

ATTACHMENT C1: GEOGRAPHIC INFORMATION SYSTEM AND DATABASE METHODOLOGIES

Attachment to the State Water Resources Control Board 2021 Drinking Water Needs Assessment Cost Assessment Methodology Appendix C

https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/2021_needs_assessment.pdf

GIS METHODOLOGY

DATA SOURCES

Table C1.1 provides a list of data sources for water system locations, boundaries, compliance status and economic status estimation utilized in the GIS components of the Cost Assessment.

Table C1.1: Data Sources for GIS Analysis

Dataset	Source Agency	Original Feature Count	Notes
Human Right to Water ¹	State Water Board	3,279	Compliance status, analyte data
Monterey County SWS Out-of-Compliance 2019-03	Environmental Justice Coalition for Water (EJCW)	233	Merged with Human-Right-to-Water compliance data
California Census Block Groups ²	U.S. Census Bureau Tiger/Line Shapefiles	23,212	GIS polygon data
Median Household Income 2013-2017	U.S. Census Bureau American Fact Finder	23,213	Joined to block groups to provide DAC statuses. Includes average MHI data for 2013-2017.

¹ [Human Right to Water](https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/needs/2021_needs_assessment.pdf)

<https://www.arcgis.com/apps/MapJournal/index.html?appid=143794cd74e344a29eb8b96190f4658b>

² [U.S. Census Bureau-Current Block Group](https://catalog.data.gov/dataset/tiger-line-shapefile-2016-state-california-current-block-group-state-based)

<https://catalog.data.gov/dataset/tiger-line-shapefile-2016-state-california-current-block-group-state-based>

Dataset	Source Agency	Original Feature Count	Notes
California Block Group ³			
California Water System Service Areas ⁴	Tracking California	4,696	Waters system boundaries
RCAC Small Water Systems ⁵	Rural Community Assistance Corporation (RCAC)	1,132	Merged with California Water System Service Areas
Monterey County Revised Water System Boundaries ⁶	State Water Board staff	6,676	Multiple parcel features per system. These corrected boundaries were used in the physical consolidation analysis.
Monterey County Small Water Systems ⁷	Environmental Justice Coalition for Water (EJCW)	2,935	Merged with California Water System Service Areas
Water System Well Locations ⁸	State Water Board's GAMA Program	22,672	Used to better locate Human Right to Water Systems without accurate boundaries
Domestic Well Locations and	State Water Board's GAMA Program	347,592	Domestic wells by square mile section and modeled water quality from the Needs Analysis GAMA Tool, from 9/21/2020

³ [U.S. Census Bureau-Median Household Income](https://data.census.gov/cedsci/table?q=B19013&tid=ACSDT1Y2018.B19013&hidePreview=false)

<https://data.census.gov/cedsci/table?q=B19013&tid=ACSDT1Y2018.B19013&hidePreview=false>

⁴ [Tracking California Water Boundary Tool](https://trackingcalifornia.org/water/map-viewer) used for Water System Service Areas was retried on July 1, 2020.
<https://trackingcalifornia.org/water/map-viewer>

⁵ RCAC Small Water System dataset contains information from the following counties; Colusa, Contra, El Dorado, Fresno, Glen, Humboldt, Kern, Kings, Lake, Mariposa, Merced, Mono, Nevada, Plumas, Riverside, San Bernardino, San Diego, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, Shasta, Solano, Sutter, Tulare, Ventura and Yolo counties. Unpublished.

⁶ Provided by the State Water Board. Unpublished.

⁷ A pdf version of the map can be viewed at [Monterey County Water System Quality](https://www.co.monterey.ca.us/home/showdocument?id=67378)
<https://www.co.monterey.ca.us/home/showdocument?id=67378>

The GIS data was provided by EJCW. Unpublished.

⁸ [GAMA Groundwater](https://gamagroundwater.waterboards.ca.gov/gama/gamamap/public/Default.asp)

<https://gamagroundwater.waterboards.ca.gov/gama/gamamap/public/Default.asp>

Dataset	Source Agency	Original Feature Count	Notes
Modeled Water Quality ⁹			
Building Footprint Method Water System Boundaries ¹⁰	Pacific Institute	56	Revised boundaries based on where buildings are within a system. Used in selected situations for the physical consolidation analysis.

WATER SYSTEM LOCATIONS AND BOUNDARIES

To support cost estimates based on potential pipeline lengths and other factors, the accuracy of water system locations and service area boundaries is important. Where available, more detailed estimates of water system locations, especially for small systems, and boundaries have been integrated into the water systems dataset.

Water system boundaries from the California Water System Service Areas serve as the starting point for this dataset. However, this dataset does not include locations or boundaries for most small systems. To incorporate small systems, multiple small system datasets have been mined, merged and joined with the California Water System Service Area dataset. As needed, the small systems have been located in GIS using the Groundwater Ambient Monitoring and Assessment (GAMA) Program Groundwater Information System's Groundwater Well Locations dataset based on water system identification, or reverse geocoded to addresses provided from the raw sources. State small water system locational data from a recent RCAC project was incorporated. Data was not available for all counties, and the data was provided in a variety of formats. Domestic well locational data is only available as a count per square mile. Each dataset has limitations and inaccuracies and pending improvements to the locations of water systems and boundaries will increase the accuracy of future analyses. These data, summarized in Table C1.2, have been integrated into the final water systems data layer.

⁹ [Needs Analysis GAMA Tool](https://gispublic.waterboards.ca.gov/portal/apps/webappviewer/index.html?id=292dd4434c9c4c1ab8291b94a91cee85) accessed 9/21/2020, <https://gispublic.waterboards.ca.gov/portal/apps/webappviewer/index.html?id=292dd4434c9c4c1ab8291b94a91cee85>

¹⁰ Shimabuku, Morgan, 2019. Pacific Institute. *Boundary Refinement Methods and Notes*. Unpublished.

