Domestic Well Areas Version 2.0

Rempel, J.*, C. Pace*, L. Cushing, R. Morello-Frosch. (2023). Domestic Well Areas Version 2.0. Drinking Water Tool metadata, prepared by the Water Equity Science Shop, UC Berkeley. *Designates shared co-first authorship.
Contact: cpace@berkeley.edu and jrempel@berkeley.edu

File name: DWA_v2_plss_040125.shp

Spatial Reference

Geographic Coordinate System	NAD 1983	Projected Coordinate System	NAD 1983 (Teale) Albers (Meters)
WKID	4269	Projection	3310
Authority	EPSG	Authority	EPSG
Angular Unit	Degree (0.0174532925199433)	Linear Unit	Meters (1.0)
Prime Meridian	Greenwich (0.0)	False Easting	0.00
Datum	D North American 1983	False Northing	-4000000.0
Spheroid	GRS 1980	Central Meridian	-120.0
Semimajor Axis	6378137.0	Standard Parallel 1	34.0
Semiminor Axis	6356752.314140356	Standard Parallel 2	40.5
Inverse Flattening	298.257222101	Latitude of Origin	0.0

Description

This shapefile contains 61,760 public land survey system (PLSS) domestic well areas. Each domestic well area is derived from an approximately 1x1 mile grid square identified by the field "MTRS" (Meridian, Township, Range and Section). Statewide, we estimate that 1.58 million people are served by approximately 292,093 domestic wells. Domestic well population is estimated for each domestic well area.

Water quality for arsenic, nitrate, hexavalent chromium (Cr[VI]), and 123-trichloropropane (123-TCP) were provided by State Water Resources Control Board 2024 Aquifer Risk Map individual contaminant layers. These layers contain summarized water quality risk per square mile section. The water quality risk is based on water quality results from public and domestic supply wells.

Methods

1. Improved domestic well locations (based on the methodology used in Rempel & Belfer, et al., in review[1])
    a. Downloaded Online System for Well Completion Reports (OSWCR) data on July 12, 2021, N= 1,032,652 well completion reports (CADWR, 2021[2])

b. Selected new domestic wells drilled on or after January 1, 1970, following the approach and time horizon used by the California State Water Resources Control Board (Water Board)[3], N=293,540
c. Used a multi-method approach to improve domestic well location estimates by:
    i. Matching OSWCR data to a statewide residential parcel dataset[4] by assessor's parcel number (APN) and county
    ii. Matching address strings in OSWCR to address strings in the statewide parcel dataset[4] using the RecordLinkage package to implement a Jaro-Winkler string distance algorithm[5]
        1. Addresses were preprocessed to normalize characters, standardize street names, and remove punctuation and special characters. We then required exact numeric matches and identical city and county names, and we used a minimum string distance threshold of >0.96, which we determined to be a conservative threshold through manual inspection of the address pairs.
    iii. Geocoding using the Google Geocoding API tool
        1. Well addresses that did not match via the above methods were retained for geocoding after pre-processing to ensure that the well record included a number in the street address but did not consist only of numbers. These addresses were preprocessed using an address cleaning script via the usaddress library in Python[6] and manual corrections to city spelling errors. We then geocoded addresses using the Google Geocoding API, accessed via a Python script. We retained only high-quality geocoding results (e.g., "rooftop"), which we determined via manual verification of a subset of the geocoding results. These results were retained if the geocoded coordinates were located within 1,200 meters of the well's initial provided coordinates in OSWCR. This distance was used because a well's provided coordinates should be no more than approximately 1,200 meters from its reported coordinates, if reported at the centroid of the PLSS section.
    iv. Matching OSWCR data to a statewide residential parcel dataset by county and assessor's parcel number (APN) with leading and trailing zeros removed
d. Used a tiered approach to select appropriate location estimate in cases where the above methods resulted in different location matches, which occurred due to the variety of methods used. First, we assigned wells to a parcel if they matched to that parcel via a greater number of methods than any other parcel. Next, we resolved wells with two different sources of matches based on the perceived accuracy of each matching method, in the following order: exact APN, exact address, string distance address, geocoding, and APN with leading and trailing zeros removed. We retained the original OSWCR coordinates for wells that we could not definitively match to a single parcel, including wells that did not match to any parcel using our analysis techniques.
e. This approach resulted in updated location estimates for 120,631 wells

