

Water Rights Demand Data Analysis Methodology Overview

STATE WATER RESOURCES CONTROL BOARD
DIVISION OF WATER RIGHTS

Version 1.0
Last Updated February 28, 2022

EXECUTIVE SUMMARY	2
BACKGROUND	2
COMMON TYPES OF DATA ERRORS.....	3
Excessive Reported Diversions	3
Duplicate Reporting	3
Missing Data	4
Water Availability Data.....	5
DEMAND DATA ANALYSIS METHODOLOGY OVERVIEW.....	5
Watershed/Area of Interest.....	6
Demand Data - Pre-Processing.....	6
Demand Data Analysis	8
Demand Data - Post-Processing	9
Results.....	9
NEXT STEPS.....	9
Additional Data Errors and Data Processing Modules	9
eWRIMS Data Deficiencies	10
Data collected from initial water rights applications.....	10
Data collected from use reports	11
APPENDIX I: DETAILED DIAGRAM.....	12

Executive Summary

The State Water Resources Control Board, Division of Water Rights (Division) has developed a methodology to standardize and improve the accuracy of water diversion and use data that is used to determine water availability and inform water management and regulatory decisions. The Water Rights Demand Data Analysis Methodology (Methodology) is a series of data pre-processing steps, R Scripts, and Data Processing Modules that identify and help address data quality issues related to both the self-reported water diversion and use data from water right holders or their agents and the Division of Water Rights electronic water rights data. The water diversion and use data submitted annually by water right holders to the Report Management System (RMS) is relied upon by Division staff and other interested parties to derive water demand estimates, along with the data in the electronic Water Rights Information Management System (eWRIMS). However, the data contains significant inaccuracies and inconsistencies related, but not limited to duplicative or missing water diversion and use reports, unit conversion errors, and other mis-reporting discrepancies. This Methodology considers and incorporates previous demand data quality efforts completed by Division staff into a universal approach to comprehensively identify incorrect water right description and diversion data. The Methodology includes automated solutions when possible and provides guidance to help standardize and document manual corrections to address data errors. The Methodology process results in a “cleaned” water demand dataset that can be used by Division staff and interested parties to inform water availability analyses and water management decisions related to drought, permitting, and public trust resources.

Background

Division staff performed an inventory and assessment of previous demand data efforts and compiled a report summarizing key desired elements to be incorporated into the standardized Water Right Demand Data Analysis Methodology. The first draft version of the Methodology was presented at a Water Boards public meeting on April 16, 2021. The draft Methodology was then applied in the Russian River watershed during the 2021 emergency regulation drought response effort to develop monthly surface water demand estimates that were used in the calculation of the water availability analysis to inform the issuance of initial water unavailability notices and subsequent curtailments. The lessons learned from implementation in the Russian River watershed have been incorporated into the Methodology. Primary changes to the April 16, 2021 draft Methodology include a more comprehensive set of guidance documents, standard operating procedures, geospatial pre-processing instructions, revised data flags, pre-calculated reference metrics, and improved visual and table-based results of data.

The Methodology will be updated regularly, as needed, based on feedback from interested parties and lessons learned through implementation.

Common Types of Data Errors

A summary of the types of errors and data statistics identified using the Data Processing Modules are categorized and described in the section below.

Excessive Reported Diversions

The excessive reported diversions category includes water rights with reported diversion and use values that exceed the water right's face value amount or are otherwise not representative of reasonable diversion and use amounts. The reporting errors flagged under this category tend to account for the most significant data errors, from a volume-based water availability analysis standpoint. For example, a simple oversight on volumetric units can result in reported volumes of water greater than 325,000 times the actual quantity of water diverted, which can significantly distort the mass balance of the supply and demand in a given watershed. The excessive reported diversions module calculates diversion and use statistics and identifies the following potential errors:

- a. Unit Conversion Errors: Provides pre-calculated metrics that assist staff in identifying reports in which incorrect units were used by the water right holder submitting their annual diversion and use values (e.g., reporting gallons of water used in units of acre feet).
- b. Diversion Exceeds Face Value or Initial Reported Diversion: Identifies water rights with reported annual diversion values that exceed the permitted or licensed face value amount or the initial reported diversion for statements of water diversion and use.
- c. Diversion Reported Outside Season of Diversion: Identifies water rights with reported diversion amounts for months outside the permitted or licensed season of diversion.
- d. Diversion and Use Data Statistics: Identifies potential data errors by calculating a suite of statistics for watershed assessment and analysis, including the average and standard deviation of diversion and use values to identify reported diversion and use value outliers.

