatter plot of the predictor variable (x-axis) and residuals(y-axis) for a simple linear regression with urban water usage as the predictor variable and agricultural water usage as the response variable.

Were our linear regression model appropriate for the data, we would see a fairly even spread of the data points around the residuals = 0 line. However, that isn't what we observed. There are several oddities:

1. Most of the points trend towards smaller values of the predictor variable.
2. There is a strong linear trend among some of the residuals.
3. Most of the points are negative. This is especially noticeable at the higher values for urban water usage.
4. There are some distinct clusters among the residuals.

5 Discussion

In this paper, we sought to determine if we could use measurements of urban water use to predict measurements of agricultural use. We started with data provided by the Department of Water Resources as part of the Sustainable Groundwater Management Plan. Then, we fit a linear regression model with urban water usage as the predictor variable and agricultural water usage as the response variable.

When examining the results, we could note a general positive relation between the two variables, confirmed by our hypothesis test. However, an examination of the residuals has shown that this is not a good model to use if we want to predict the agricultural water usage with the urban water usage.

In Figure 3, we noted many oddities, such as the significant number of negative residuals and a linear trend among the negative residuals. What is extremely odd, however, is that the scatter plot of the residuals against the predictor variable almost looks like a skewed version of Figure 1.

We did also observe some clusters. In the future, we can add variables to our model to try to explain some of the grouping in the residuals. Research into the potential relation between urban water usage and agricultural usage has brought up several potential variables not included in this dataset, such as population and attitudes and legislation related to drought. Trends in urban water use and agricultural water use have been found in relation to both population and cultural attitudes post-drought. Urban water use has stayed relatively constant, despite growing populations (Mount, Hanak, and Peterson 2019). Agricultural water use also hasn't seen as much change as expected (Peterson et al. 2023).

In conclusion, the measured urban water usage alone is not enough to predict agricultural water use. However, we may be able to leverage other variables such as drought-related attitudes to create a model that can give us a more accurate estimate of the agricultural water usage we can expect to see.

References

- ABC10, and Rosemary Knight. 2023. “California’s Groundwater Systems, Explained | Extended Interview.” *YouTube*. ABC10. <https://youtu.be/RylAOmJfMHk?si=CNx3xjbP6RJyrPem>.
- “California Water Use.” 2018. USGS. <https://www.usgs.gov/centers/california-water-science-center/science/california-water-use>.
- Cooley, Heather. 2020. “Urban and Agricultural Water Use in California, 1960-2015.” Pacific Institute. https://pacinst.org/wp-content/uploads/2020/06/PI_Water_Use_Trends_June_2020.pdf.
- “Groundwater Sustainability Plans.” 2018. water.ca.gov. <https://water.ca.gov/Programs/Groundwater-Management/SGMA-Groundwater-Management/Groundwater-Sustainability-Plans>.
- Mount, Jeffrey, Ellen Hanak, and Caitlin Peterson. 2019. “Water Use in California.” Public Policy Institute of California. <https://www.ppic.org/publication/water-use-in-california/>.
- PBSNewsHour. 2025. “California Farms Face Pressure to Boost Efficiency as Water Supply Declines.” *Youtube*. PBS NewsHour. <https://youtu.be/9NvxwnhJS4s?si=iXps00QrI6dUPfJ1>.
- Peterson, Caitlin, Alvar Escriva-Bou, Josue Medellin-Azuara, and Spencer Cole. 2023. “Water Use in California’s Agriculture.” Public Policy Institute of California. <https://www.ppic.org/publication/water-use-in-californias-agriculture/>.
- R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.