Estimating U.S. subnational freshwater withdrawals by water use category from 1995 to 2021

Sarah Rehkamp, USDA, sarah.rehkamp@usda.gov
James Chandler Zachary, USDA, james.zachary@usda.gov

Selected Paper prepared for presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

This research was supported by the U.S. Department of Agriculture, Economic Research Service.

Introduction

Freshwater is a natural resource that is increasingly stressed, thus motivating research on the spatial and temporal patterns of water use in the United States (Lee, Xu, Daystar, Elgowainy, and Wang 2019) and coverage in the popular press (e.g., Uncharted Waters Series, New York Times 2023). Data from the World Resources Institute show that the United States currently faces medium-high water stress and is projected to face this level of stress in the future, measured as the ratio of demand over renewable water supply (Kuzma, Saccoccia, and Chertock 2023). Globally, the United States is a top user of freshwater (Gleick et al. 2014), both in total and per capita use. This water is used throughout the U.S. economy, dominated by irrigation and thermoelectric uses (DOI-USGS 2023).

The U.S. Geological Survey's National Water Information System (NWIS) is a hub for water data in the United States and a source for understanding water use (DOI-USGS 2023). Data on water use are published in five-year intervals by water use category. The last data release comprehensively covering the categories was 2015 and the delay in data publication limits their ability to inform current water issues.

In this paper, we aim to address this gap by producing a U.S. subnational, annual dataset from 1995-2021 by nine water use categories; a dataset that has not been available previously along the same dimensions. This work speaks to several of the key challenges in environmental and resource economics put forth by Bretschger and Pittel (2020), but specifically the dynamics of the economic-ecological system. In this context, these data may be useful in research applications such as annualizing water use estimates linked to U.S. food production (Rehkamp, Canning, and Birney 2021), or understanding virtual water flows which have been looked at for the United States (Mubako et al. 2013) and internationally (Han et al. 2023). Furthermore, more frequent and recent time series data could inform decisionmakers where and by whom water is being used, indicating areas of potential regional stress or scarcity and, thus, appropriate management practices.

The paper is organized as follows. First, we describe background research that informs our work. Secondly, we present the data and our methodological approach. We organize our approach in three steps: 1) develop regression models to better understand factors that may be correlated with water use, 2) forecast annual water use to generate initial estimates, and 3) adjust the forecasts with a mathematical programming model to generate final estimates. We then present our results, compare the final estimates to other relevant data in the discussion section, and, finally, conclude.

Background

The U.S. Geological Survey (USGS) provides data on freshwater withdrawals in their eight main categories from both groundwater and surface water sources. Table 1 summarizes these data for the five most recent years available, aggregated to the national-level. Over time, thermoelectric and irrigation are the major water use categories, combined making up over three-fourths of total water withdrawals.

Table 1 Freshwater withdrawals by USGS water use category, cubic kilometers (km^3) per year

	1995	2000	2005	2010	2015
Aquaculture	4.49	8.00	12.20	12.35	10.42
Domestic	4.66	4.94	5.15	4.87	4.49
Industrial	27.56	25.16	23.31	20.99	19.34
Irrigation	179.97	191.93	175.94	159.54	162.77
Livestock	3.15	3.26	2.96	2.75	2.75
Public Supply	54.89	59.05	60.34	56.60	52.69
Mining	3.28	3.57	3.14	2.71	2.61
Thermoelectric	182.39	185.62	197.48	162.88	131.44
Total	460.38	481.53	480.51	422.70	386.51

Source: DOI-USGS (2023) and authors' calculations

In 2002, the National Research Council (NRC) published an evaluation of the USGS National Water-Use Information Program. The NRC's report documents the history of the water data (i.e., data are both estimated and provided by the agencies and parties that USGS works with), provides a review, and makes recommendations going forward in their estimation and publication.

Dziegielewski et al. (2002) analyze water use from 1950 to 1995, although the data are more extensive between 1985 and 1995. The authors estimate eight structural models with ordinary least squares (OLS) using total water withdrawals (fresh plus saline) from each category as dependent variables and relevant subsets of water use determinants as independent variables. In some cases, log-linear models utilize water-use rates as independent variables to predict per-capita water withdrawals.

The analysis in Dziegielewski et al. (2002) uses state-level data on socio-economic, demographic, weather, water access rights, labor force, land use, and water source variables, some of which are publicly available through federal sources and others are available only by purchase such as the price of water or the price of pumping. The results of Dziegielewski et al. (2002) document that water demand can be predicted using OLS. The results show that the main drivers of water (i.e., population, employment, irrigated area, and power generation) explain most of the variation in per unit water withdrawals in most categories. The outlier analysis further demonstrates that state and state-year combinations can also be significant predictors. The water-use trends documented in the analysis show increasing aggregate water use in most categories for the time period studied, with some leveling off after 1980.

Brown (2000) and Brown et al. (2013) offer projections of future water withdrawals using differing methods. First, they create projections by multiplying total demand units, such as population or irrigated acres, by a single per-unit use factor for each category. Growth in each demand factor is modelled using an annual rate of change and a rate of change decay. They do not model the demand factors or the underlying relationships to withdrawals.

Using these data and other sources on water, there has been research estimating demand or forecasting future use of U.S. freshwater. Climate scenarios are explored in several papers as a driver of water demand in the United States (Miller et al. 2021; Warziniack et al. 2021) and particularly irrigation demand (Nie et al. 2020), the nation's largest user of consumptive water (Sowby and Dictaldo 2022). Sun et al. (2008) focus on water stress in the Southeastern United States and explore the effects of population and land use/land cover, in addition to climate, on water availability. Franczyk and Chang (2009) conduct a spatial analysis water for counties in Oregon between 1985 to 2005. They conclude that the spatial patterns of total water withdrawals across the state are determined, in large part, by climate variability.

As of the writing of this paper, the USGS published new data and modeling approaches that generated the data for selected water use categories (irrigation published October 31, 2023, public supply published November 1, 2023, and thermoelectric published October 31, 2023), including data for 2020 and earlier years at an increased frequency (monthly and annually) and detail (e.g., power plant level). These data represent a reanalysis of historical water use data meaning USGS undertook a "process of reevaluating and recalculating water-use data using updated or refined methods, data sources, models, or assumptions" (Galanter 2023).