Duplicate Reporting

The duplicate reporting category includes water rights with submitted annual reports that contain duplicate diversion and use values under several different types of misreporting scenarios. The reporting errors flagged under this category can artificially inflate the estimated demand as a result of inadvertent double reporting diversion and use of the same volume of water. To help assess potential duplicate reporting errors this module identifies the following:

- a. Duplicate Points of Diversion: Identifies when there are multiple water rights with the same point of diversion, indicating that a duplicate active water right application or statement may exist.
- b. Duplicate Diversion and Use Values Reported for Multiple Water Rights: Identifies when a diverter has more than one water right and reports the same diversion and use amounts for more than one water right. This may indicate inaccurate reporting due to water not being split properly amongst multiple rights that are associated with a single diversion operation.
- c. Duplicate Diversion and Use Values Reported for Multiple Months or Years: Identifies water rights with the same diversion and use values reported for multiple months and/or years which may require verification for accuracy. Typically, actual diversion patterns vary from month to month and year to year based on available water supply and the application of water for beneficial use purposes.
- d. Duplicate Diversion, Storage, and Use Values: Identifies water rights with the same storage, use, or diversion values reported throughout the year, which may indicate inaccurate reporting and verification may be needed. Typically, seasonal storage is used to offset the time water is diverted to storage from the time period when water is used for beneficial purposes (e.g., diversion to storage in the wet season for use in the dry season). Data are also flagged when total use of water significantly diverges from the total volume of water diverted over time (e.g., volume of water diverted to storage significantly exceeds the reported volume of water used).

Missing Data

The missing data category includes water rights with missing information or missing annual reports. When water diversion and use information is not reported, it is difficult for the Division to predict what total annual diversions are, let alone the monthly timing of those diversions. Based on an analysis of data submitted for the majority of water rights, assuming that the maximum authorized amount is being diverted would artificially inflate estimated demand. For example, using a potentially inflated estimate of demand could lead to a scenario where the Division would determine that not enough water supply is available to meet all demand in a watershed, which could negatively impact junior water right holders due to curtailment or other required water use reduction measures. To avoid this, the simplest and most conservative approach would be to assume that if there are not any reported diversions it means that no diversion has occurred. It should be noted however, that this approach may have the opposite effect and result in an allocation of an inflated estimate of available water that could result in impacts to senior water right holders. These modules flag the following missing data records:

- a. Missing Contact Information: Identifies water rights that have missing contact information in eWRIMS, which reduces the capacity for Division staff to follow-up on reports and other outreach.
- b. Missing RMS Reports: Identifies water rights that are missing an annual report(s) from 2014-present.

Water Availability Data

The water availability data category includes a series of calculations using the existing data in RMS and eWRIMS to provide further context when analyzing the demand data or determining water allocation during times of shortage when projected supply is insufficient to meet projected demand. These modules identify the following information:

- a. Assign Priority Date: Uses existing information related to a water right application or initial statement filing date to standardize the way a priority date is assigned to all pre-1914, riparian, and appropriative water rights. It also provides placeholder solutions when data is not available to assist in priority-based water availability analyses.
- b. Account for Water Returned Back to System: Accounts for any water being returned to the surface water system by irrigation runoff or non-consumptive beneficial uses (e.g., power and aquaculture). The flow can either be added back into the supply flow, when feasible, or, in most cases subtracted from the estimated demand.
- c. Multiple Points of Diversion in Different Sub Watersheds: Identifies water rights that have multiple active points of diversion in multiple sub watersheds with different Hydrologic Unit Codes and standardizes the method for selecting a single point of diversion to ensure the demand is attributed to the appropriate sub watershed.
- d. Primary Beneficial Use: Identifies all potential beneficial uses and standardizes the assignment of the dominant beneficial use type for all water rights. This is required to determine when diversions are for non-consumptive uses, as discussed above.

