

# California Surface Water Data:

## Details of data sources and processing choices

Nick Hagerty

December 23, 2021

### 1 Deliveries, diversions, and allocations

#### 1.1 State Water Project (SWP)

**Maximum entitlements.** Maximum contract (Table A) amounts are taken from Table B-4 of Bulletin 132-18 Appendix B, downloaded from the DWR at <https://water.ca.gov/Programs/State-Water-Project/Management/Bulletin-132>. They are available by user, sector, and year from 1962 through 2018. For a time-invariant baseline amount, I choose maximum entitlements in 1990, because that is the first year all currently-existing sections of the SWP were completed, and maximum Table A amounts were stabilized.

**Allocation percentages.** Percentage allocations by year and sector are available from 1970 through 2018. For 1996-2018 they are taken from published notices to SWP contractors (i.e., users holding contracts), downloaded from the DWR at <https://water.ca.gov/Programs/State-Water-Project/Management/SWP-Water-Contractors>. For 1970-1995 they are taken from Table 2-3 of the Monterey Plus Draft Environmental Impact Report, downloaded from [http://www.water.ca.gov/environmentalservices/monterey\\_plus.cfm](http://www.water.ca.gov/environmentalservices/monterey_plus.cfm).

**Deliveries.** Deliveries by user, sector, and year are taken from Table B-5B of Bulletin 132-18 Appendix B, downloaded from DWR at <https://water.ca.gov/Programs/State-Water-Project/Management/Bulletin-132>. They are available from 1962 through 2018 and include Table A amounts plus non-project, surplus, and Article 21 water deliveries.

#### 1.2 Central Valley Project (CVP)

**Maximum entitlements.** Maximum contract volumes are downloaded from the U.S. Bureau of Reclamation (USBR) at <https://www.usbr.gov/mp/cvp-water/water-contractors.html>. I sum contract volumes by user and sector (municipal & industrial vs. agricultural). For three contracts that are shared among multiple users, I split maximum volumes among the users by their ratio of average deliveries, if available, or evenly otherwise. For two observations whose municipal & industrial

volume exceeds maximum contract volume, I set municipal & industrial volume to equal maximum contract volume.

In order to calculate user-specific allocation percentages, I also sum contracts by base vs. project supply. Base supply is contracts for delivery of water based on water rights pre-dating the CVP, while project supply is contracts for delivery of new water made available by the CVP. Contracts in the category “South of Delta Water Rights Contracts” (exchange contractors) are base supply, as are contracts held by Oakdale Irrigation District, South San Joaquin Irrigation District, and wildlife refuges. Sacramento River settlement contractors hold both base and project supply; these amounts are downloaded from [http://baydeltaoffice.water.ca.gov/modeling/hydrology/CalSim3/documentation/ReleaseReady112917/IndividualChapters/CS3\\_VolI\\_14\\_ContractsandWaterRights.pdf](http://baydeltaoffice.water.ca.gov/modeling/hydrology/CalSim3/documentation/ReleaseReady112917/IndividualChapters/CS3_VolI_14_ContractsandWaterRights.pdf). All other users are project supply. For the City of West Sacramento, base/project split is not available, so I assume its entire contract volume is project supply. The sector split is not given for Sacramento River settlement contractors, so I assume all are agricultural (their names and internet searches suggest the vast majority, if not all, are indeed farm operations).

**Allocation percentages.** Percentage allocations by year and contract category are downloaded from the USBR at [https://www.usbr.gov/mp/cvo/vungvari/water\\_allocations\\_historical.pdf](https://www.usbr.gov/mp/cvo/vungvari/water_allocations_historical.pdf). They are available for each contract year (the 12 months from March of the named year through February of following year) from 1977 to the present. Percentage allocations are determined separately for each of 13 contract categories (North of Delta Agricultural Contractors, North of Delta Urban Contractors (M&I), North of Delta Wildlife Refuges, North of Delta Settlement Contractors/Water Rights, American River M&I Contractors, In Delta - Contra Costa, South of Delta Agricultural Contractors, South of Delta Urban Contractors (M&I), South of Delta Wildlife Refuges, South of Delta Settlement Contractors/Water Rights, Eastside Division Contractors, Friant - Class 1, Friant - Class 2). Some categories are combined in earlier years; when “American River M&I Contractors” (i.e., municipal and industrial) and “In Delta – Contra Costa” (i.e., the Contra Costa Water District, which serves suburban areas east of the San Francisco Bay) are not specified separately, I impute the value for “North of Delta Urban Contractors”. In addition to these, allocations are always 100% for the Hidden and Buchanan Units of the Friant Division. For “uncontrolled season” releases in the Friant Division in 2016 and 2018, I assume the entire available volume is destined for Class 2 users and divided equally among them.