USGS's reanalysis uses different approaches for each category of water withdrawals and predicts volumes at the watershed boundary level for 87,020 12-digit hydrologic unit codes (HUC12) in the conterminous U.S. over different time periods. For example, public supply volumes are predicted using the XGBoost algorithm and ensemble learning methods with socio-economic, demographic, and climate variables, as well as variables related to housing characteristics.

The above literature contributes to our understanding of water withdrawals in the United States over time. We build on work that has been done to estimate U.S. freshwater withdrawals between 1995-2021.

Materials and Methods

This section documents our empirical analysis.

Data

This paper uses data from the U.S. Geological Survey's NWIS (DOI-USGS 2023). We use water data for the water use categories, or synonymously, water withdrawal categories: aquaculture, domestic, industrial, livestock, mining, public supply, and thermoelectric and water use subcategories for irrigation: crop irrigation and golf irrigation. The subnational data (either at the county- or state-level) not only account for heterogeneity in water use, but, practically, these data also increase our observations compared to using the national data. These data are published every five years and we use data from 1995 to 2015, as 2015 is the most recent year of comprehensive data published. These data are our dependent variables.

To maintain transparency and reproducibility, we exclusively use publicly available data. Data choices were limited to where we had annual, subnational data over time. The data sources we chose for the regressors in each model are presented in Table 2. Transformations to the data inputs are described below.

The Quarterly Census for Employment and Wages (QCEW) is a rich dataset that is available along the dimensions of our analysis, but there are data suppressions for employment, especially at narrower North American Industry Classification System (NAICS) levels or geographies. However, the QCEW establishment data are complete. For mining, we record the data source as eQCEW, indicating that these employment data are enhanced to estimate the suppressions that exist even at the 2-digit NAICS level for states. There are suppressions in four states (Alaska, Delaware, Maine, and Rhode Island) and we estimate these suppressions by taking the average of the two "bookend" values, so the datapoints for the year before and after the suppression.

Net electricity generation by thermoelectric fuel source data are derived from DOE-EIA (2023). The thermoelectric fuel sources included are coal, geothermal, nuclear, other, other biomass, other gases, petroleum, solar thermal/photovoltaic, wood/wood-derived fuels.

Since the regional, real gross domestic product (GDP) data are NAICS-based, we impute the data for 1995 and 1996 where the data are based on the Standard Industrial Classification (SIC) system. We multiply the real GDP in chained 2012 dollars in 1997 by a quantity index for 1995 and 1996 when using real GDP of industrial-classified industries.

The population variable, *Pop*, is computed in two steps. First, using the USGS data on population served by public supply and total population, a percent of population served is computed for every five-year epoch. Next, this percentage is applied to the Census estimate of intercensal population for each of the four years following the reference year (see DOC-USCB references for each year of data). For example, the 1995 percentage as computed from the 1995 USGS data is applied to the Census estimates of population for the years 1996–1999. This is repeated for each five-year epoch in the data.

Table 2 Data Sources Summary Table

Regression	Geographical level	Variable (unit)	Variable in equations below	Source
Aquaculture	State	None		
Domestic	State	None		
Irrigation – Crop	State (excluding Hawaii)	Temperature (average degrees Celsius)	*temp*	NOAA (2023)
		Precipitation (millimeters)	*precip*	NOAA (2023)
Irrigation – Golf	State	Establishments (number by NAICS 713910)	*estab*	BLS-QCEW (2023)
		Employees (number by NAICS 713910)	*emp*	BLS-QCEW (2023)
Industrial		Regional real GDP for industrial industries (NAICS 23- and 3-; millions of chained 2012 dollars)	*GDP*	BEA (2023)
Livestock	State	Cattle inventory	*cattle*	USDA-NASS (2022)
	State	Establishments (number by NAICS 112-, excluding 1125-)	*estab*	BLS-QCEW (2023)
Mining	State	Employees (number by NAICS 21-)	*emp*	BLS-eQCEW (2023)
Public Supply	County	Population served by public supply sources	*pop*	Census Bureau and authors' calculations
Thermoelectric	State	Net electricity generation by thermoelectric fuel source (MWh/year)	*elec*	DOE-EIA (2023)
		Precipitation (millimeters)	*precip*	NOAA (2023)
All	State or county	Water withdrawals (cubic kilometers)	Left-hand side variables	USGS (2023)

Notes: All variables are annual and the years of data used for the x-variables are 1995, 2000, 2005, 2010, and 2015 for the regression analysis, corresponding with the water data available from USGS that are used as dependent variables. The years of data used for the independent variables are 1995 through 2021 for the forecasting. Temperature and precipitation were converted from their original units of average degrees Fahrenheit and inches, respectively, to average degrees Celsius and millimeters (mm). Water withdrawals were converted from their original units of million gallons per day to cubic kilometers per year.

Methodological steps

We summarize our methodological steps below:

1. We first estimate regressions for each of the water use categories where freshwater withdrawal is the dependent variable. For each regression, we tried several model specifications. The chosen model specification for each water use category is presented in the section below.
2. Once the regressions are estimated, we developed forecasts using annual data between 1995 and 2021 for each water use category which we call initial estimates.
3. We then adjust the forecasts using a mathematical programming model for final estimates by water use category.

Step 1: Regression analysis

To develop the nine regression specifications presented below, we experimented with many variables and data scales following the framework for estimating water use in the United States outlined in the 2002 NRC publication. We also consider and build on what others have done in the literature (Dziegielewski et al. 2002; Franczyk and Chang 2009; Nie et al. 2020; Warziniack 2022).

We used OLS regressions for ease of interpretation of the results and generally level variables for ease of forecasting, rather than a per capita ratio as others have used (Dziegielewski et al.

2002; Warziniack 2022). For each regression, we included different variables and functional forms based on economic significance, choosing the models we determined generated the best results based on review of the statical significance and forecast reliability.

