

Predicting Agricultural Water Usage in California*

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

Catharina Castillo

September 23, 2025

“More specifically, a good recipe for an abstract is: first sentence: specify the general area of the paper and encourage the reader; second sentence: specify the dataset and methods at a general level; third sentence: specify the headline result; and a fourth sentence about implications”

1 Introduction

Historically, it has been harder to measure agricultural water usage than urban water usage. The infrastructure used in urban settings means water was able to be measured directly using metering. The agricultural sector was harder to measure, as it could not be done directly. Instead, land use surveys were 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 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)

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

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 accuracy of those measurement methods.

In Section 2, we will explain more about the data. In Section 3, we will introduce the model used. Finally, in Section 4 we will lay out our results and in Section 5 we will discuss the implications and future steps. While we have not found that measured urban water usage is a good predictor of measured agricultural water usage, this experiment does still provide interesting revelations.

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. These reports were originally submitted to the Department of Water Resources through the Sustainable Groundwater Management Act's Portal.

The data provided is sourced through Groundwater Sustainability Agencies and Alternative Agencies as part of the Groundwater Sustainability Plan. 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.

Of course, the limitation is that these are measured values rather than the true values, so we are immediately limited by the accuracy of the measurement tools.

The link to the data is: https://data.ca.gov/dataset/groundwater-sustainability-plan-annual-report-data/resource/d361ff20-b713-4f57-b3a4-65400933b0a1?inner_span=True

Each record represents one annual report from one basin. These annual reports are filed through the SGMA portal: <https://sgma.water.ca.gov/portal/>. 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. Each individual entry represents the amount for the report year in the basin area.

"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. As with "WUS_URBAN", each individual entry represents the amount for the report year in the basin area.

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 from the final model, as there were no values inputted for water use in any sector. 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: <https://data.ca.gov/dataset/groundwater-sustainability-plan-annual-report-data/resource/36b030d4-9fc1-43dd-86e8-5451b0e41374> Entries with measurements for water use here but not in our original 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, and 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.

We did also have some outliers with high values for urban water usage, and extremely low values for agricultural water usage. Upon examination, it was discovered that all of these records were from subbasin 8-001, which is a coastal, urbanized area. All the records for this basin have been excluded as it's not relevant to our research question.

The subbasin numbering system referenced is standard in California.

After removing these entries, we still have 450 reports from 2017-2024, though most are from the last 4 years.

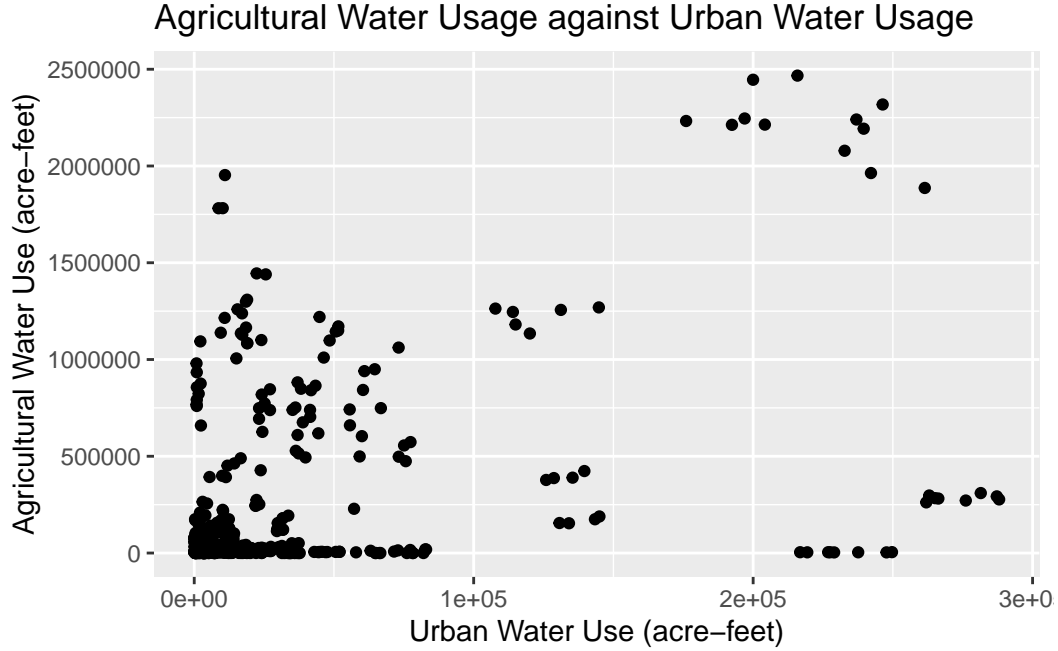


Figure 1: Scatter plot of urban water usage (x-axis) and agricultural water usage (y-axis). Both are measures of volume in units of acre-feet. The plot shows that there are many points in a vaguely linear fashion, though there are also many points at higher urban water use that may give our fit some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

3 Model

We will be using the [normal errors?] 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. Finally, 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:

$b_0 = 145890.6$ and $b_1 = 3.518$

From our model, we would expect to see the agricultural water usage measurement around 145890.6 acre-feet when we measure no urban water usage. Then, for every acre-foot of measured water usage in the urban sector, we should expect to see an additional 3.518 acre-feet of measured water usage in the agricultural sector.

