A Panel Analysis of Groundwater Use in California

Jing Ke, Helcio Blum*

Energy Analysis & Environmental Impacts Division, Lawrence Berkeley National Laboratory,

1 Cyclotron Rd., Berkeley, CA 94720, USA

* Corresponding author. E-mail address: hblum@lbl.gov (H. Blum)

ABSTRACT

Groundwater is a relevant source of drinking and agricultural water in many regions of the world, but many aquifers have been unsustainably over-drafted and polluted. This has significant environmental, health, and economic implications. We rely on panel analysis, with small-sample corrections for cluster-robust variance estimation and hypothesis testing, to investigate the dynamics of groundwater extraction. We focus on California, yet our approach could be helpful to analyze the dynamics of groundwater extraction in other groundwater-reliant regions of the world. In California, over-reliance on groundwater has led to significant overdraft, affecting long-term water supply reliability and groundwater pumping costs. It further caused subsidence and infrastructure damage, harmed groundwater-dependent ecosystems, and threatened the sustainability of groundwater resources in the state. We use panel data of the 56 California Water Plan planning areas over the 1998–2015 period. We concentrate on agricultural and urban water use and the major water projects in the state, to provide a better understanding of the relationships between groundwater extraction and water use and supply. Results suggest that reducing agricultural water in Central California and urban water in Southern California could reduce groundwater extraction in these regions by approximately the same amount of the reduced water. Other opportunities to reduce the stress on the groundwater resources in the state are available for other regions, yet with lower benefits. Results also suggest that a decrease in deliveries from the Central Valley Project to the southern part of the Central Valley would increase groundwater extraction by approximately the same proportion. Changes in deliveries from the major water projects in the state, as well as from other sources of surface water, would also have some, yet lower impacts on groundwater extraction in the Central and Southern California.

Keywords:

Groundwater resources, water management, sustainability, panel data, small-sample methods

---

NOTICE: This is the authors' version of the article accepted for publication in the Journal of Cleaner Production. Changes resulting from the publishing process, such as copy-editing, formatting, technical enhancements and pagination may not be reflected in this document. The final version of this article can be accessed at https://doi.org/10.1016/j.jclepro.2021.129357.

This is the accepted manuscript of the article published at https://doi.org/10.1016/j.jclepro.2021.129357.

A Panel Analysis of Groundwater Use in California

1. Introduction

Groundwater is one of the most important natural resources in the world and vital for drinking water and agricultural production. The clean production and effective use of groundwater are crucial for environmental protection and sustainable development. However, groundwater resources have been overexploited in many major agricultural and urban areas in the world. Also, climate change, characterized by more frequent and intense hydrologic extremes, amplifies groundwater use and affects groundwater recharge (Lall et al., 2020). Overdraft, depletion, and pollution of groundwater have become a global sustainability concern (Lall et al., 2020) and have environmental, health, and economic implications. In the United States, the rate of groundwater depletion has increased significantly since the 1950s, and the total groundwater depletion during the 1900-2008 period was estimated to be approximately 1,000 cubic kilometers (km^3) (Konikow, 2013). India and Mexico, two of the largest users of groundwater in the world, both face critical overdraft challenges (Scott and Shah, 2004). Groundwater overexploitation is also a critical issue in North China Plain, with significant adverse impacts on the environment (Changming et al., 2001). Other large groundwater users in the world, such as Pakistan (Watto and Mugera, 2016), are also facing groundwater depletion challenges.

In the United States, groundwater is a relevant source of fresh water in several states, particularly in California. California is the most populated state and is the one with the largest economy in the United States, with a total population of approximately 40 million and a gross state product of about 3 trillion dollars as of 2018 (BEA, 2019). Water resources play an important role in California's social and economic development (DWR, 2019b). California relies on both surface water and groundwater for its water supplies, and groundwater is vital for sustaining the state's environmental, social, and economic conditions (DWR, 2016). Groundwater contributes approximately 38 percent to the state's total water supply during an average year and up to 46 percent (or even more) during dry years (DWR, 2019c). Groundwater also supports California's 46 billion dollars agricultural economy (Mehta et al., 2018). Many rural and municipal communities depend on groundwater for up to 100 percent of their water supply (DWR, 2019c). While groundwater use varies across the state and over time, it has overall largely increased from approximately 11 km^3, or 9 million acre-feet[1] (MAF) in 1947 to about 24 km^3 (20 MAF) per year from 2005 to 2009 (Mehta et al., 2018).

Groundwater serves as a critical buffer against the severe impacts from droughts in California, contributing up to 46 percent (or even more) of the state's annual supply during dry years (DWR, 2019c; Lund et al., 2018). However, the ability to mitigate those impacts has significant implications: Over-reliance on groundwater has led to significant overdraft, which reduces water supply reliability in the long term, increases groundwater pumping costs, causes subsidence and infrastructure damage, harms the

[1] One acre-foot is equivalent to approximately 1233 cubic meters.

groundwater-dependent ecosystems, and seriously threatens the sustainability of state's groundwater resources (Mehta et al., 2018; Moran et al., 2014). The Central Valley in California, with an area of approximately 52,000 km^2 (Famiglietti et al., 2011) and a population of 6.5 million, cultivates more than 250 different types of crops and claims more than 70% of state's groundwater supply (Ojha et al., 2018). Ojha et al. (2018) found that droughts significantly exacerbated the stress on the aquifer systems that serve the region, when over-drafting and low rates of natural recharge resulted in an accelerated decline of groundwater levels across the area. By investigating the depletion and degradation of the aquifer-system during the 2007–2010 severe drought in the region, they found that about 2% of the total aquifer-system storage in the area was permanently lost due to irreversible compaction of the system. They estimated that a total of 21.3 ± 7.2 km^3 of groundwater storage was lost from December 2006 to January 2010. Ojha et al. (2019) further estimated that a total volume of 24.2 ± 9.3 km^3 of groundwater storage was lost in the San Joaquin Valley, part of the Central Valley, during the so-called California's millennium drought of 2012–2015.

Given California's heavy reliance on groundwater, a more unreliable and unsustainable groundwater supply to the state will unavoidably lead to greater environmental damages and economic losses in the long run. According to Gleeson et al (2010), groundwater residence times greater than 11,000 years are common in the United States, and the social and economic benefits from large volume withdrawals may not make up for the significant depletion of aquifers that are non-renewable on human timescales. The exploitation of such slowly renewed aquifers should therefore follow sustainability goals set on a multigenerational time horizon. Mays (2013) suggests that the development of groundwater sustainable solutions requires "both holistic and multi-objective approaches" that include the economics of overexploitation and sustainability indices. Pandey et al (2011) propose a "groundwater sustainability infrastructure index" as a framework to measure and evaluate progress of groundwater sustainability. In their framework, infrastructure refers to the knowledge, practices, and institutions that contribute to achieving groundwater sustainability. Elshall et al (2020) review the concept of groundwater sustainability from the scientific and policy perspectives. They conclude that the effective implementation of groundwater sustainability policies requires (a) the engagement of stakeholders through collaborative modeling and social learning; (b) an improved understanding of the coevolving surface water-groundwater systems, ecosystems, and human activities; and (c) addressing the uncertainty in our scientific knowledge and diversity of societal preferences.

In 2014, during the severe drought of 2012–2015, California passed the Sustainable Groundwater Management Act (SGMA). For the first time in its history, the state has implemented a framework to guide the sustainable management and use of groundwater. The Act established a new structure for managing groundwater resources at the local level and by local actors.[2] SGMA mandates the formation of

[2] This is consistent with the sustainability strategy proposed by Gleeson et al (2010), who suggest that community involvement is essential for the success of long-term management strategies. Aquifer- or watershed-based communities should understand the fragility of the resource and therefore be involved in setting specific goals for groundwater use.

groundwater sustainability agencies for all over-drafted groundwater basins across the state.[3] The agencies are required to develop and implement groundwater sustainability plans to mitigate overdraft and avoid the undesirable results within 20 years. The development of such plans requires reliable data and appropriate tools, and one important aspect of SGMA is the support the Act provides to data and tools that inform groundwater management decision-making (DWR, 2019d).

Indeed, groundwater data collection, modeling and analysis have been playing an important role in planning, implementing, monitoring, and evaluating groundwater supply and management. Commonly used groundwater data can be roughly classified into three categories: time series, cross-sectional and panel data. The latter is also referred to as longitudinal data and includes both the cross-sectional and time series dimensions. This study uses panel data for its analysis.

Panel data has received increasing attention from research in the environmental, energy, resources, climate, and sustainability areas. Chen et al. (2004) used maximum likelihood panel data estimates to examine the potential effects of climate change on crop yield variance for the major agricultural crops in the United States. Asici (2013) employed a panel data regression method to investigate the environmental sustainability of economic growth in 213 countries. Chakraborty and Mukherjee (2013) used panel data to analyze how trade and investment flows affect environmental sustainability in 114 countries. Hao et al. (2016) investigated the influence of climate change on carbon dioxide emissions using Chinese provincial panel data. Fan et al. (2017) used panel data of 31 Chinese provinces over the period 2000 to 2014 to investigate the relationship between energy production and water resource utilization. Altintas and Kassouri (2020) employed panel data methods and data on 14 European countries to investigate whether the environmental Kuznets Curve hypothesis is related to the per-capita ecological footprint or CO_2 emissions. Omojolaibi and Nathaniel (2020) used panel data econometric techniques to assess the impact of environmental regulations, trade, economic growth, and energy consumption on ecological footprint in Middle East and North Africa countries. Sadik-Zada and Gatto (2021) employed panel data methods to assess the significance of their proposed theoretical framework, where they relied on a three-sector decision model to assess the interaction of the structural and institutional factors that affect environmental pollution in oil-reliant economies. There are several reasons for this increasing attention (Dougherty, 2011). First, panel data may provide a solution to the problem of omitted variable bias caused by unobserved heterogeneity. Second, panel data record the timing of various events and can provide information of changes, trends and durations of events. As a consequence, exploiting panel data may reveal dynamics that are difficult to detect with cross-sectional data only. Finally, panel data often includes a large number of observations[4] (Dougherty, 2011).

