Clear Lake Watershed Measurement Files Evaluation (2018-2022)

Telemetry Research Unit, Division of Water Rights, State Water Resources Control Board

March 2025

Executive Summary

The State Water Resources Control Board's Division of Water Rights launched a Telemetry Pilot Project[1] in 2023 to evaluate data collection and reporting processes and explore telemetered water monitoring practices. For water monitoring data, telemetry is the process of automated collection by sensors and automated transmission using communication systems like cellular towers and satellites. Telemetry enables real-time, remote water data collection that improves efficiency and accuracy by reducing field visits and reporting errors. A typical telemetry system includes a sensor, data logger, antenna, and power source.

As part of the Telemetry Pilot Project, the Division evaluated water diversion data files submitted by reporters in the Clear Lake watershed[2] between 2018 and 2022 in compliance with the Board's Water Measurement and Reporting Regulation[3] (also called Senate Bill 88 or SB 88 reporting requirements).[4] These files are not telemetered and current regulation does not require telemetry by any reporters. The purpose of the evaluation was to understand how reporters format their data files and if those files are suitable for automated data analysis.

The Clear Lake watershed was selected because it is identified as a priority watershed in the Board's Strategic Work Plan[5] and because of recent efforts to protect the Clear Lake hitch (*Lavinia exilicauda*). The Clear Lake watershed also serves as a representative small, rural, and agricultural area that can be used to identify trends and challenges common to similar watersheds statewide.

The Clear Lake Watershed Measurement Files Evaluation (2018-2022) is one of several reports developed by the Division on this topic. The results of these reports form a baseline understanding of data collection and formatting practices and identify potential

[1] https://www.waterboards.ca.gov/telemetry
[2] Hydrologic Unit Code (HUC) 1802011603
[3] https://www.waterboards.ca.gov/waterrights/water_issues/programs/diversion_use/water_measurement.html
[4] California Code of Regulations 23:931-938
[5] https://waterboards.ca.gov/board_info/priorities/docs/workplan_2023.pdf

barriers to compliance. Insights from the reports will inform future projects aimed at reducing reporting burdens for water users and increasing real-time water availability data for improved water management.

Between 2018 and 2022, reporters in the Clear Lake watershed submitted 23 data files, which represented reporting compliance of less than 23%. The Division's evaluation of these files found that reporters submitted different types of water data, namely volume, flow rate, and evaporation, using a variety of data formats.

While all the files were in tabular, digital form, these variations in data types and format limit data usability, as the data requires considerable manual staff processing before any automated analysis. These barriers limit the Division's ability to accurately assess water availability in a timely manner. A major finding was that data format sometimes changed even when the device didn't change, highlighting that devices can produce a wide range of data formats and the Water Board may need to establish a reporting format in order to receive standardized data. Additionally, low reporting compliance provides an incomplete picture of water use, further hindering effective water management.

This technical report details the findings of the Clear Lake Watershed Measurement Files Evaluation (2018-2022).

Table of Contents

1. Reporting Characteristics

The 23 submitted files came from 6 different application identifiers (i.e., "AppID"), where each AppID roughly equates to a water right. For this 5-year review period, 4 of the 6 AppIDs submitted fewer than 5 reports, meaning that reports were not submitted for every year (Figure 1). For the other 2 AppIDs, more than 5 submissions per AppID were submitted for the 5-year period. Multiple submissions per year may be caused by a water right being used at multiple points of diversion or by a reporter switching monitoring equipment mid-year. The submissions from the 6 AppIDs reported 14 different measurement devices (i.e., "DeviceIDs") (Figure 2). Comparing Figures 1 and 2 suggests that changes in equipment may be the primary reason reporters submitted multiple files in a year.