WATER SYSTEM BOUNDARY REFINEMENT METHODOLOGY

BACKGROUND

This methodology was developed to update the boundaries of drinking water systems to facilitate a more accurate estimation of the potential for physical consolidation of Human Right to Water (HR2W) list systems, At-Risk public water systems, state small water systems (SSWSs), and domestic wells. This is a sub-step of the larger Drinking Water Needs Assessment Cost Assessment Model.

OBJECTIVE

The objective of the methodology presented below is to adjust the shapefiles of public water system boundaries so that the boundaries are more reflective of the extent of each system's existing infrastructure.

METHOD OVERVIEW

The Water Boundary Tool contains a database of customer service area boundaries for public water systems in California. This tool was first developed and hosted by TrackingCalifornia.org. In July 2020, the tool was transferred to the State Water Board's Division of Drinking Water.¹¹

This database contains the most comprehensive digital record of California's public water system customer service areas that is freely available. However, the shapefiles associated with customer service area boundaries do not always accurately reflect the extent of the physical infrastructure of the systems. Multiple types of errors are possible such as: shapefiles may reflect the planned boundaries of the systems that will be achieved at some future date, or in some cases, systems have no boundaries uploaded to the public database. To understand actual costs of potential consolidations, it is necessary to have more accurate measurements of the distances between the physical infrastructure of the potential joining and receiving systems' physical extent. The physical extent of a water system can be estimated from the extent of built residential and commercial buildings, which can be assumed to include water system connections. Building footprint data created by Microsoft and freely licensed through Github.com provides this kind of information and therefore can be used to refine water system boundaries.

This Attachment records a description of the data processing steps taken to refine the water system boundaries based on a building footprint data set. Note that this work is focused on modifying a subset of the potential consolidating systems (n=57), as determined by previous work by Corona Environmental on systems within 0-3 miles of a water system on the HR2W

¹¹ [California Drinking Water System Area Boundaries](https://gis.data.ca.gov/datasets/waterboards::california-drinking-water-system-area-boundaries)

<https://gis.data.ca.gov/datasets/waterboards::california-drinking-water-system-area-boundaries>

list.

Detailed Method

Table C1.2 presents the complete list and description of data and files that were used in this method development. Note that actual file names are noted in throughout the method and refer to a set of files delivered along with this methodology to Corona Environmental in Fall of 2019.

Table C1.2: Description (including file name, as provided to Corona Environmental), notes, and source of the data and files used in the method development

Description	Notes	Source
Shapefile of the potential receiving systems (R_Bounds.shp)	Includes the customer service area boundaries for 606 public water systems in California.	This file is a modified version of the Water Boundary Tool, provided by Corona Environmental to Pacific Institute, Fall 2019.
Shapefile of the potential joining systems (J_Bounds.shp)	Includes the customer service area boundaries for 3,800 public water systems in California.	This file is a modified version of the Water Boundary Tool, provided by Corona Environmental to Pacific Institute, Fall 2019.
Building footprint data for California (California.geojson)	10,988,525 individual buildings included. Data vintage: variable. No dates provided, but publication date was late 2018.	GitHub: U.S. Building Footprints ¹² Retrieved October 2019.

To process these data, the following steps were developed in ArcGIS Pro 2.4.2.

- 1. Import potential receiving system shapefile (R_Bounds.shp) and building footprint file (California.geojson)**
- 2. Convert building footprint file to a shapefile to allow for further processing.**

Tool: Json to Feature Class
Input: California.geojson
Output: CaliforniaBF.shp

- 3. Reduce the receiving system shapefile to the 57 systems of interest.**

Tool: Layer from Selection.

¹² [GitHub U.S. Building Footprints](https://github.com/microsoft/USBuildingFootprints)
<https://github.com/microsoft/USBuildingFootprints>

Input: R_Bounds.shp
Output: R_Bounds selection.shp

4. Export R_Bounds selection.shp to a new feature class file and save within the geodatabase.

Tool: Feature Class to Feature Class
Input: R_Bounds selection.shp
Output: R_Bounds_subset.shp

5. Clip the new building footprint shapefile to the areas just within the 57 systems.

Tool: Clip
Input: CaliforniaBF.shp & R_Bounds_subset.shp
Output: R_BoundsBF.shp

6. Create a random raster to cover extent of the area contained by the 57 systems.

Tool: Create random raster
Input: None
Output: raster2
Other settings: Distribution: Integer, 1-100; Output extent: R_Bounds_subset.shp; cell size: 0.0006

7. Clip the raster to only cover the building footprints.

Tool: Clip raster
Input: raster2
Output extent: R_Bounds_subset.shp
Output: raster2_Clip2
Other settings: checked "Use input features for clipping geometry;" checked "maintain clipping extent"

8. Convert the new raster to a polygon shapefile for further processing.

Tool: Raster to Polygon
Input: raster2_Clip2
Output: R_Bound_polyraster.shp
Other settings: Field: Value; Unchecked "simplify polygons;" unchecked "create multipart features."

9. Join the new polygon raster shapefile with the building footprint clip to create a shapefile with coverage only of cells that contain building footprints.

Tool: Spatial Join
Target features: R_Bound_polyraster.shp

Join Features: R_BoundsBF.shp
Output: R_Bound_polyrasterjoin.shp
Other settings: Operation = one to many, unchecked “keep all target features,”
match option = intersect, search radius = 100 ft.
Note: The search radius is used to simply reduce the number of “islands” or parts
in the multi-part polygon that is created.

10. Remove raster cells from the polygon by dissolving all boundaries.

Tool: Dissolve
Input: R_Bound_polyrasterjoin.shp
Output: R_Bound_Dissolve.shp
Other settings: Dissolve field = Join_Count (all values were 1), checked “create
multipart feature,” Unchecked “unsplit lines”

11. To re-connect the new polygon with 57 unique water systems so that the modified system boundary shapefile would contain all of the original attributes of the original system boundary (R_Bounds_subset.shp).