---

[3] This is in line with the notion of territorial social responsibility introduced by Del Baldo (2013). Territorial social responsibility is a form of governance based on the concept of Corporate Social Responsibility (Dahlsrud 2008; Freeman and Hasnaoui, 2011) and sustainability-oriented strategies. The idea is promoted by networks of public and private, for and non-profit local actors who share the same territory and whose policies are oriented toward sustainable development (Rusciano, 2019).

[4] More specifically, a panel data set of N units of observation (entities) and T time periods may potentially include observations consisting of time series of length T on N parallel units. This may potentially lead to data sets with as many as $N \times T$ data points (Dougherty, 2011).

Recent research has used panel data analysis methods to investigate groundwater supply and use and demonstrated the potential of these methods to analyze large and complex groundwater data sets. Izady et al. (2012) used a panel-data model to predict groundwater levels in the Neishaboor plain, Iran, and found that the two-way fixed-effects model was superior to a model based on artificial neural network. Hendricks and Peterson (2012) employed the fixed-effects approach to decompose and estimate the price elasticity of irrigation water demand, using a 16-year panel dataset of more than 14,000 individual fields in Western Kansas, United States, overlying the High Plains Aquifer. The estimates of the price elasticity of water demand were used to evaluate the cost of reducing the use of irrigation water through three classes of water management policies: water pricing, irrigation cessation, and intensity reduction. Pfeiffer and Lin (2014) utilized panel data from groundwater-irrigated fields in Western Kansas over the period of 1996–2005 to investigate the effect of conversion to efficient irrigation technology on groundwater extraction. They found that the intended reduction in groundwater use did not occur and the amount of groundwater extraction in the area increased along with the shift to more efficient irrigation technology during the study period. They attributed the increase in groundwater extraction to changes in cropping patterns. Balasubramanian (2015) examined the impact of climate variables on groundwater sources used for irrigation and on agricultural income in Tamil Nadu, India. They used panel data of 11 districts observed over a 40-year period and found that increases in rainfall had a significant effect in reducing the depth to water table and increases in maximum temperature significantly reduced groundwater availability. Kishore et al. (2020) used panel regression to analyze the determinants for groundwater decline and depletion in 41 districts in India where groundwater level declined more than 4 meters between 2002 and 2016. They found that increases in irrigated area and water intensive crops led to groundwater level declines.

This study investigates the dynamics of groundwater extraction in California, and the impact that water use and supply have on that extraction. The study is motivated by the SGMA's focus on tools that can inform groundwater management. We rely on a panel of data estimated by the California Department of Water Resources (DWR) for the 56 California Water Plan planning areas over the period of 1998–2015. Our emphasis is on the agricultural and urban water use and major water projects in the state. We employ panel regression and robust inference to quantitatively assess the relationship between groundwater extraction and water use and supplies in the state. To the best of our knowledge, this is the first study to rely on panel analysis with small-sample corrections for inference to estimate these relationships. Our estimates allow for inter- and intraregional assessment of how changes in water supply and use impact groundwater extraction. The econometric, data-oriented approach we propose can therefore be used to inform water management and policy at the macro- and meso-level in California, where it can be periodically reused, with new data, to evaluate the progress of measures designed to reduce stress on aquifers in the state. Our approach can also support groundwater management in other regions of the world.

The remaining of the paper is organized as follows. Section 2 briefly describes the water resources and major water projects in California. Section 3 describes the water data used in the study. Section 4 describes the panel regression analysis and inference methods we used. Section 5 presents main results and a discussion of those results. Finally, Section 6 concludes with final remarks.

2. Water resources and major water projects in California

2.1 Water resources and hydrologic regions

The spatial and temporal distribution of water resources is highly uneven in California. California contains multiple climate zones and hydrologic conditions in the state vary greatly from place to place and time to time (Bureau of Reclamation, 2008). The volume of annual average precipitation in California is approximately 240 km^3, or 200 MAF. Much of this water is lost through evaporation, particularly in the hot and dry areas of the state, and the remaining part of the water, known as “unimpaired runoff” and averaging approximately 93 km^3 (75 MAF) per year, flows into streams and groundwater basins, and becomes water resources available for management and use (Hanak et al., 2011). Note that the available water resources for the state could vary greatly year to year and decrease significantly during dry years. For example, in 2015, a recent critically dry year for California, the volume of statewide precipitation was only about 81 km^3 (66 MAF), approximately one third of the precipitation in an average year. Furthermore, the state’s variable topography and hydrologic conditions contribute to the large variation in the amount of regional precipitation (DWR, 2016). Most of California’s precipitation falls in the northern part of the state during the winter, while much of the water is used in the central and southern parts of the state during the spring and summer (Bureau of Reclamation, 2008).

California’s regional differences make it necessary to divide the state into regions for the purpose of statewide planning and operation. The DWR divides the state into 10 hydrologic regions that characterize the large watersheds in the state (DWR, 2016) and further into 56 planning areas. Figure 1 shows these hydrologic regions and planning areas. For the purpose of discussing water demand and supply, especially the impacts of major water projects in California, we aggregate these 10 hydrologic regions into five geographic zones based on geography and hydrologic conditions (Bureau of Reclamation, 2008; DWR, 2016). The five zones are the North Coast and Lahontan zone, the San Francisco (SF) Bay and Central Coast zone, the Northern Central Valley zone, the Southern Central Valley zone, and the Southern California zone. The five zones are shown in Figure 2.

2.2 Major water projects in California

California has invested great efforts into water supplies by implementing many water projects. The projects move water from its source to where it is used, sometimes over thousands of kilometers away (Bureau of Reclamation, 2008).

2.2.1 State Water Project (SWP)

The California State Water Project (SWP) is a water storage and delivery system that extends from Northern California to Southern California. It includes 36 storage facilities, 21 pumping plants, five hydroelectric power plants, four pumping-generating plants, and about 1100 kilometers (700 miles) of canals, tunnels, and pipelines (DWR, 2019e). The SWP was planned and constructed and is operated by DWR. It is the nation's largest state-built, multi-purpose, user-financed water project. Its primary purpose is to supply water, although the project was designed to also provide additional benefits, such as flood control, power generation, recreation, and fish and wildlife habitat, while balancing the needs of water delivery and environmental protection. The SWP was designed to deliver about 5.2 km^3 (4.2 MAF) of water per year to farms, homes, and industries through 29 long-term SWP Water Supply Contractors (DWR, 2019e). The project supplies water to almost 27 million people in northern California, the San Francisco Bay Area, the San Joaquin Valley, the Central Coast and southern California. The project also irrigates about 300 thousand hectares (750 thousand acres) of farmland, mainly in the San Joaquin Valley (DWR, 2019e). SWP's water supply depends on rainfall, snowpack, runoff, stored water, pumping capacity,[5] and multiple environmental and operational constraints (DWR, 2019e).

2.2.2 Central Valley Project (CVP)

The Central Valley Project (CVP) is a federal water management project in California that is under the supervision of the United States Bureau of Reclamation (USBR). The project consists of a series of dams, reservoirs, canals, aqueducts, and pump plants. The CVP extends through 640 kilometers (400 miles) in central California, providing flood protection for the Central Valley and supplying domestic and industrial water to the valley (Bureau of Reclamation, 2017). The project also supplies water to major urban areas in the Greater Sacramento and San Francisco Bay areas and provides water to restore and protect fish and wildlife, and to enhance water quality (Bureau of Reclamation, 2017). The CVP manages some 11 km^3 (9 MAF) of water, with an annual delivery of about 8.6 km^3 (7 MAF) of water for agricultural, urban and wildlife use (Bureau of Reclamation, 2019b).

2.2.3 Colorado River supplies

The Colorado River is one of the major water sources for Southern California and is critical for sustaining Southern California's municipal and agricultural water supplies.

2.2.3.1 Colorado River Aqueduct

The Colorado River Aqueduct (CRA), a 389-kilometer (242-mile) water conveyance system in Southern California, is built and operated by the Metropolitan Water District of Southern California (MWD). The CRA includes two reservoirs, five pumping plants, 101 kilometers (63 miles) of lined canals, 148 kilometers (92 miles) of tunnels, 89 kilometers (55 miles) of covered canals and 46 kilometers (29 miles) of inverted siphons (Zetland, 2011). The project takes water out of the Colorado River, from Lake Havasu at the California-Arizona border to Lake Mathews in Riverside, California. The delivery capacity of the CRA is over 1.5 km^3 (1.2 MAF) a year. Along with the SWP, the CRA is one of two sources of drinking water imported to Southern California (MWD, 2019).

[5] From the Sacramento-San Joaquin Delta.