For each user and year, I obtain overall allocation percentages by calculating a weighted average across each user’s contract types, where the weights are the maximum contract volumes.

**Deliveries.** Deliveries by user and month are downloaded from USBR at <https://www.usbr.gov/mp/cvo/deliv.html>. They are available from 1993 to the present; deliveries for 1985-92 are also online but not yet digitized. I sum deliveries across months within both calendar year (for consistency with other data sources) and contract year (i.e., Mar-Feb, for consistency with allocation percentages). The difference between these year definitions is relatively small; only 8.6% of water

deliveries occur in January and February.

Sector is given in the maximum contract volume dataset. Most users are classified as fully agricultural or fully municipal. For users that hold contracts for both sectors, I assume deliveries are divided by sector in the same proportion as the maximum contract volume. For delivery recipients who do not hold contracts (and therefore do not appear in the contract dataset), I assign to agricultural those whose name includes one of several keywords (farm, I.D., irrigation, land, ranch, vineyard), and to municipal those whose name includes one of several keywords (city of, construction, golf, inc., properties, P.U.D., university). I then assign to municipal several users known as such (La Grange W.D., Lakeside W.D., and Los Banos Gravel) and the remainder to agricultural.

### 1.3 Lower Colorado Project

**Maximum entitlements.** Maximum entitlements and sector of each user are constructed from lists of Lower Colorado River water entitlements in California, downloaded from <https://www.usbr.gov/lc/region/g4000/contracts/entitlements.html>, and Appendix E of the Final Environmental Impact Statement of 2007 for the Colorado River Interim Guidelines for Lower Basin Shortages and Coordinated Operations for Lakes Powell and Mead, downloaded from <https://www.usbr.gov/lc/region/programs/strategies/FEIS/index.html>. Maximum entitlements do not appear to have changed in the last couple of decades.

**Allocation percentages.** Percentage allocations are all 100% because a shortage had never been declared on the Colorado River prior to 2021.

**Deliveries.** Diversions by user and month are available for 2003-2018 in the annual Colorado River Accounting and Water Use Reports for Arizona, California, and Nevada, and for 1980-2002 in the annual Compilation of Records in Accordance with Article V. All are downloaded from USBR at <https://www.usbr.gov/lc/region/g4000/wtracct.html>. Diversions are summed across months within calendar year. When diversions are attributed to one user but transferred to another user, I classify them as deliveries to the receiving user.

### 1.4 Surface Water Rights

Surface water rights are not directly measured, so I calculate average annual diversions during 2010-14 from self-reports on file with the State Water Resources Control Board (SWRCB). These average diversions represent the best available data on both maximum entitlements and actual diversions for holders of surface water rights. (The SWRCB holds records of the face value of some rights, but these have been repeatedly shown to be wildly unreliable by researchers and journalists, and the SWRCB's own water supply models use reported diversions, not face values. Face values are also not available for riparian or pre-1914 appropriative rights.) Water rights have changed little since 1980 and are almost never curtailed, so it is reasonable to approximate them as permanent, fixed entitlements with a 100% allocation percentage.

**Reporting.** Users holding post-1914 appropriative rights are required to submit annual reports of use. Riparian and pre-1914 appropriative rights were not systematically tracked by any government agency prior to 2010. However, since 2010 these rights holders must submit Statements of Diversion & Use, with civil penalties for noncompliance. From 2010 to 2016 this reporting requirement was once every three years; since 2016 rights holders must report every year. This means from 2012 onward, the SWRCB had at least one report of quantity diverted of nearly every water right claimed in California. In addition to these regular reporting requirements, in 2015 the SWRCB required major rights-holders in the Central Valley to report their diversions from 2014 (Informational Order WR 2015-0002-DWR).

