Standard Operating Procedure for Data Verification and Validation

October 31, 2024

STI-8077

APPROVED:

_____

Sonoma Technology date

_____

Fenceline Monitoring Refinery Representative date

Contents

1. Scope and Application

This standard operating procedure (SOP) document describes the verification and validation of fenceline monitoring data, including daily checks and quarterly analyses. Additional information is provided in the corresponding air monitoring plan (AMP) and Section D.2 of the quality assurance project plan (QAPP).

2. Introduction and Overview

At all fenceline monitoring sites, a Data Acquisition System (DAS), or data logger, performs basic quality control (QC), averages data to 5-min resolution, and aggregates data into a desired file format. Data are then transmitted from each sampling site to a cloud-based file storage service via cellular modem, where they are stored and available for retrieval as needed. Data from the cloud are ingested into Sonoma Technology's Insight© data management system (DMS), where a robust automated QC (AutoQC) logic and subsequent DMS-level screening checks assign data flags in real time based on instrument diagnostics, local meteorological measurements, and concentrations of target compounds. Data are stored within the DMS and any changes to data are recorded via chain-of-custody logs. These preliminary data are displayed on the public website within 10–15 minutes of raw data collection.

Data are reviewed daily by air quality data analysts to assess system operations, confirm the automated data flagging is correct, and ensure any corrections to data flagging are propagated to the public website immediately. Extended analyses are performed every calendar quarter and reviewed by the project Quality Assurance (QA) Manager. Figure 1 illustrates the general data flow and QC schematic.

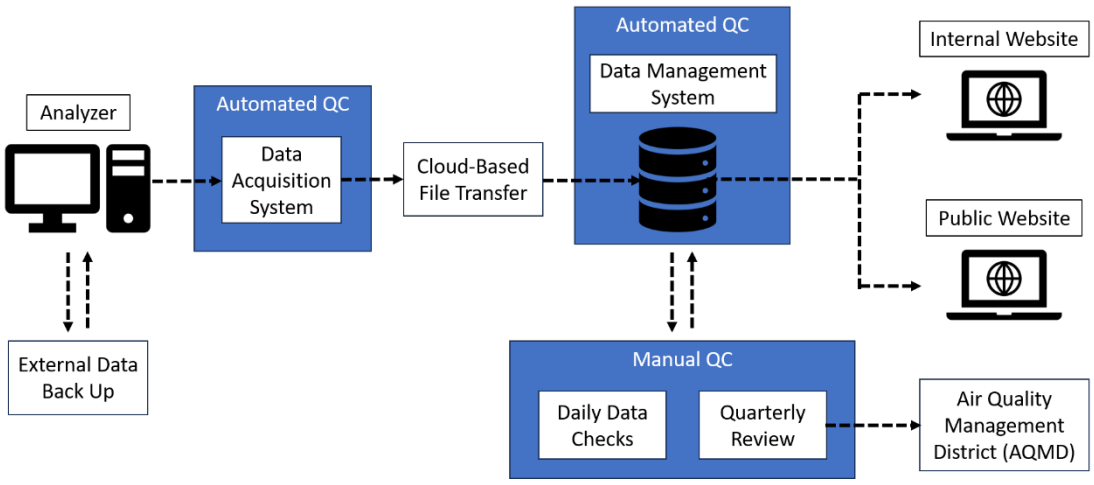

Figure 1. Data flow and QC schematic.

3. Definitions

Data verification is a process of comparing how the data were gathered with the procedures established by the project QAPP and SOPs. It is a data-review technique that evaluates the conformance of data-collection practices with established methods, procedures, or specifications. Data verification consists of checking that SOPs were followed and that QC activities were performed.[1]

Data validation is a process of confirming that reported values meet the quality objectives of the project. It is a data-review technique that examines whether the particular requirements for a specific, intended use are fulfilled. Data validation examines whether acceptance criteria outlined in the QAPP were achieved.[1]

Acceptance criteria are the defined performance criteria for the project that data must meet to be considered acceptable for reporting. Additional information regarding acceptance criteria is provided in Section A.7 of the QAPP.

Data flags are assigned to each data point based on instrument diagnostics and local meteorological measurements. Within the Sonoma Technology Insight DMS, data flags are composed of QC and operational (OP) codes. Additional information regarding data flags is provided in Section D.2 of the QAPP.

The **minimum detection limit (MDL)** is the lowest concentration of each parameter that can be measured by each analyzer at each sampling path. Additional information regarding MDLs is provided in Section B.2 of the QAPP.

