California Urban and Agricultural Water Usage: an examination of the use of urban water usage to predict agricultural water usage*

Subtitle

Catharina Castillo

September 20, 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"

Introduction

Historically, it has been harder to measure agricultural water usage than urban water usage. The infrastructure used in urban settings makes metering easy, while in the agricultural sector it has been a bit harder. Irrigation alone has been noted as one of the more unreliable measurements[3], relying on satellite images to estimate land usage.

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 [2]

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. [sgma]

While technological improvements have led to increased accuracy in measuring agricultural water usage, this paper seeks to explore if we can use the measured urban water usage to estimate the agricultural water usage. While this can't replace direct measurements of agricultural water usage, an estimate can help assess the confidence in the data collected.

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

[concise explanation of the data?]

We will then fit a linear regression model to model a way to estimate the agricultural water use from the urban water use.

Once we have our results, we will discuss any conclusions we can make

Data

- provide graphs showing data behavior
- tables providing summary statistics can be useful as well
- all figures and graphs should be self-contained

We will begin by using the total water use dataset provided by the Groundwater Sustainability Plan Annual Report datasets submitted to data.ca.gov by Groundwater Sustainability Agencies and Alternative Agencies. These reports are originally submitted to the Department of Water Resources through the Sustainable Groundwater Management Act's Portal.

This data also

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. All water usage and water sources are measured in acre-feet. More information can be found in the data dictionary on the page.

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 for the basin in question. X_i represents the measured value for the urban water usage in acre-feet for the basin. 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).

Results

After building our linear regression model, we end up with the following results for b_0 and b_1: #### SUBJECT TO CHANGE ONCE FINISHED CLEANING THE DATA $b_0 = 192974$ $b_1 = 1.64$

From our model, we would expect to see the agricultural water usage measurement around 192974 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 1.64 acre-feet of measured water usage in the agricultural sector.

Before relying on our model, however, we need to examine the accuracy.

[yada yada look at standard deviation? other summary statistics]

We can look at a few graphs to determine the accuracy of our model. By plotting the model residuals against X_i, we receive the following graph. [insert graph]

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. First, most of the points trend towards smaller values of the predictor variable. This does indicate a trend towards lower levels of water usage in the urban sector.

Second, there is a strong linear trend among the lower residuals.

Third, most of the points are negative. This shows that while there may be an overall trend of an increase in agricultural water usage with an increase in urban water, most of our errors are still negative. Also notable is that the residuals for higher water usage measurements in the urban sector are all negative.

Finally, we may notice some distinct clusters. This suggests there may be at least one other variable that could account for the residuals and the visibly grouping.

We can look at a plot of the residuals against time, in case there is a pattern there that shows up. [insert time plot]

We can see from the plot that the mean of the residuals stays fairly consistent. The only notable pattern is that for 2017 and 2018, there aren't as many high residuals. However, this can be explained by the fact that there are less annual reports for this year than other years. [check data to cite specific numbers]

We can also look at a boxplot of the residuals and a quantile plot to see how they are distributed [insert box plot of residuals and qq plot]

From the graph we can see that while the residuals do trend towards a mean, we do have a very wide spread, especially towards high positive values. This not only shows that our errors are not normal, but that they are also very large in a significant number of the cases.

From the diagnostic graphs we draw the following conclusions: the relation between Agricultural and Urban water usage is not linear (without maybe considering/controlling for other variables first). The error terms we obtain from a linear regression model built from the two do not have constant variance and are not normal.

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 correlation[check corr] 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.

While no concrete statement can be made whether or not there is a relation between agricultural water usage and urban water usage, by plotting the residuals against the predictor variable we did observe some clusters. These clusters do imply that we may be able to add some variables to our model to explain some of the variance in the residuals. Further research is needed to determine what additional predictor variables might be useful.

Some of the possible explanation for the agricultural water usage not accounted for with this model may be accountable with data not available in this dataset. Research into the potential relation between urban water usage and agricultural usage has brought up several potential variables.

Two potential variables that we might want to examine are population and the affect of drought years and attitudes and behaviors related to droughts. Trends in urban water use and agricultural water use have been found in relation to both these variables. Urban water use has stayed relatively constant, despite growing populations[5]. Agricultural water usage also hasn't seen much change[6].

In conclusion, the urban water usage alone is not enough to predict the expected agricultural water usage.

References

[1] "Groundwater Sustainability Plans," Ca.gov, Apr. 18, 2018. https://water.ca.gov/Programs/Groundwater-Management/SGMA-Groundwater-Management/Groundwater-Sustainability-Plans

[2] Youtu.be, 2025. https://youtu.be/9NvxwnhJS4s?si=xGyeBn9BYvI4z0aS (accessed Sep. 21, 2025).

[3] "California Water Use," USGS, Nov. 2023. https://www.usgs.gov/centers/california-water-science-center/science/california-water-use#overview (accessed Sep. 21, 2025).

[4] Youtu.be, 2025. https://youtu.be/RylAOmJfMHk?si=ONCDCzWZZzFKPulO (accessed Sep. 21, 2025)

[5] E. Hanak and J. Mount, “Water Use in California - Public Policy Institute of California,” Public Policy Institute of California, May 2019. https://www.ppic.org/publication/water-use-in-california/

[6] C. Peterson, A. Escriva-Bou, J. Medellín-Azuara, and S. Cole, “Water Use in California’s Agriculture,” Public Policy Institute of California, Apr. 2023. https://www.ppic.org/publication/water-use-in-californias-agriculture/

[7] J. Mount and E. Hanak, “Water Use in California,” May 2019.