#266312 v7 – TR-61 (Field Monitoring and Data Collection) Chapter 4 - Text
200-1100
LKH/AWO/JEB/KMH/NJN/mid
5/8/23, 4/18/23, 4/5/23, 3/13/23, 1/6/23

SEWRPC Technical Report No. 61

FIELD MONITORING AND DATA COLLECTION FOR THE CHLORIDE IMPACT STUDY

Chapter 4

DATA MANAGEMENT AND DOCUMENTATION

4.1 INTRODUCTION

Due to the large quantity of data collected for the Chloride Impact Study, significant planning and effort were invested in the organization, storage, accessibility, and the quality assurance and quality control of the project datasets. Quality assurance (QA) includes the planning and processes established to ensure data quality and prevent issues that could arise from various aspects of the Study including field work activities, data collection, and data management. Quality control (QC) includes examining the data collected to ensure the quality of the datasets and final products. A wide array of data management strategies were established to effectively manage the transfer of data from field collection to the desktop for data interpretation and analysis.

The primary objective of data management for the Chloride Impact Study was to adhere to the following principles.

- Data Security – included the storage and maintenance of the datasets on Commission servers and protection against data loss

- Data Accessibility – included the systematic organization and documentation of data such that the data were accessible as needed and available in a format that was both usable and understandable

- Data Consistency – included the strategies and standardized processes that were performed at regular intervals to develop the datasets in a clear and repeatable manner

- Data Quality – included the processes employed throughout the data collection period and during post-processing to discover and resolve data issues and ensure that the quality of the data collected was adequate for Study purposes

Several protocols and procedures were developed for the Chloride Impact Study to support data management objectives and principles. The protocols and procedures cover various aspects of the Study from field work and data collection to data management. Field work procedures were utilized throughout the Study to address pre-deployment equipment preparation and testing, monitoring site installation and maintenance, and monitoring site decommissioning. Field data collection procedures were employed to standardize water quality sample collection and documentation. Data management procedures facilitated the organization and handling of project-related data. Additionally, record-keeping procedures and checklists allowed for tracking project assets, workflows, and data.

This Chapter describes the general data management QA/QC protocols employed for the Chloride Impact Study along with the data management processes specific to the datasets collected and maintained by Commission staff. Furthermore, this Chapter discusses the post-processing of the continuous datasets collected at the stream monitoring sites.

4.2 DATA MANAGEMENT PROCESSES AND DOCUMENTATION

General best practices for data organization and storage were established early in the Study. These practices include frequent data transfers to the Commission network where data are stored on servers that are routinely backed up to protect against data loss; preserving a copy of the raw, unaltered datasets and download files; reviewing the data and preserving a copy of the data used in the QC review process; and summarizing the data to create workable datasets for further evaluation and analysis. Standardized naming conventions were established for data stored on the Commission network to aid in data accessibility and organization. Additionally, detailed documentation of field work activities, data collection, data summaries, and data management were maintained throughout the Study.

Monthly checklists and workflow processes were developed to track and standardize data collection and data management throughout the Study from site installation through site decommissioning. To promote efficiency and reduce errors/typos, templates were developed for some elements of data collection and data management. The incremental development and review of project datasets at regular intervals allowed for the identification of issues as they arose. Because the project generated substantial amounts of data on a daily basis, investing time on assembling and proofing datasets throughout the Study rather than waiting until the data collection phase was complete, was instrumental to producing quality datasets. The following sections describe the data management practices employed for various datasets collected throughout the Study, including the continuous datasets collected at stream monitoring sites, the handheld sonde data, and the water quality sample data and associated laboratory analysis results.

Continuous Datasets Collected at Stream Monitoring Sites

The equipment and methods used to collect in-stream data from the stream monitoring sites are presented in Chapter 3. Following the completion of the water quality sample field collection each month, the data recorded at every stream monitoring site were downloaded from the ZENTRA Cloud, a cloud-based data storage system discussed in Chapter 3. To ensure the integrity of the data stored on the ZENTRA Cloud platform, the dates of each data download were specified to overlap several days with the previous download at each site to verify that the data for overlapping periods were consistently matching. Any differences between the overlapping data could indicate an issue with the cloud-based data storage system. The monthly download frequency allowed for reasonably efficient download speeds and helped limit the time spent retrieving data for every monitoring site in the Region. The raw data download files were preserved with Read-Only permissions on the Commission network and organized by download date and monitoring site.

Raw continuous datasets for the stream monitoring sites were constructed incrementally from the periodic downloads, starting from site installation through decommission. However, before the downloaded data were added to the raw continuous dataset for each site, the data were reviewed and evaluated for acceptability. The continuous monitoring data download and review process described in the following paragraphs is illustrated in Figure 4.1.

During review, the datasets were examined to identify data gaps, repeated data, data with timestamps out of chronological order, or suspect data that might indicate an issue with the equipment. The continuous datasets were subject to two different types of data gaps: sensor gaps occurred when the in-stream sensor was unable to communicate with the data logger, while telemetry gaps occurred when the data loggers were unable to either record data or upload data. Telemetry gaps could be caused by the loss of battery power or an equipment problem, preventing the device from recording data. Telemetry gaps could also occur when a device was unable to upload data to the ZENTRA Cloud platform due to an issue with the communication components or problems with ZENTRA Cloud. For sensor gaps, the timestamps are present in the dataset during the sensor gap periods, but “NA” is displayed in place of the missing sensor data. In contrast, for telemetry gaps both the sensor data and timestamps are missing from the continuous record for each 5-minute interval over the data gap period. In rare cases, data recorded during a telemetry gap were saved on the data logger and were recovered by downloading the data directly from the data logger device in the field. The results of the reviews were documented, and separate QC files were maintained for project records. Problematic or unacceptable data were discussed with the equipment manufacturer and were typically resolved by the manufacturer's technical support team.[1] Periods of equipment malfunction, and other situations for which the questionable data could not be resolved, were flagged and noted.

If the downloaded data were considered acceptable, they were formatted and appended to the raw continuous dataset for each site. The raw continuous datasets did not adhere to the biannual time changes for Daylight Savings, and all of the timestamps were aligned with Central Daylight Time (CDT), which is 5 hours behind Universal Coordinated Time (UTC -5:00). To maintain consistency throughout the year, data collected during the winter months when the local time was aligned with Central Standard Time (UTC -6:00) were manually adjusted by one hour. This convention required careful coordination and documentation of the data downloads and manual adjustments used for timestamp conversions to ensure proper interpretation of the data and alignment with other datasets that are recorded using local time, which was observed using both standard time and daylight savings time. The ZENTRA Cloud platform evolved throughout the Study and programming updates included a timezone override function, allowing the user to select which timezone should be applied to the data. Commission staff utilized this function later in the Study to align the continuous data with UTC -5:00 prior to download, rather than manually adjusting the data downloaded for each monitoring site.

The raw continuous datasets for the stream monitoring sites were not modified, with the exception of adjusting the timestamps by one hour for data that were recorded during local standard time. A parallel dataset was developed for each monitoring site using the R programming language with the "tidyverse" package to fill-in missing timestamp data.[2,3] This dataset provided a continuous data record with a complete set of date and time to be used for analysis. This dataset was referred to as the continuous CTD data with shared time intervals because every 5-minute interval during the study period was represented in the

---

[1] *The equipment manufacturer, METER Group, was also the host of the ZENTRA Cloud data management platform as discussed in Chapter 3.*

[2] *R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. cran.r-project.org.*

[3] *H. Wickham, et al.,* "Welcome to the Tidyverse." *Journal of Open Source Software, 4 (43), 1686, 2019.*

continuous dataset for each stream monitoring site. The timestamps that were filled in during telemetry gaps were assigned “NA” to represent the missing data. The format of the shared time interval datasets was similar to the raw continuous dataset, except the start date for every dataset was October 1, 2018, which was the first month of the routine monthly water quality sampling period for the Study.

Commission staff collaborated with other agencies such as the U.S. Geological Survey (USGS) and the Milwaukee Metropolitan Sewerage District (MMSD) to discuss issues with field deployment as well as data interpretation. Continuous data post-processing procedures are discussed later in this Chapter.