We generally tried precipitation and temperature data in the regressions, operating under the supposition that weather affects withdrawal rates. We also tried to include employment, establishment, or real GDP to get at the size of the industry over time, or similarly population for the public supply model. The expectation is that the bigger the industry, the more water used. Because establishment sizes can vary, employment or real GDP may be a better measure to capture this, although the employment data are peppered with suppressions while the establishment data are not (BLS-QCEW, 2023).

In all of the following model specifications,

> *c* is a subscript for county (n = 3139 where c = {1, …, 3139} representing all U.S. counties)
>
> *s* is a subscript for state (n=50 where s = {1,…,50} representing U.S. states)
>
> *t* is a subscript for year (n=27 where t = {1,…,27} representing years 1995-2021)
>
> *year* is a time trend
>
> *D* represents state dummy variables

In all of the models presented below, we include a linear time trend to capture the relationship between water withdrawals and time. We also include state indicators to account for within-state heterogeneity that is unobservable and constant over time, using Alabama as the base

(Dziegielewski et al. 2002; Suits 1984). The other variables included and their data sources are defined in Table 2.

Equation 1 Regression model specification for aquaculture freshwater withdrawals

$$\text{aquaculture}_{s,t} = \alpha + \beta year + \gamma year^2 + \sum \delta_s D_s + \varepsilon_{s,t}$$

Aquaculture data were sparse along the dimensions needed for forecasting, so the only variables we included in this regression were dummy variables for the states and time trends. We tried other data such as the employment or establishment data, but we would have had an unbalanced panel due to the growth of this industry over the time period of analysis.

Equation 2 Regression model specification for domestic self-supplied freshwater withdrawals

$$\text{domestic}_{s,t} = \alpha + \beta year + \gamma year^2 + \sum \delta_s D_s + \sum \lambda_s (D_s * year) + \sum \theta_s (D_s * year^2) + \varepsilon_{s,t}$$

Following Dziegielewski et al. (2002), we used state dummies and linear time trends, along with a quadratic trend to capture the curvature observed in the USGS data.

Equation 3 Regression model specification for crop irrigation freshwater withdrawals

$$\text{irrigation\_crop}_{s,t} = \alpha + \beta temp_{s,t} + \gamma precip_{s,t} + \delta year + \sum \lambda_s D_s + \sum \theta_s (precip_{s,t} * D_s) + \varepsilon_{s,t}$$

Weather variables precipitation and temperature were included and precipitation was also interacted with the state dummy variables. There are 245 observations in this regression instead of 250 (50 states by five years); there are no precipitation data for Hawaii (NOAA 2023).

Equation 4 Regression model specification for golf irrigation freshwater withdrawals

$$\text{irrigation\_golf}_{s,t} = \alpha + \beta estab_{s,t} + \gamma emp_{s,t} + \delta year + \lambda year^2 + \sum \theta_s D_s + \varepsilon_{s,t}$$

We used establishments and employment for NAICS 713910: golf courses and country clubs to represent to represent the size of the industry over time. There are 245 observations in this regression instead of 250 (50 states by five years); there are no precipitation data for Hawaii (NOAA 2023).

Equation 5 Regression model specification for industrial freshwater withdrawals

$$\text{industrial}_{s,t} = \alpha + \beta GDP_{s,t} + \gamma year + \sum \delta_s D_s + \sum \lambda_s (GDP_{s,t} * D_s) + \varepsilon_{s,t}$$

We used regional real GDP to represent the size of the industrial-classified industries over time: NAICS that begin with 23 (construction) and 3 (manufacturing).

Equation 6 Regression model specification for livestock freshwater withdrawals

$$\text{livestock}_{s,t} = \alpha + \beta cattle_{s,t} + \gamma estab_{s,t} + \delta year + \sum \lambda_s D_s + \varepsilon_{s,t}$$

We used cattle inventory from the NASS surveys. Other livestock inventory was tried, but had incompatible datasets (i.e., grouping states with lower inventories together) and cattle have the highest water intake rate (Lovelace 2009). We were also able to use establishments for NAICS beginning with 112 (animal production and aquaculture), but removed the data associated with NAICS 1125 (aquaculture) since these water withdrawals are reported in the aquaculture water use category.

Equation 7 Regression model specification for mining freshwater withdrawals

$$\text{mining}_{s,t} = \alpha + \beta emp_{s,t} + \gamma year + \sum \delta_s D_s + \sum \lambda_s (emp_{s,t} * D_s) + \sum \theta_s (year * D_s) + \varepsilon_{s,t}$$

We used employees for NAICS 21 (mining) and also employees interacted with the state dummy variables in this regression. Additionally, we interacted the time trend and the state dummy variables.

Equation 8 Regression model specification for public supply freshwater withdrawals

$$\text{supply}_{c,t} = \alpha + \beta pop_{c,t} + \gamma year + \delta year^2 + \sum \lambda_s D_s + \theta(pop_{c,t} * year) + \mu(pop_{c,t} * year^2) + \varepsilon_{c,t}$$

We estimated models using county-level data where possible to reduce RMSE and capture local water-use characteristics (Dziegielewski et al. 2002), and this was achievable in the public supply model.

Equation 9 Regression model specification for thermoelectric freshwater withdrawals

$$\text{thermoelectric}_{s,t} = \alpha + \beta elec_{s,t} + \gamma precip_{s,t} + \delta year + \sum \theta_s (elec_{s,t} * D_s) + \varepsilon_{s,t}$$

Here, we used the net electricity generation by thermoelectric fuel sources, precipitation, and net electricity interacted with the state dummy variables. There are 244 observations in this regression instead of 250 (50 states by five years); there are no precipitation data for Hawaii (NOAA 2023) and there is an N/A value for $\text{thermoelectric}_{Washington,1995}$.

Step 2: Forecasts (initial estimates)

In the second step of our analysis, we predict the y-variables at the subnational level. The predictions rely on our estimated coefficients (Table 3) and on annual data between 1995 and 2021 for the same x-variables used to fit the models described above.