Spectral data refers to the individual absorbance spectra collected at each open-path monitoring site. The analytic software evaluates raw spectral data in real time to report concentration data and can be reviewed if data-validation efforts identify potential issues that require investigation. Additional information regarding spectral data is provided in Section B.10 of the QAPP.

A **spectral match parameter** is used to identify parameters of interest and potential interferences present with open-path sampling. Spectral subtraction is used in cases with overlapping absorbance features, and the subtraction technique is proprietary to the instrument manufacturer. Additional information regarding spectral match parameters is provided in Section B.2 of the QAPP.

4. Safe Work, Hazard Identification, and Training for Data Review

Data verification and validation is performed remotely in an office environment away from field sampling sites. Common office hazards include slips and trips, sprains and strains, poor workstation

[1] EPA Quality Assurance Handbook Vol II, https://www.epa.gov/sites/default/files/2020-10/documents/final_handbook_document_1_17.pdf.

ergonomics, indoor air-quality problems, insufficient or excessive lighting, noise, and electrical hazards. These hazards are identified and mitigated through safety walkthroughs, formal reporting of unsafe conditions, and regular training sessions on correcting safety hazards.

Training for data review is conducted by senior staff with at least one year of experience with refinery fenceline monitoring systems, as well as analyzer manufacturers. Project personnel are provided copies of the QAPP and SOPs and receive updated versions when available. Initial training is provided prior to personnel performing work, and refresher trainings are conducted on an annual basis. The QA Manager will identify specific training requirements for all project personnel and will determine when trainees are qualified to work independently. Training records will be maintained by the Field Staff Manager.

5. Data Review and Quality Control

Data for the fenceline monitoring network appear on both public and internal websites (Figure 1). The internal website allows for detailed QC and flagging of the data, which are checked daily and finalized quarterly as outlined in Section D.2 of the QAPP. The following section describes how to perform daily and quarterly data verification and validation.

5.1 Daily Checks

Data are manually reviewed on a daily basis by qualified analysts. An internal, non-public field operations website (Sonoma Technology's Insight DMS) is used for customized data queries and visualizations, and the public-facing website is reviewed to ensure the real-time display is current and accurate. Instrument performance issues or data gaps are escalated to the field operations team, and the Project Manager determines the appropriate actions to resolve issues in a timely manner.

Data analysts are qualified to perform daily checks after a training period of at least two weeks. Training includes education on monitoring rules and requirements, instrumentation, automated QC processing, expected concentration ranges, diagnostic thresholds, and internal procedures to escalate a data issue. Trainees shadow an experienced data analyst during daily checks for a week and are then overseen by an experienced analyst while they perform their first week of data checks.

This daily data review allows the analyst to visualize fenceline concentrations and instrument diagnostics, thereby identifying potential instrument performance issues and confirming the current operational status of the network. Instances of missing or invalid data, instrument signal strength below defined thresholds, repeated (stuck) data values, rapid changes (spikes or dips) in data values, or negative concentrations are recorded in a spreadsheet that tracks observations across a calendar quarter. This form provides a template for data analysts to check relevant parameters against acceptance criteria and commonly expected values for routine reference. Daily data check spreadsheets from past quarters are maintained in an archive with other project documentation.

The following procedure details the daily data check for an individual monitoring network:

1. Log in to the Sonoma Technology Insight DMS with the facility-specific credentials for the check being performed. Credentials are stored in secure password management software.
2. Navigate to the Dashboard, which contains configurable widgets showing time-series plots of relevant parameters.
3. Record any periods of **missing data** and verify that (1) missing data alerts were sent and (2) resolution is being pursued by the field operations team. If this protocol cannot be verified, escalate the finding to the Project Manager.
4. Review any **detection events** and confirm the validity determination made by the after-hours on-call team through review of instrument diagnostic parameters and spectral validation (see Section 5.3 of this SOP).
5. Review recent **visibility measurements** and note any periods with visibility of under 2.5 miles on the spreadsheet.
    a. Periods with visibility of under 2.5 miles may impact performance of open-path analyzers and often correlate with observations of invalid data at multiple sampling paths.
6. Check that **meteorological instrumentation** is functional and that measured values are within reason (e.g., directional measurements are between 0 and 360 degrees and speeds generally match local weather conditions). Record any instances of missing or abnormal meteorological measurements in the daily observation spreadsheet and escalate to the field operations team.
7. Compare the current **open-path analyzer diagnostic parameters** with the acceptance criteria (see Section A.7 of the QAPP) to assess operational status of the monitoring network.
    a. Periods with diagnostic parameters (e.g., signal strength, integration time) outside of acceptance criteria will result in invalid data and may result from open-path analyzer misalignment. Record these periods in the spreadsheet and escalate to the field operations team for necessary site visits.
8. Because data flagging for each compound is determined by the analyzer diagnostic parameters, review one **representative compound concentration time series** for each analyzer at each sampling site and note the following information:
    a. Any periods of missing or invalid data.
    b. Any periods of data flagged for additional review (e.g., automated checks for range, rate-of-change, and sticking).
    c. For periods where data were flagged as invalid due to environmental conditions, confirm that (1) analyzer diagnostic parameters are outside of acceptance criteria, and (2) visibility is under 2.5 miles.