Tool: Clip
Input: R_Bounds_subset.shp
Clip Features: R_Bound_Dissolve.shp
Output: R_Bounds_modified.shp

The original boundary of a water system is shown in Figure C1.1. An example of the results of performing steps 1-11 to update the boundaries is shown in Figure C1.2.

Figure C1.1: Unmodified boundaries of a potential receiving systems, in green

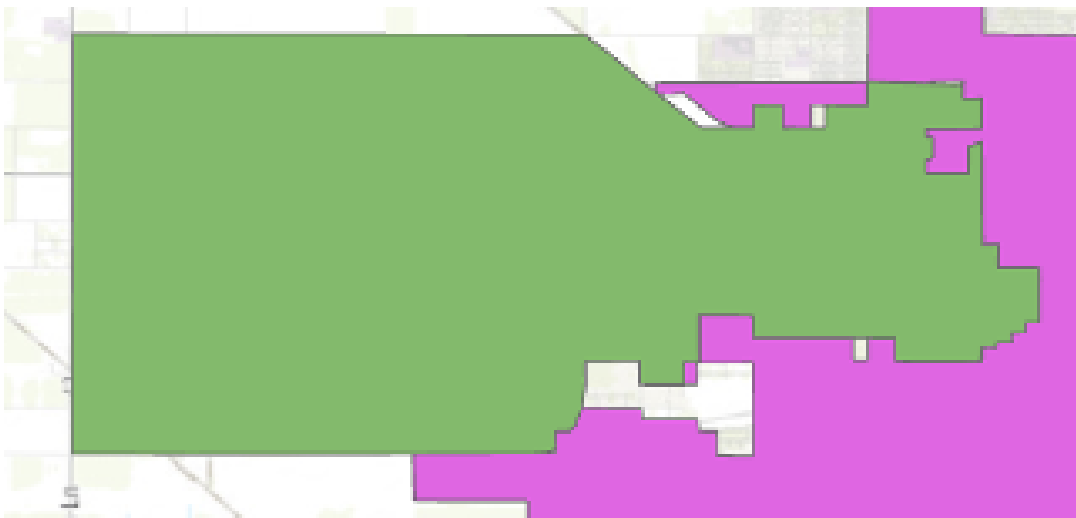
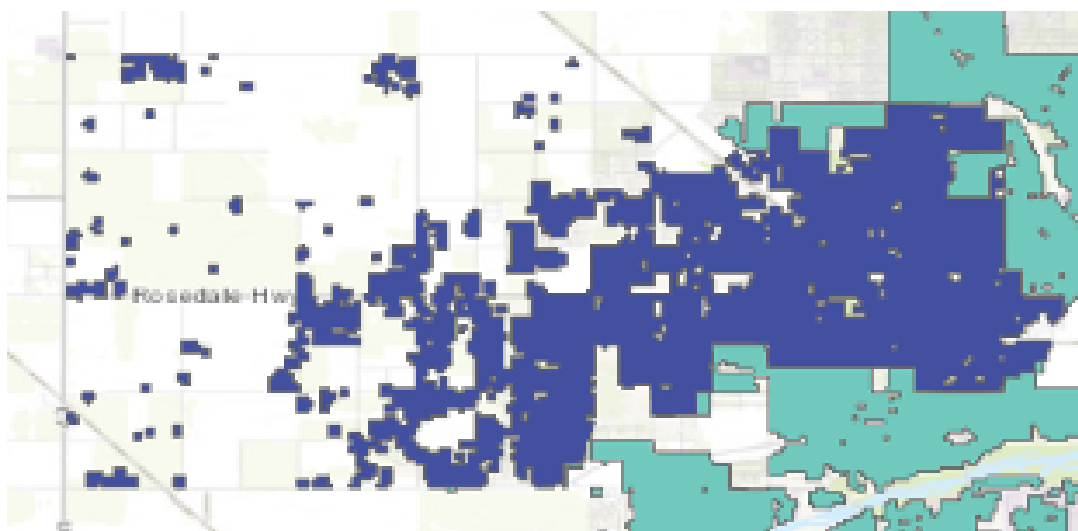


Figure C1.2: Updated boundaries of the same potential receiving system, in dark blue



Additional steps were developed to further modify system boundaries where potential joining and potential receiving system boundary shapefiles overlapped. This was done because it is possible that the water system infrastructure from the potential consolidating systems do not extend into areas where there is overlap with a potential joining system.

The steps below describe the process taken to clip the newly modified water system boundaries file (R_Bounds_modified.shp) with the potential joining systems file (J_Bounds.shp).

12. Import the potential joining water systems boundaries shapefile (J_Bounds.shp) and then save it within the geodatabase.

Tool: Feature Class to Feature Class
Input: J_Bounds.shp
Output: J_BoundsFull.shp

13. Erase all portions of the modified receiving system boundary shapefile that directly overlap with potential joining systems.

Tool: Erase
Input: R_Bounds_modified.shp
Erase features: J_BoundsFull.shp
Output: R_Bounds_modified_2.shp

14. Delete three water system boundaries from R_Bounds_modified_2.shp that are significantly reduced from the Erase tool.

No tool required – deleted individual systems in the attribute table.
Systems deleted: San Bernadino City, Yucaipa Valley WD, and City of Modesto.

The results from steps 12-14 offer a more conservative estimate of the extent of water system infrastructure; however, in some cases this process may under-estimate (i.e., inaccurately reduce) the area of a potential receiving system's water infrastructure. Therefore, the final output from these steps (R_Bounds_modified_2.shp) should be used in comparison to the output from the steps above (R_Bounds_modified.shp). Also, in cases where a modified boundary of a potential receiving system is reduced to such a small area as to seem implausible, the implausible version can be removed. In this data set, there were three system boundaries removed because of this: San Bernadino City, Yucaipa Valley WD, and City of Modesto.