This is the accepted manuscript of the article published at https://doi.org/10.1016/j.jclepro.2021.129357.

2.2.3.2 All-American Canal System

The All-American Canal System is located in the southeastern corner of California. It includes the Imperial Diversion Dam and Desilting Works, the 129-kilometer (80-mile) All-American Canal, and the 198-kilometer (123-mile) Coachella Canal (Bureau of Reclamation, 2019a). The system diverts water from the Colorado River, at the Imperial Dam, to irrigate about 214 thousand hectares (530 thousand acres) of fertile land in the Imperial Valley and about 32 thousand hectares (79 thousand acres) in the Coachella Valley (Bureau of Reclamation, 2019a).

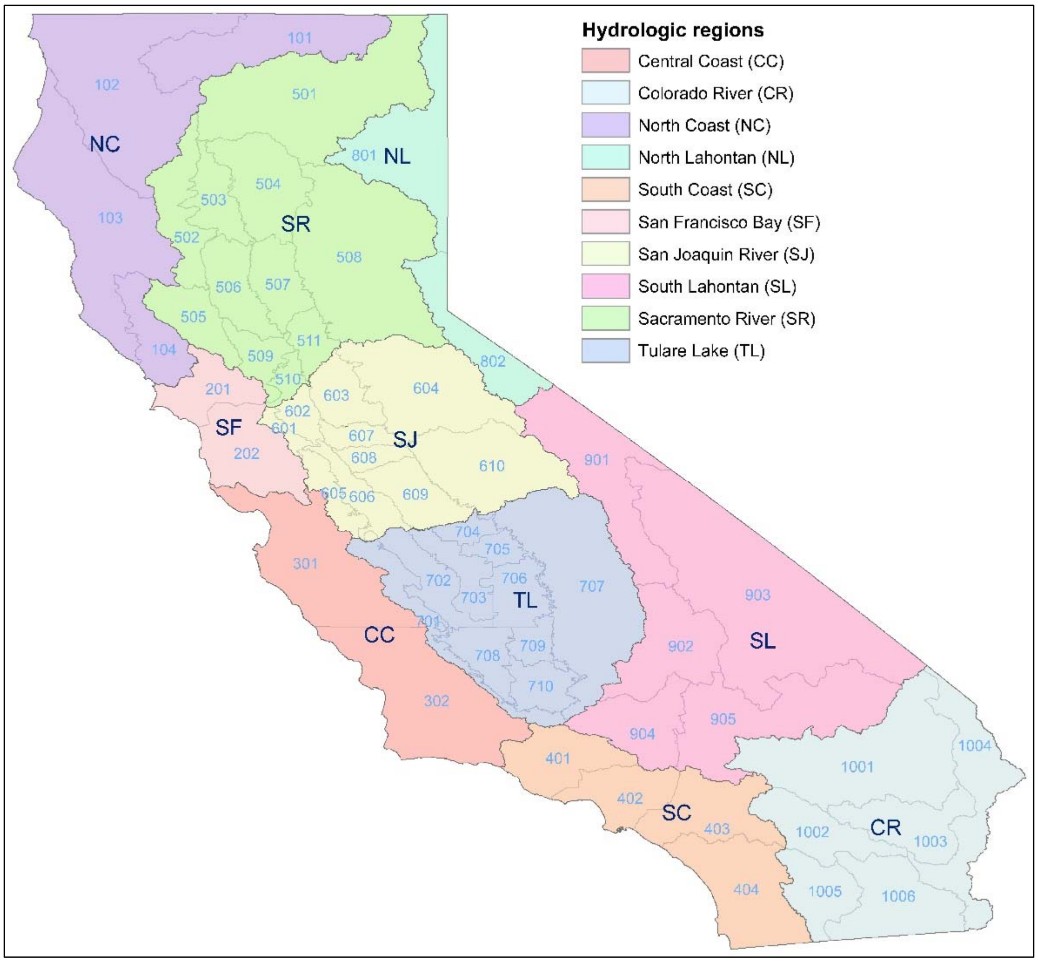

Fig. 1. California hydrologic regions and planning areas
Sources: (DWR, 2016, 2019a).

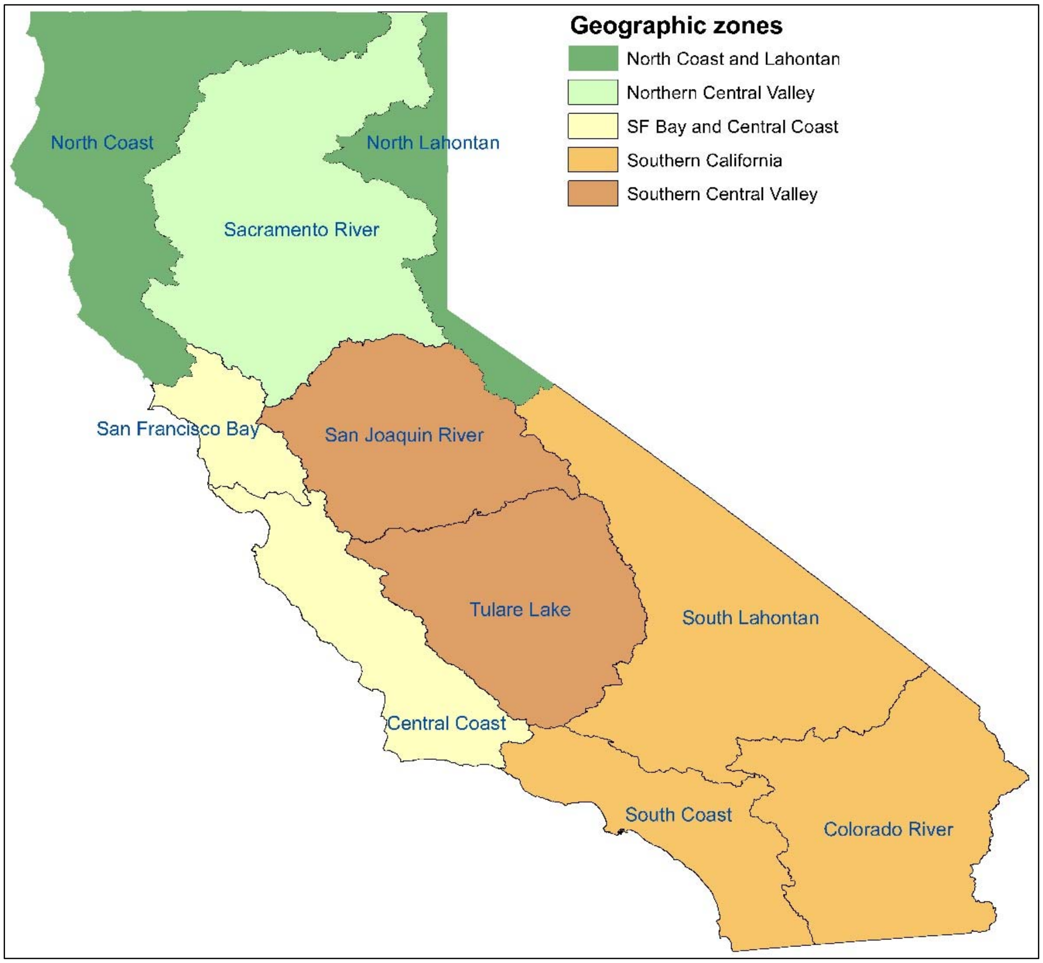

Fig. 2. California geographic zones and hydrologic regions
Sources: (Bureau of Reclamation, 2008; DWR, 2016, 2019a)

3 Data

This study relies mainly on the data that underlie the 2018 California Water Plan (CWP), developed by the DWR (2019a). The CWP is the state's "strategic plan for sustainably managing and developing water resources for current and future generations." The plan documents the development and status of California's water resources and demands and is updated every 5 years. The data we rely on include (water-year based) annual water balances from 1998 to 2015. The balances comprise water use and supply. Water use data are disaggregated into several types of application organized in three major categories of use: urban, agriculture, and environmental. Water supply data are provided by source, and include supplies from multiple sources of surface water, such as local sources, the SWP, the CVP, Colorado River supplies, and other federal deliveries. The balances further include supplies from groundwater extraction and reused/recycled water. All data are provided at the level of planning area.

In addition to the data provided by the CWP, we rely on groundwater data from the California's Groundwater Update 2013 (DWR, 2016), developed for the 2013 CWP. The data include, for each hydrologic region, information on groundwater supply for urban and agricultural use at the level of planning area. Table 1 summarizes the data used in this study according to the five geographic zones represented in Figure 2.