Demand Data Analysis Methodology Overview

This Methodology consists of a series of data pre-processing steps, R Scripts, and data processing modules that flag and correct, when possible, data errors to produce a cleaned water demand dataset. The overall approach is based on a compilation of decision trees developed by Division staff that both identify the most significant and common errors found in the RMS and eWRIMS databases and provide applicable data clean-up solutions. An overview of the Demand Data Analysis Methodology Workflow is shown below in Figure 1.

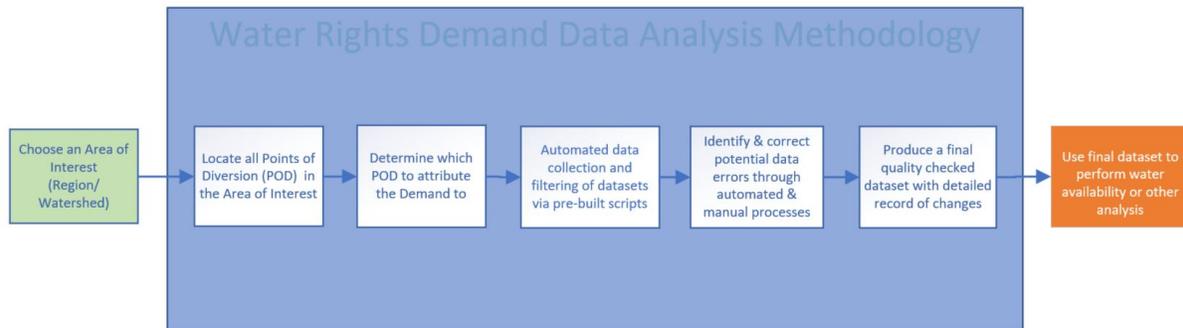


Figure 1. Overview of the Water Rights Demand Data Analysis Methodology Workflow.

Watershed/Area of Interest

The user should first consider the scope of their intended analysis and any associated time constraints (e.g., emergency regulation). If staff resources are limited, then targeting a selection of the demand data is recommended. Currently, there are over 9.4 million water diversion and use records in the eWRIMS RMS database; the number of records included in the demand data analysis can be greatly reduced by identifying an area of interest at a region or watershed geospatial scale. Additionally, trends in beneficial use types, methods of diversion, and other similarities tend to apply at a watershed level which provides context when reviewing and/or interpreting the water diversion and use data. Alternatively, the user can also prioritize using the Data Processing Modules that have the most pertinence and impact to the purpose of their analysis.

Demand Data - Pre-Processing

Once the scope of the analysis has been determined, the user will complete a series of data pre-processing steps to convert the raw data from the eWRIMS and RMS flat files into a workable input dataset for the Data Processing Modules. The user will run an R script that is provided to generate a set of input tables with pre-selected data fields, apply pre-defined universal filters, and generate the reduced list of eWRIMS and RMS data fields to assess water rights within the area of interest. A subset of the data input tables is then passed through a series of GIS pre-processing steps to identify water rights that could divert water from within the boundary of the area of interest. The water rights are assessed for accuracy by confirming the physical location of the PODs for each water right. A geospatial comparison of the recorded latitude/longitude coordinates and the Public Land Survey System (PLSS) section for each POD revealed location discrepancies for over 4% of water rights in the eWRIMS database system. To address potential location discrepancies, the user will need to use ArcGIS Pro and follow a set of standard operating procedures, pre-built templates, and automated data input scripts to compare the location of each POD by the recorded latitude/longitude coordinates and the PLSS section recorded for each water right and revise their dataset accordingly. The input dataset can then be exported from ArcGIS Pro to a CSV file to

be used in the Data Processing Modules. The user then inputs the revised dataset generated in the prior step into the Data Processing Modules that have been developed in either R or an Excel Workbook format. The Data Processing Modules guide the user through a series of modules that help identify the errors and calculate the data statistics categories discussed above (i.e., excessive reported diversions, duplicate reporting, missing data, and water availability data). The Data Processing Modules are described below:

- a. Priority Date: Assigns a single priority date for each water right or assigns a date code for riparian and/or pre-1914 water rights when a priority date is not available. Note: the calculations in the other modules are dependent upon the results of the Priority Date module, therefore it must be completed first.
- b. Beneficial Use and Return Flow: Assigns a primary beneficial use for each water right and calculates the percentage of return flow for water rights with a non-consumptive primary beneficial use.
- c. Missing RMS Reports: Identifies water rights that have missing annual RMS reports and calculates the number of missing annual RMS reports versus the number of reports expected.
- d. Duplicate Values – Months and Years: Identifies water rights with duplicate values reported for multiple months in a single annual water diversion and use report.
- e. Duplicate Diversion for Multiple Water Rights: Identifies water rights with the same annual diversion amount reported across multiple water rights by a single water right owner.
- f. Duplicate Points of Diversion: Identifies when multiple water rights might be sharing an identical reported point of diversion
- g. Diversion Out of Season Part A and B: Identifies water rights reporting diversion amounts outside of the diversion season authorized by water right.
- h. Expected Monthly Demands, Diversion Exceeds Face Value, Unit Conversion, Duplicate Diversion/Storage/Use, and Diversion Statistics: Restructures the diversion and use data into more understandable tables at annual, monthly, and individual report scales and does the following:
 - Provides specific metrics for assessing whether diversions exceed expected total annual amounts using Face Value or Initial Reported Diversion data;
 - Pre-calculates common potential unit conversion reference values;
 - Flags potential errors related to duplicate reported diversion, storage, or use; and

- Generates statistics on average and standard deviation across monthly and annual reported values to assist the user in identifying outlier reported data.

This module will produce results based on diversion and use values, as reported by users. To increase transparency and better facilitate tracking of any changes made to diversion and use data, any diversion related changes will be made in the Demand QAQC Working File (Working File).

After all of the Modules have been run, the user merges the results into a single “working file” dataset in data-only format. The Working File consolidates all relevant metrics into one place for the user to review. This file is also used to memorialize any changes made to the water rights data to improve transparency and reproducibility of the final diversion and use data.

Demand Data Analysis

The next step is to conduct an analysis of the Demand Data output dataset. The Demand Data Analysis Suggested Review Guidelines Document was developed to help the user determine which water rights records and/or water use reports contain data errors that should either be modified to improve accuracy or be excluded from the final dataset. To confirm whether data was properly flagged as an error, the user must have a good understanding of the intent of the Data Processing Modules, the source of the data fields used, and the basis and intent of the calculations. A fundamental understanding of water rights principles and the way water rights data is characterized in the eWRIMS database is also important. More detailed review of the water right and reported data may be needed to fully verify whether the data was flagged due to user reported error, water rights database data entry error, or if it was just improperly or unnecessarily flagged by the modules. This typically includes review of additional information such as the individual water right’s corresponding electronic/physical records, related water rights, GIS mapping and point(s) of diversion, and reporting history.

Once flagged records have been confirmed to be data errors, the data should be corrected based on the recommended course of action. Some types of errors will have automated recommended solutions, such as assigning a priority date when no data is available or assigning a primary beneficial use most representative of the largest use of water. Others may require a more nuanced solution based on a more thorough understanding and characterization of the water right(s) involved. Alternative solutions may include removing the flagged records, setting them to zero values, updating missing or errant water rights database records, contacting the stakeholder for clarification, or recommending no action at this time. The user should memorialize any changes to the dataset in the Working File, documenting the data correction action taken and justification for correcting or removing records from the final demand dataset to provide transparency.