Sonde Data

The collection of handheld sonde data in the field is detailed in Chapter 3. The sonde data were downloaded at the end of each day of data collection. The daily frequency of data downloads from the mobile device used to the collect sonde data was initiated to reduce the risk for accidental data loss during field work. Copies of the raw data downloaded from the device were preserved and organized according to the collection date. The sonde data were reviewed and extracted from the raw data files and input into a summary spreadsheet. A sonde data summary spreadsheet was developed for each sampling period using a standardized format that facilitated further usage and analysis of the dataset. Sonde calibration reports, generated during the equipment calibration performed at the beginning of each day of water quality sampling, were maintained throughout the Study.

Water Quality Sample Collection and Laboratory Analysis Results

As discussed in Chapter 3, water quality samples were collected from streams and lakes across the Region and sent to the Wisconsin State Laboratory of Hygiene (WSLH) for chemical analysis. Lab datasheets and sample bottle labels were filled out consistently for each sampling period (see Appendix D). The lab datasheets were printed on water-resistant paper to minimize the risk of water damage in the field, and templates were created to utilize pre-filled fields that increased documentation efficiency and reduced the chance for errors. The forms were typically completed in the field after sample collection and were reviewed for completeness and compared to the sample collection bottle labels to ensure consistency. Any missing or suspect data were checked against the field book and other project documentation. Following review, the forms were scanned prior to submission to the WSLH, and the electronic copies were preserved for project records to support the tracking of water quality samples collected for the Study. A consistent file naming convention was established using the Sample ID which included information related to the site and the sample date and the files were organized by sampling period.

After collection, surface water quality samples were stored in a dedicated refrigerator at the Commission for no more than 7 days, as described in Chapter 3. Submissions to WSLH were timed to accommodate the lab schedule and to avoid exceeding the maximum holding time requirements for lab analyses. The samples were delivered directly to WSLH intake staff, at which time the individual lab datasheets and sample bottle labels were reviewed to ensure they matched, and custody of the water quality samples was officially transferred to WSLH.

The lab normally analyzed the water quality samples in batches, and the full set of samples from a given sampling period were often processed in multiple batches. In addition, individual analysis batches sometimes consisted of samples from stream monitoring sites along with lake samples. The lab analysis results for each sample batch were summarized in a single PDF file and transmitted via email by WSLH. The original WSLH PDF files were preserved on the Commission network. For each individual sampling period, the analysis results for each monitoring site were extracted from the separate WSLH-generated PDF files and consolidated into a single PDF per sampling period. The lab results and comments were reviewed for errors and suspect data. Any issues discovered during review were discussed with WSLH staff for further investigation and correction when warranted.

The lab results were also transmitted directly from WSLH to the Wisconsin Department of Natural Resources (WDNR) Surface Water Integrated Monitoring System (SWIMS) database. Typically, these data were available for download from SWIMS the day after the lab released the PDF results. The lab data were routinely downloaded from SWIMS and the raw downloads were preserved on the Commission network. Each download was evaluated to check for missing samples and other issues. The individual downloads were used to construct a readily useable dataset. Commission staff collaborated with WDNR and WSLH staff to correct typographic errors and ensure that all of the lab results samples collected for the Study would be available for download from SWIMS using the project identification code (Project ID) to query the database.[4]

Chapter 3 describes the procedures used to collect water quality samples and the measures taken to avoid contamination or introducing bias during field collection. These quality assurance measures included the collection of field blank samples and replicate samples. As discussed in Chapter 3, field blanks were samples of distilled water, subject to the same conditions and sample collection processes as the routine water quality samples and submitted to the lab for analysis. The detection of analytes in a field blank sample could

---

[4] *The SWIMS database Project ID for the Chloride Impact Study for the Southeastern Wisconsin Region is GLPF2018_LM1802_CS*

be an indication of contamination resulting from the sample handling process or through atmospheric exposure. Analytes were detected in a small number of the Study field blank samples. The detects were dispersed among several sites over different dates, showing no pattern temporally or spatially that would indicate a quality issue with the field procedures. The concentrations of analytes detected in field blanks were very low, typically between the limit of detection and limit of quantification. Furthermore, when compared to the routine monthly water quality samples, the detected analyte levels were insignificant and on the order of one percent or less when compared to the measured concentrations in the field samples. While the detection of analytes in a field blank sample could be an indication of contamination, the concentrations were so low that they would not have had an appreciable effect on the lab results and were considered acceptable for quality control purposes.

Similar to field blank samples, replicate samples were part of the QA/QC framework for the Study. Replicate samples were collected simultaneously with and in close proximity to routine field samples and were subject to the same conditions and processes, as detailed in Chapter 3. Because the field sample and replicate samples were not drawn or split from a common collection vessel, there could be differences in chemical composition between the two samples, given the heterogeneity of surface water flowing in the stream. The results for the replicate samples were compared to the corresponding field samples, and the replicate samples were considered acceptable if the relative percent difference (RPD) was less than or equal to 20 percent. Only one replicate sample failed to meet the acceptance criteria throughout the entire Study, with an RPD of approximately 30 percent for one analyte. For this particular replicate sample, the measured analyte concentration was approximately five times the limit of quantification, and with such a low concentration the elevated RPD of the replicate sample was considered acceptable. Another sample had five different analytes with RPD values ranging between 10 and 20 percent, and while the RPD levels were in the acceptable range, this sample was considered unique because of the number of analytes involved and was reviewed further. Documentation recorded in the field book for that sample revealed that high water levels at the monitoring site hampered the sample collection effort and prevented Commission staff from entering the stream. Under these circumstances, the samples were collected using the assisted sampler as described in Chapter 3 of this Report. As a result, the replicate sample could not be taken simultaneously with the field sample but instead was collected a short time later. This deviation in the sample collection time is likely the cause of the differences in the sample analysis results and was considered an isolated occurrence rather than an indication of a problem with the sample collection process. Furthermore, there were no discernable patterns or trends identified in the lab results for field blank samples or replicate samples and thus no indication of a widespread quality issue.

Additional Documentation

Additional documentation maintained throughout the Study included a comprehensive accounting of field work activities and logs, data collection, data summaries, and data management. Field work documentation included maintaining a field visit log, which was the electronic version of the field book. The field visit log summarized all the site visit work notes for water quality sampling, equipment cleaning and maintenance, battery changes, water depth measurements, and any other relevant site or equipment information. A separate water sampling log was maintained to summarize the detailed sampling data collected during each sampling period. Data were regularly transferred from the field books into these logs, which prevented accidental data loss and provided readily accessible field information for Commission staff that may not otherwise have access to the physical field books. Additionally, weather logs were maintained during each winter season throughout the Study to document winter weather events and provide anecdotal information related to the regional distribution, type of precipitation, and estimated precipitation quantities when available. These weather logs were useful for data interpretation.

Commission staff maintained detailed equipment records tracking the equipment purchased and deployed for the Study. The equipment logs included information related to the equipment types, identification numbers, configuration specifications, and deployment locations and dates. The equipment logs also tracked equipment issues and replacements. Data recorded on a data logger remains associated with that specific device on the ZENTRA Cloud platform; therefore, it was necessary to track equipment movements, particularly for equipment deployed at multiple sites over the course of the Study, to ensure that the data downloaded from individual data loggers were attributed to the correct monitoring site. In order to construct a continuous dataset for a particular monitoring site, a record of the different equipment deployed at each site was necessary. A separate battery status log was maintained for the data logger devices to monitor battery conditions at certain sites where drastic drops in battery power created problems for the equipment and data collection. As discussed in Chapter 3, battery loss posed a risk for data loss, especially during the colder months of the year when the batteries drained more rapidly.