Table 1. Descriptive statistics of the data by geographic zone

Geographic zone	Variable (TAF)	Observations*	Mean	Std. Dev.	Min	Max
North Coast and Lahontan	Groundwater extraction	108	85	61	13	285
	Agricultural water use	108	194	169	19	664
	Urban water use	108	30	25	8	94
	Misc. surface water supply	108	3347	5350	98	22358
	Precipitation (MAF)	108	9	8	1	27
Northern Central Valley	Groundwater extraction	198	243	209	1	822
	Agricultural water use	198	653	691	3	2538
	Urban water use	198	75	109	1	474
	Environmental water use	198	52	100	0	372
	Misc. surface water supply	198	995	2034	2	12026
	CVP deliveries	198	350	635	0	2224
	Precipitation (MAF)	198	4	5	0	27
SF Bay and Central Coast	Groundwater extraction	72	349	260	4	863
	Agricultural water use	72	290	253	14	776
	Urban water use	72	335	339	105	1010
	Misc. surface water supply	72	225	234	13	669
	SWP deliveries	72	40	43	0	168
	CVP deliveries	72	44	55	0	202
	Recycled water supply	72	9	11	0	40
	Precipitation (MAF)	72	4	2	1	11
Southern Central Valley	Groundwater extraction	360	492	549	0	2447
	Agricultural water use	360	844	716	0	2765
	Urban water use	360	63	59	0	307
	Environmental water use	360	29	93	0	456
	Misc. surface water supply	360	330	379	0	2844
	SWP deliveries	360	53	144	0	810
	CVP deliveries	360	167	316	0	1511
	Precipitation (MAF)	360	1	3	0	15
Southern California	Groundwater extraction	270	164	220	0	972
	Agricultural water use	270	304	590	2	2665
	Urban water use	270	298	482	0	1908
	Misc. surface water supply	270	48	85	0	463
	SWP deliveries	270	88	147	0	794
	Colorado River deliveries	270	323	673	0	3056
	Recycled water supply	270	14	27	0	109
	Precipitation (MAF)	270	1	1	0	8

* Our panel data dataset is balanced. Variations in the number of observations are due to variations in the number of planning areas included in each geographic zone. TAF means thousand acre-feet, and MAF million acre-feet.

4. Methodology

4.1 Model and estimation

We rely on panel regression modeling for our analysis (Greene, 2018; Pesaran, 2015; Wooldridge, 2015). Consider the following unobserved-effects regression model:

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta} + \alpha_i + \delta t + \varepsilon_{i,t} \quad (1)$$

where $y_{i,t}$ is the dependent variable, $\mathbf{x}_{i,t}$ is a vector of observed explanatory variables, $\boldsymbol{\beta}$ is a vector of regression coefficients, α_i represents unobserved entity-specific effects, where the index *i* refers to the entity or unit of observation (in this case, planning area), and *t* refers to the time period. The error (disturbance) term $\varepsilon_{i,t}$ is assumed to satisfy the usual regression model conditions. The trend term *t* is introduced here to allow for a shift of the intercept over time and could be replaced by a set of dummy variables if the implicit assumption of a constant trend seemed too strong (Dougherty, 2011). Assuming the unobserved entity-specific effects, represented by α_i, are time-invariant, the model in (1) could be estimated using the fixed-effects estimator (also known as the *within estimator*) or the least square dummy variable (LSDV) estimator approaches (Dougherty, 2011).

For our study, the dependent variable $y_{i,t}$ is the total groundwater use (extraction) estimated for planning area *i* in year *t*, denoted below as $U_{i,t}^{(\mathrm{gw})}$. We further define:

$$U_{i,t}^{(\mathrm{urb})} = U_{i,t}^{(\mathrm{lnd})} + U_{i,t}^{(\mathrm{com})} + U_{i,t}^{(\mathrm{ind})} + U_{i,t}^{(\mathrm{enr})} + U_{i,t}^{(\mathrm{int})} + U_{i,t}^{(\mathrm{ext})} \quad (2)$$

$$S_{i,t}^{(\mathrm{sf})} = S_{i,t}^{(\mathrm{msf})} + S_{i,t}^{(\mathrm{swp})} + S_{i,t}^{(\mathrm{cvp})} + S_{i,t}^{(\mathrm{col})} \quad (3)$$

$$S_{i,t}^{(\mathrm{msf})} = S_{i,t}^{(\mathrm{loc})} + S_{i,t}^{(\mathrm{imp})} + S_{i,t}^{(\mathrm{fed})} \quad (4)$$

to express, respectively, urban water use (*U*) and surface water supplies (*S*).

Using (2)–(4) we define panel data regression models that are specific to each of the five geographic zones we analyze:[6]

$$U_{i,t}^{(\mathrm{gw})} = \beta_{\mathrm{ag}} U_{i,t}^{(\mathrm{ag})} + \beta_{\mathrm{urb}} U_{i,t}^{(\mathrm{urb})} + \beta_{\mathrm{msf}} S_{i,t}^{(\mathrm{msf})} + \beta_{\mathrm{prp}} S_{i,t}^{(\mathrm{prp})} + \alpha_i + \delta t + \varepsilon_{i,t} \quad (5)$$

[6] Note that the dummy variable trap is avoided in the fixed-effects estimations by omitting one of the values that the dummy variable represents or dropping the overall constant (Greene, 2018). This is implemented in the statistical software we used.

$$U_{i,t}^{(\mathrm{gw})} = \beta_{\mathrm{ag}} U_{i,t}^{(\mathrm{ag})} + \beta_{\mathrm{urb}} U_{i,t}^{(\mathrm{urb})} + \beta_{\mathrm{msf}} S_{i,t}^{(\mathrm{msf})} + \beta_{\mathrm{swp}} S_{i,t}^{(\mathrm{swp})} + \beta_{\mathrm{cvp}} S_{i,t}^{(\mathrm{cvp})} + \beta_{\mathrm{rec}} S_{i,t}^{(\mathrm{rec})} + \alpha_i + \delta t + \varepsilon_{i,t} \tag{6}$$

$$U_{i,t}^{(\mathrm{gw})} = \beta_{\mathrm{ag}} U_{i,t}^{(\mathrm{ag})} + \beta_{\mathrm{urb}} U_{i,t}^{(\mathrm{urb})} + \beta_{\mathrm{env}} U_{i,t}^{(\mathrm{env})} + \beta_{\mathrm{msf}} S_{i,t}^{(\mathrm{msf})} + \beta_{\mathrm{cvp}} S_{i,t}^{(\mathrm{cvp})} + \alpha_i + \delta t + \varepsilon_{i,t} \tag{7}$$

$$U_{i,t}^{(\mathrm{gw})} = \beta_{\mathrm{ag}} U_{i,t}^{(\mathrm{ag})} + \beta_{\mathrm{urb}} U_{i,t}^{(\mathrm{urb})} + \beta_{\mathrm{env}} U_{i,t}^{(\mathrm{env})} + \beta_{\mathrm{msf}} S_{i,t}^{(\mathrm{msf})} + \beta_{\mathrm{swp}} S_{i,t}^{(\mathrm{swp})} + \beta_{\mathrm{cvp}} S_{i,t}^{(\mathrm{cvp})} + \alpha_i + \delta t + \varepsilon_{i,t} \tag{8}$$

$$U_{i,t}^{(\mathrm{gw})} = \beta_{\mathrm{ag}} U_{i,t}^{(\mathrm{ag})} + \beta_{\mathrm{urb}} U_{i,t}^{(\mathrm{urb})} + \beta_{\mathrm{msf}} S_{i,t}^{(\mathrm{msf})} + \beta_{\mathrm{swp}} S_{i,t}^{(\mathrm{swp})} + \beta_{\mathrm{col}} S_{i,t}^{(\mathrm{col})} + \beta_{\mathrm{rec}} S_{i,t}^{(\mathrm{rec})} + \alpha_i + \delta t + \varepsilon_{i,t} \tag{9}$$

where equation (5) refers to the North Coast and Lahontan zone, (6) to the SF Bay and Central Coast zone, (7) to the northern Central Valley zone, (8) to the southern Central Valley zone, and (9) to the Southern California zone, and:

i index of the planning area,

t water-year index ($t = 1, 2, \ldots, 18$, corresponding to water-years from 1998 to 2015),

$U_{i,t}^{(\mathrm{ag})}$ on-farm applied water,

$U_{i,t}^{(\mathrm{urb})}$ total urban applied water,

$U_{i,t}^{(\mathrm{env})}$ water applied for environmental purposes (wetland management only),

$U_{i,t}^{(\mathrm{lnd})}$ water applied for large urban landscape irrigation,

$U_{i,t}^{(\mathrm{com})}$ water applied for commercial purposes,

$U_{i,t}^{(\mathrm{ind})}$ water applied for industrial purposes,

$U_{i,t}^{(\mathrm{enr})}$ water applied for energy production,

$U_{i,t}^{(\mathrm{int})}$ water applied for indoor residential purposes,

$U_{i,t}^{(\mathrm{ext})}$ water applied for outdoor residential purposes,

$S_{i,t}^{(\mathrm{sf})}$ total surface water deliveries,

$S_{i,t}^{(\mathrm{swp})}$ deliveries from the California State Water Project,

$S_{i,t}^{(\mathrm{cvp})}$ deliveries from the Central Valley Project,

$S_{i,t}^{(\mathrm{col})}$ deliveries from the Colorado River,

This is the accepted manuscript of the article published at https://doi.org/10.1016/j.jclepro.2021.129357.

$S_{i,t}^{(\mathrm{msf})}$ deliveries from other sources of surface water (excluding SWP, CVP, and Colorado River),

$S_{i,t}^{(\mathrm{loc})}$ deliveries from local sources of surface water,

$S_{i,t}^{(\mathrm{imp})}$ deliveries from imported water,

$S_{i,t}^{(\mathrm{fed})}$ deliveries from other federal projects,

$S_{i,t}^{(\mathrm{rec})}$ deliveries from recycled water, and

$S_{i,t}^{(\mathrm{prp})}$ precipitation,

with all applied water and water supplies in thousand acre-feet (TAF), and precipitation in MAF.