Demand Data - Post-Processing

Finally, the user compiles the fields that contain the results of the Demand Data Analysis and corrected demand data into a “master” version of the dataset. The Working File spreadsheet includes the Master Demand Table tab with most fields pre-populated using formulas that pull data from the relevant Processing Module tabs where data was previously copied and pasted as data only. The final dataset in the Master Demand Table tab will contain the fields necessary to characterize demand in the watershed or area of interest, including diversion information, water right priority date, owner information, POD location, beneficial uses, and whether any records were altered in the analysis process. These data fields best characterize water demand in a watershed or area of interest, but the user may want to add additional fields to the final dataset that are relevant to their specific project.

Results

The compiled Master Demand Table is the final product resulting from the Methodology and can be used with the Drought Water Rights Allocation Tool or other methods of determining water availability. The dataset can also be used generally to characterize water demand within a watershed and can be input into data visualization applications such as Tableau or Power BI.

A diagram detailing each step in the Demand Analysis Methodology Workflow covered above is provided in Appendix I.

Next Steps

This Methodology will be updated regularly, as needed, based on feedback from interested parties and lessons learned through implementation. Division staff broadly recommend the following next steps to improve the current Methodology:

- The Data Processing Modules are currently in an Excel Workbook format and should be converted to Python or R Scripts to enhance the scalability to larger areas of interest and watersheds.
- Establish a feedback loop between Division staff and the public to help address the underlying causes of the data errors associated with data reported by water right holders.
- Additional data processing techniques were considered to bolster the analysis, and may be useful for future versions, but a significant portion of the relevant fields currently within the eWRIMS database are unreliable or otherwise unusable in its current state. These are discussed in more detail below.

Additional Data Errors and Data Processing Modules

The following approaches and data processing modules were considered by staff to identify additional data errors, however additional research needs to be conducted or

improved data collection and storage measures implemented in order for them to provide meaningful information.

- Verify POD/POU mapped locations - using coordinates and Assessor Parcel Numbers (APNs). Currently, APN data is an inconsistently populated field.
- Approximating diversion amount for non-reporters - If reported diversion and use data is not available for appropriate water rights, develop alternative methods to approximate demand. For example, identify and develop trends based on other water rights with reported use, such as a proportion of face value used based on size and/or beneficial use type.
- Identify infeasible diversion and use reporting for domestic water rights by calculating the gallons of water being used per day per person based on the reported diversion and use. Currently, "number of people served" is an inconsistently populated field.
- Identify infeasible diversion and use reporting for "irrigation" water rights based on estimated crop duty. Currently, acreage and crop types are inconsistently populated fields in the water rights databases.

eWRIMS Data Deficiencies

The following items have been identified by staff as areas where data within eWRIMS is inadequate to optimally calculate and/or quality check water use and demand estimates.

Data collected from initial water rights applications:

- Data quality control for POD locations: POD locations are now reported on applications using latitude and longitude, but there are no validation or quality checks to ensure those points were entered into eWRIMS correctly. Even though there is corroborating information on the application (i.e., APN or PLSS location), they are not consistently recorded in eWRIMS or are derived from latitude/longitude.
- Comprehensive spatial record of Place of Use (POU): eWRIMS POU data is not recorded, which makes it difficult to assess ownership and use information in addition to accounting for any water returned to the water system.
- Standard GIS format for spatial data: Standard geographic coordinate systems and projections are not used, which require staff to complete additional interpretation and conversions.
- Standard water source reference: The preferred source to identify surface waters of the State is the National Hydrography Dataset (NHD) layer. The latest NHD version has established an extensive surface water dataset down to the ephemeral stream type level. The water source name field the applicant provides is unusable if they are not using a standard reference.

- Different fields used to establish priority dates: Date fields used to establish water right seniority dates are different for various water right types.

Data collected from use reports:

- Water use is reported by water right without the granularity to see the amount diverted from individual PODs;
- Water use is not reported by beneficial use;
- POU location and amount used at POU is not a requirement of reporting;
- Water use reports don't have consistent units (i.e., acre feet vs cubic feet);
- Missing reports makes it difficult to determine diversions and use.