An example of the results of performing steps 12-14 is shown in Figure C1.4.

Figure C1.3: Example before the application of steps 12-14

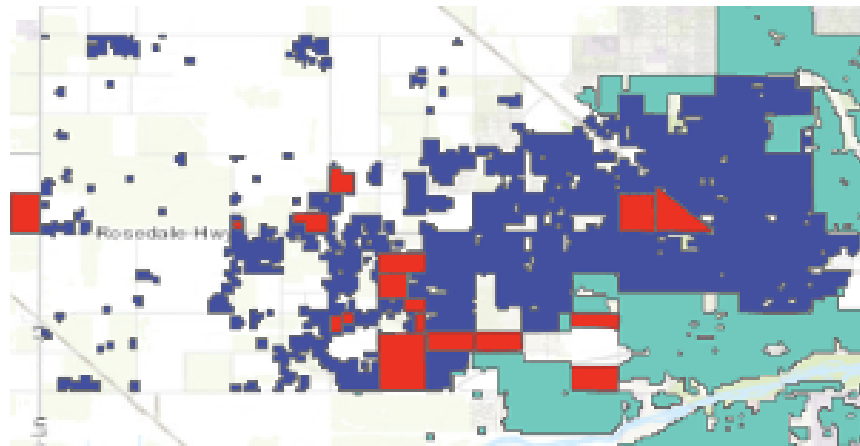
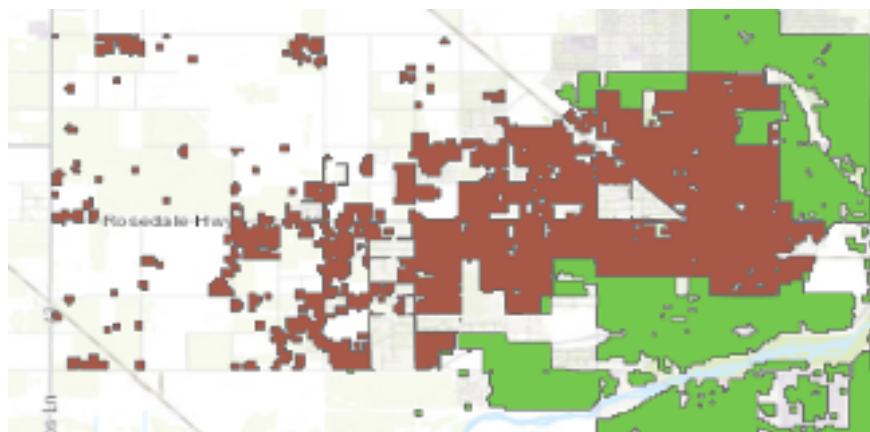


Figure C1.4: Example after the application of steps 12-14



PHYSICAL CONSOLIDATION GIS METHODOLOGY

The methodology for the potential physical consolidation analysis of the Cost Assessment Model was implemented in previous work performed by Corona for the Water Foundation¹³ and was initially based on the method developed through a project at UC Davis Center for Regional Change, London et al.¹⁴

Using “Network Analysis” in GIS, the shortest path along roadways from a Joining system to the nearest Receiving system was determined.

General assumptions for consolidation include:

- < 1 mile is favorable, 1 – 3 miles may be possible, and > 3 miles is not feasible for consolidation
- 1000-foot buffer was added for each path
- 1000-foot buffer was added for systems that intersect with an existing receiving system boundary
- For SSWS pickups (routes and intersects)
 - Maximum distance to merger route = 0.38 mi (~2000 ft)
 - No addition of 1000 ft for route
- Regional solutions consider up to 3 miles for both SWS and SSWS

Steps to perform the analysis are included in detail below.

- Software requirements
 - ArcGIS Pro
 - Network Analysis Extension
 - Street Map Premium for California
- Update location information, boundaries, and system attributes (e.g., population, connections, compliance status, system type)
 - The boundary file and associated attributes used in this analysis were compiled from multiple sources
 - Tracking CA boundaries
 - GAMA DDW PWS Well locations plus DDW Siteloc
 - EPA SDWIS Active CA systems
 - HR2W site OOC, RTC, IC
 - RCAC and County datasets for SSWS
 - Any improvement to water system boundaries should be included in the shapefile of California water systems (e.g., ca_system_layer)

¹³ Henrie, Tarrah and Chad Seidel, 2019. [Cost Analysis of California Drinking Water System Mergers](https://waterfdn.org/wp-content/uploads/2019/08/COSTAN1.pdf). Water Foundation. <https://waterfdn.org/wp-content/uploads/2019/08/COSTAN1.pdf>

¹⁴ London, J.; Fencil, A.; Watterson, S.; Jarin, J.; Aranda, A.; King, A.; Pannu, C.; Seaton, P.; Firestone, L.; Dawson, M.; & Nguyen, P., 2018. The Struggle for Water Justice in California's San Joaquin Valley: A Focus on Disadvantaged Unincorporated Communities. UC Davis Center for Regional Change.