Although these diversion statements are self-reported, it is reasonable to treat them as the full, legally defensible value of present water rights. This is because appropriative rights are based on documented continuous beneficial use, and these statements are public information, so they could be used in future legal disputes. Therefore, users have incentives to neither report less than they would like to use in the future nor more than other evidence would support.

**Data.** All of SWRCB's records – water right permits, licenses, and Statements of Diversion & Use – are publicly available in the SWRCB's Electronic Water Rights Information Management System (eWRIMS). Prior to November 2021, the online eWRIMS interface made it difficult to view or download details for many records at once. Instead, I use a full extraction of the eWRIMS database as of February 26, 2015 that was posted online as an exhibit in a 2016 administrative civil liability hearing for the Byron-Bethany Irrigation District. This was downloaded from [http://www.waterboards.ca.gov/waterrights/water\\_issues/programs/hearings/byron\\_bethany/docs/exhibits/pt/wr70.csv](http://www.waterboards.ca.gov/waterrights/water_issues/programs/hearings/byron_bethany/docs/exhibits/pt/wr70.csv). This dataset contains records of all water rights in the state and includes reported diversions for 2010 through 2013. It also indicates the face value of rights (for post-1914 rights), types of beneficial uses, year of first diversion, and latitude & longitude of the point of diversion (POD).

I supplement this file with another dataset that contains reported diversions in 2014. This dataset, the 2015 Informational Order Demand Dataset, was developed by SWRCB for a water availability analysis in 2015 and posted online. It was downloaded from: [https://www.waterboards.ca.gov/waterrights/water\\_issues/programs/drought/analysis/docs/info\\_order\\_demand.xlsx](https://www.waterboards.ca.gov/waterrights/water_issues/programs/drought/analysis/docs/info_order_demand.xlsx).

**Cleaning.** I follow the data cleaning and quality control procedures described by SWRCB in another exhibit ("Exhibit WR-11: Testimony of Jeffrey Yeazell", [http://www.waterboards.ca.gov/waterrights/water\\_issues/programs/hearings/byron\\_bethany/docs/exhibits/pt/wr11.pdf](http://www.waterboards.ca.gov/waterrights/water_issues/programs/hearings/byron_bethany/docs/exhibits/pt/wr11.pdf)), adding a number of further checks and corrections. I drop rights that are canceled, inactive, removed, or revoked, and those not yet active, and minor types of water rights (such as stock ponds and livestock), leaving only appropriative rights and statements of diversion and use.

The dataset has 95,535 observations at the level of right by point of diversion (POD) by beneficial use type, with a few duplicates. I drop duplicate observations so that the combination of these three variables form a unique key, then I reshape to the level of right by point of diversion, resulting

in a dataset of 56,508 observations. For rights with multiple PODs, I keep only one so that a right is a unique record. SWRCB chooses the POD by alphabetical order on watershed name; I instead choose the POD from the watershed, source within watershed, and 12-digit hydrologic unit within source with the most PODs for that right; if there duplicates within 12-digit hydrologic unit, I keep the POD with the earliest number.

To construct the year a right first began, I use the year of first use when available (nearly all pre-1914 and riparian rights holders, and some post-1914 rights holders), followed by original permit issue date when available, license original issue date when available, and record status year when available. To construct the year a right ended, I take the first year a right was canceled, closed, inactive, rejected, or revoked.

I remove non-consumptive diversions by power-only and aquaculture-only, following SWRCB procedure. For rights that report no diversion to storage, I set diversions to zero. For diversions that do report diversion to storage, I subtract the amount used from the amount diverted, censoring negative values at zero.

I correct for over-reporting following SWRCB procedure. For post-1914 rights, most observations include the face value of the rights, so if reported diversions in a year exceed the face value, I scale down that year's monthly reports so their total equals the face value. For pre-1914 and riparian rights, face value is not available but some report irrigated acres, so if reported diversions exceed 8 acre-feet per acre, I scale down that year's monthly reports so their total equals this limit. I add one more correction not performed by SWRCB: For post-1914 rights for which the face value is unavailable but irrigated acres is available, I apply the same acres-based correction, but conservatively only for observations whose total diversions exceed 80 acre-feet per acre.