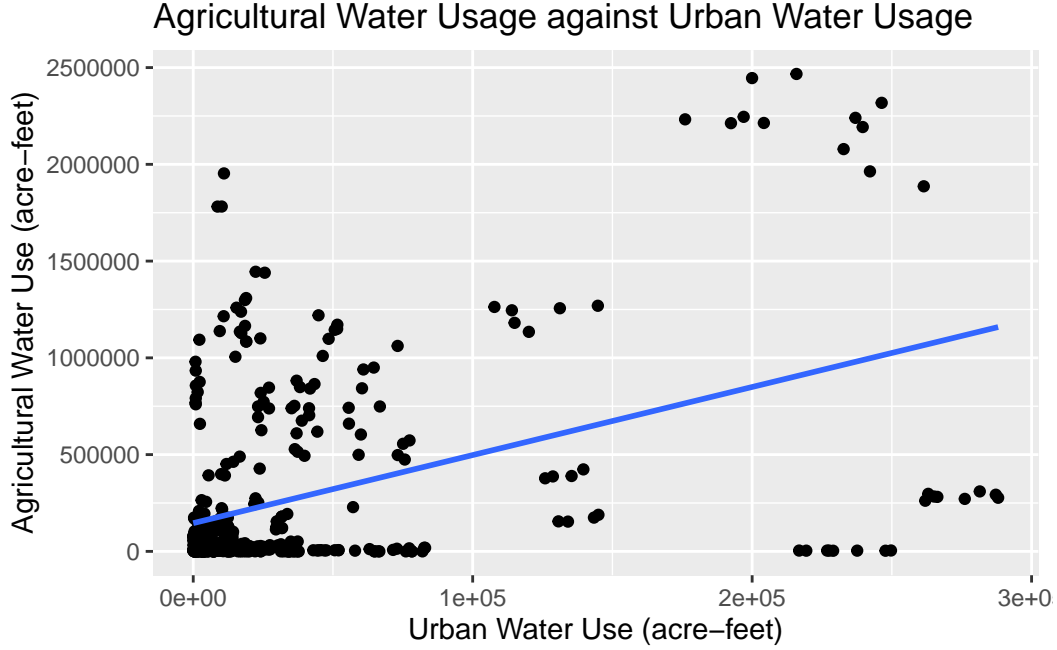


Figure 2: Scatter plot of urban water usage (x-axis) and agricultural water usage (y-axis) once again, with the linear equation found by the regression plotted. Both axes measure volume in units of acre-feet.

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, it's worth performing a hypothesis test using the null hypothesis $\beta_1 = 0$ and alternate hypothesis $\beta_1 \neq 0$. We'll use the test statistic $t^* = \frac{b_1}{s\{b_1\}} = \frac{3.518}{0.346} = 10.175$. We'll use 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 $|t^*| > 0.835$, so we can reject the null hypothesis $\beta_1 = 0$.

We can look at a few graphs to determine the accuracy of our model. By plotting the model residuals against X_i , we receive:

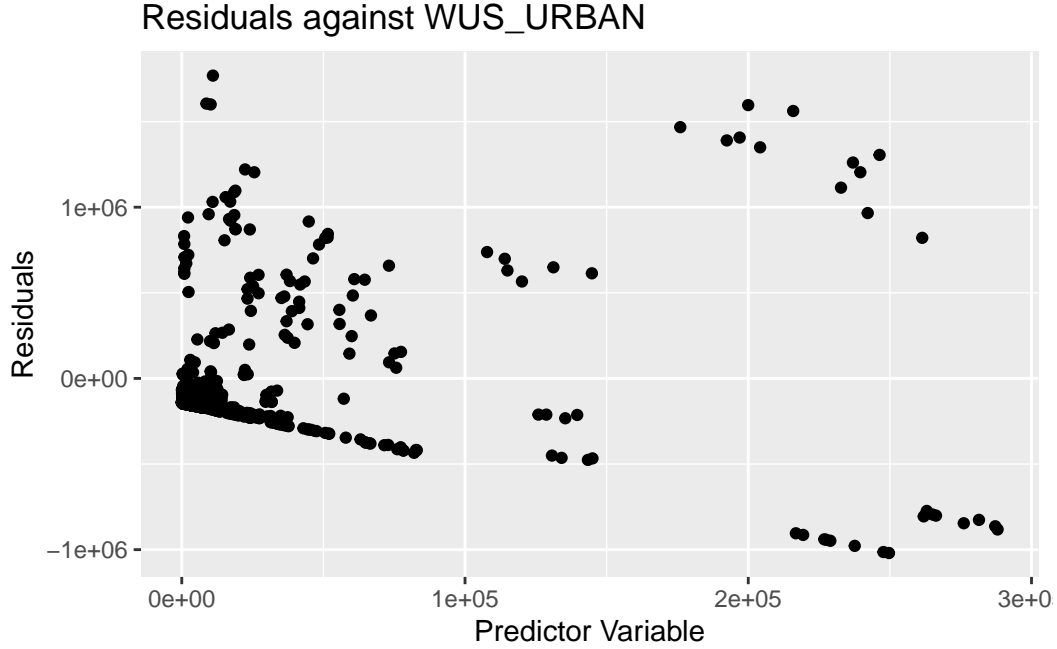


Figure 3: Scatter plot of the predictor variable (x-axis) and residuals(y-axis) for a simple linear regression with urban water usage as the predictor variable and agricultural water usage as the response variable.

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 that have implications about the appropriateness of this linear regression model on our data:

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.

Upon examination of the results, while we could note a general positive relation between the two variables, 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.

By plotting the residuals against the predictor variable, 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 our scatter plot of the observed agricultural water usage and urban water usage.

We did also observe some clusters, which might mean we can add variables to our model 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 the effect of drought years on attitudes and legislation. 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 usage also hasn't seen much change (Peterson et al. 2023).

In conclusion, the measured urban water usage alone is not enough to predict the expected agricultural water usage. 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=CNx3xjbP6RJrPem>.
- "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/>.