During the field data collection phase of the Study, documentation was maintained to verify equipment operation and performance. Following data collection, these data summaries were also useful for data interpretation. As described in Chapter 3, the depth of water above the CTD-10 sensor was measured during each field visit. A spreadsheet was developed to compare the field depth measurements with the simultaneous depth reading from the CTD-10 sensor and evaluate the performance of the pressure sensors throughout the Study. These summaries allowed Commission staff to identify large differences or diverging trends that would lead to the replacement of the CTD-10 sensors when necessary. During each field visit, a handheld sonde was used to collect specific conductance data, as detailed in Chapter 3. To evaluate the CTD-10 performance, a spreadsheet was created to compare the sonde data with the specific conductance simultaneously recorded by the CTD-10. Differences greater than the CTD-10 sensor accuracy (plus or minus 10 percent) were an indication of potential sensor fouling, and the CTD-10 sensors were prioritized for cleaning and maintenance. The documentation kept for cleaning and maintenance visits provided additional information related to equipment performance. The specific conductance and water depth data observed by the CTD-10 sensor immediately before and after the sensor cleaning were summarized in a spreadsheet and the percent difference was computed. Large changes in specific conductance before and after a sensor cleaning were an indication that the CTD-10 had been fouled, buried, or was performing poorly prior to the cleaning. The before and after cleaning data summary was instrumental for post-processing of the continuous datasets, described later in this Chapter.

Some of the project documentation included data summaries that would be useful for further evaluation and analysis. One of these documents was the River Sample Master Table, which summarized three datasets: the lab analysis results for water quality samples, the corresponding data collected in-stream by the CTD-10 sensor, and the handheld sonde data recorded at the time of the sample. A single datapoint was manually selected from the continuous specific conductance dataset to pair with the lab data. Typically, the specific conductance data was chosen to match the timing of the water quality sample collection; however, site visits during which the sensor was cleaned or adjusted, the specific conductance datapoint was selected after the sensor disturbance. The River Sample Master Table summary also served as a QC review tool to ensure consistency across the datasets.

4.3 CONTINUOUS DATA POST-PROCESSING

Chapter 3 described how the Chloride Impact Study used CTD-10 sensors at stream monitoring sites to collect specific conductance, temperature, and water level data measurements at five-minute intervals. Rivers and streams are dynamic environments where water quality conditions can fluctuate rapidly. Using sensors to collect continuous data can provide researchers with a high-resolution dataset to capture these rapidly changing conditions; however, aquatic environments can have deleterious effects on electronic sensors. The growth of biofilms, algae, resident invertebrates, and the accumulation of sediment on the sensor and in the sensor housing, are examples of sensor fouling that erode the ability of the sensor to record accurate measurements over time. Consequently, CTD-10 sensor data needed to be reviewed carefully and, when appropriate, adjusted for fouling to produce the high quality datasets to be used for future Study analyses.

This Study used methods that were modified from the USGS Guidelines and Standard Procedures for Continuous Water-Quality Monitors to adjust data due to sensor fouling.[5] This is an accepted practice to adjust for sensor issues that can arise between monitoring site visits. These methods assume that fouling takes place at a constant rate over time. The data adjustment method used for this Study also assumed that a CTD-10 sensor was recording accurate readings after being cleaned during a field maintenance or monthly water sampling visit. This assumption was necessary because the CTD-10 sensors cannot be calibrated in the field. Occasionally, a large disparity was recorded between sensor readings before and after cleaning. These differences were an indication of potential sensor fouling. This section describes the methods and procedures used to identify and adjust CTD-10 sensor data affected by sensor fouling during the Chloride Impact Study.

Examination of Continuous Datasets

Commission staff initially examined the record of each stream monitoring site individually to locate common signatures in the data. Each record was plotted using the dygraph package in R.[6] This package creates interactive plots that allow the user to adjust the scaling of the axes, in order to view the entire time series or any portion of the time series in detail. The plots were constructed by graphing water level and specific conductance data that were collected by the CTD-10 sensors on separate vertical axes and time on the horizontal axis for the entire period of record.

Simultaneous visualization of specific conductance and water level highlighted the relationship between the two variables and was critical to the process of identifying data potentially affected by fouling. Specific conductance measures the ability of water to conduct electricity. Water level is related to the amount of water flowing in the stream. In many situations, there is an inverse relationship between specific conductance and water level. As the amount of water flowing in a stream increases, water levels rise. Under some conditions this increase in water volume will dilute the number of ions in the water, causing a decrease in specific conductance. Figure 4.2 shows an example of this dilution effect that was observed during a precipitation event at a stream monitored for this Study. When the amount of water entering the stream due to the rainfall and runoff decreased, water levels receded, and specific conductance levels typically

---

[5] *R.J. Wagner, R.W., Boulger, Jr., C.J. Oblinger, and B.A., Smith,* Guidelines and Standard Procedures for Continuous Water-Quality Monitors: Station Operation, Record Computation, and Data Reporting*, U.S. Geological Survey Techniques and Methods 1-D3, 2006.*

[6] *D. Vanderkam, J.J. Allaire, J. Owen, D. Gromer, and B. Thieurmel,* Dygraphs: Interface to 'Dygraphs' Interactive Time Series Charting Library. *R package version 1.1.1.6, 2018.*

rebounded. Additionally, a different signal may occur under other conditions. For example, during spring thaw in an urbanized area, water from melting snow may transport additional chloride to waterways. Under these conditions, specific conductance may increase as water levels increase.

The examples given in the previous paragraph highlight the importance of using other types of information to interpret the specific conductance and water level signatures shown in the visualizations. During examination of the CTD-10 sensor data, Commission staff consulted other references as needed. These references included field books containing notes on the condition of the sensors, a weather logbook describing localized precipitation events, and National Weather Service records.

Identification of Data Signatures

Each monitoring site dygraph was initially examined to identify and define common signatures in the data for each site. The most common data signatures and the frequency at which each occurred were recorded in a spreadsheet. The most commonly observed signatures throughout the CTD-10 continuous datasets for all monitoring sites were referred to as spikes, noise, teeth, and dampened data.

A spike in the continuous dataset is characterized by an unusually high, short duration increase in specific conductance that rapidly returns to its prior level (see Figure 4.3). Spikes that had greater than 25 percent divergence from the initial reading were investigated further, as such spikes could result from legitimate increase in conductance caused by the first flush of chloride-laden runoff into the stream due to a precipitation event. The majority of the specific conductance spikes examined were explained by changes in water level or in the meteorological records and did not require any adjustment. When a water level increase or weather event did not precede the spike, it was flagged as potentially erroneous data. Given their short duration, spikes are likely to have little effect on future analyses for this Study. Because of this, no adjustments were made to specific conductance spikes in the Study dataset.

Noise in the data was defined by repeated high frequency fluctuations in specific conductance (see Figure 4.4). Some noise events in the CTD-10 dataset lasted for a few hours, while others lasted for weeks or months. The cause of these fluctuations could not be identified. Potential explanations include radio frequency interference or invertebrates interfering with the electrodes in the sensor, but these hypotheses could not be substantiated. The short duration and relatively small magnitude of individual fluctuations during noise events suggested that noise would have limited impact on analyses for this Study. Because of this, no adjustments were made to noise in the continuous specific conductance datasets.

A tooth in the data was characterized by sudden decrease or increase in specific conductance followed by a period during which conductance remains at the lower or higher level. This period is followed by a sudden return to the original level (see Figure 4.5). Tooth-type signatures were identified when changes in measured specific conductance of at least 10 percent occurred over periods of 15 minutes or less. These were less common than some of the other signatures and only five instances of this example were recorded. The cause of this data signature could not be identified. Because of their rarity, no adjustments were made to the continuous specific conductance datasets for these instances.

Dampened data was characterized by magnitudes of specific conductance that were unusually low given the other data at the site and the specific hydrologic context (see Figure 4.6). The amplitude of fluctuations in specific conductance was also often reduced during dampening. Three signatures in the specific conductance data suggested that dampening might be affecting readings. First, a large increase in specific conductance following cleaning of a CTD-10 sensor was considered a sign of dampening. Second, failure of the specific conductance readings to recover to the pre-event level following a hydrologic event was an indication that the CTD-10 sensor data were dampened. Third, specific conductance readings taken by the handheld sonde that were substantially higher than those taken simultaneously by the CTD-10 sensor were signs of potential dampening.[7] The presence of more than one indicator was considered stronger evidence of potential dampening. The dampened data was likely the result of CTD-10 sensor fouling that affected the performance of the CTD-10 sensor and lowered specific conductance readings. Since this artificial lowering of specific conductance could lead to underestimates of calculated instream chloride concentrations, instances of dampened data due to sensor fouling were investigated as described in the following section.