Step 3: Mathematical programming (final estimates)

Upon inspection of our prediction results, there were cases when the predicted values were negative or did not predict the published USGS data exactly. Since negative results are nonsensical in this context and we wanted to ensure the integrity of the published data, we used a nonlinear programming model with linear constraints that: 1) enforce the benchmark year estimates to the USGS data, and 2) enforce all the estimates to be greater than zero. The objective function is a least squares problem where we minimize the adjustments to year-over-year percentage change estimates from the first stage regressions and the posteriors are scaled by a ratio of the priors (i.e., our initial year-over-year-change from the regression forecasts).

Equation 10 Mathematical programming model specification for final estimates

$$minimize \sum_{g,t1,t2}\left(\frac{w1_{g,t2} - w1_{g,t1} \times ratio0}{w1_{g,t1}}\right)^2 = \sum_{g,t1,t2}\left(\frac{w1_{g,t2}}{w1_{g,t1}} - ratio0\right)^2$$

subject to $\quad w1_{g,t} = w_{g,t}\ \forall\ t \in B$ and $w1_{g,t} \geq 1$

where the subscripts *g* represents the geography, *t* represents the year = {1995,…,2021}, and *t1* and *t2* represent sequential years.

Then, $w0$ represents the prior estimates, $w1$ are the posterior estimates, w are the published data, $ratio0 = \frac{w0_{g,t2}}{w0_{g,t1}}$, and B = the benchmark years of water data = {1995, 2000, 2005, 2010, 2015}.

This step was done using GAMS (General Algebraic Modeling Software) and using the CONOPT3 solver given the nonlinearities. We ran this same model at the state-level for each of the water use categories, aggregating the initial estimates at the county-level from public supply model. We also ran these models in the water data's source unit of million gallons (Mgal) and then converted the final results to cubic kilometers.

The result of using the mathematical programming is to adjust our initial estimates so that, when there are discrepancies between our forecasts and the published USGS data, our final estimates match the published data in the benchmark years and are greater than zero in all years while minimizing the change between our initial estimates in subsequent years.

Results

Table 3 shows the coefficient estimates for the primary predictor variables for each of the regressions. Robust standard errors were used and reflected in the confidence intervals.

Table 3 Regressions summary

	Aquaculture	Domestic	Irrigation–Crop	Irrigation–Golf	Industrial	Livestock	Mining	Public Supply	Thermoelectric
temp			0.4351^{+}						
			(0.2412)						
precip			0.0002						-4.7448*
			(0.0010)						(1.9608)
emp				0.0000			-0.0000***		
				(0.0000)			(0.0000)		
estab				-0.0004^{+}		0.0001**			
				(0.0002)		(0.0000)			
GDP					-0.0000^{+}				
					(0.0000)				
cattle						0.0000**			
						(0.0000)			
pop								0.0002***	
								(0.0000)	
elec									0.0002^{+}
									(0.0001)
pop × year								0.0000	
								(0.0000)	
pop × year²								-0.0000*	
								(0.0000)	
year	0.0252***	-0.0035	-0.0425*	0.0020*	-0.0076**	-0.0008**	-0.0007		-198.8887***
	(0.0069)	(0.0026)	(0.0182)	(0.0009)	(0.0028)	(0.0002)	(0.0006)		(52.8732)
year²	-0.0009**	0.0001		-0.0001*					
	(0.0003)	(0.0001)		(0.0000)					
Constant	-0.0524	0.0945***	-7.2281^{+}	0.05508**	1.7612***	0.0033	0.1377***	0.0052***	27808.9839*
	(0.0535)	(0.0109)	(4.1955)	(0.0185)	(0.4694)	(0.0127)	(0.0340)	(0.0012)	(11686.6460)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. Obs.	250	250	245	245	250	250	250	15693	244
Std. Errors	HC1	HC1	HC1	HC1	HC1	HC1	HC1	HC1	HC1
R2	0.88	0.95	0.98	0.95	0.97	0.95	0.92	0.75	0.98
R2 Adj.	0.85	0.88	0.97	0.94	0.95	0.94	0.81	0.75	0.96

Note: Robust standard errors for the regressions are in parenthesis. *** = 0.001, ** = 0.01, * = 0.05, $^{+}$ = 0.1.