d. Any QC and OP codes that are inconsistent with the expected values, based on analyzer diagnostic parameters.

e. Instances of apparent baseline drift.

f. Reported concentrations (valid data) not within the specified range for each compound.

    i. Verify that concentration patterns are reasonable with respect to the time of day, season, and current meteorological conditions, and whether they correlate across multiple sampling sites. This check is for information purposes only and should not itself result in invalidation of data.

    ii. Verify that concentrations are within the limits of what can be measured by the instrument.

9. Edit the QC and OP codes to reflect the findings of the daily data check. Additional information regarding propagation of these changes to the public website is provided in Section D.3 of the QAPP.

    a. Changes to **concentration values** are extremely infrequent and should only be made if a problem with the analytical concentration determination or the averaging scheme is discovered. Changes to QC and OP codes are often sufficient to correctly reflect data quality.

    b. Changes to **QC and OP codes** are infrequent and are only made if a clear reason to do so is discovered.

        i. For example, invalidate data if a violation of the acceptance criteria that was not captured by AutoQC logic is discovered.

        ii. Similarly, validate data that were flagged for additional review if they meet acceptance criteria.

    c. All edits to data and data quality flags are captured by chain-of-custody logs within the DMS.

10. Verify detection events with elevated concentrations (>5-10 times the MDL) by performing spectral validation according to Section 5.3 of this SOP.

11. Navigate to the public website for the facility, review the 5-min time series plot for each individual compound, and verify that:

    a. Data displayed on the public site match the data seen on the non-public data review website (Sonoma Technology's Insight DMS). Note any discrepancies in the spreadsheet and escalate to the Project Manager.

    b. Periods of missing or invalid data are correctly reflected on the public website.

    c. Detection events are correctly reflected on the public website, and suspect/questionable data flags are not shown for legitimate detection events.

12. Report any observed anomalies to the field operations team and Project Manager and update the daily data check spreadsheet in the project documentation archive. Additional information regarding project documentation is provided in Section A.9 of the QAPP.

Typical observations for select diagnostic parameters and compounds monitored under BAAQMD Rule 12-15 are listed in Table 1. The full list of measured compounds is provided in Section A.5 of the QAPP.

Table 1. Typical observations for select compounds monitored under BAAQMD Rule 12-15.

Parameter	Observational Notes
Visibility	The maximum value measured by the sensor is 30 miles. Values under 10 miles may affect instrument performance, and values under 2.5 miles may result in invalid data due to environmental conditions if analyzer diagnostics are also not within acceptance criteria.
Integration Time (UV-DOAS)	This parameter is commonly between 20–250 ms for most sampling sites. Values stuck at 300 ms should be recorded and escalated to the field operations team. A pattern of elevated integration time and low signal strength can indicate poor instrument alignment. This parameter should be anticorrelated with visibility during fog events.
Signal Strength (UV-DOAS)	This parameter is commonly above 70%. A pattern of elevated integration time and low signal strength can indicate poor instrument alignment.
Signal Power (TDLAS)	This parameter is commonly above 0.1 mW. A pattern of low signal power can indicate poor instrument alignment.
Signal Strength (FTIR)	This parameter is commonly above 2%. A pattern of low signal strength can indicate poor instrument alignment.
Winds	Wind speeds are typically 1-20 mph, and wind direction should generally follow expected patterns due to local terrain (e.g., sea and land breezes along the coast).
Benzene	This parameter is typically below the detection limit of the instrument.
Toluene	This parameter is typically below the detection limit of the instrument.
Ethylbenzene	This parameter is typically below the detection limit of the instrument.
o-xylene	This parameter is typically below the detection limit of the instrument.
m-xylene	This parameter is typically below the detection limit of the instrument.
p-xylene	This parameter is typically below the detection limit of the instrument.