We note that models (6)–(9) do not include precipitation as an independent variable. This is partially because the number of variables in these models is relatively large for the size of their samples. After performing a variable selection process, we concluded that precipitation is not a statistically significant driver to the dynamics of groundwater extraction in the corresponding geographic zones.

We also note that fixed-effects estimations relax the strict exogeneity assumption and can address endogeneity problems under much weaker assumptions when compared to common ordinary least squares (OLS) estimations (Collischon and Eberl, 2020; Roberts and Whited, 2013). Generally speaking, fixed effects estimations are more credible and preferable to common OLS estimations, despite not being perfect and presenting limitations or drawbacks (Collischon and Eberl, 2020). It should also be noted that "in most cases, it is not correct to interpret the coefficients of any standard OLS, fixed-effects, or comparable model as causal effects but rather as partial correlations" (Collischon and Eberl, 2020).

4.2 Inference

A distinctive feature of panel data is that they are clustered. In panel data studies, the clusters are composed of the repeated measurements obtained from a single entity (or unit of observation) at different occasions. The measurements on units within a cluster are usually more similar than the measurements of units in different clusters, or more precisely, measurements within a cluster will typically be correlated, and the degree of clustering can be described using the correlation among the measurements on units within the same cluster. Because this correlation invalidates the assumption of independence that is crucial for many standard statistical techniques, statistical models for clustered data need to account for this correlation. As panel data are a special case of clustered data, this correlation must be accounted for in panel data analysis (Fitzmaurice et al., 2011).

For policy studies at the regional (e.g., state or provincial) level, inference is typically conducted based on standard errors clustered at the regional level to account for correlation in the variables across entities or units (Jones et al., 2015). The commonly used procedure is the cluster-robust variance estimator, or CRVE (Cameron et al., 2008; MacKinnon, 2019). However, CRVE is asymptotic consistent, i.e., the consistency of CRVE is asymptotic in the number of independent clusters (Pustejovsky and Tipton, 2018). Recent methodological work and simulation studies have demonstrated that CRVE can be significantly biased downward when there are only few clusters and lead to high Type I error rates for associated hypothesis tests (Cameron et al., 2008; Matthew, 2014; Pustejovsky and Tipton, 2018). Bell and McCaffrey (2002) proposed a bias-reduced linearization (BRL) method to improve the small-sample properties of CRVE. Tipton and Pustejovsky (2015) and Pustejovsky and Tipton (2018) further developed the BRL method and extended it for general application. Simulations and empirical studies showed that the small-sample methods proposed by Tipton and Pustejovsky (2015) and Pustejovsky and Tipton (2018) maintained promising performance over a wide range of scenarios.

For this study, we investigated five geographic zones in California. As some of these zones include few planning areas,[7] the issue of small number of clusters needs to be addressed. Therefore, for inference, we adopt the small-sample method proposed by Tipton and Pustejovsky (2015) and Pustejovsky and Tipton (2018) for cluster-robust standard errors and *t*-tests (Marcelo et al., 2017). We also compare the results from the small-sample method with those from the conventional asymptotic CRVE.

5. Results and analysis

We used the models described in (5)–(9) to evaluate the panel data we organized for each of the five geographic zones represented in Figure 2. Fixed-effect variables control time-invariant unobserved effects associated with planning areas, and trend terms capture any deterministic trend in regional (geographic zone) groundwater use.[8] We present our panel analysis results for each geographic zone.

5.1. North Coast and Lahontan zone

Table 2 shows statistical results for model (5), where groundwater extraction in the North Coast and Lahontan zone is evaluated from agricultural and urban water use, deliveries from miscellaneous surface water sources, precipitation, and time trend.

[7] For instance, the North Coast and Lahontan zone is comprised of six planning areas, and the SF Bay and Central Coast zone includes only four planning areas.

[8] It is common, in studies that rely on univariate time series, that the time-varying data are tested for stationarity (unit root test). We note that the test usually applies to when the time series are relatively long, which is not the case of the time series we rely on for this study.

Table 2. Fixed effects model estimation results for the North Coast and Lahontan zone

Dependent variable	Groundwater extraction							
Method	Asymptotic CRVE				Small sample method			
Independent variables	Coef.	SE	*p*	95% CI	Coef.	SE	*p*	95% CI
Agricultural water	0.224	0.114	0.1066	[–0.069, 0.518]	0.224	0.165	0.3351	[–0.697, 1.146]
Urban water	0.756	0.919	0.4484	[–1.61, 3.119]	0.756	1.354	0.6351	[–5.323, 6.835]
Misc. surface water	–0.000562	0.001	0.4845	[–0.0025, 0.00135]	–0.000562	0.001	0.7084	[–0.0123, 0.0112]
Precipitation	–1.227	0.465	0.0462	[–2.42, –0.0305]	–1.227	0.593	0.2321	[–5.399, 2.946]
Time	0.705	0.664	0.3370	[–1.00, 2.411]	0.705	0.720	0.3782	[–1.225, 2.635]
R–squared	0.907				0.907			
Observations	108				108			
Cluster number	6				6			

Note: SE stands for standard error, and CI stands for confidence interval.

As shown in Table 2, none of the estimated coefficients on agricultural water use, urban water use, miscellaneous surface water, precipitation, and time is statistically significant at the conventional significance levels, if the tests are based on the small sample method. If, however, the test is based on the asymptotic CRVE method, the coefficient estimated for precipitation becomes statistically significant at the 5% level ($p < 0.05$), and the one for agricultural water use becomes statistically significant just shy of 10% level ($p = 0.11$).

We should note that the four independent variables used for the estimation are correlated, and such multicollinearity makes it somewhat difficult for the estimation process to detect and estimate the partial effect of each variable (Wooldridge, 2015). In an attempt to still assess the effects of those variables on groundwater extraction in this geographic zone, we try alternative, similar fixed-effects models where we combine the four independent variables in different ways. Table 3 lists estimation results from the alternative fixed effects models, solved with the small-sample method. As shown in the table, the *R*–squared of the alternative model (5a) is 0.897. This suggests that nearly 90% of the variation of groundwater extraction could be explained by the alternative model (5a). In other words, agricultural water use could explain – although not statistically significant at the conventional levels – nearly 90% of the variation in groundwater extraction, while the other variables seem to have little effect on that extraction. Note that the estimated coefficient of agricultural water is quite consistent when adding or removing other variables to the model, and it suggests that one extra TAF of agricultural water would likely lead to an increase in groundwater extraction by approximately 0.2 TAF. Note also that the coefficient estimated for precipitation, which is the close to being statistically significant at the 20% level (with an average *p*-value of around 0.2), is consistent with the notion that increased precipitation reduces the pressure on groundwater extraction. The estimates suggest that one extra MAF of precipitation in the zone would likely lead to a decrease of approximately 1.3 ± 1.1 TAF in groundwater extraction.

Table 3. Alternative fixed-effects models for the North Coast and Lahontan zone

Dependent variable: groundwater extraction						
Independent variables	(5a)	(5b)	(5c)	(5d)	(5e)	(5f)
Agricultural water	0.228	0.218	0.215	0.211		0.227
Standard error	0.170	0.174	0.171	0.163		0.166
p-value	0.347	0.369	0.368	0.359		0.341
Urban water		0.407	0.477	0.618		
Standard error		1.563	1.555	1.500		
p-value		0.824	0.794	0.726		
Misc. surface water			–0.00111	–0.00093		
Standard error			0.001	0.001		
p-value			0.480	0.552		
Precipitation				–1.360	–1.225	–1.181
Standard error				0.645	0.556	0.449
p-value				0.222	0.218	0.178
R–squared	0.897	0.899	0.900	0.904	0.871	0.900
Observations	108	108	108	108	108	108
Cluster number	6	6	6	6	6	6

5.2. SF Bay and Central Coast zone

Table 4 shows statistical results for model (6), where groundwater extraction in the SF Bay and Central Coast zone is evaluated from agricultural and urban water use, deliveries from miscellaneous surface water sources, recycled water, SWP and CVP, and time trend.

Table 4. Fixed effects model estimation results for the SF Bay and Central Coast zone

Dependent variable	Groundwater extraction							
Method	Asymptotic CRVE				Small sample method			
Independent variables	Coef.	SE	*p*	95% CI	Coef.	SE	*p*	95% CI
Agricultural water	0.978	0.025	0.0000	[0.90, 1.058]	0.978	0.035	0.0123	[0.675, 1.282]
Urban water	0.785	0.042	0.0003	[0.65, 0.919]	0.785	0.084	0.0311	[0.228, 1.342]
Misc. surface water	–0.171	0.131	0.2832	[–0.59, 0.246]	–0.171	0.147	0.3701	[–0.837, 0.495]
SWP deliveries	–0.408	0.095	0.0232	[–0.71, –0.106]	–0.408	0.196	0.2421	[–1.961, 1.144]
CVP deliveries	–0.0707	0.093	0.5020	[–0.37, 0.225]	–0.0707	0.178	0.7543	[–1.924, 1.783]
Recycled water	–0.315	0.200	0.2126	[–0.95, 0.320]	–0.315	0.200	0.2687	[–1.269, 0.638]
Time	2.320	0.806	0.0635	[–0.24, 4.883]	2.320	1.017	0.1081	[–0.943, 5.583]
R-squared	0.995				0.995			
Observations	72				72			
Cluster number	4				4			

Note: SE stands for standard error, and CI stands for confidence interval.