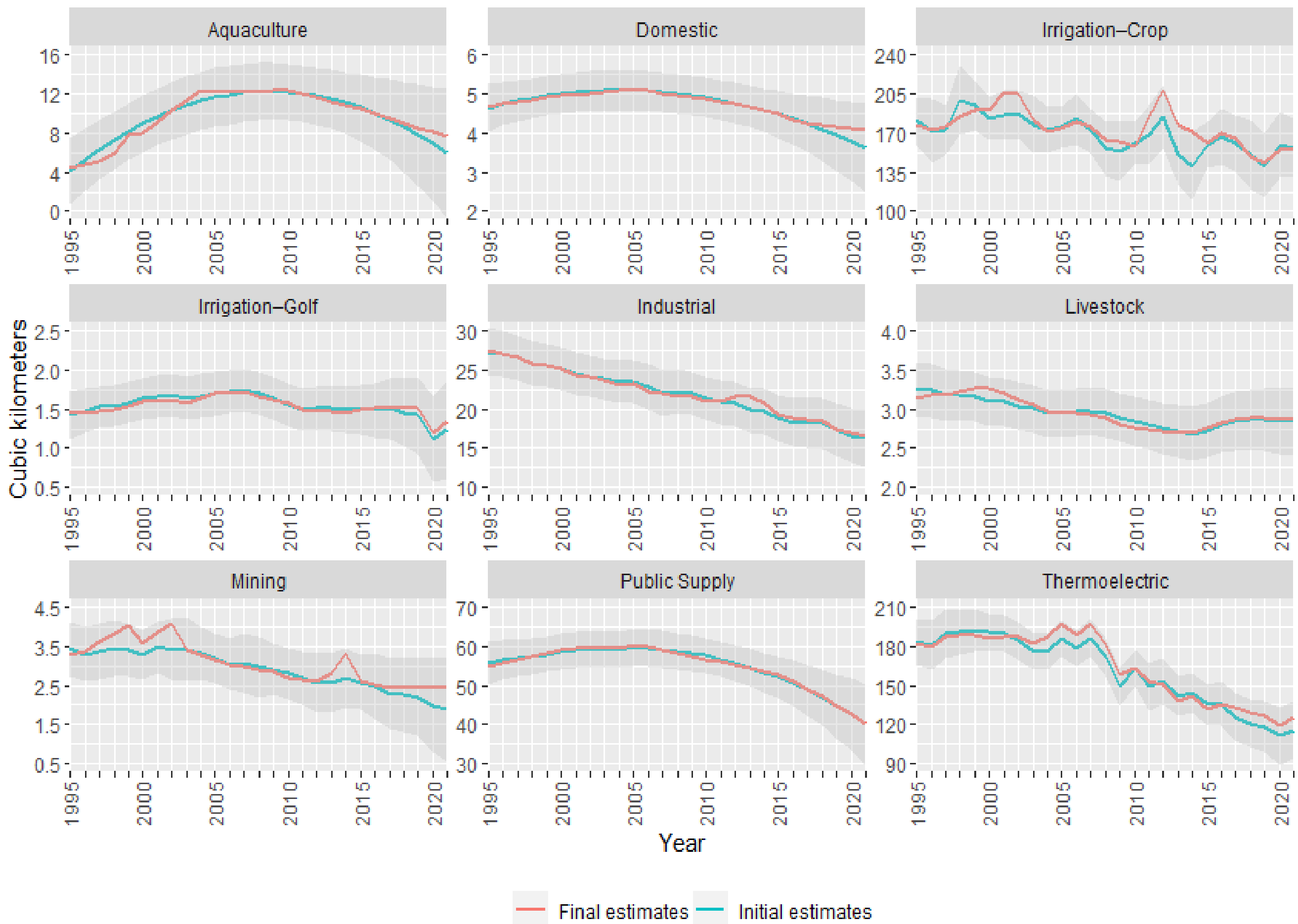
Statistics on the predictions (initial estimates from step 2) for each water use category are presented in Table 4. We generate the root mean square error (RMSE) and mean absolute percentage error (MAPE) to evaluate accuracy.

Table 4 Statistics on the predictions for each water use category

	Root mean square error (RMSE)	Mean absolute percentage error (MAPE)
Aquaculture	0.1578	0.86
Domestic	0.0194	1.00
Irrigation – Crop	0.8727	1.23
Irrigation – Golf	0.0145	0.70
Industrial	0.1269	0.63
Livestock	0.0176	0.62
Mining	0.0232	0.35
Public Supply	0.035	0.98
Thermoelectric	2448.6338	0.88

The prediction results (initial estimates from step 2) and the mathematical programming results (final estimates from step 3) are shown below in Fig. 1. for all water use categories. Although we generate subnational estimates, we present the results at the national level for conciseness. To present the predicted results at the national level, we appropriately adjust the prediction intervals of the state-level results. This adjustment is made by using the estimated coefficients from the models presented to generate predictions, then adjusting the variance of the predictions with the residuals from the estimated model. This adjusted variance is then used to generate a t-distribution for computing new prediction intervals around the aggregate (i.e., national) water withdrawals for each model.