- Any new systems should be added including all relevant data fields (e.g., name, pwsid, population, connections, system type...etc.)
 - Compliance information should be updated from the HR2W dataset
 - System type should be included from EPA SDWIS, HR2W, DDW data sources, or any other available source
 - Population and connections should be updated from the DDW database, if available, otherwise from EPA SDWIS, or any updated files from other sources (e.g., population and connection information is included in RCAC data for most counties). If either population or connections is blank/null, set to 0.
- For any system with a PWSID, County is populated based on the County code in the PWSID number. For any system without a PWSID for which location information is available, County is populated based on a spatial join with California County boundaries.
- ClassID is populated as follows:
 - Population > 3300 = NON-SWS
 - Population < 3301 = SWS
 - Population < 26 and Connections < 15 and no PWSID = SSWS*
- PWSID_Name field is filled in concatenating PWSID + " " + the system name (this is because some systems (e.g., SSWS) do not have PWSIDs and must be identified by system name).
- Based on the updated population, connection, and compliance information, potential receiving and potential joining systems are selected from the fully updated ca_system_layer
 - Receiving system criteria: CWS, population > 3300, HR2W IC/RTC
 - Joining system criteria: All system types, population <= 3300, includes SSWS
 - Repair polygons: [Toolbox: Repair geometry]
- Develop Inputs for the ArcGIS Pro Network Analysis Closest Facility
 - Receiving systems:
 - Boundary layer from ca_system_layer based on above criteria
 - Facilities layer from intersection of Receiving Boundary layer with roads layer
 - Intersect the boundaries with the road 1k layer, then explode multipoint to get all the points (edit, features, modify, modify features, divide, explode)
 - Joining systems:
 - Boundary layer from ca_system_layer based on above criteria
 - Incidents layer from centroids/point locations of Joining Boundary layer. Add CentroidX and CentroidY fields (double) to attribute table, then [Toolbox: Calculate geometry attributes], then export as table and display XY data

- Prior to performing the network analysis to find the shortest path between receiving and joining systems, INTERSECTING systems need to be identified.
 - Check for any Joining systems that intersect with Receiving system boundaries.
 - Spatial join was used to find any points within boundaries, pulling the PWSID_Name of each system, then, in the intersection layer, delete any null matches.
 - Move these to the potential consolidation list and remove them from the closest facility analysis incidents layer (since these joining systems are within a receiving system boundary, we do not want to include them in the path analysis).
 - To delete the intersecting small systems in the incidents layer (to exclude from the closest facility analysis), select by location with boundaries and delete from the incidents table.
- Network Analysis
 - Using Facilities/Receiving systems and Incidents/Joining systems layers from above (with intersecting Joining systems removed).
 - Make sure Analysis tab - Network Analysis - Drop-down Network Data Source: Should be Routing_ND (North America geodatabase)
 - Analysis: Network Analysis: Closest Facility
 - Mode: Change from Driving time to Driving distance
 - Change Travel Settings properties (the small arrow lower right of the Travel Settings section of the menu bar)
 - Costs: change km to mi
 - Restrictions: Adjust as needed. Unchecked avoid private roads, avoid unpaved roads, under construction prohibited, and through traffic prohibited.
 - Cutoff: 3 (miles)
 - Import Facilities – Receiving System points that intersected with 1k Roads: selected 500 ft search tolerance, select PWSID_Name for Name (this is how the system will be identified once the analysis is complete)
 - Import Incidents – Joining Systems: used 2000 ft search tolerance with 7550 located
 - FOR ALL IMPORTS MAKE SURE Name is set to the water system id or PWSID_Name as specified above. This is needed to identify which systems are associated with each route.
 - For first file import of facilities (and again for incidents), uncheck append (then make sure append is checked for the second layer if using more than one layer for facilities or incidents).
 - Then hit “Run” and once complete, the routes layer can be exported and results can be analyzed in Excel, for cost estimations, etc.
- Route Elevations
 - The elevation changes along pipeline routes were determined to assess pumping costs. The elevation statistics for the pipeline routes requires a (1) pipeline route layer, a (2) geographic information system (GIS) application for editing, and (3) USGS 3D Elevation Program (3DEP) raster data. For this analysis, QGIS 3.16.2

was the preferred GIS application used. The first step in the process included transforming the pipeline route from a line-type feature to points. For QGIS 3.16.2, the process tool *Points Along Geometry* was used for this, which creates points at set intervals along a line. The pipeline layer formerly as a line-type, but now as points can be spatially joined to the USGS 3DEP elevation raster data. For QGIS 3.16.2, the *Point Sampling Tool* extracts the elevation data from the USGS 3DEP layer for each specific point and joins the elevation to each point. The final step requires joining the elevation data to the original pipeline line-type layer. For QGIS 3.16.2, the *Join Attributes by Location (Summary)* process tool was used. This tool joins the pipeline line-type layer to the pipeline point layer with the elevation data. The elevation field from the points layer is summarized; the maximum, minimum and mean for each route is calculated. The *Join Attributes by Location (Summary)* tool reads each point that intersects a pipeline route, extracts the elevation data for each point, and creates a new layer that includes the pipeline routes with the maximum, minimum and mean elevation for each route as attribute data.

- Additional notes:
 - Note on wholesalers: Inclusion of wholesalers with a population of at least 3,300. We had discussed removal of wholesalers; however, systems marked as wholesalers in the EPA SDWIS data were not specifically removed. This was because some regular water systems were also marked as wholesalers, so removal of wholesalers would exclude systems that should be included. Many wholesalers are listed with a population of 0 or 1, so they were ultimately removed by the population screening for receiving systems (minimum population of 3,300 people). Systems identified as wholesalers in SDWIS data WITH population < 25 were removed--> removed 24 systems. AVEK was separately removed as a potential receiving system in Kern County; this is a wholesaler that was not screened out by population.
- SSWS Pickups
 - Remove SSWS direct routes from the All SWS scenario
 - Generate points along routes
 - Perform network analysis as above
 - Facilities: Points along "Receiving path" – either AllSWS (excluding SSWS) or HR2W
 - Incidents: SSWS (non-intersects)
 - Maximum distance: 0.38 mi (~2000 ft)
- Domestic Well Pickups
 - Intersect layer with domestic wells by PLSS with All physical consolidation routes to identify which sections have a route passing through. If any route passes through the PLSS, the domestic wells in that PLSS are considered as potential pickups. The water quality of the PLSS is used to determine classification as At-Risk; grades 4, 5, and 6 are considered At-Risk.