Identification and Interpretation of Sensor Fouling

Considering the challenges inherent to long-term, in-stream water quality monitoring, sensor fouling was expected as the CTD-10 sensors were exposed to conditions that can affect the accuracy of the recorded measurements over time. Several steps were taken during continuous data post-processing to verify sensor fouling. First, periods of potential sensor fouling were identified in the continuous specific conductance datasets. Further investigation was performed to estimate the magnitude and duration of the sensor fouling,

---

[7] *As noted in Chapter 3, the handheld sonde was calibrated daily before field visits and the sonde data had higher accuracy than the CTD-10 data. In some instances, the locations at which sonde readings were taken were some distance from the CTD-10 sensor. In those instances, a difference between the sonde and CTD-10 readings was not considered an indicator of fouling.*

and to determine a cause. Finally, this information was evaluated to determine whether it would be appropriate to adjust the data affected by sensor fouling using the methodology described later in the Data Adjustment Calculation and Application Procedures section.

Sensor fouling was typically identified by reviewing data from site visits during which the CTD-10 sensor was replaced, cleaned, or otherwise maintained. As previously mentioned, large differences in specific conductance values measured by the sensor before and after maintenance were an indication that the sensor had been fouled or was performing poorly prior to the site visit. Figure 4.7 illustrates some typical data signatures observed in the continuous datasets during a sensor cleaning, showing how the specific conductance and water level data abruptly decreased to near zero while the sensor was removed from the stream for maintenance. The figure also demonstrates how sensor maintenance can affect specific conductance measurements, comparing small and large differences in specific conductance values before and after cleaning.

The extent of sensor fouling was estimated based on the magnitude of the change in specific conductance values prior to and following CTD-10 sensor replacements and equipment maintenance. Specific conductance changes of less than 10 percent fall within the range of sensor accuracy and were not considered for data adjustment, even if there was additional evidence in the field notes that the sensor had been fouled. When the difference in specific conductance before and after CTD-10 sensor cleanings was greater than 10 percent, the specific conductance data were assumed to be affected by significant sensor fouling and were considered for data adjustment using the procedure describe in the next section. Typically, when the change in specific conductance resulting from a sensor cleaning was greater than 50 percent, the data prior to the cleaning were not considered for data adjustment as these data were deemed too greatly impacted to be adjusted. In these situations, the affected data were flagged such that concerns about the accuracy of the specific conductance readings could be reviewed for subsequent Study analyses. Furthermore large, abrupt changes in specific conductance that occurred between site visits were not considered candidates for data adjustment.

Following the identification of suspected sensor fouling, the specific conductance and water level data records preceding the site visit were inspected in order to estimate the duration of sensor fouling. The inspection sought to identify a potentially significant fouling event that occurred between site visits to establish a starting point for the sensor fouling. For example, the start of sensor fouling occasionally appeared to coincide with a large flooding event. If a specific event causing fouling could not be clearly identified, sensor fouling was assumed to begin after the previous sensor cleaning or deployment. The periods of data suspected to be affected by sensor fouling were further investigated using several resources to try to determine a cause of the sensor fouling. Notes recorded in the field book during site visits provided information related to the condition of the sensor and relevant site conditions. Meteorological data and streamflow data, where available, were useful for identifying initiating conditions for sensor fouling. In general, when a cause of sensor fouling could not be determined or explained, the affected data were not considered for data adjustment.

Information obtained through the investigation of suspected periods of sensor fouling was evaluated in context with other site data and individual site characteristics to determine whether portions of the continuous specific conductance datasets should be considered for data adjustment. For the periods of suspected sensor fouling that were considered candidates for data adjustment, the estimated magnitude and duration of sensor fouling were used in the data adjustment procedure described in the next section.

Data Adjustment Calculation and Application Procedures

Modifications to the continuous specific conductance datasets were considered only after rigorous review and verification using the evidence that was described in the previous two sections. Specific conductance data were considered for adjustment in those cases where there was strong evidence that the CTD-10 sensor had been fouled.

This Study used methods to adjust data affected by sensor fouling that originated in the guidelines and standard procedures for continuous water quality monitoring developed by the USGS in 2006 and a separate study that expanded upon these methods in 2011.[8, 9] An R programming package named driftR was developed based on these methods.[10] This software package provides a free, publicly available open-source option for adjusting data affected by sensor drift, which can be caused by fouling. Commission staff used driftR to assist with adjustments of dampened datasets collected for this Study. The same adjustment procedure was used for this Study regardless of the severity or type of sensor fouling.

---

[8] *Wagner, Boulger, Jr., Oblinger, and Smith, 2006,* op. cit.

[9] *E.A. Hasenmueller,* "The hydrology and geochemistry of urban and rural watersheds in east-central Missouri," *Washington University in St. Louis, 2011.*

[10] *A.R. Shaughnessy, C.G. Prener, and E. A. Hasenmueller,* "An R Package for Correcting Continuous Water Quality Monitoring Data for Drift," *Environmental Monitoring and Assessment, 191:1-10, 2019.*

The data adjustment procedure makes two assumptions. First, the procedure assumes that fouling happens linearly over time. For any given specific conductance value in a data record:

$$f_t = \left(\frac{t}{\sum t}\right)$$

Where:

f_t = adjustment factor

t = interval of time from the start of the fouling

$\sum t$ = total duration of fouling

This equation creates a time-weighted factor that adjusts the data for increased error due to fouling over the period affected. Because of the time-weighting, the data at the beginning of the fouling period are adjusted less than the data at the end of the fouling period.

Second, the procedure assumes that the values read by the CTD-10 sensor after being cleaned were accurate, in lieu of field calibration.[11] This post-cleaning specific conductance value was used as a calibration standard to create a one-point adjustment using the factor calculated in the previous equation. The adjusted specific conductance is expressed as:

$$C = m + f_t\,(s_i - s_f)$$

Where:

C = adjusted value of the specific conductance

M = raw specific conductance data value

s_i = specific conductance value from the cleaned sensor

s_f = specific conductance value from immediately before the sensor was cleaned

The calculations for the data adjustment were performed by the driftR program. Data for the affected duration and the specific conductance values before and after the cleaning of the sensor were entered into the program. The software calculated an adjusted value for each reading in the portion of the data record that was affected by fouling which can be seen represented in orange in Figure 4.8.

---

[11] *The CTD-10 sensors were factory-calibrated and could not be calibrated in the field; therefore, a cleaned sensor was assumed to be reading as accurately as a calibrated sensor.*

The adjusted data were reviewed visually using dygraphs that show both the raw and corrected specific conductance data. This review examined whether the adjusted data followed patterns similar to those in data records not affected by fouling. These patterns included whether specific conductance fluctuations occurred in response to hydrological events and whether patterns were similar to those observed elsewhere in the data record. Table 4.1 summarizes the portions of the specific conductance datasets for which adjustments were made.