Fig. 1 National initial and final estimates by water use category, 1995-2021

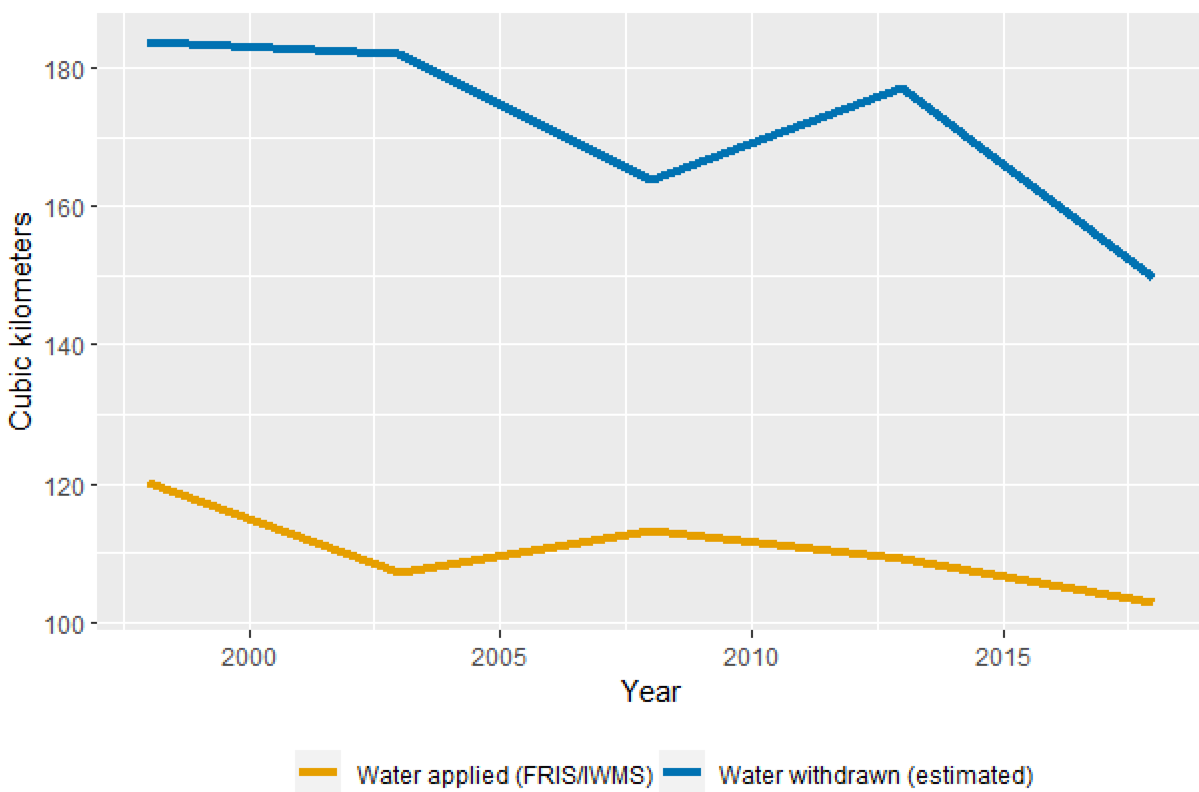


Source: Authors' calculations

Discussion

To evaluate our estimates of crop irrigation water withdrawals, we compare our estimates to published data on water applied from the Census of Agriculture's follow-along survey, the Irrigation Water Management Survey (USDA-NASS 2019b), previously called the Farm and Ranch Irrigation Survey (FRIS) in earlier years (USDA-NASS 2004; USDA-NASS 2014). Water withdrawals are expectedly higher than water applied in each year due to conveyance loss and other factors. Looking at the trends for each series, estimated withdrawals follow applications between 1998-2003 and 2013-2018, but not in 2003-2008 and 2008-2013; this may not be a discrepancy as things such as technology or reductions in loss may contribute to the differences in applications versus withdrawals.

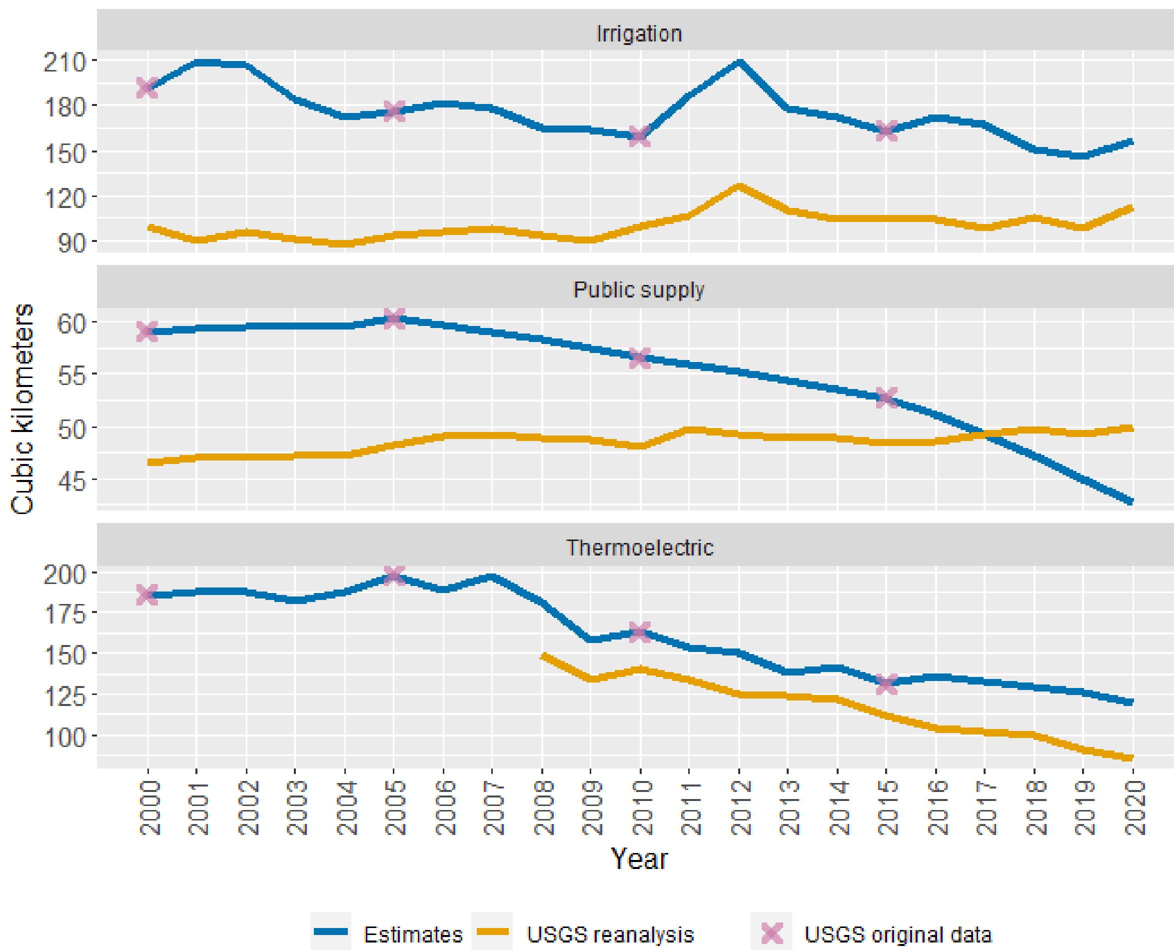
Fig. 2 Comparison between published water applied and estimated water withdrawn in the United States



Sources: Authors' estimates for water withdrawn; USDA-NASS 2019b, USDA-NASS (2004), and USDA-NASS (2014) for water applied.

Next, in Fig. 3., we compare our estimates which overlap with the USGS original data published by NWIS and the recently published USGS reanalysis of water use data for three water use categories using updated methods.

Fig. 3 Comparison between estimated water withdrawals, USGS original data, and USGS reanalysis for three water use categories



Source: Authors' estimates; DOI-USGS, 2023 for USGS original data; Galanter (2023), Luukkonen et al. (2023), and Martin et al. (2023) for USGS reanalysis.

Notes. The USGS original data are the data published in 5-year intervals and exactly match our estimated values due to model construction. The estimated series for irrigation freshwater comparison in the first panel are a sum of our estimated crop and golf irrigation estimates since there was not a breakout in the USGS reanalysis. The thermoelectric withdrawals in the third panel are based on the USGS data with the once-through saline cooling type removed. There was no indication of water type (freshwater or saline water) for the other cooling types.

Data were aggregated to the national-level from their respective geographic scales for comparison in Fig. 3, but the comparisons are not apples-to-apples. First, the USGS reanalysis of irrigation data represent consumptive use, not withdrawals as the original USGS data and our estimates do. Also, these data are an abbreviated time horizon (irrigation 2000-2020; public supply 2000-2020; thermoelectric 2008-2020) compared to the original USGS data. Although there is a difference in magnitude between the new USGS series and our estimates, they follow similar trends.