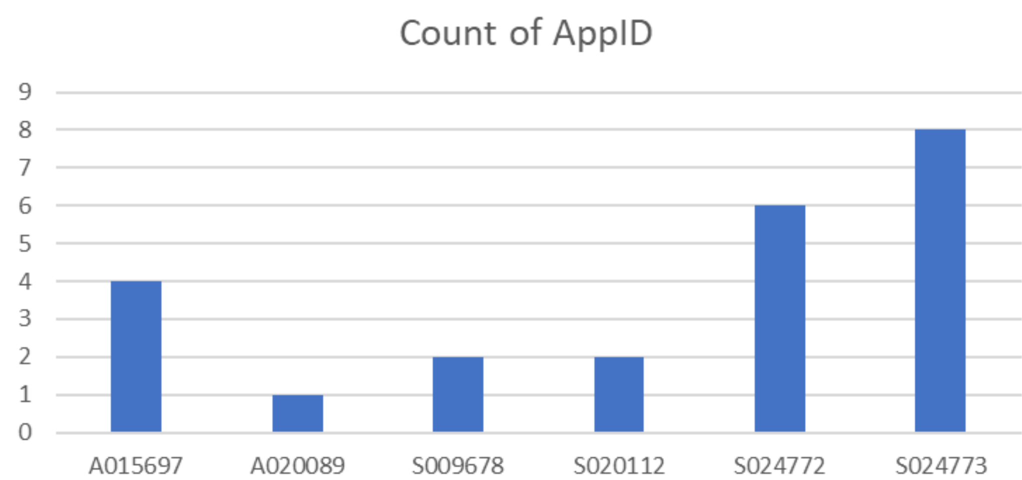

Figure 1. Vertical bar graph showing application identifiers (i.e., AppIDs) on the x-axis, and number of files submitted under that AppID on the y-axis, for the period 2018-2022.

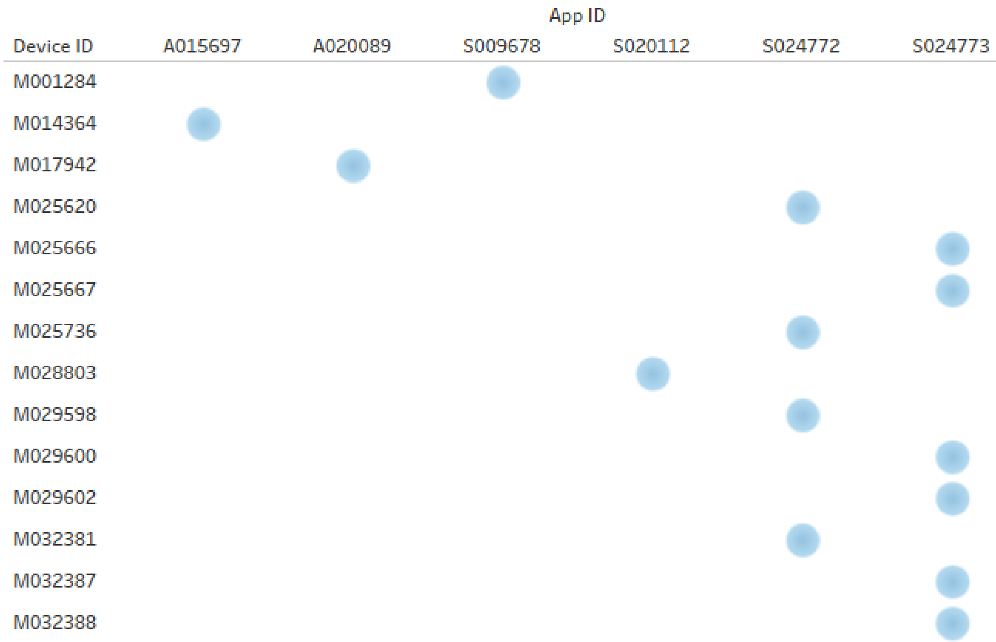

Figure 2. Blot graph showing application identifiers (i.e., AppIDs) on the x-axis, and measurement device identifiers (i.e., DeviceIDs) on the y-axis, for files submitted for the period 2018-2022.

2. Equipment Manufacturer Characteristics

Among the 23 submitted files, 14 contained water data collected by equipment manufactured by Seametrics, 2 by McCrometer, and 7 submissions did not specify the equipment manufacturer (referred to as "empty_make") (Figure 3).

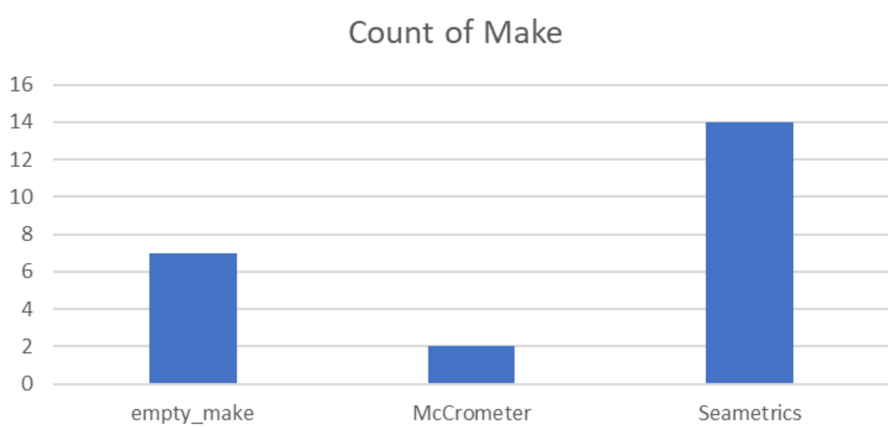

Figure 3. Vertical bar graph showing device manufacturer on the x-axis, and number of files submitted under that manufacturer on the y-axis, for the period 2018-2022.