Parameter	Observational Notes
SO_2	This parameter is typically below the detection limit of the instrument, but ambient background concentrations are detected occasionally.
H_2S	This parameter is typically below the detection limit of the instrument.

5.2 Quarterly Validation

Quarterly validation is performed after each calendar quarter to ensure quality and accordance with Rule 12-15 requirements. The 5-min-resolution data for all relevant parameters are exported from the Sonoma Technology Insight DMS and imported into an R analysis workbook. Here, analysts perform additional analyses, visualizations, and assessments to verify AutoQC, confirm determinations made during daily data checks, and perform completeness calculations (see Section A.7 of the QAPP).

Data analysts are qualified to perform quarterly analysis after a training period of at least three months. Training includes the daily data check steps outlined in Section 5.1, experience conducting daily data checks for a period of three months, and shadowing an experienced data analyst during at least one complete quarterly analysis process. The trainee's work is reviewed by the experienced analyst while they perform their first round of quarterly analysis, and they are released for independent work during their second round.

The following procedure details the quarterly analysis process for an individual monitoring network:

1. Log in to the Sonoma Technology Insight DMS with the facility-specific credentials for the check being performed. Credentials are stored in a secure password management software.
2. Navigate to the Data Export page, where configurable export queries can be programmed and executed. Export all relevant parameters collected during the quarter as .csv files.
3. Import these exported data files into an R workspace using the most recent quarterly analysis script(s), maintained in the Sonoma Technology Bitbucket code repository.
4. Execute the quarterly analysis script(s), which performs the following tasks:
    a. Assign data flags using the same AutoQC logic as real-time data flagging (see Sections B.10 and D.2 of the QAPP).
    b. Flag negative outliers as invalid, using "–3 x MDL" as the threshold for comparison.
    c. Generate a series of output tables, figures, and .csv files for additional review.
5. Review the quarterly data set in detail. Note that findings pertaining to missing or invalid data are usually identified and corrected during the quarter, meaning the quarterly analysis process is commonly a redundant check on previously identified issues and resolutions. Any

new findings identified during quarterly analysis must be escalated to the Project Manager for investigation, and proposed resolutions must be approved by the QA Manager.

a. If more than 5% of diagnostic parameters (signal strength, integration time, spectral match) are missing, escalate the finding to the Project Manager for additional investigation.

b. If more than 25% of missing or invalid data points are associated with one analyzer or sampling site, escalate the finding to the Project Manager for additional investigation.

c. If concentration data were collected but coincident diagnostic information was missing or outside of the expected range, AutoQC determinations are manually inspected to ensure accurate data flagging.

    i. If visibility data are missing during a period of invalid concentration data, manual data review confirms that concentration data **are not** flagged as "invalid due to environmental conditions." In the absence of visibility data to confirm low-visibility conditions, data count against completeness (see Section A.7 of the QAPP).

    ii. If instrument diagnostic parameters are missing but concentration data are reported, a malfunction of the analyzer software (which reports both fields simultaneously) has occurred and data are flagged as invalid.

    iii. If spectral match parameters are reported outside of the expected range (e.g., below 0 or greater than 100 for percent match), a malfunction of the data acquisition script has occurred and data are flagged as invalid.

d. Review periods where visibility is under 2.5 miles, ensuring that **only** concentration data with diagnostic parameters outside of acceptance criteria are flagged as "invalid due to environmental conditions" and are not counted against completeness.

e. Review periods of invalid or missing data lasting longer than two hours and escalate findings to the Project Manager. Create null records in the place of missing data according to Section C.2 of the QAPP.

f. Review periods of planned and unplanned maintenance and compare them with information in field logbooks and maintenance records. Ensure that routine maintenance was performed as scheduled, logs reflect a sufficient level of detail, and all bump tests results are within acceptance criteria.

    i. Any failed bump tests should have been followed by additional maintenance and retesting as outlined in the individual instrument SOPs.

    ii. Maintenance logs may explain data anomalies and justify adjustments to QC and OP codes, following discussion with the Project Manager.

g. If applicable, review any suspect/questionable data for which investigation was not performed and QC and OP codes were not adjusted during the daily data checks. Adjust QC and OP codes according to Section D.2 of the QAPP and escalate the finding(s) to the Project Manager.

h. Review any and all changes to QC and OP codes, including their justification, with the Project Manager.