Our estimate of thermoelectric freshwater water withdrawals in 2008 is 160 km^3. In Averyt et al. (2013), the median estimate thermoelectric water withdrawals in 194 km^3 (converted from the billion gallons reported) and the authors report that 86 percent of these withdrawals were from freshwater sources. Applying the 86 percent to 194 km^3, the resulting 168 km^3 withdrawals is quite close to our estimate.

There are three notable spikes in our mining water estimates (Fig. 1, panel g) in 1999, 2002, and 2014 that are an artifact of our mathematical programming model. Because we match the USGS estimates, but also maintain the year-over-year change from the initial estimates, these cases result in knowingly unrealistic estimates that drive the spikes. For example, the initial estimates of water withdrawals in West Virigina were quite small in 1995 (only 745,559 cubic meters/year), so when the model adjusted this to the actual volume of withdrawals in 1995 from the USGS data (18,516,225 cubic meters/year) and applied *ratio0*, the following for estimates before the next benchmark year (1996, 1997, 1998, 1999) are much higher than what they reasonably were. Beyond 2000, the series smooths out for the state's estimates. We accept this model byproduct because it happens in few cases and mining water use in minor compared to the other primary categories (Fig. 1). As a future improvement, we could consider adjusting model constraints such as imposing a limit on the magnitude of change from the benchmark years.

A limitation of our analysis presented in this paper is data availability for the right-hand side variables. The price of water for relevant categories (e.g., public supply) was not publicly available for the dimensions that we would need for our analysis but could potentially add explanatory power. Aquaculture was a particularly difficult water use category to find data on for our entire time period; there were many zeros when looking at potential variables in the 1990s. Another example is the scale of the irrigated crop water regression; certainly state-level, annual averages obscure weather variations at a local scale and variations within a shorter time interval.

We recognize the challenges and acknowledge the limitations of using a regression approach that are presented by the NRC (2002). There are complex relationships that are not captured or perhaps incorrectly captured (e.g., a time trend actually capturing the effects of a policy change instead of technological progress, explanatory variables that are a function of water use themselves). For example, Chen et al. (2013) show the decreasing freshwater withdrawals per unit of electricity generated over time, due to newer thermoelectric plants and shifts towards different cooling systems.

We use water defined by political geographical boundaries as USGS publishes the data and to correspond with the other variables used in the regression analysis. Also, data defined by geographical boundaries may facilitate a broader set of future research applications. However, geographical units defined by hydrological unit code (HUC) may more accurately capture the physical attributes of water resources.

Future work may consider ways to better capture the changes that occur over time influencing water use such as structural shifts in the economy, water efficiency improvements, and population growth or movement.

Conclusion

Estimating water use is difficult work with complex underlying relationships. This work aims to better understand the drivers of water use and create a useful time series for research applications. We develop regression estimates, forecasts, and mathematical programming models to generate U.S. subnational estimates for nine water use categories between 1995-2021 while maintaining the integrity of the benchmark water use data published by USGS. Having annual data on freshwater withdrawals in the United States may help facilitate time series analysis that previously was not possible with the USGS data in five-year intervals and brings forward dated resource data. These data may inform where and by whom water is being used to identify regions of relative stress or scarcity and, thus, appropriate management practices.

References

Bretschger L, Pittel K (2020) Twenty key challenges in environmental and resource economics. Environmental & Resource Economics 77:725–750. https://doi.org/10.1007/s10640-020-00516-y

Brown TC (2000) Projecting U.S. freshwater withdrawals. Water Resources Research 36(3):769–780. https://doi.org/10.1029/1999WR900284

Brown TC, Foti R, Ramirez JA (2013) Projected freshwater withdrawals in the United States under a changing climate. Water Resources Research 49:1259–1276. https://doi.org/10.1002/wrcr.20076

Bureau of Economic Analysis (BEA) (2023) Regional economic accounts, SAGDP9N real GDP by state. https://www.bea.gov/data/economic-accounts/regional. Accessed 22 August 2023

Bureau of Labor Statistics, Quarterly Census of Employment and Wages (BLS-QCEW) (2023). QCEW NAICS-based data files, annual averages by area. https://www.bls.gov/cew/downloadable-data-files.htm. Accessed 23 January 2023 and 11 April 2023

Campbell AM, Conti SN, O'Dell I (2013). Water resources data, Idaho, 2003, conversion factors. Water-Data Report ID-03-3, U.S. Geological Survey Publication. https://pubs.usgs.gov/wdr/wdr-id-03-3/ID03v3-InsideBackCover.pdf. Accessed 1 February 2014

Chen L, Roy SB, Goldstein RA (2013). Projected freshwater withdrawals under efficiency scenarios for electricity generation and municipal use in the United States for 2030. Journal of the American Water Resources Association 49(1):231-246. https://doi.org/10.1111/jawr.12013

Dziegielewski et al (2002) Analysis of water use trends in the United States: 1950–1995. Southern Illinois University, Carbondale, Illinois, Report on Grant No. 99HQGR0222, USGS National Competitive Grants Program, United States Geological Survey, Reston, Virginia.