- Intersect PLSS layer with Receiving system boundaries to determine the potential for domestic well pickups as intersects.
- Regional Solutions
 - To convert multiple branching routes to one continuous route to a Receiving system
 - Copy the Routes layer, include Receiving system ID and any other details about the merging systems
 - Use the Integrate tool to merge any overlapping routes that are not perfectly overlapping, XY tolerance of 10 feet
 - Use the Dissolve tool to merge the single lines into one line
 - Dissolve Field (to aggregate) = R_PWSID
 - Statistics Fields
 - J_ID Count, Ttl Mile Sum, J_conn Sum, J_pop Sum
 - CHECK “Create multipart features”
 - Uncheck “Unsplit lines”
 - Provides distance per receiving system for regional solutions.
 - This results in routes merged by Rpwsid, but each merged route (of overlapping route) has its own distance. Add new field CalcNewLen and Calculate Geometry to get updated merged path distance.
 - THEN do a spatial join of J systems from CA_Routes with new merged routes and create a field in CA_Routes that has the object id of the new merged route. This way we can link all the J systems to the correct merged cluster route.
 - A regional flag was also set manually for each HR2W system. “Regional” was set to Y for any HR2W route that overlapped all or part of a potential consolidation route for another system, regardless of compliance status. If a SSWS could be picked up by a HR2W route, the HR2W route was also flagged as a regional route.

IDENTIFICATION OF AT-RISK SSWSS & DOMESTIC WELLS

The GAMA Needs Analysis Tool¹⁵ was developed by the Division of Water Quality Groundwater Ambient Monitoring and Assessment (GAMA) Unit of the State Water Resources Control Board to identify At-Risk SSWSSs and domestic wells.¹⁶ The dataset includes the domestic well count in one square mile sections by Public Land Survey System (PLSS) sections from Department of Water Resources Online System of Well Completion Reports. Water quality information for nitrate, arsenic, hexavalent chromium, uranium, 1,2,3 trichloropropane (123-TCP), and perchlorate was downloaded from the GAMA tool (September 2020) to assess the incidence of these contaminants individually and as co-contaminants. This water quality information informs the assessment of costs for POU and POE treatment systems needed for impacted domestic wells and SSWSSs. For the purpose of the Cost Assessment, hexavalent chromium was not included, since there is no current regulation. For SSWS and local small water systems in Monterey County, actual water quality results were used for the risk analysis.

The GAMA tool provides water quality data by grade based on the ratio of the average section detection to the MCL for a given constituent; the water quality grades also factor in MCL exceedances. SSWS and domestic wells were classified At-Risk if they mapped into a grade 4, 5, or 6 area. Grades are defined as follows:

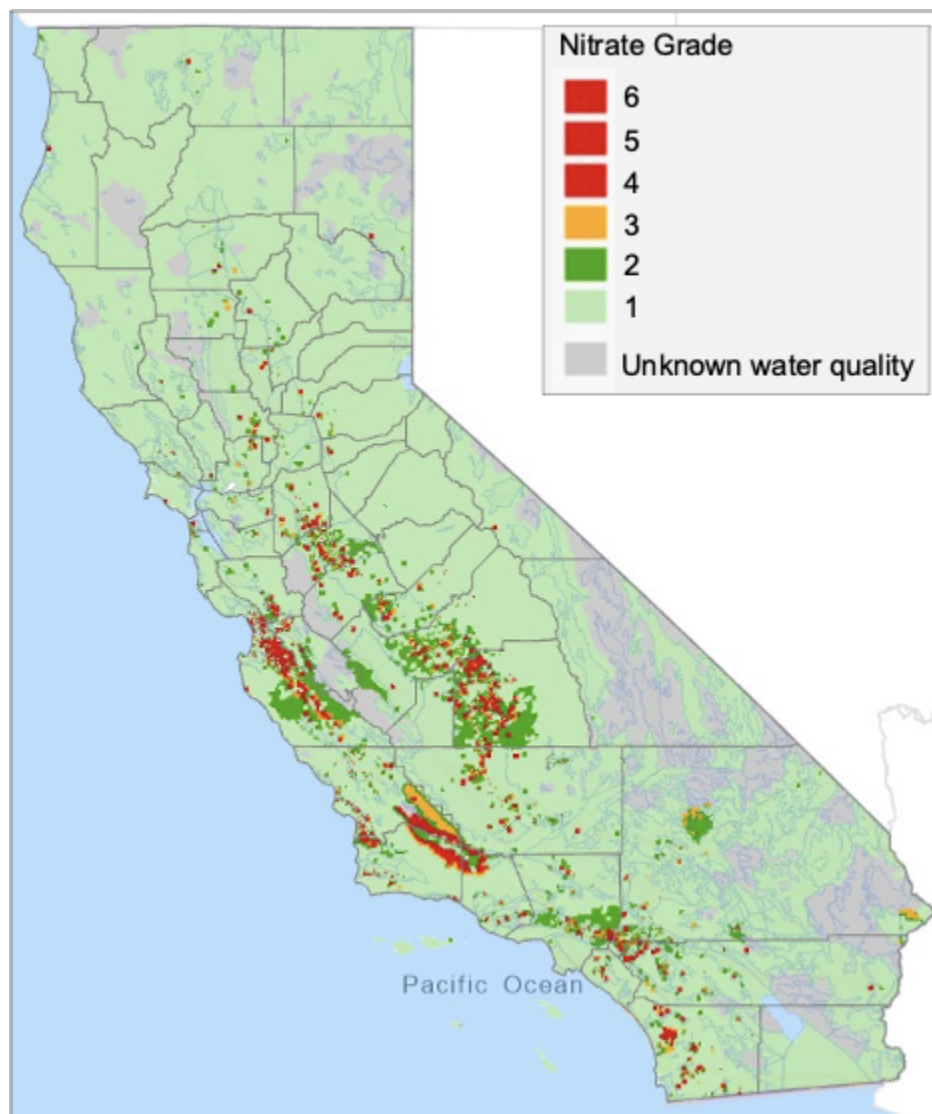
- 6: Recent MCL exceedances > 0, average section detection > MCL
- 5: Recent MCL exceedances = 0, average section detection > MCL
- 4: Recent MCL exceedances > 0, average section detection < MCL
- 3: Recent MCL exceedances = 0, average section detection 80 – 100% of MCL
- 2: Recent MCL exceedances = 0, average section detection between 50 – 80% of MCL
- 1: Recent MCL exceedances = 0, average section detection < 50% of MCL

For example, nitrate grades by PLSS section are mapped in Figure C1.5.