SEWRPC Technical Report No. 61

FIELD MONITORING AND DATA COLLECTION FOR THE CHLORIDE IMPACT STUDY

Chapter 4

DATA MANAGEMENT AND DOCUMENTATION

TABLES

Table 4.1
Summary of Data Adjustments

Site ID	Site Name	Adjustment Start (CDT)[a]	Adjustment End (CDT)[a]	Specific Conductance Difference[b] (µS/cm)	Adjustment Span (days)
3	Mukwonago River at Mukwonago	2019-03-14 12:00:00	2019-04-09 10:50:00	86	26.0
4	Sugar Creek	2018-10-01 00:00:00	2018-10-19 11:00:00	122	18.5
		2019-03-12 18:00:00	2019-04-10 11:10:00	80	28.7
		2019-09-10 07:00:00	2019-09-19 10:20:00	169	9.1
8	Pewaukee River	2019-05-23 15:55:00	2020-05-12 11:10:00	262	354.8
9	Oak Creek	2020-08-02 15:20:00	2020-10-08 15:55:00	397	67.0
10	Pike River	2018-10-30 18:05:00	2018-11-12 13:10:00	127	12.8
		2019-04-15 12:30:00	2019-06-11 11:55:00	109	57.0
11	Bark River Upstream	2019-03-13 12:40:00	2019-04-09 13:55:00	134	27.1
		2019-04-22 23:20:00	2019-06-13 11:50:00	107	51.5
13	Ulao Creek	2020-03-28 22:20:00	2020-04-07 09:55:00	435	9.5
		2020-05-17 09:35:00	2020-06-15 13:35:00	185	29.2
14	Sauk Creek	2019-06-12 19:40:00	2019-06-14 13:00:00	180	1.7
		2019-09-13 17:35:00	2019-09-20 14:10:00	170	6.9
		2020-07-09 20:25:00	2020-10-09 09:20:00	104	91.5
15	Kilbourn Road Ditch	2018-10-10 17:00:00	2018-12-12 13:00:00	103	62.8
		2019-03-14 04:45:00	2019-04-15 11:25:00	344	32.3
		2020-09-08 16:20:00	2020-10-08 11:45:00	290	29.8
16	Jackson Creek	2018-10-06 01:35:00	2018-10-12 12:00:00	72	6.4
18	Oconomowoc River Upstream	2019-02-01 01:30:00	2019-04-09 14:25:00	101	67.5
		2020-03-09 16:55:00	2020-05-12 10:15:00	185	63.7
20	Oconomowoc River Downstream	2020-03-08 23:25:00	2020-07-16 12:40:00	203	129.6
21	East Branch Milwaukee River	2019-03-17 07:15:00	2019-06-03 11:25:00	119	78.2
23	Milwaukee River Downstream of Newburg	2019-03-13 17:00:00	2019-04-08 12:00:00	59	25.8
25	Root River Canal	2018-10-01 00:20:00	2018-10-24 14:50:00	376	23.6
		2018-12-01 18:00:00	2018-12-12 13:35:00	121	10.8
28	East Branch Rock River	2018-10-30 16:30:00	2018-12-07 11:50:00	135	37.8
		2019-07-20 12:25:00	2019-09-11 11:25:00	97	53.0
29	Root River near Horlick Dam	2018-10-03 00:00:00	2018-10-16 12:10:00	316	13.5
30	Des Plaines River	2019-10-27 06:20:00	2019-11-20 11:40:00	235	24.2
32	Turtle Creek	2018-12-21 00:35:00	2019-04-12 11:55:00	248	112.5
		2020-03-09 14:40:00	2020-04-14 11:20:00	162	35.9
33	Pebble Brook	2020-06-22 21:45:00	2020-10-07 10:35:00	131	106.5
36	Honey Creek Downstream of East Troy	2018-10-01 12:30:00	2018-10-19 11:40:00	237	18.0
38	North Branch Milwaukee River	2019-07-20 12:20:00	2019-09-24 11:00:00	83	65.9
40	Stony Creek	2019-03-14 06:30:00	2019-04-08 13:00:00	86	25.3
41	Milwaukee River near Saukville	2019-03-13 14:00:00	2019-04-08 11:10:00	246	25.9
		2020-08-25 07:00:00	2020-10-09 10:20:00	80	45.1
47	Fox River at Rochester	2019-03-14 11:40:00	2019-04-10 10:35:00	92	27.0
51	Rubicon River	2019-09-10 02:00:00	2019-09-11 12:10:00	155	1.4
		2019-10-01 09:20:00	2020-01-14 15:20:00	215	105.3
		2020-06-20 12:25:00	2020-06-25 11:45:00	104	5.0
53	Honey Creek at Wauwatosa	2019-08-26 08:40:00	2019-09-09 09:25:00	376	14.0
		2020-05-17 04:00:00	2020-07-06 13:35:00	206	50.4

Table continued on next page.

Table 4.1 (Continued)

Site ID	Site Name	Adjustment Start (CDT)[a]	Adjustment End (CDT)[a]	Specific Conductance Difference[b] (µS/cm)	Adjustment Span (days)
54	Whitewater Creek	2019-02-03 14:20:00	2019-04-12 10:55:00	130	67.9
		2020-07-13 11:05:00	2020-07-22 10:35:00	161	9.0
		2020-07-22 10:45:00	2020-10-19 11:30:00	112	89.0
55	Bark River Downstream	2019-03-09 17:00:00	2019-04-09 12:20:00	343	30.8
		2020-07-09 19:15:00	2020-10-06 13:40:00	80	88.8
57	Menomonee River at Wauwatosa	2020-05-17 03:50:00	2020-06-15 10:20:00	226	29.3
		2020-08-02 21:30:00	2020-08-10 10:35:00	-80	7.5
		2021-04-11 22:50:00	2021-04-15 11:50:00	257	3.5
		2021-05-04 11:55:00	2021-05-12 11:20:00	702	8.0
58	Milwaukee River at Estabrook Park	2021-05-03 20:00:00	2021-05-12 11:55:00	76	8.7
59	Root River near Horlick Dam	2020-08-10 22:45:00	2020-10-08 14:05:00	266	58.6
87	Underwood Creek	2021-04-10 17:10:00	2021-04-26 16:55:00	464	16.0

[a] *Central Daylight Time (CDT) or UTC -5:00.*

[b] *The specific conductance difference is calculated by subtracting the specific conductance value immediately before the sensor was cleaned (s_f) from the specific conductance value immediately after the sensor was cleaned (s_i).*