6. Assess the data reasonableness through statistical analyses, review of exceptional conditions off-refinery, and comparisons with outside data.

    a. For all sampling sites and parameters, calculate the monthly and quarterly 5-min concentration mean, standard deviation, and minimum and maximum values. Nondetections are assigned a value of "0 ppb" for statistical analysis.

        i. Compare observations with previous quarters and typical observations (Table 1) and escalate any anomalies to the Project Manager.

        ii. Confirm that known detection events observed during the quarter are correctly reflected in the calculated monthly and quarterly maxima.

        iii. Verify that the calculated monthly and quarterly minima are > "–3 x MDL," and escalate the finding of any negative outliers to the Project Manager.

    b. In the event of anomalous data, review exceptional conditions off-refinery in conjunction with meteorological data to assess potential impacts to fenceline measurements. Data should not be invalidated based on the potential emission source being identified as off refinery property.

    c. In the event of anomalous data, retrieve available outside data sources for comparison according to Section B.9 of the QAPP.

7. Review any detection events and periods where reported concentrations are greater than 5-10 times the MDL, according to Section 5.3 of this SOP.

8. For all sites and compounds, calculate data completeness according to Section A.7 of the QAPP.

    a. Review the resulting completeness tables and compare against logbooks and maintenance logs.

    b. Escalate findings of data completeness less than 95% to the Project Manager for additional investigation.

9. For all sites and compounds, manually calculate the rolling hourly average data from quality-controlled 5-min data, following the same logic used in real time by the public website.

    a. Investigate the cause of any discrepancies between hourly values reported on the website and hourly values calculated during quarterly analysis and escalate findings to the Project Manager.

5.3 Spectral Validation

Measurements made in real time by open-path analyzers can be replicated with manufacturer-provided data-processing software. This process enables analysts to visualize and confirm each real-time analytic determination made by the analyzer during daily and quarterly analysis to ensure data accuracy and quality.

The following procedure details the spectral validation process:

1. Retrieve the relevant (raw) data files from the analyzer of interest, which correspond to the event requiring spectral validation.
    a. These files consist of:
        i. The daily data summary file corresponding to the day of interest.
        ii. Individual absorbance spectra (including applicable background spectra) corresponding to the time period of interest.
        iii. Instrument libraries from the analyzer of interest.
        iv. The analytic software configuration file from the analyzer of interest.
2. Open the manufacturer-provided data-processing software and adjust the analysis parameters to match the configuration file retrieved from the analyzer of interest.
3. Where applicable, load the instrument libraries and verify that the available compounds match the expected values for the monitoring network.
4. Since this is a differential measurement, concentrations are reported by comparing observations in detection to observations of clean air. Examine the daily data summary file from the date and time period of interest to determine which individual absorbance spectra are of interest and, where applicable, which dynamic background was loaded for each absorbance spectrum.
5. Load the applicable background and absorbance spectra. The software will automatically duplicate the analysis, according to the configuration parameters (see Step 2 above), and generate a differential absorbance spectrum overlaid with the applicable instrument library.
6. Select the compound of interest and visually inspect the plot, confirming the curve fitting to the library reference spectrum, and associated spectral match determination relative to the programmed analysis regions.
7. Report any discrepancies or incorrect peak fits to the Project Manager for additional investigation.

5.4 Quality Assurance Management

Following the completion of quarterly analysis, an independent review of post-QC quarterly data sets is completed by the QA Analyst. This independent review is conducted on a representative subset of data using similar methods to those described in Section 5.2 of this SOP. Data are reviewed without input from the data analyst who prepared the post-QC quarterly data set.

Findings of the independent review are compared with those of the quarterly analysis in a joint meeting between the Data Analyst, QA Analyst, and QA Manager. The QA Manager conducts an informal interview process to ensure that quarterly analysis was conducted according to the QAPP and that field documentation reflects the procedures outlined in the QAPP. New findings or deviations are reported to the Project Manager by the QA Manager for additional investigation, and any necessary resolutions must be approved by the QA Manager before quarterly data are submitted (see Section C.2 of the QAPP).

6. Public Website Display

Data are displayed on the public website according to the data flags (QC and OP codes) for each record. Data flags are assigned by the AutoQC logic and further evaluated according to Section D.2 of the QAPP. Display behavior is the same for both 5-min data and rolling hourly average data.

Valid detections (e.g., QC=0, OP=0) are shown on the website (**Figure 2**). Concentrations are reported as recorded by the analyzer, with numeric values shown in the time series detail panel (left), the map detail panel (middle), and the map marker (right). For valid detections with concentrations below the MDL (e.g., QC=0, OP=17), numeric values are shown but data are flagged as "<MDL." For hourly data only, the outer ring of the map marker visualizes the concentration relative to an hourly threshold, with green indicating values below the threshold and orange indicating values above the threshold.