¹⁵State Water Resources Control Board. (2020). [Needs Analysis GAMA Tool](https://gispublic.waterboards.ca.gov/portal/apps/webappviewer/index.html?id=292dd4434c9c4c1ab8291b94a91cee85). GAMA Groundwater Ambient Monitoring and Assessment Program.
<https://gispublic.waterboards.ca.gov/portal/apps/webappviewer/index.html?id=292dd4434c9c4c1ab8291b94a91cee85>

¹⁶State Water Resources Control Board. (2020). Methodology to Estimate Groundwater Quality Accessed by Domestic Wells in California, Draft 2/14/2020. Division of Water Quality, Groundwater Ambient Monitoring and Assessment Unit.

Figure C1.5: Map of GAMA Needs Analysis Tool Nitrate Grade



DATABASE METHODOLOGY

The database houses all relevant data for the project, including information required for and generated by the GIS and cost evaluation efforts. The database is a PostgreSQL (Postgres) database managed using pgAdmin, an open source administration and development platform for Postgres. The open source software R for statistical computing is used as needed for data analysis and formatting data tables ahead of uploading to the PostgreSQL database. The following sources have been incorporated into the database:

- Safe Drinking Water Information System (SDWIS) federal reports data¹⁷
- State Water Board's Division of Drinking Water (DDW) water quality data¹⁸
- Water system economic status from the GIS analysis
- Human Right to Water (HR2W) list data¹⁹
- Water system demand calculations data

The SDWIS federal reports for Water System Summary, Water System Detail, Facilities, and Violations were downloaded as csv files from the SDWIS Federal Reports Advanced Search online portal.²⁰ The csv files were then uploaded to the database as individual data tables.

The DDW water quality data tables were downloaded from the California Water Boards Electronic Data Transfer (EDT) Library and Water Quality Analyses Data and Download Page.²¹ The data tables that were downloaded include Chemical.dbf, Chemhist.dbf, Chemarch.dbf, Chemxarc.dbf, Siteloc.dbf, Storet.dbf, and Watsys.dbf. The tables were read into R and a quality assurance and quality check (QA/QC) process was performed to ensure that all sample results had a valid collection date and sample location indicated by the primary station code, all sample locations had a valid primary station code associated with a unique and valid source name and a unique and valid system number, and all system names and source names used valid encoding. A data field for the public water system identification number (PWSID) was added for each system number. The PWSIDs were formed by adding a "0" at the beginning of any water system number with six characters to ensure all system numbers were seven characters in length and then adding "CA" at the beginning of all water system numbers such that each PWSID was nine characters in length and consistent with the U.S. EPA PWSIDs. After the QA/QC process was performed, the R script exported the data tables as csv files that were then uploaded to the Postgres database.

A SQL query was developed and run using pgAdmin to create a data table of DDW water quality data for samples collected between January 1, 2009 and the present (currently up to April 29, 2019), including:

- Water system ID (system_no)
- Water system name (system_nam)
- Sample collection date (samp_date)
- Sample collection time (samp_time)
- Primary station code or state source number/sample location ID (prim_sta_c)
- Source/sample location name (source_nam)
- Status of the source/sample location

¹⁷ [USEPA. SDWIS Federal Reporting Services System. https://ofmpub.epa.gov/apex/sfdw/f?p=108:200::NO::](https://ofmpub.epa.gov/apex/sfdw/f?p=108:200::NO::) Accessed December 5, 2019.

¹⁸ [California SWRCB. EDT Library and Water Quality Analyses Data. https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html](https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html) Accessed March 17, 2020.

¹⁹ California SWRCB. [Human Right to Water Portal: Water System Drinking Water Data](https://www.waterboards.ca.gov/water_issues/programs/hr2w/). Accessed October 28, 2019. https://www.waterboards.ca.gov/water_issues/programs/hr2w/

²⁰ [SDWIS Federal Reports Advanced Search online portal https://ofmpub.epa.gov/apex/sfdw/f?p=108:1::NO:1](https://ofmpub.epa.gov/apex/sfdw/f?p=108:1::NO:1)

²¹ [Electronic Data Transfer \(EDT\) Library and Water Quality Analyses Data https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html](https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html)

- AB = abandoned
- DS = destroyed
- IR = inactive raw
- IT = inactive treated
- IU = inactive untreated
- SR = standby raw
- ST = standby treated
- SU = standby untreated
- AR = active raw
- AT = active treated
- AU = active untreated
- MW = monitoring (not a drinking water source)
- AG = agricultural/irrigation well (not a drinking water well)
- DT = distribution system sample point, treated
- DR = distribution system sample point, raw
- CT = combined treated
- CU = combined sources which are not treated
- CR = combined raw
- CM = combined mixed (combined sources)
- PN = pending (source not yet established)
- PR = purchased raw
- PT = purchased treated
- PU = purchased untreated
- WW = wastewater (not for drinking)
- Water type/source of water:
 - G = well/groundwater
 - M = mixed (mixture of surface and ground water, i.e. well/river)
 - S = surface water
 - W = waste (wastewater generator)
- Chemical/analyte
- Detection level for purposes of reporting (DLR)
- US EPA STORET number for chemical/parameter (store_num)
- Modifier for chemical finding (xmod)
 - "<" = Not Detected
 - "F" = False Positive confirmed with two or more follow-up samples
 - "I" = Invalid
 - "Q" = Questionable
 - "-" = for Langelier Index findings
- Numerical finding/result of analysis
- Reporting unit for chemical/analyte