Source: SEWRPC

SEWRPC Technical Report No. 61

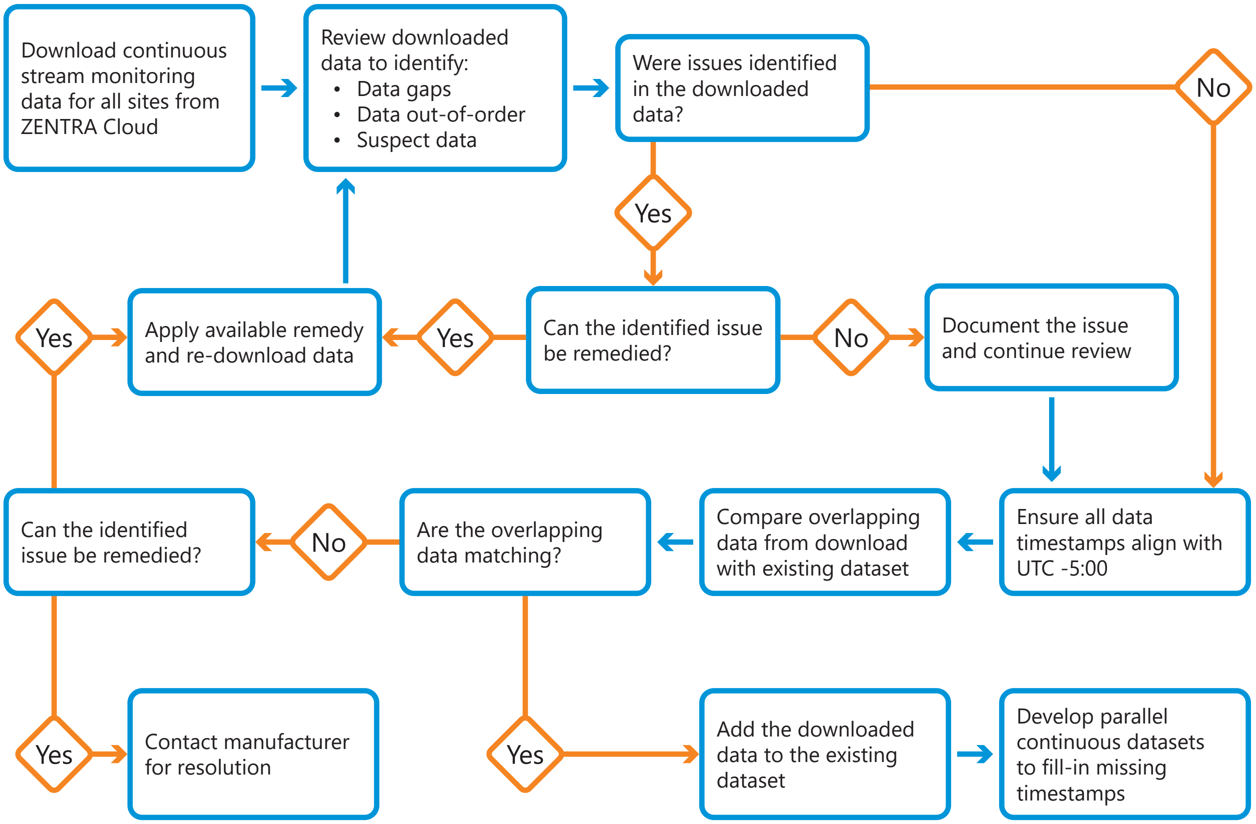
FIELD MONITORING AND DATA COLLECTION FOR THE CHLORIDE IMPACT STUDY

Chapter 4

DATA MANAGEMENT AND DOCUMENTATION

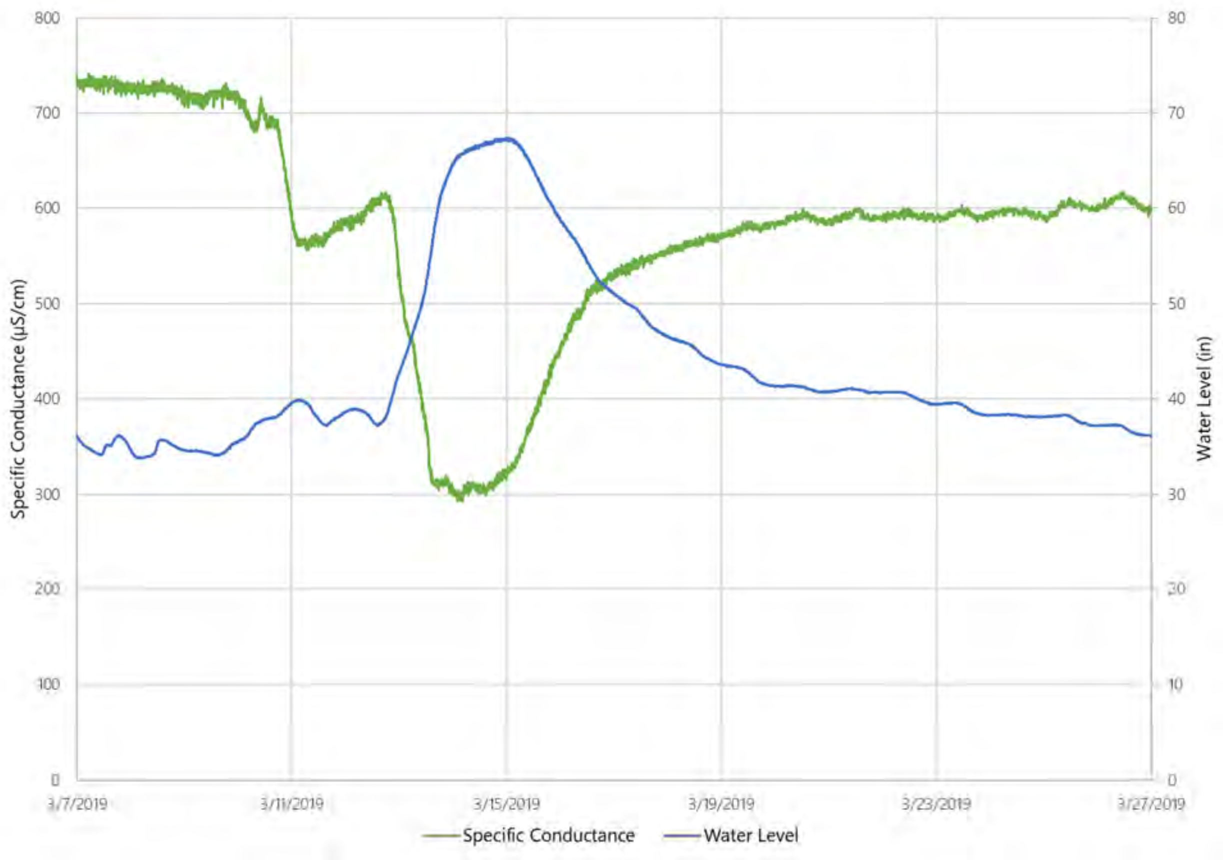
FIGURES

Figure 4.1
Continuous Stream Monitoring Data Download and Review Process



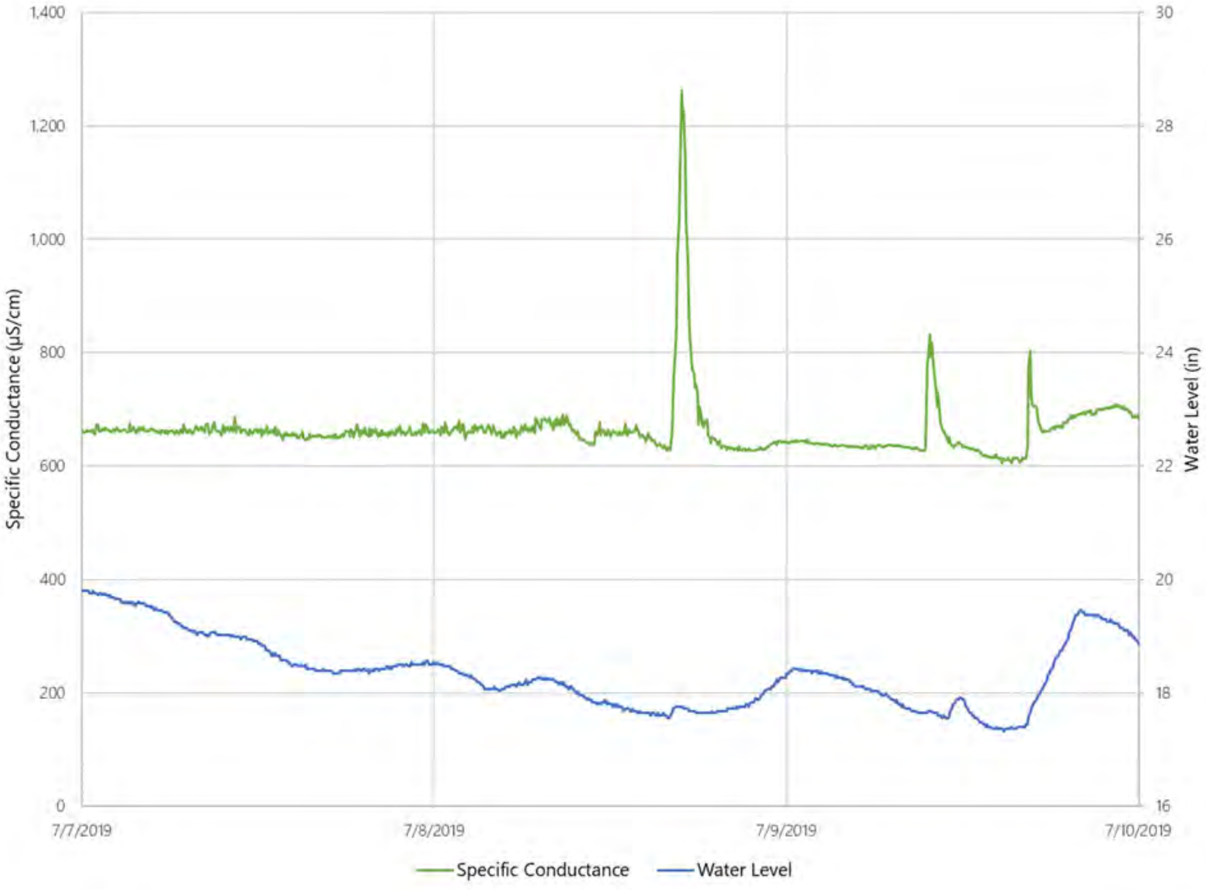
Source: SEWRPC

Figure 4.2
Stream Water Level and Specific Conductance During a Precipitation Event