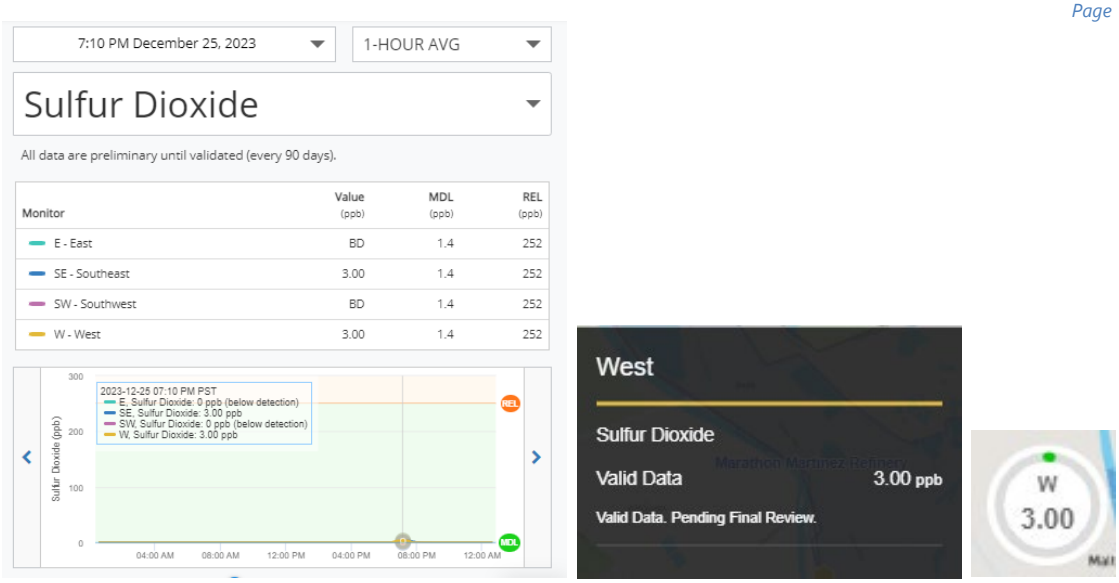

Figure 2: Example of valid detections as displayed on the public website.

Valid nondetections (e.g., QC=0, OP=74) are shown on the website (**Figure 3**). Concentrations are reported as "0 ppb" on the time series detail panel (left) according to Section D.2 of the QAPP. On the map detail panel (middle) and map marker (right), valid nondetections are flagged as "<MDL."

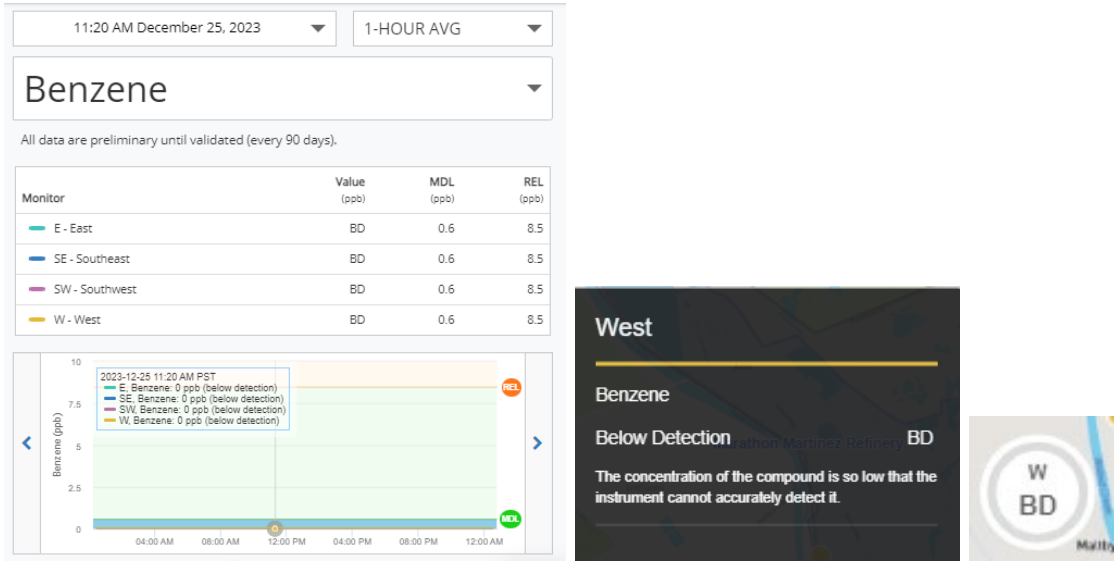

Figure 3: Example of valid nondetections as displayed on the public website.