    f. At this stage we removed 1,276 duplicate records, where the legacy log number or the permit number and the date, county, township, range, and section were the same (and were not listed as NA or Not Available)
2. Assigned domestic wells to domestic well areas
    a. Separated domestic well point location data into two datasets based on data accuracy
        i. Centroid resolution – 172,909 wells located at the centroid of a PLSS section
        ii. Parcel resolution – 120,631 wells with parcel resolution
    b. Centroid-resolution wells
        i. Selected PLSS sections with wells
        ii. Intersected PLSS section geography with census block geography
        iii. Excluded unpopulated census blocks
    c. Higher-accuracy wells (e.g. parcel resolution)
        i. Intersected well points with census blocks
        ii. Excluded blocks with no wells
        iii. If block contained <10 people, expanded the block area to include the PLSS section containing the well
        iv. If well was located in unpopulated block, included the PLSS section containing the well
    d. Combined areas from steps 2b and 2c
    e. Use identity function to assign PLSS section ID to domestic well areas (N=61,760)
3. Assigned water quality from the State Water Resources Control Board 2025 Aquifer Risk Map individual contaminant layers[3] to domestic well areas by PLSS section ID
    a. **Description:** These layers contain summarized water quality risk per square mile section. The water quality risk is based on water quality results from public and domestic supply wells. The methodology used to determine water quality risk is outlined here and data layers are available here.
    b. The 20-year averages for each of the four contaminants was calculated for each domestic well area by multiplying the "20-Year Average (Comparison Concentration Index)" by each contaminant's Maximum Contaminant Level (MCL).
4. Estimated the population served by domestic wells using a tiered approach
    a. Areas served exclusively by wells (i.e. areas that do not overlap with community water system service area boundaries) were assigned a population using gridded 2020 population estimates developed by Nick Depsky[7]. We intersected the Depsky (2022) population estimates with PLSS section geographies and summed the population in each PLSS section.
    b. Areas that contain one or more domestic well(s) and are within the service area of a community water system[8] were assigned population estimates by first summing the wells in each PLSS section and then multiplying the number of wells by the estimated population served by each well (i.e., 1 person per well).
        i. We considered a matrix of options for calculating the number of people served by each well across the state. Factors that we considered were the number of wells that are active/in use (i.e. activity weight), the number of wells that may be missing from the OSWCR dataset (i.e. completeness weight), and the number of people in each household (i.e. population weight). We tested a range of options for each variable and evaluated the plausibility of each option. Based on

our results and consultation with our technical advisory committee, we assigned a weight of 1 person to each domestic well.

Attribute Table

Field heading	**Field Description**
OBJECTID_12*	Field ID
Shape*	polygon
MTRS	PLSS section identifier; Meridian (M), Township (T), Range (R), Section (S)
Well_pop_1	Total population served by domestic wells in PLSS section Well_Pop_1 = 1,579,365
Total_well	Number of domestic wells in each PLSS section Total N=292,093 Note: original point parcel dataset contained 293,540 wells; 1,447 wells at the centroid resolution were in areas that are no longer populated according to census data and were dropped in step 2,b,iii.
As_ugL	Arsenic (As) water quality concentration in µg/L (MCL=10 µg/L) -999 = missing data
N_mgL	Nitrate as N water quality concentration in mg/L (MCL=10 mg/L) -999 = missing data
Cr6_ugL	Hexavalent Chromium (Cr[VI]) water quality concentration in µg/L (MCL=10 µg/L) -999 = missing data
123-TCP_ugL	1,2,3-Trichloropropane water quality concentration in ug/L (MCL=.005 µg/L) -999 = missing data
Av_depth	Average total completed depth of wells in section (4-3,150 ft.)
SD_depth	Standard deviation of total completed depth for wells in section (0-2,990.7 ft.)

Shape_Length	GIS generated – length in meters
Shape_area	GIS generated – area in square meters

References

1. Rempel, J., Belfer, E., Ray, I., Morello-Frosch, R. (2023).Access for Sale? Overlying Rights, Land Transactions, and Groundwater Access in California. (in review).
2. CADWR. (2021). Online System for Well Completion Reports. California Department of Water Resources. Available from https://data.ca.gov/dataset/well-completion-reports. Accessed July 2021.
3. CSWRCB. (2025). *2025 Aquifer Risk Map Methodology*. California State Water Resources Control Board - Division of Water Quality. https://www.waterboards.ca.gov/water_issues/programs/gama/docs/armmethods25.pdf
4. DMP Lightbox. (2018). *SmartParcels®*. DMP Lightbox. Available from https://www.digmap.com/platform/smartparcels/.
5. Winkler, W.E. (2014). Matching and record linkage. *WIREs Computational Statistics, 6(5),* 313-325. Available from https://doi.org/10.1002/wics.1317.
6. Frontiera, P. (2019). Usaddress_cleaning [Python]. dlab-geo. https://github.com/dlab-geo/usaddress_cleaning
7. Depsky, N., Cushing, L., Morello-Frosch, R. (2022). *High-resolution gridded estimates of population and sociodemographics from the 2020 census in California*. PLOS ONE. Available from https://doi.org/10.1371/journal.pone.0270746.
8. Pace, C., Bangia, K., Fisher, E., Subramanian, A., Cushing, L., Morello-Frosch, R. (2023). Water System Boundaries Version 2.0. Drinking Water Tool metadata, prepared by the Water Equity Science Shop, UC Berkeley.