An R script was developed to format the resulting DDW data table prior to uploading to the Postgres database as follows:

1. Remove invalid, questionable, and false positive data as indicated by the xmod data field (xmod = "F", "I", or "Q")

2. Create data fields for the 'method detection limit' and 'below detection' indication. For non-detect data records as indicated by the xmod data field (xmod = "<"), set the 'method detection limit' field equal to the value in the 'finding' field, set the 'below detection' field equal to "Y" for yes, and replace the value in the 'finding' field with "0". By doing so, all non-detect data are set equal to zero for data analysis purposes. For data records with detected results, the 'method detection limit' field is set to "NA" and the 'below detection' field is set to "N" for no.
3. Nitrate data are currently reported in mg/L as N, but previously in California, nitrate data were reported in mg/L as NO₃. As a result, the DDW data contains two different chemicals: "NITRATE (AS N)" and "NITRATE (AS NO₃)". Data for "NITRATE (AS NO₃)" are converted from mg/L as NO₃ to mg/L as N by multiplying by the molecular weight of nitrogen divided by the molecular weight of nitrate, 14.0067/62.0037. The chemical name for "NITRATE (AS NO₃)" data is then changed to "NITRATE (AS N)".
4. Create a data field for the maximum contaminant level (MCL). First create a MCL data table including data fields: chemicals, reporting units, and detection limits for reporting purposes from the DDW data, and MCLs from the California Water Boards²² last updated March 13, 2019, in R and upload to the Postgres database. The MCL data table is merged with the DDW data to include the MCL data field. For unregulated chemicals, the MCL data field is set to "NA".

The formatted DDW data table was then written to a comma-separated values (csv) file and uploaded to the Postgres database, using the table name: "ddw_wqdata". This data table can be used to find detailed sample results data for a given water system, a given source/sample location, or a given analyte as needed.

For the Cost Assessment Model, it is important to use the DDW data to identify water quality concerns that could impact treatment processes. An R script was developed to create a table summarizing data for analytes that could be a concern for some treatment technologies. For each analyte and each source/sample location with data for the given analyte, the sample results for that analyte are summarized by the minimum result, median result, mean result, 95th percentile result, and the maximum result. Note that non-detect data are treated as zero for the purpose of averaging. The table was then uploaded to the Postgres database using the table name: 'ddw_wq_treatment'. The analytes that are included in this table are arsenic, chloride, iron, manganese, nitrate, pH, sulfate, total alkalinity, total dissolved solids (TDS), total hardness, and total organic carbon (TOC).

The DDW water quality data can also be used to prioritize systems by identifying systems that may not yet have a violation but have water quality data close to and/or approaching an MCL exceedance. In order to identify systems meeting this criterion, an R script was developed to create a table summarizing data for chemicals that have been detected at 80% of the MCL or greater. For each chemical and each source/sample location with a detected result at 80% of the MCL or greater for the given chemical, the chemical results are summarized by the minimum result, median result, mean result, 95th percentile result, and the maximum result. Note that non-detect data are treated as zero for the purpose of averaging. Additionally, a

²² [MCLs from the California Water Boards](https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/mclreview/mcls_dlrs_phgs.pdf)

https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/documents/mclreview/mcls_dlrs_phgs.pdf

'trend' data field is created. For each chemical and source/sample location included in the data table where there are available data results for the given chemical at the given source/sample location for each year from 2009 through 2018, the Mann Kendall statistical test was applied to test for a monotonic increasing or decreasing trend over time. If the resulting p-value is less than 0.05 and the test statistic is positive, the given chemical at the given source/sample location is identified as increasing. If the p-value is less than 0.05 and the test statistic is negative, the given chemical at the given source/sample location is identified as decreasing. If the p-value is greater than 0.05, there is not sufficient evidence to identify a trend over time. If there was not sufficient data to apply the Mann Kendall test, the 'trend' data field is set to "NA". The resulting table is uploaded to the Postgres database using the table name: 'ddw_wq_80permcl'.

The DAC/SDAC status for each water system will be identified in the GIS effort. The output of this effort will include a table identifying the DAC/SDAC status for each water system. The table will be uploaded to the database.

The HR2W data was used to identify systems with health-based violations. The data includes information regarding the contaminant resulting in a violation for each out of compliance system. The excel spreadsheet of Exceedance/Compliance Status of PWSs Data available on the California Water Boards website²³ was downloaded, saved as a csv file, and then uploaded to the database.

The water system demand calculations were developed for the cost assessment process. The calculations, which include ADD, maximum daily demand under different operational scenarios, and peak hourly demand, were developed in an excel spreadsheet. This spreadsheet was then saved as a csv file and uploaded to the Postgres database as a data table.

A data table containing all relevant data for the cost assessment was then created by running a SQL query which pulled desired information from the various data tables described above. For each water system, the table includes system information (PWSID, system name, county, population served, and number of connections), the analyte for which the system is in violation, the percent of the system that is a severely disadvantaged community (SDAC), disadvantaged community (DAC), not a disadvantaged community, and unknown DAC status, the ADD estimate, and the maximum daily demand estimate. For each system and violation analyte, there are then rows of data for each primary station code where the violation analyte has been detected at 80% of the MCL or greater. For each primary station code, the table contains data for the source name, status, water type, and the number of data records, mean result and max result for the violation analyte and other analytes that may impact treatment, including arsenic, chloride, TOC, alkalinity, pH, iron, manganese, sulfate, TDS, nitrate, and hardness.

²³ [California Water Boards website](https://www.waterboards.ca.gov/water_issues/programs/hr2w/)

https://www.waterboards.ca.gov/water_issues/programs/hr2w/