Suspect/Questionable data (e.g., QC=5) are shown on the website (**Figure 4**). Concentration values are reported on the time series detail panel (left), either as recorded (e.g., OP=72) or as "0 ppb" (e.g., OP=76), depending on whether they correspond to detection events. The map detail panel (middle) and map marker (right) do not display values and indicate that further review is needed. This additional review is completed according to Section D.2 of the QAPP.

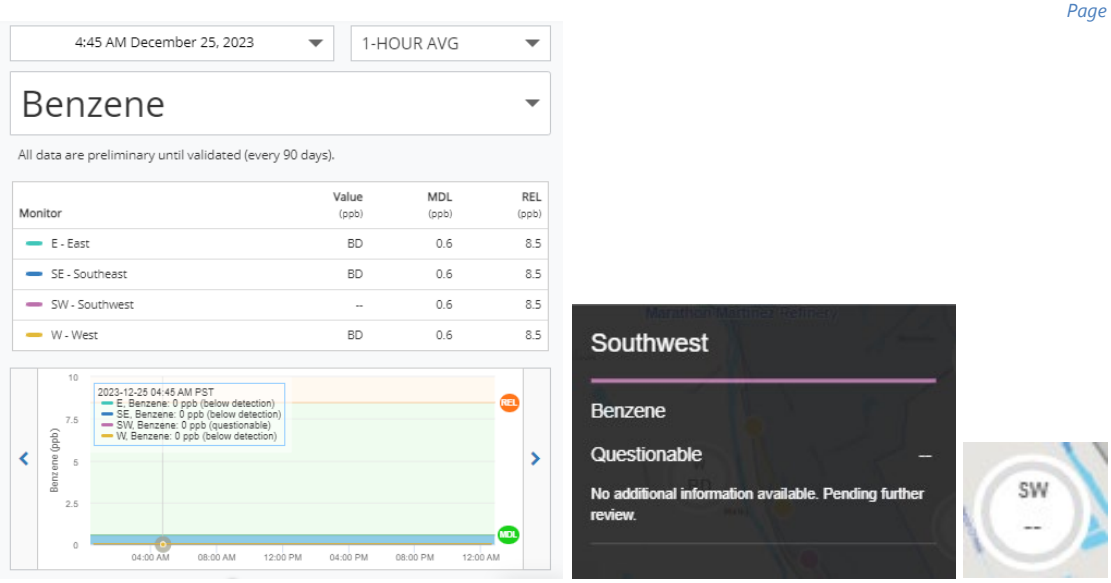

Figure 4: Example of questionable data as displayed on the public website.

Invalid data (e.g., QC=9) are not shown on the website (Figure 5). Concentration values are not reported on the time series detail panel (left), the map detail panel (middle), or the map marker (right) because the data do not meet acceptance criteria.

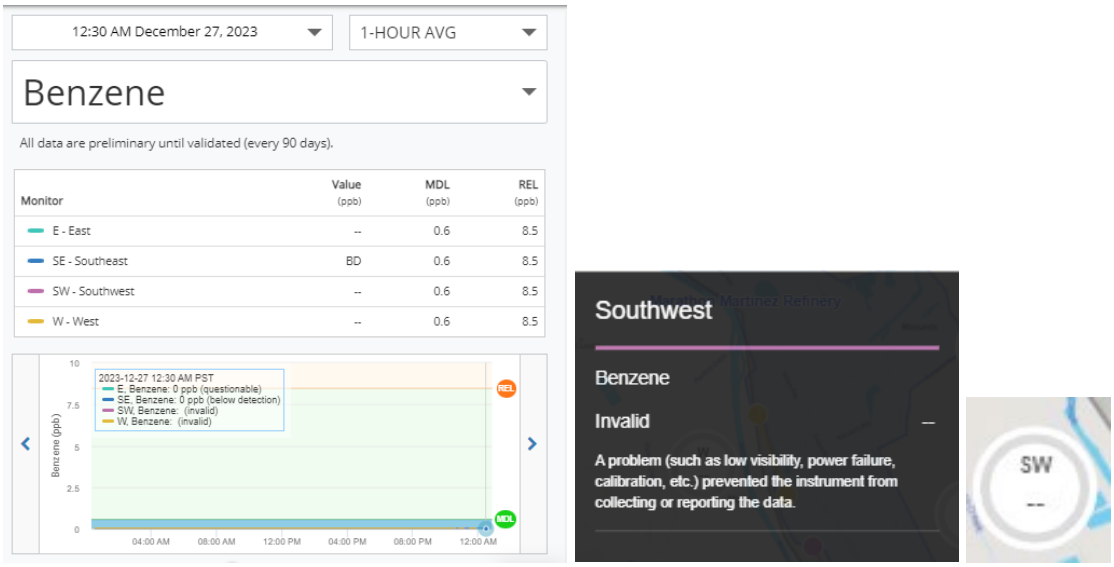

Figure 5: Example of invalid data as displayed on the public website.