3. Data Format Characteristics

Most submissions (21 of 23) were xlsx file types; the remaining two were xls file types. For all submissions, monitoring data were available in the file as tabular values that are potentially machine-readable, while some files had metadata included as images and tabular values. No monitoring data was included as images or other inaccessible content.

3.1 Header Row Location

Submissions varied greatly in the format of their headers and content, representing a sizable barrier to automated data analysis. For example, the header row, describing the data in each column, appeared in row 1 to row 9 of the submissions and often was spread among multiple rows (Figure 4). When comparing the header location to DeviceID for all 23 files, 11 out of 14 of the DeviceIDs consistently had the same header location in the submitted files, while the remaining 3 had 2 different header locations among the submitted files (Figure 4).

When considering manufacturer categories, among the 14 Seametrics files there were 2 header locations while among the 2 McCrometer files there were also 2 header locations. For the 7 files missing manufacturer information (i.e., “empty_make”), the header row was always spread across multiple rows and appeared in 4 different locations. Among the three manufacturer categories, Seametrics had the least variation in the header row locations. Both due to its consistency and due to the header being located in a single row instead of spread among several rows, the Seametrics files in this data set would be most amenable to automated data analysis.

When the DeviceID for a specific AppID was compared to the header location, there did not appear to be a relationship between the values. For example, AppID S024773 had a change in DeviceID that did not lead to a change in the header format, while AppID S024772 had a header change while the DeviceID remained the same. It’s unclear from the data if this relationship for S024772 is due to an unexplained change in format or an unreported change in the device, but regardless highlights the difficulty in automatically handling unstandardized reporter data.

3.2 Unit Row Location

The results of the analysis for the unit row location were similar to the results for the header row location. There were also 2 unit locations for the 14 Seametrics files, 2 for the 2 McCrometer files, and 4 for the 7 empty_make files (Figure 5). Similar to the analysis of header row location, DeviceID for an AppID was not related to unit location. Seametrics files still showed the most consistent unit location and unit location did not split among multiple rows, suggesting these files would be most amenable to automated data analysis.

For some files, unit information was entered in the same row as the other header information. The Seametrics files frequently, though not always, show this combined formatting (compare rows in Figures 4 and 5). However, for other files, the unit information was contained in a separate row or multiple rows. Even for the Seametrics files, the unit was sometimes combined with the parameter name in one column, and other times separated into its own column. This inconsistent formatting among the files is a barrier to automated analysis of the entire data set.

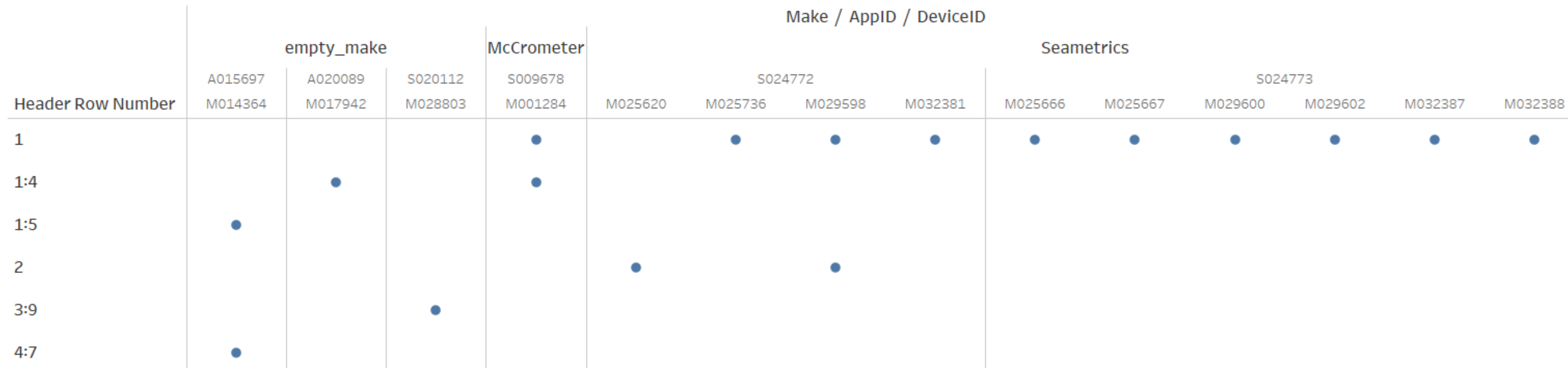

Figure 4. Blot graph showing manufacturer (i.e., Make), application identifiers (i.e., AppIDs), and measurement device identifiers (i.e., DeviceIDs) in that order on the x-axis and header row number on the y-axis for files submitted for the period 2018-2022.