Franczyk J, Chang H (2009) Spatial analysis of water use in Oregon, USA, 1985-2005. Water Resources Management 23:755-744. https://doi.org/10.1007/s11269-008-9298-9

Galanter AE, Gorman Sanisaca LE, Skinner KD, Harris MA, Diehl TH, Chamberlin CA, McCarthy BA, Halper AS, Niswonger RG, Stewart JS, Markstrom SL, Embry I and Worland S (2023) Thermoelectric-power water use reanalysis for the 2008-2020 period by power plant, month, and year for the conterminous United States: U.S. Geological Survey data release. https://doi.org/10.5066/P9ZE2FVM. Accessed 29 November 2023

Gleick et al (2014) The world's water: The biennial report on freshwater resources. Volume 8, Table 2. Island Press. Available at: http://worldwater.org/water-data/

Han A, Liu A, Guo Z, et al (2023) Measuring gains and losses in virtual water trade from environmental and economic perspectives. Environmental & Resource Economics 85:195–209. https://doi.org/10.1007/s10640-023-00763-9

Hayes AF, Cai L (2007) Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods 39:709-722

Kuzma S, Saccoccia L, and Chertock, M (2023, August 16) 25 countries, housing one-quarter of the population, face extremely high water stress. *World Resources Institute*. https://www.wri.org/insights/highest-water-stressed-countries. Accessed 13 December 2023

Lee U, Xu H, Daystar J, Elgowainy A, and Wang M (2019) AWARE-US: Quantifying water stress impacts of energy systems in the United States. The Science of the Total Environment 648:1313–1322. https://doi.org/10.1016/j.scitotenv.2018.08.250

Lovelace JK (2009) Methods for estimating water withdrawals for livestock in the United States, 2005. Scientific Investigations Report 2009-5041, U.S. Geological Survey Publication. https://doi.org/10.3133/sir20095041

Luukkonen CL, Alzraiee AH, Larsen JD, Martin DJ, Herbert DM, Buchwald CA, Houston NA, Valseth KJ, Paulinski S, Miller LD, Niswonger RG, Stewart JS, and Dieter CA (2023) Public supply water use reanalysis for the 2000-2020 period by HUC12, month, and year for the conterminous United States: U.S. Geological Survey data release. https://doi.org/10.5066/P9FUL880. Accessed 29 November 2023

Martin DJ, Regan RS, Haynes JV, Read AL, Henson WR, Stewart JS, Brandt JT, Niswonger RG (2023) Irrigation water use reanalysis for the 2000-20 period by HUC12, month, and year for the conterminous United States: U.S. Geological Survey data release. https://doi.org/10.5066/P9YWR0OJ. Accessed 29 November 2023

Miller OL, Putman AL, Alder J, Miller M, Jones DK, Wise DR (2021) Changing climate drives future streamflow declines and challenges in meeting water demand across the southwestern United States. Journal of Hydrology X 11:100074.

Mubako S, Lahiri S, Lant C (2013) Input-output analysis of virtual water transfers: Case study of California and Illinois. Ecological Economics 93:230-238.

National Research Council (2002) Estimating water use in the United States: A new paradigm for the National Water-Use Information Program. Washington, DC: National Academy Press.

New York Times (2023) Unchartered waters: A series on the causes and consequences of disappearing water. https://www.nytimes.com/series/uncharted-waters. Accessed 28 November 2023

Nie W, Zaitchik BF, Rodell M, Kumar SV, Arsenault KR, Badr HS (2021) Irrigation water demand sensitivity to climate variability across the contiguous United States. Water Resources Research 57(3): 2020WR027738.

National Oceanic and Atmospheric Administration, National Centers for Environmental Information (NOAA) (2023). Climate at a glance: Statewide time series, average temperature and precipitation. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series. Accessed 14 June 2023

Regan RS, Markstrom SL, Hay LE, Viger RJ, Norton PA, Driscoll JM, LaFontaine JH (2018) Description of the national hydrologic model for use with the precipitation-runoff modeling system (PRMS): U.S. Geological Survey techniques and methods, book 6, chap B9, 38 p., https://doi.org/10.3133/tm6B9.

Rehkamp S, Canning P, Birney C (2021) Tracking the U.S. domestic food supply chain's freshwater use over time. Economic Research Report No. (ERR-288), U.S. Department of Agriculture, Economic Research Service.

Sowby RB, Dicataldo E (2022) The energy footprint of US irrigation: A first estimate from open data. Energy Nexus 6:100066.

Suits DB (1984) Dummy variables: Mechanics v. interpretation. The Review of Economics and Statistics 66(1):177-180. https://doi.org/10.2307/1924713

Sun G, McNulty SG, Moore Myers JA, Cohen EC (2008) Impacts of multiple stresses on water demand and supply across the Southeastern United States. Journal of the American Water Resources Association 44(6):1441-1457.

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) (2022) Cattle inventory surveys, first of January. https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Cattle_Inventory/index.php. Accessed 22 May 2023

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) (2019a, April 11 release date) Census of agriculture data. Volume 1, Chapter 1: U.S. national level data, Table 1. https://www.nass.usda.gov/Publications/AgCensus/2017/. Accessed 11 December 2023

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) (2019b, November release date) Irrigation water management survey data (2018 and 2013), Table 7. https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Farm_and_Ranch_Irrigation_Survey/index.php. Accessed 26 September 2023

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) (2014, November issue date) Farm and ranch irrigation survey data (2008), Table 6. Accessed

September 26, 2023 from https://agcensus.library.cornell.edu/census_parts/2012-2013-farm-and-ranch-irrigation-survey/

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) (2004 November, issue date) Farm and ranch irrigation survey data (1998 and 2003), Table 8. https://agcensus.library.cornell.edu/census_parts/2002-farm-and-ranch-irrigation-survey/. Accessed 26 September 2023

U.S. Department of Commerce, Census Bureau (DOC-USCB) (1995–2021) Intercensal population and housing unit estimates. https://www2.census.gov/programs-surveys/popest/. Accessed 12 June 2023

U.S. Department of Energy, U.S. Energy Information Administration (DOE-EIA) (2023) Electricity, data, generation and thermal output, State-level generation and fuel consumption data, Annual (back to 1990). https://www.eia.gov/electricity/data.php. Accessed 12 June 2023

U.S. Department of the Interior, U.S. Geological Survey (DOI-USGS) (2023) USGS water use information system: Web interface. https://waterdata.usgs.gov/nwis/wu?. Data accessed 13 November 2018 and 12 May 2022.

Warziniack T, Arabi M, Brown TC, Froemke P, Ghosh R, Rasmussen S, Swartzentruber R (2022) Projections of freshwater use in the United States under climate change. Earth's Future 10(2): e2021EF002222.

Wooldridge JM (2013) Introductory econometrics: A modern approach. 5th edition. South Western, Cengage Learning. Mason, Ohio.