The estimated coefficient of agricultural water use is statistically significant at the 5% level ($p < 0.05$) and indicates that one extra TAF of agricultural water use is expected to increase groundwater extraction by an average of 0.98 TAF. As for urban water use, the estimated coefficient is also statistically significant at the 5% level ($p < 0.05$) and suggests that, on average, one extra TAF of urban water use is expected to increase groundwater extraction by 0.79 TAF. The coefficient associated with time suggests that —at the statistical significance of approximately 10% level ($p < 0.11$)—groundwater extraction has been increasing by an annual amount of approximately 2.3 TAF.

The coefficients estimated for the other variables are not statistically significant at the conventional significance levels. If, however, the test is based on the asymptotic CRVE method, the coefficients estimated for SWP deliveries and time are statistically significant at approximately the 5% level.

5.3. Northern Central Valley zone

Table 5 shows statistical results for model (7), where groundwater extraction in the Northern Central Valley zone is evaluated from agricultural, urban and environmental (managed wetlands) water use, deliveries from miscellaneous surface water sources and CVP, and time trend.

Table 5. Fixed effects model estimation results for the Northern Central Valley zone

Dependent variable	Groundwater extraction							
Method	Asymptotic CRVE				Small sample method			
Independent variables	Coef.	SE	*p*	95% CI	Coef.	SE	*p*	95% CI
Agricultural water	0.141	0.068	0.0647	[–0.010, 0.292]	0.141	0.076	0.1899	[–0.153, 0.435]
Urban water	0.484	0.086	0.0002	[0.29, 0.676]	0.484	0.089	0.0708	[–0.146, 1.115]
Environmental water	0.657	0.087	0.0000	[0.46, 0.850]	0.657	0.116	0.0504	[–0.00348, 1.317]
Misc. surface water	–0.000917	0.003	0.7401	[–0.0069, 0.00508]	–0.000917	0.005	0.8677	[–0.0348, 0.0330]
CVP deliveries	–0.00908	0.010	0.4019	[–0.032, 0.0140]	–0.00908	0.042	0.8617	[–0.401, 0.383]
Time	2.614	0.929	0.0183	[0.55, 4.683]	2.614	0.916	0.0186	[0.548, 4.681]
R-squared	0.952				0.952			
Observations	198				198			
Cluster number	11				11			

Note: SE stands for standard error, and CI stands for confidence interval.

The coefficient estimated for urban water use is statistically significant at the 10% level ($p < 0.1$) and indicates that one extra TAF of urban water use is expected to increase groundwater extraction by an average of 0.48 TAF. As for environmental water use, the coefficient estimated suggests that—at the statistical significance of 10% level ($p < 0.1$)—one additional TAF of environmental water use is expected to increase groundwater extraction by approximately 0.66 TAF. Further, the coefficient associated with time is statistically significant at the 5% level ($p < 0.05$) and indicates that groundwater extraction exhibits an increasing trend of 2.6 TAF per year.

The coefficients estimated for agricultural water use and for deliveries from miscellaneous surface water sources and CVP are not statistically significant at the conventional significance levels. The coefficients associated with deliveries from miscellaneous surface water sources and CVP are very small (in absolute term), which implies that the impact of those deliveries on groundwater extraction in this zone is likely insignificant. Note that, if the test is based on the asymptotic CRVE method, the coefficient estimated for agricultural water use becomes statistically significant at the 10% level ($p < 0.1$).

5.4. Southern Central Valley zone

Table 6 shows statistical results for model (8), where groundwater extraction in the Southern Central Valley zone is evaluated from agricultural, urban and environmental (managed wetlands) water use, deliveries from miscellaneous surface water sources and CVP, and time trend.

Table 6. Fixed effects model estimation results for the Southern Central Valley zone

Dependent variable	Groundwater extraction							
Method	Asymptotic CRVE				Small sample method			
Independent variables	Coef.	SE	p	95% CI	Coef.	SE	p	95% CI
Agricultural water	1.079	0.114	0.0000	[0.84, 1.317]	1.079	0.147	0.0001	[0.741, 1.416]
Urban water	0.349	0.314	0.2807	[–0.31, 1.006]	0.349	0.389	0.4331	[–0.854, 1.551]
Environmental water	0.398	0.843	0.6417	[–1.37, 2.162]	0.398	1.286	0.7892	[–5.863, 6.660]
Misc. surface water	–0.200	0.166	0.2431	[–0.55, 0.147]	–0.200	0.313	0.5997	[–1.850, 1.450]
SWP deliveries	–0.521	0.067	0.0000	[–0.66, –0.381]	–0.521	0.086	0.0360	[–0.949, –0.0933]
CVP deliveries	–1.011	0.114	0.0000	[–1.25, –0.773]	–1.011	0.142	0.0068	[–1.477, –0.545]
Time	1.229	1.700	0.4783	[–2.33, 4.787]	1.229	1.892	0.5249	[–2.777, 5.235]
R-squared	0.975				0.975			
Observations	360				360			
Cluster number	20				20			

Note: SE stands for standard error, and CI stands for confidence interval.

The coefficient estimated for agricultural water use is statistically significant at the 0.1% level ($p < 0.001$) and indicates that one additional TAF of agricultural water use is expected to increase groundwater extraction by an average of 1.1 TAF. The large and highly significant coefficient on agricultural water use demonstrates—and is consistent with—the high reliance of agriculture on groundwater in the Southern Central Valley zone. The coefficient estimated for SWP deliveries is statistically significant at the 5% level ($p < 0.05$) and suggests that one additional TAF of SWP deliveries is expected to decrease groundwater extraction by an average of 0.52 TAF. As for deliveries from CVP, the coefficient estimated is statistically significant at the 1% level ($p < 0.01$) and indicates that one additional TAF of CVP deliveries is expected to decrease groundwater extraction by 1.0 TAF. The large absolute values of the coefficients associated with deliveries from SWP and CVP, along with their statistical significance, confirms the relevance of deliveries from the two projects to reduce the pressure, mainly from agricultural activities, on groundwater extraction in the zone. The coefficients estimated for the other variables are not statistically significant at the conventional significance levels, even if the test is based on the asymptotic CRVE method.

5.5. Southern California zone

Table 7 shows statistical results for model (9), where groundwater extraction in the Southern California zone is evaluated from agricultural and urban water use, deliveries from miscellaneous surface water sources, recycled water, SWP and the Colorado River, and time trend.

Table 7. Fixed effects model estimation results for the Southern California zone

Dependent variable	Groundwater extraction							
Method	Asymptotic CRVE				Small sample method			
Independent variables	Coef.	SE	*p*	95% CI	Coef.	SE	*p*	95% CI
Agricultural water	0.577	0.188	0.0083	[0.17, 0.981]	0.577	0.220	0.0618	[–0.0470, 1.202]
Urban water	0.635	0.081	0.0000	[0.46, 0.808]	0.635	0.096	0.0046	[0.350, 0.920]
Misc. surface water	–0.577	0.067	0.0000	[–0.72, –0.434]	–0.577	0.080	0.0103	[–0.867, –0.287]
SWP deliveries	–0.592	0.160	0.0024	[–0.93, –0.249]	–0.592	0.181	0.0295	[–1.089, –0.0946]
Colorado River	–0.573	0.131	0.0006	[–0.86, –0.292]	–0.573	0.143	0.0111	[–0.946, –0.201]
Recycled water	–0.531	0.151	0.0035	[–0.86, –0.206]	–0.531	0.180	0.0891	[–1.249, 0.188]
Time	–0.184	1.116	0.8716	[–2.58, 2.210]	–0.184	1.142	0.8746	[–2.647, 2.279]
R-squared	0.969				0.969			
Observations *N*	270				270			
Cluster number	15				15			

Note: SE stands for standard error, and CI stands for confidence interval.

The coefficient estimated for agricultural water use is statistically significant at the 10% level ($p < 0.1$) and indicates that one extra TAF of agricultural water use is expected to increase groundwater extraction by an average of 0.58 TAF. The coefficient estimated for urban water use is statistically significant at the 1% level ($p < 0.01$) and suggests that one extra TAF of urban water use would increase groundwater extraction by 0.63 TAF. The Greater Los Angeles and Greater San Diego areas, which are among the most populated areas in the nation, are in this zone. Urban water use accounts for a large fraction of the total water used in the zone.

The coefficient estimated for deliveries from miscellaneous surface water sources is statistically significant at the 5% level ($p < 0.05$) and indicates that one extra TAF of delivery is expected to decrease groundwater use by an average of 0.58 TAF. The large absolute value and statistical significance of this result shows the practical impact that the availability of local water resources and other supplemental surface water sources (excluding SWP and Colorado River) have on groundwater extraction in this zone.

The coefficient estimated for deliveries from SWP is statistically significant at the 5% level ($p < 0.05$) and indicates that one additional TAF of delivery from SWP is expected to reduce groundwater extraction by an average of 0.59 TAF. The coefficient estimated for deliveries from the Colorado River is also statistically significant at the 5% level ($p < 0.05$) and suggests that one extra TAF of water delivered from the Colorado River is expected to reduce groundwater extraction in the zone by an average of 0.57 TAF. The large absolute value and statistical significance of these coefficients demonstrate the relevance of deliveries from SWP and the Colorado River on groundwater extraction in the Southern California zone. In addition, the coefficient associated with recycled water is statistically significant at the 10% level ($p < 0.1$) and suggests that one additional TAF of recycled water would reduce groundwater extraction by an average of 0.53 TAF. This demonstrates the practical effects that increasing water recycling and reuse in the Southern California zone have on reducing the pressure on groundwater extraction in the zone.