Appendix I: Detailed Diagram

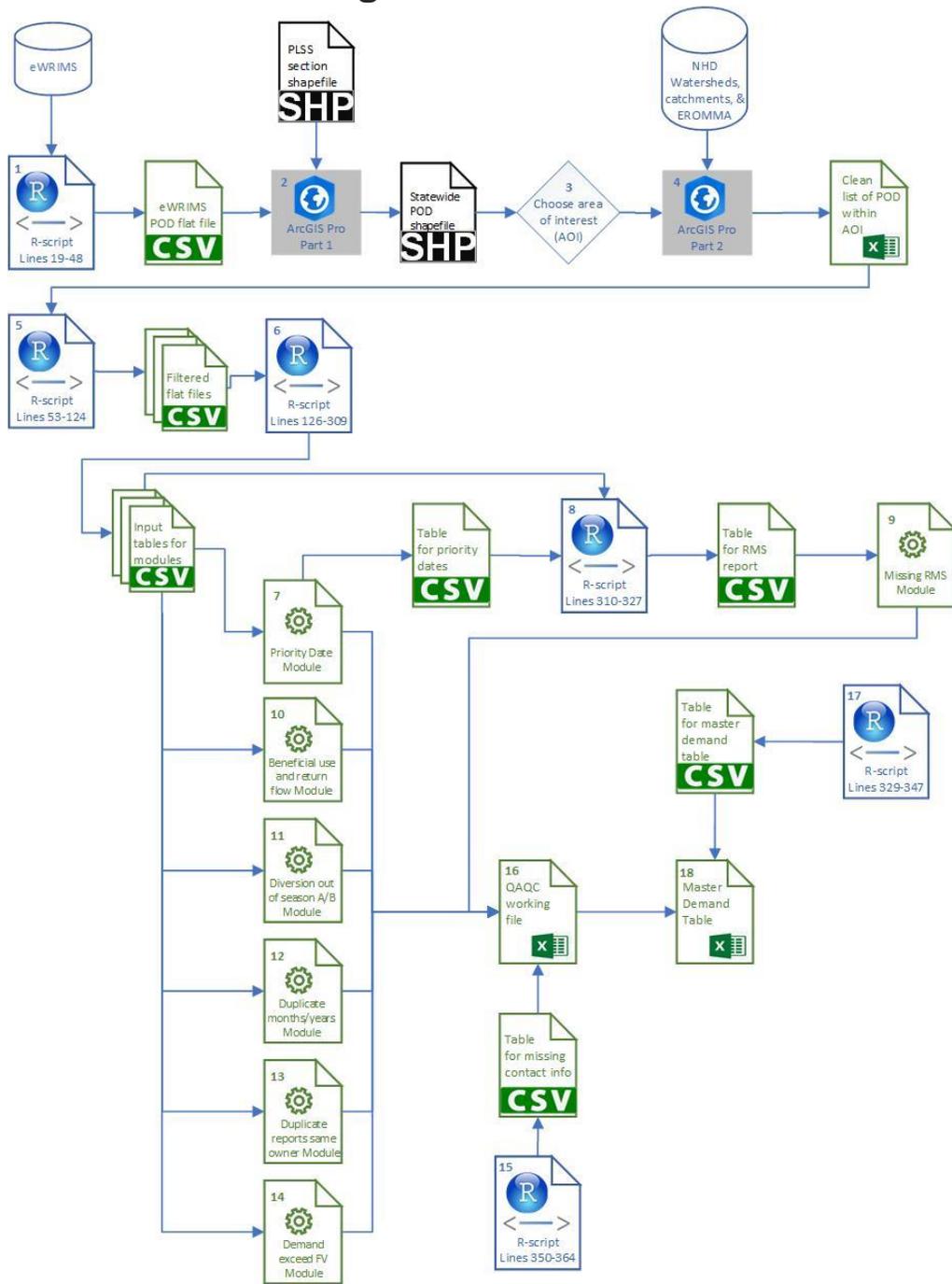


Figure 2. Detailed Diagram of the Demand Data Analysis Methodology Workflow