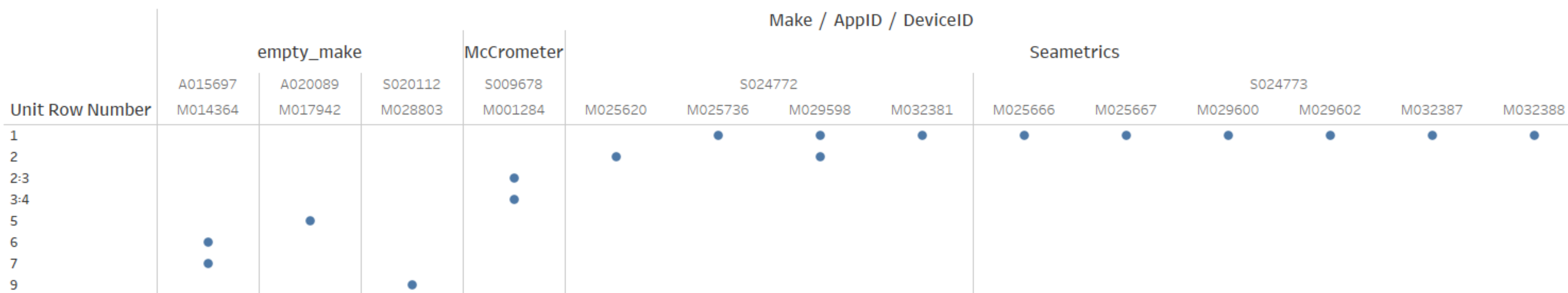

Figure 5. Blot graph showing manufacturer (i.e., Make), application identifiers (i.e., AppIDs), and measurement device identifiers (i.e., DeviceIDs) in that order on the x-axis and unit row number on the y-axis for files submitted for the period 2018-2022.

4. Data Content Characteristics

Among the 23 files, 4 types of parameters were reported: volume of water, water level, flow rate, and evaporation (Table 1).

4.1 Volume of Water

Volume of water was the most frequent parameter category and also exhibited the greatest variation in how it was represented. Of the 23 files, 21 files reported the volume of water with names such as *storage*, *Total Pos*, *diverted*, and *quantity*. The terms *diverted* and *storage* were simultaneously used in two empty_make files to describe the volume of diverted water and the volume of stored water in a lake. The units for volume of water in the empty_make files were *acre-feet* or *KGL* (meaning kilo gallon), while the unit of water volume in all 14 Seametrics files was *ccf (*meaning one hundred cubic feet).

4.2 Water Level

The term *stage* represented the water level in two empty_make files. These two files also included *storage* (representative of water volume), suggesting that the water level in these files was used to estimate the volume of stored water in the lake.

4.3 Flow Rate

Flow rates were reported in these files in two different ways. In the two McCrometer files, the daily volume of a reservoir was calculated from the daily change in flow using units of *gallon per day*. The volume difference between two consecutive days can yield the volume of water that was discharged from or recharged to the reservoir. In the two empty_make files, flow rate was reported as *Total Neg* using units of *gallon per minute*.

4.4 Evaporation

One empty_make file contained measurements for evaporation rate named as *EVP PAN INCH*.

Table 1. Characteristics of data content in Clear Lake SB88 files that shows Make; Number of files reporting this parameter; and Name of parameter in files (Name of unit).*

Measured Parameter	Format 1	Format 2	Format 3	Format 4
Water Level	empty_make; 2; stage (feet)			
Volume	Seametrics; 14; quantity (ccf)	empty_make; 5; storage (Acre-feet)	empty_make; 2; Total Pos (KGL)	empty_make; 2; diverted (acre-feet)
Flow rate	empty_make; 2; total Neg (Gal/m)	McCrometer; 2; flow (MGD)		
Evaporation rate	empty_make; 1; EVP PAN INCH (inch)			

*This table reflects the exact names of parameters and units as displayed in the files.

5. Manual File Evaluation and Other Staff Observations

TRU staff qualitatively assessed files to determine if each submitted file was manually generated and any concerns about file usability. Files that showed simple, consistent formatting were determined to be potentially machine-generated, while files that displayed inconsistent internal formatting or added pictures or tables were determined to be likely hand-generated. Inconsistently formatted files require more staff time to ingest and analyze, leading to increased business costs and delay when making management decisions. For these 23 files, TRU determined that 79% (18 out of 23) were likely generated manually.

Staff also documented a usability concern about the inconsistent location of unit data in Seametrics files. Unit data varied between being embedded in the parameter header (e.g., “Quantity_ccf”) or in a separate column. This inconsistency even among a single manufacturer is a barrier to automatic analysis of the files.