The coefficient associated with time is not statistically significant and, therefore, nothing can be said about a potential time trend associated with groundwater extraction in the zone. All results are consistent across the two inference methods used.

5.6. Further discussion

The results presented above point to opportunities that could potentially reduce the pressure from water use and supplies on the groundwater resources in the state. In the North Coast and Lahontan zone, we estimate (with a moderate level of uncertainty) that reducing agricultural water by a certain amount would reduce groundwater extraction by approximately 20 percent of that amount. In the SF Bay and Central Coast zone we estimate that reducing urban water use would avoid groundwater extraction by approximately 80 percent of that reduction and reducing agricultural water could reduce groundwater extraction by almost the same amount. We also find that groundwater extraction could be increasing over time. The latter aspect could be associated with a scaling effect (e.g., an increase in population or irrigated land) and/or with an intensity effect (e.g., an increase in the gallons per capita per day, or a shift towards more water intensive crops).

For the Northern Central Valley zone, the results suggest that urban and environmental water use are the main drivers to variations in groundwater extraction. A reduction in the former could reduce groundwater extraction by approximately 50 percent of that reduction, and a reduction in the latter by approximately 66 percent. We also find that groundwater extraction has been increasing over time. The increase could be associated with an increase in population or in the gallons per capita per day, or with an increase in the environmental use of water.

The Southern Central Valley zone is an agriculture intensive area, where groundwater plays a relevant role. We estimate that a decrease in agricultural water in the zone reduces groundwater extraction by approximately the same amount. We also estimate that a decrease in deliveries from the SWP and CVP can increase groundwater extraction, respectively, in approximately 50 percent and 100 percent of that decrease.

Finally, in the Southern California zone, we estimate that reducing agriculture and urban water use can reduce groundwater extraction, respectively, by approximately 58 percent and 64 percent of the reduced water. We also estimate that a decrease in deliveries from the SWP, the Colorado River, and from local and other sources of surface water available in the zone, could increase groundwater extraction in approximately 57 to 59 percent of the decreased deliveries. In addition, increasing water recycling and reuse in the zone could contribute to reducing groundwater extraction by approximately 53 percent of the amount of water recycled or reused.

While the above results and panel analysis methodology based on robust inference can be useful in several ways, it is important to recognize the limitations of the results and methodology. We note the following major limitations. First, the water balance data we rely on are estimated mainly from modeling. Due to the very limited availability of groundwater pumping data, particularly when the pumping is conducted on farm, the amount of groundwater extracted in each planning area could be either under- or overestimated. Nevertheless, we believe that analyzing the data at the level of the geographic zones we defined, should partially compensate for the differences between the estimated and the actual groundwater extractions. Second, the cross-sectional and time dimensions in our panel data are short. Our dataset, despite being a balanced panel data, includes only 18 years of observations, and in some cases as few as four clusters of data. We attempt to reduce the bias that could result from analyzing a small sample of panel data using the conventional CRVE method. We employ a method with small-sample corrections for inference. The method we employ leads to larger confidence intervals that are adjusted to if our samples were larger and, therefore, to more appropriate results.

We assumed that panel data models satisfy the usual regression model conditions (Dougherty, 2011), which mainly include (1) linearity—the relationship between the dependent variable and the independent variables is linear; (2) homoscedasticity—the disturbance term is homoscedastic; (3) independence—the values of the disturbance term are independent of each other; and (4) normality—the disturbance term is normally distributed. However, it is unlikely that real-world applications strictly satisfy all these conditions. Therefore, it is often required to check the model against the assumptions (Dougherty, 2011; Wooldridge, 2015) and make trade-offs between simplicity and complexity in the modeling (Zellner et al., 2001). Specifically, we note two considerations for time series and panel data modeling: stationarity and endogeneity. Stationarity testing is often applied in time series modeling to avoid spurious regression. Generally speaking, stationarity testing does not work well with short time series due to distortions from the size of the sample and low power of the testing. A few stationarity testing methods have been developed for when the time and cross-section dimensions of the panel data are sufficiently large (Pesaran, 2015). Endogeneity problems could potentially exist in any complex real-world applications and there is no perfect solution for this problem (Roberts and Whited, 2013). Although, as mentioned above, fixed-effect estimations relax the strict exogeneity assumption and are more credible and preferable to common OLS estimations (Collischon and Eberl, 2020; Roberts and Whited, 2013), as pointed out by Roberts and Whited (2013), "fixed effects cannot remedy any arbitrary endogeneity problem and are by no means an endogeneity panacea".

6. Conclusions

California relies on both surface water and groundwater for its water supplies. In some regions of the state, groundwater is the main—if not the only—source of water. Population growth and economic activity, particularly agricultural activity, have exerted increasing pressure on groundwater resources, with long term implications. A sustainable approach for groundwater management and use is therefore necessary. An important aspect for a sustainable management of groundwater resources is the understanding of what factors put pressure on groundwater resources and what factors can help to alleviate some of that pressure.

This study investigates the dynamics of groundwater extraction and the impact that water use and surface water supplies have on groundwater extraction in California. We rely on a panel with data from the state's 56 Water Plan planning areas over the 1998–2015 period. We emphasize in our analyses agricultural and urban water use and the major water projects in the state. Our results, despite being somehow intuitive, quantify the trade-offs between groundwater extraction and water use and surface water supplies in five geographic zones in the state. They further shed light on the potential opportunities to reduce pressure on groundwater resources in the state and can, therefore, inform policy making. Policies that increase irrigation water efficiency or lead growers to switch to less water intensive crops would reduce agricultural water requirements in central California and, consequently, reduce the pressure on groundwater resources in the region by the same magnitude as the reduced requirements of agricultural water. Similarly, policies that limit the amount of urban indoor and outdoor (landscape irrigation) water use in Southern California would contribute to reducing groundwater extraction in the region by the same extent as the water use is reduced. Additional policy measures targeting other regions in the state can also be informed by our findings.

Care should be taken, however, when using our results. First, some of the water data we rely on are modeled, not observed, and the amount of groundwater extracted could be either under- or overestimated. Second, longer time-series would better reflect the long-term dynamics of groundwater extraction and use, and therefore provide more accurate results. Third, the linear regression approach we used does not capture the non-linear aspects of the dynamics of the systems we investigated. Despite those limiting factors, the approach we used can provide valuable information for groundwater resource planning and management. Future research should combine our results with energy and economic parameters to evaluate the cost-effectiveness of policies that would reduce groundwater extraction and potential overdraft. New studies can also rely on this approach to evaluate the impacts from water supply and use on groundwater resources usage and sustainability in other regions of the world.

Finally, in addition to contributing to a better understanding of the relationships between groundwater extraction and water use and supply in California, our study calls attention to the potential sensitivity of results from panel data analysis to the inference method used. Particularly, we show that in case of small samples, like the ones we relied on for some of the geographic zones we analyzed, and which can be found in several other research subjects and studies, the choice of the inference method can significantly affect the statistical significance of some regressors, and thus change what can be concluded from the model.

Acknowledgements

This work was supported by the California Energy Commission (CEC) under Agreement Number EPC-15-035. It does not necessarily represent the views of the CEC, its employees, or the State of California. The CEC, the State of California, its employees, contractors, and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this article; nor does any party represent that the uses of this information will not infringe upon privately owned rights. We thank Joe O'Hagan, CEC, for his feedback on a draft version of this paper, and Jennifer D. Kofoid-Stricklin, California Department of Water Resources, for providing state water balance and groundwater data. We also thank three anonymous reviewers for their helpful and constructive comments and suggestions which helped improve this manuscript. This manuscript has been authored by authors at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

References

Altıntas, H., Kassouri, Y., 2020. Is the environmental Kuznets Curve in Europe related to the per-capita ecological footprint or CO_2 emissions? Ecological Indicators 113, 106187.

Asici, A.A., 2013. Economic growth and its impact on environment: A panel data analysis. Ecological Indicators 24, 324-333.

Balasubramanian, R., 2015. Climate Sensitivity of Groundwater Systems Critical for Agricultural Incomes in South India. South Asian Network for Development and Environmental Economics (SANDEE), SANDEE Working Paper No. 96–15, Kathmandu, Nepal. http://www.sandeeonline.org/uploads/documents/publication/1061_PUB_Working_Paper_96_Balu.pdf.

BEA, 2019. Gross Domestic Product by State, Fourth Quarter and Annual 2018. US Department of Commerce, Bureau of Economic Analysis (BEA), Washington, DC, USA. www.bea.gov.

Bell, R.M., McCaffrey, D.F., 2002. Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology 28, 169-181.

Bureau of Reclamation, 2008. Water Supply and Yield Study. Bureau of Reclamation's Mid-Pacific Region, Sacramento, CA, USA. https://www.usbr.gov/mp/cvp/docs/water-supply-and-field-study.pdf.

Bureau of Reclamation, 2017. Central Valley Project (CVP). Bureau of Reclamation's Mid-Pacific Region, Sacramento, CA, USA. https://www.usbr.gov/mp/cvp/.