I make further corrections to high outliers in a process not separately conducted by SWRCB. Many of these are likely errors in unit selection; there may also be low outliers, but I cannot detect them effectively. I calculate the standard deviation of the natural log of all monthly diversion values. For observations for which this standard deviation is greater than 2 and the average annual diversion exceeds the face value of the rights by more than 100 acre-feet, for years in which the total diversion exceed the smallest annual total by more than 100 times, I scale down each monthly value proportionally so that that year's total equals the smallest annual total. Although this correction process affects only 82 observations, it changes the total statewide reported diversions by more than 12 orders of magnitude. I also drop one riparian right held by an individual that implausibly reports an annual diversion of more than 100,000 acre-feet.

**Further sample restrictions.** I drop water held by federal and state projects, which are accounted for in separate datasets. I drop non-consumptive rights: those whose beneficial use is aesthetic, aquaculture, fish & wildlife, incidental power, power, recreational, or snow-making; several known environmental or recreational users (California Department of Fish & Wildlife; California Department of Forestry & Fire Prevention; California Department of Parks & Recreation; Nature Conservancy; Pine Mountain Lake Association; Tuscany Research; U.S. Bureau of Land Manage-

ment; U.S. National Park Service; U.S. Forest Service; U.S. Department of Fish & Wildlife; White Mallard, Inc.; Woody's on the River, LLC); two known electricity-generating users (Pacific Gas & Electric Co., Southern California Edison Company); and those whose name includes one of several keywords (duck club, gun club, power, preservation, shooting club, waterfowl, wetlands). I drop a small number of rights (151) whose point of diversion is unknown.

**Sector.** I categorize each right as agricultural if its record lists irrigation or stockwatering as a beneficial use, and municipal & industrial otherwise. I then set to municipal all users whose names include "city of" or "golf", and several users in Orange County found to be municipal in internet searches (Irvine Ranch W.D., Orange County W.D., Santa Margarita W.D., Serrano W.D.).

**Final variables.** For each right, I average across reported annual diversions from 2010 through 2014. I then sum across rights within user and sector, keeping location information for the point of diversion with the largest volume. Finally, CVP settlement and exchange contractors likely have the same rights reported in both CVP and rights datasets. So as not to double-count these, I subtract the maximum contract volume for base supply from diversion volumes.

## 1.5 Reallocation of subcontracts

Several water users that hold especially large water entitlements function as pass-through entities for other, smaller water districts. The wholesale agency receives water from the project and allocates it to its member districts on the basis of long-term contracts. For users like these that I can identify and obtain information about, I reallocate entitlement and delivery volumes to their members.

For the Kern County Water Agency, I reallocate volumes to 15 member units based on the table found at <http://www.wakc.com/wp-content/uploads/2016/01/SWP-Contracts-in-Kern-County.pdf>. I assume that deliveries are passed through in proportion to each member agency's contract amount, within sector.

For the Joint Water Districts Board, I reallocate volumes to 4 member units based on information found in the 2015 Agricultural Water Management Plans, downloaded from [https://wuedata.water.ca.gov/awmp\\_plans](https://wuedata.water.ca.gov/awmp_plans). A 1969 agreement allocates the surface water rights held by the the Board in the following proportions: 24 percent to the Butte Water District, 29 percent to the Biggs West Gridley Water District, 27 percent to the Richvale Irrigation District, and 20 percent to the Sutter Extension Water District.