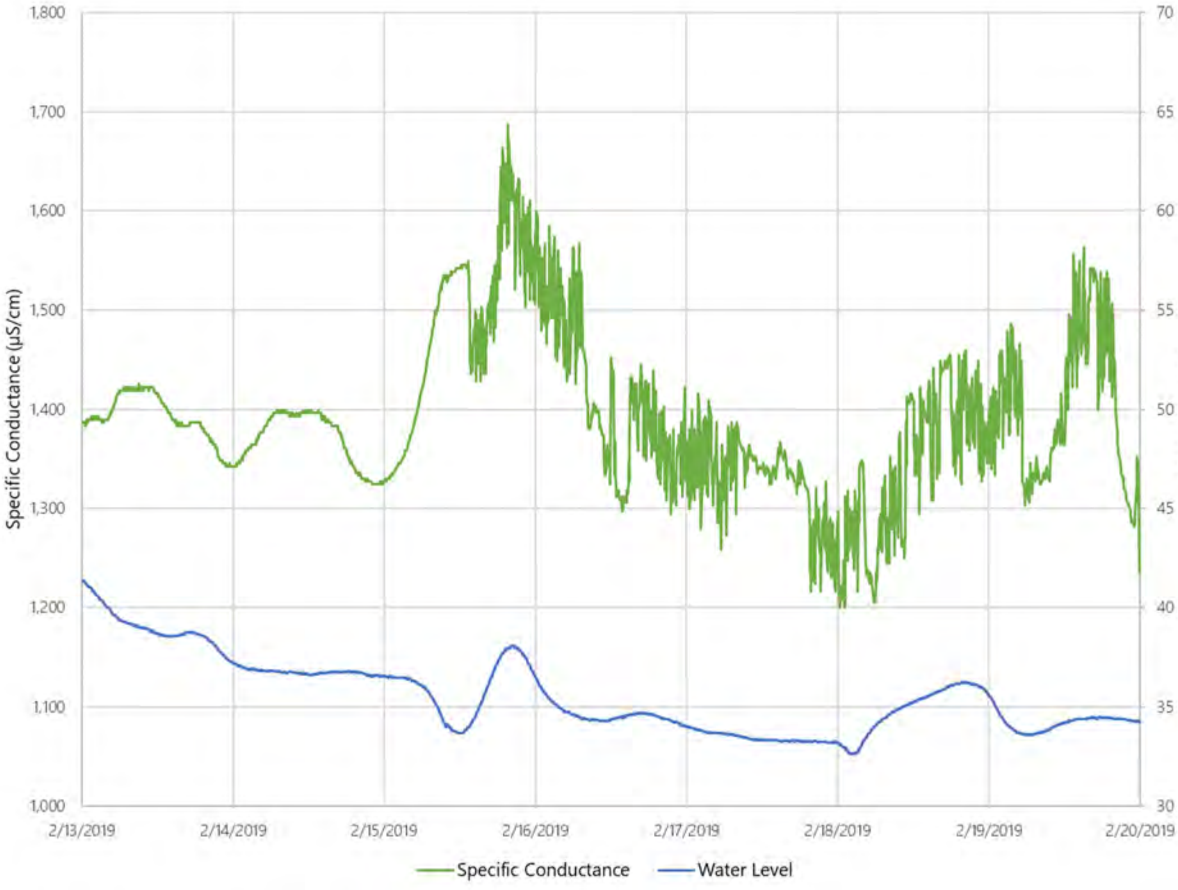
Source: SEWRPC

Figure 4.3
Example of Specific Conductance Spike Data Signature



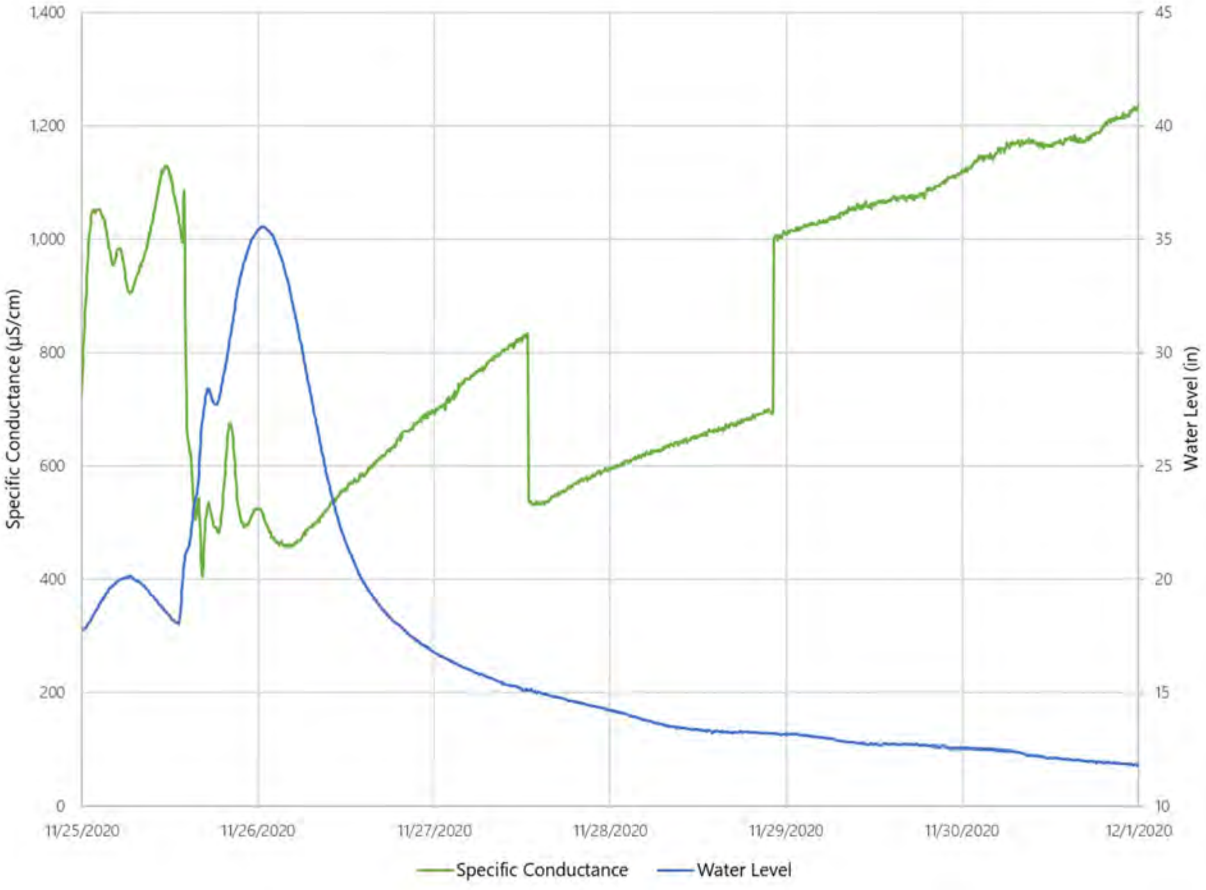
Source: SEWRPC

Figure 4.4
Example of Specific Conductance Noise Data Signature



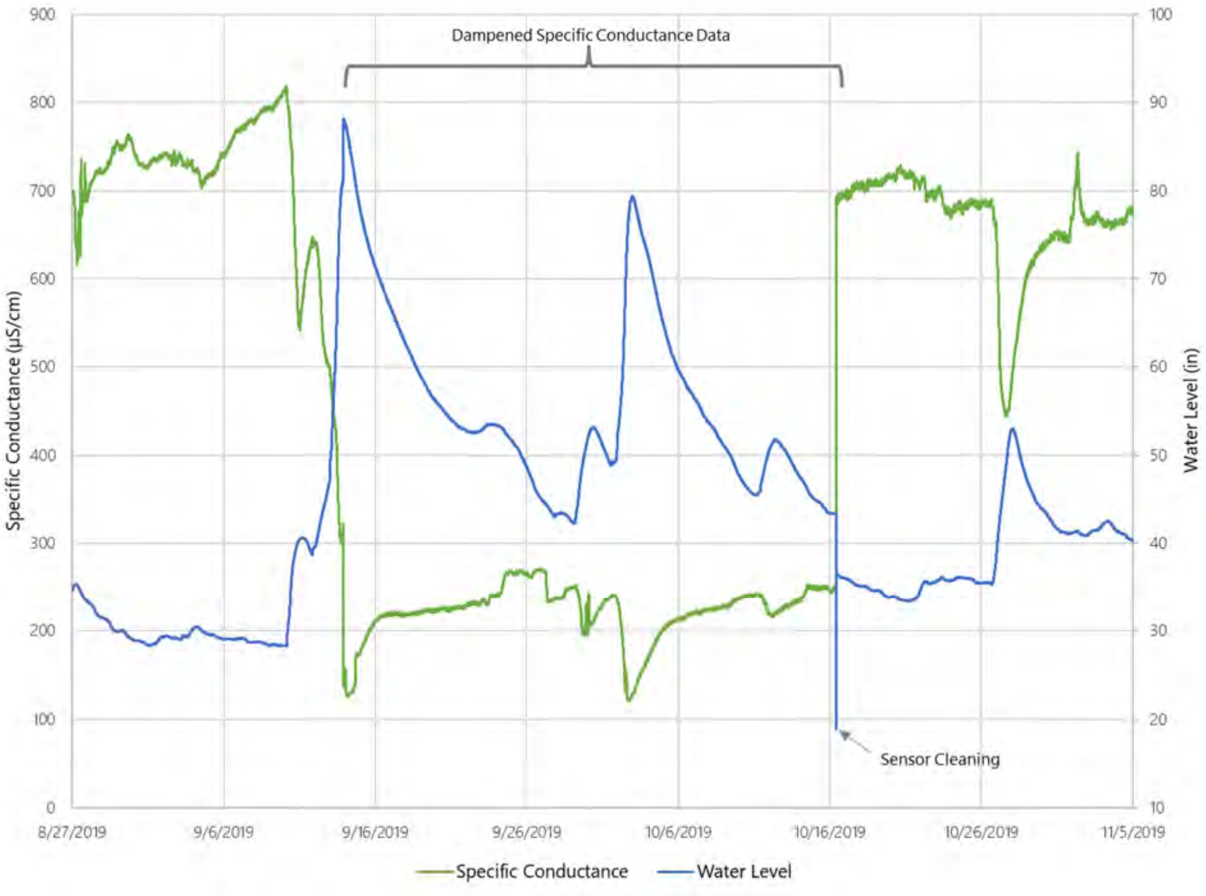
Source: SEWRPC

Figure 4.5
Example of Specific Conductance Tooth Data Signature



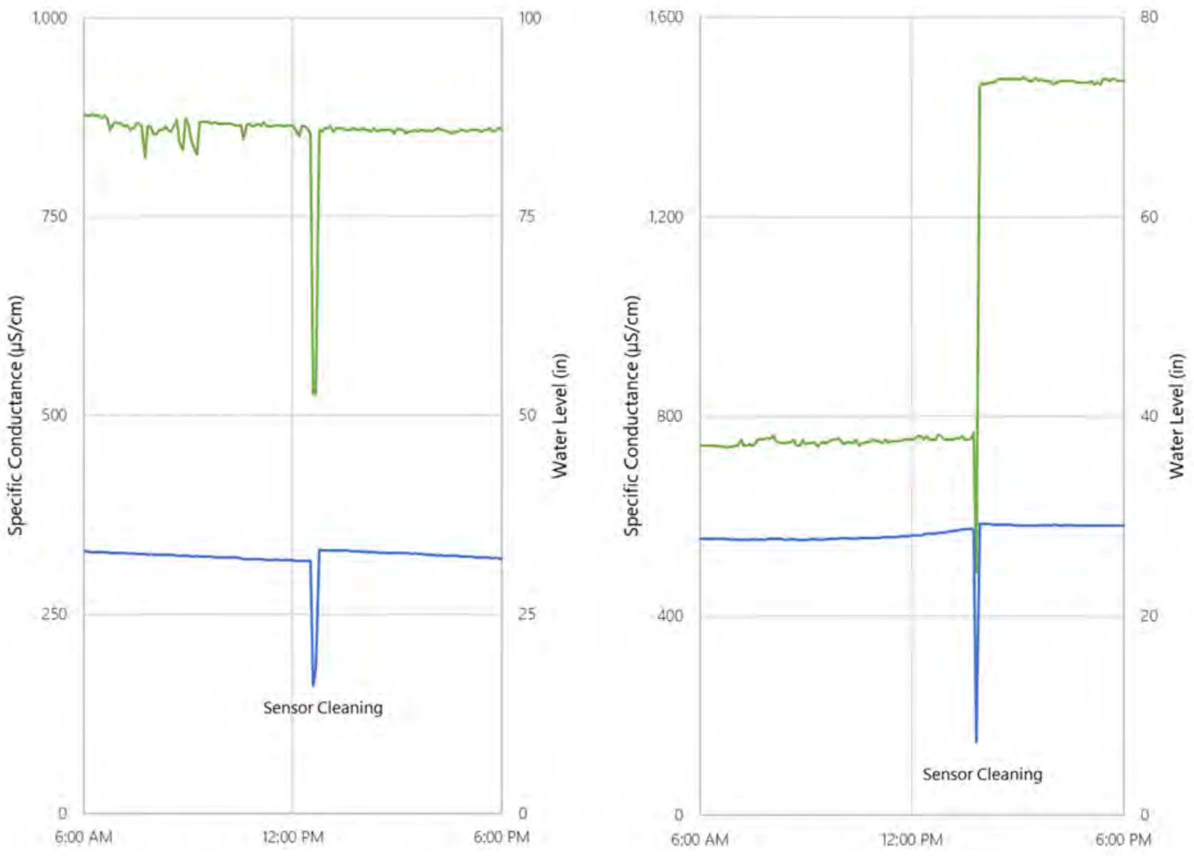
Source: SEWRPC

Figure 4.6