Bureau of Reclamation, 2019a. Boulder Canyon Project - All-American Canal System. Bureau of Reclamation, Boulder City, NV, USA. https://www.usbr.gov/projects/index.php?id=514.

Bureau of Reclamation, 2019b. Central Valley Project. Bureau of Reclamation's Mid-Pacific Regional Office, Sacramento, CA, USA. https://www.usbr.gov/projects/index.php?id=506.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 90, 414-427.

Chakraborty, D., Mukherjee, S., 2013. How do trade and investment flows affect environmental sustainability? Evidence from panel data. Environmental Development 6, 34-47.

Changming, L., Jingjie, Y., Kendy, E., 2001. Groundwater exploitation and its impact on the environment in the North China Plain. Water International 26, 265-272.

Chen, C.-C., McCarl, B.A., Schimmelpfennig, D.E., 2004. Yield variability as influenced by climate: A statistical investigation. Climatic Change 66, 239-261.

Collischon, M., Eberl, A., 2020. Let’s talk about fixed effects: Let’s talk about all the good things and the bad things. KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie 72, 289-299.

Dahlsrud, A., 2008. How corporate social responsibility is defined: An analysis of 37 definitions. Corporate Social Responsibility and Environmental Management 15 (1), 1-13.

Del Baldo, M, 2013. Territorial social responsibility and territorial small and medium-sized enterprises. In Encyclopedia of Corporate Social Responsibility (2515-2523). Springer, Berlin, Heidelberg.

Dougherty, C., 2011. Introduction to econometrics, 4th ed. Oxford University Press, Oxford ; New York.

DWR, 2016. California’s Groundwater Update 2013. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Water-Basics/Groundwater.

DWR, 2019a. California Water Plan. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Programs/California-Water-Plan.

DWR, 2019b. The California Water System. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Water-Basics/The-California-Water-System.

DWR, 2019c. Groundwater. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Water-Basics/Groundwater.

DWR, 2019d. SGMA Groundwater Management. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Programs/Groundwater-Management/SGMA-Groundwater-Management.

DWR, 2019e. State Water Project. California Department of Water Resources (DWR), Sacramento, CA, USA. https://water.ca.gov/Programs/State-Water-Project.

Elshall, A.S. et al, 2020. Groundwater sustainability: A review of the interactions between science and policy. Environmental Research Letters 15 093004.

Famiglietti, J.S., Lo, M., Ho, S.L., Bethune, J., Anderson, K.J., Syed, T.H., Swenson, S.C., de Linage, C.R., Rodell, M., 2011. Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophys Res Lett 38 (3).

Fan, J.-L., Hu, J.-W., Kong, L.-S., Zhang, X., 2017. Relationship between energy production and water resource utilization: A panel data analysis of 31 provinces in China. J Clean Prod 167, 88-96.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2011. Applied longitudinal analysis, 2nd ed. Wiley, Hoboken, N.J.

Freeman, I., Hasnaoui, A., 2011. The meaning of corporate social responsibility: The vision of four nations. Journal of Business Ethics 100 (3), 419-443.

Gleeson, T. et al, 2010. Groundwater sustainability strategies. Nature Geoscience 3, 378-379.

Greene, W.H., 2018. Econometric Analysis, 8th ed. Pearson Education, New York, NY.

Hanak, E., Lund, J., Dinar, A., Gray, B., Howitt, R., Mount, J., Moyle, P., Thompson, B., 2011. Managing California's Water: From Conflict to Reconciliation. Public Policy Institute of California, San Francisco, CA, USA. https://www.ppic.org/publication/managing-californias-water-from-conflict-to-reconciliation/.

Hao, Y., Chen, H., Wei, Y.M., Li, Y.M., 2016. The influence of climate change on CO_2 (carbon dioxide) emissions: an empirical estimation based on Chinese provincial panel data. J Clean Prod 131, 667-677.

Hendricks, N.P., Peterson, J.M., 2012. Fixed effects estimation of the intensive and extensive margins of irrigation water demand. Journal of Agricultural and Resource Economics 37, 1-19.

Izady, A., Davary, K., Alizadeh, A., Ghahraman, B., Sadeghi, M., Moghaddamnia, A., 2012. Application of "panel-data" modeling to predict groundwater levels in the Neishaboor Plain, Iran. Hydrogeology Journal 20, 435-447.

Jones, L.E., Milligan, K.S., Stabile, M., 2015. Child Cash Benefits and Family Expenditures: Evidence from the National Child Benefit. NBER Working Papers 21101, National Bureau of Economic Research (NBER).

Kishore, P., Singh, D.R., Chand, P., Prakash, P., 2020. What determines groundwater depletion in India? A meso level panel analysis. Journal of Soil and Water Conservation 19, 388-397.

Konikow, L.F., 2013. Groundwater Depletion in the United States (1900−2008). U.S. Geological Survey, Scientific Investigations Report 2013-5079, Reston, VA.

Lall, U., Josset, L., Russo, T., 2020. A snapshot of the world's groundwater challenges. Annual Review of Environment and Resources 45, 171-194.

Lund, J., Medellin-Azuara, J., Durand, J., Stone, K., 2018. Lessons from California's 2012–2016 drought. Journal of Water Resources Planning and Management 144, 04018067.

MacKinnon, J.G., 2019. How cluster-robust inference is changing applied econometrics. Economics Department, Queen's University. https://ideas.repec.org/p/qed/wpaper/1413.html.

Marcelo, T., James, E.P., Elizabeth, T., 2017. REG_SANDWICH: Stata module to compute cluster-robust (sandwich) variance estimators with small-sample corrections for linear regression. Statistical Software Components S458352, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s458352.html.

Matthew, D.W., 2014. Reworking Wild Bootstrap Based Inference For Clustered Errors. Economics Department, Queen's University, Kingston, Ontario, Canada. http://qed.econ.queensu.ca/working_papers/papers/qed_wp_1315.pdf.

Mays, L.W., 2013. Groundwater resources sustainability: Past, present, and future. Water Resources Management 27, 4409-4424.

Mehta, V.K., Young, C., Bresney, S.R., Spivak, D.S., Winter, J.M., 2018. How can we support the development of robust groundwater sustainability plans? California Agriculture 72, 54-64.

Moran, T., Choy, J., Sanchez, C., 2014. The Hidden Costs of Groundwater Overdraft. Stanford University, Stanford, CA, USA. http://waterinthewest.stanford.edu/groundwater/overdraft/index.html.

MWD, 2019. Colorado River Aqueduct. Metropolitan Water District of Southern California (MWD), Los Angeles, CA, USA. http://www.mwdh2o.com/AboutYourWater/Sources%20Of%20Supply/Pages/Imported.aspx.

Ojha, C., Shirzaei, M., Werth, S., Argus, D.F., Farr, T.G., 2018. Sustained groundwater loss in California's Central Valley exacerbated by intense drought periods. Water Resources Research 54, 4449-4460.

Ojha, C., Werth, S., Shirzaei, M., 2019. Groundwater loss and aquifer system compaction in San Joaquin Valley during 2012–2015 drought. Journal of Geophysical Research: Solid Earth 124, 3127-3143.

Omojolaibi, J.A., Nathaniel, S.P., 2020. Assessing the potency of environmental regulation in maintaining environmental sustainability in MENA countries: An advanced panel data estimation. Journal of Public Affairs, e2526.

Pandey, V.P. et al, 2011. A framework for measuring groundwater sustainability. Environmental Science and Policy 14, 396-407.

Pesaran, M.H., 2015. Time Series and Panel Data Econometrics. Oxford University Press, Oxford.

Pfeiffer, L., Lin, C.Y.C., 2014. Does efficient irrigation technology lead to reduced groundwater extraction? Empirical evidence. Journal of Environmental Economics and Management 67, 189-208.

Pustejovsky, J.E., Tipton, E., 2018. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. Journal of Business & Economic Statistics 36, 672-683.

Roberts, M.R., Whited, T.M., 2013. Chapter 7 - Endogeneity in Empirical Corporate Finance. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. Elsevier, pp. 493-572.

Rusciano, V., Scarpato, D., Civero, G., 2019. Territorial social responsibility: A cluster analysis on a case study. Quality-Access to Success 20 (S2), 543-548.

Sadik-Zada, E.R., Gatto, A., 2021. The puzzle of greenhouse gas footprints of oil abundance. Socio-Economic Planning Sciences 75, 100936.

Scott, C.A., Shah, T., 2004. Groundwater overdraft reduction through agricultural energy policy: insights from India and Mexico. International Journal of Water Resources Development 20, 149-164.

Tipton, E., Pustejovsky, J.E., 2015. Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. Journal of Educational and Behavioral Statistics 40, 604-634.

Watto, M.A., Mugera, A.W., 2016. Groundwater depletion in the Indus Plains of Pakistan: Imperatives, repercussions and management issues. International Journal of River Basin Management 14, 447-458.

Wooldridge, J.M., 2015. Introductory Econometrics: A Modern Approach, 6 ed. Cengage Learning, Boston MA, USA.

Zellner, A., Keuzenkamp, H.A., McAleer, M., 2001. Simplicity, inference and modeling : keeping it sophisticatedly simple. Cambridge University Press, Cambridge.

Zetland, D., 2011. Colorado River Aqueduct. In: Danver, S.L., Burch, J.R. (Eds.), Encyclopedia of Water Politics and Policy in the United States. CQ Press, Washington, D.C., pp. 420-421.