For the Kings River Water Association, I reallocate volumes to 16 member districts listed in the Kings River Handbook, downloaded from [http://www.centralvalleywater.org/wp-content/uploads/2017/12/Kings\\_River\\_Handbook\\_2009.pdf](http://www.centralvalleywater.org/wp-content/uploads/2017/12/Kings_River_Handbook_2009.pdf). A total of 28 districts are listed in this document, but 10 have no further information either on the internet or in other datasets, and 2 do not use Kings River water. The remaining 16 members are: Alta Irrigation District, Clark's Fork Reclamation District, Consolidated Irrigation District, Corcoran Irrigation Company, Empire West Side Irrigation District, Fresno Irrigation District, James Irrigation District, John Heinlen Mutual Water Com-

pany, Kings River Water District, Laguna Irrigation District, Lemoore Canal & Irrigation Company, Liberty Canal Company a.k.a. Liberty Water District, Reed Ditch Company a.k.a. Murphy Slough Association, Riverdale Irrigation District, Stinson Canal and Irrigation Company a.k.a. Stinson Water District, Stratford Irrigation District, Tranquillity Irrigation District, and Tulare Lake Basin Water Storage District. Data on allocations among these members is unavailable, so I assume volumes are allocated evenly on a per-acre basis, using calculations of cropland area within the service area boundaries of each member district.

## 1.6 Combined dataset

I combine the four sources above to create a full user-by-year dataset of water supplies and allocations. I merge all sources on name and year, matching users via the crosswalk file (see Section 2), and sum supplies and maximum entitlements across sources. I restrict to years beginning in 1981, when data is available from all four sources, and ending in 2018. The result is a nearly balanced panel of 7,188 users over 38 years, for a total of 273,144 observations.

## 1.7 Geolocations

I attach geographical location information to the data for nearly all users. When possible, I use centroids from the district boundary file (see Section 3). For users not available in the district boundary file, I attach the point of diversion listed in the rights dataset. (Note the point of diversion may be different from the place of use, which is unobserved.) For users not available in either the user location file or rights dataset, I merge to a dataset of 65 manually geolocated users. For these users, I generate coordinates based on addresses, towns, or maps found via user websites and other publicly available documents. Only remaining 15 users could not be geolocated; they account for less than 0.01% of statewide entitlement volume.

Shapefiles for counties and other geographical divisions were downloaded in 2015 from DWR's webpage for the California Water Plan Update 2013; as of 2019 they were no longer available on the DWR website, but they remain available from the State of California Geoportal, <http://portal.gis.ca.gov/geoportal/>.

## 2 User crosswalk file

The crosswalk file that links water users by name across all datasets. To create it, I export raw names from each dataset and append them together. I strip punctuation and correct misspellings and other typos. I standardize common terms into acronyms (e.g., I.D. for irrigation district; M.W.C. for mutual water company; F.C.W.C.D. for flood control and water conservation district). For names of individual people, I match full names to entries with the same last name but only first initial(s) available. For agencies, when names are closely but not precisely similar I use agency websites and other publicly available documents to determine whether (a) one agency has changed its name,

(b) one name is erroneous, or (c) they are indeed distinct agencies. I use footnotes and notes in original data sources to link users with name changes over time, keeping the most recent name. When a merger has occurred, I roll users up into the most aggregate version to maintain consistent definitions. The exception is companies with service in multiple noncontiguous locations, for which I treat each location as a separate user. The final crosswalk file has 28,765 entries (input names) pointing to 14,830 targets (output names). Excluding transactions data, the file has 17,738 entries and 13,912 targets.

### 3 User polygon boundaries

By combining all the relevant and publicly available georeferenced digital maps I can find, I create a dataset of the most accurate locations, areas, and boundaries for as many water users as possible. I combine the following datasets and link them via the crosswalk file. For each user, I keep one shape (feature) according to the following priority order:

1. DWR's Water Districts Boundaries, downloaded via the Query link found at <https://gis.water.ca.gov/arcgis/rest/services/Boundaries/WaterDistricts/FeatureServer>
2. Federal, State, and Private Water Districts shapefiles maintained by USBR and DWR, downloaded from the California Atlas at <http://www.atlas.ca.gov/download.html>.
3. Mojave Water Agency Water Companies, downloaded at <https://www.mojavewater.org/geospatial-library.html>.
4. California Environmental Health Tracking Program's Water Boundary Tool, downloaded at <http://www.cehtp.org/page/water/download>.

Before I append (merge) sources, I combine noncontiguous shapes for the same user (dissolve to create multipart features). After selecting one shape per user, I calculate the user's centroid (restricted to within shape), area, and cropland area (via zonal statistics). Cropland area comes from the 2015 cropland mask from the USDA's Cropland Data Layer.