Example of Dampened Specific Conductance Data Signature



Source: SEWRPC

Figure 4.7
Examples of Data Signatures Associated with CTD-10 Sensor Cleanings



Small Specific Conductance Change

Large Specific Conductance Change

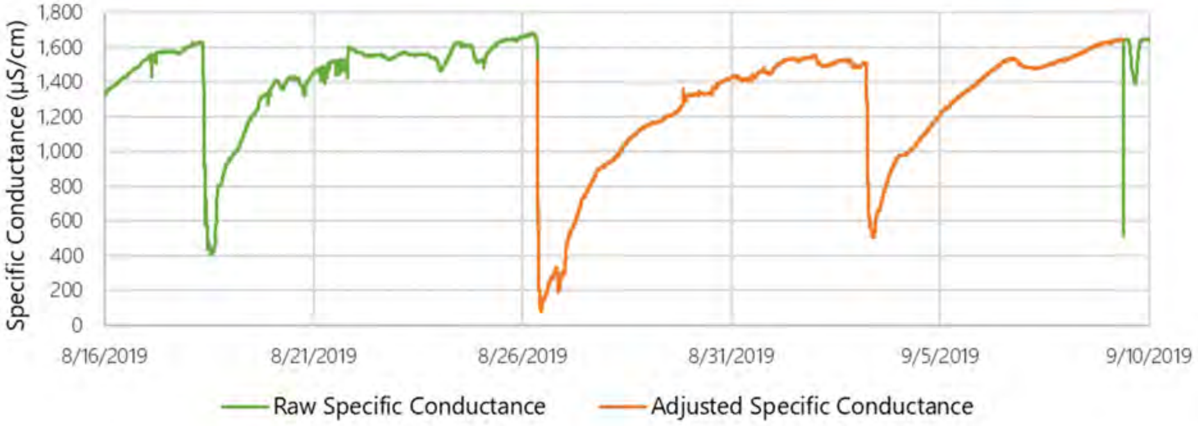
Source: SEWRPC

Figure 4.8
Specific Conductance Data Adjustment Example

a. Raw Specific Conductance Data:

b. Raw and Adjusted Specific Conductance Data:

c. Adjusted Specific Conductance Data:



Source: SEWRPC