Missing data (e.g., QC=8) are not shown on the website (Figure 6). Concentration values are not reported on the time series detail panel (left), the map detail panel (middle), or the map marker (right) because data were not collected. Missing data result in internal alerts to the project team (see Section B.6 of the QAPP).

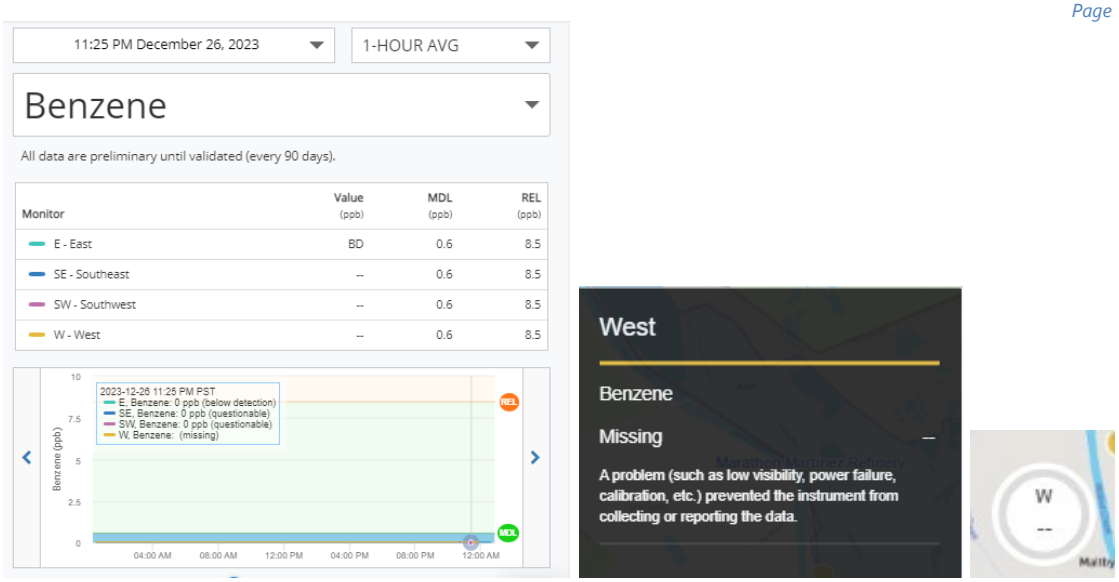

Figure 6: Example of missing data as displayed on the public website.

7. Daily Data Check Spreadsheet

Complete the checks described below and record observations in the space provided. Note any corrective action or issues that were escalated to the field operations team or Project Manager. Additional information regarding acceptance criteria is provided in Section A.7 of the QAPP.

Data Analyst: _____ Date and Time: _____

Parameter	Reference	Observations
Internal Website Review (5-min data)		
UV-DOAS		
Signal Intensity	Confirm that values ≥70%	
Integration Time	Confirm that values ≤250 ms	
Representative Parameter (e.g., benzene)	Record any instances of invalid data	
	Record any instances of missing data	
	Record maximum concentration	
TDLAS		
Signal Power	Confirm that values ≥0.1 mW	
H_2O Spectral Match	Confirm that values >0.95	
Real-Time MDL (rtMDL)	Record any instances of rtMDL >25 ppb	
Representative Parameter (e.g., H_2S)	Record any instances of invalid data	
	Record any instances of missing data	
	Record maximum concentration	
FTIR		
Signal Intensity	Confirm that values ≥70%	
Representative Parameter (e.g., hexane)	Record any instances of invalid data	
	Record any instances of missing data	
	Record maximum concentration	
Met		
Visibility	Confirm that values >2.5 mi	
Wind Direction	Confirm that values are reasonable (0-360 degrees)	
Wind Speed	Confirm that values are within typical values for this site	
Public Website Review (5-min data)		
UV-DOAS	Record any instances of invalid data	
	Record any instances of missing data	
	Record any benzene detection events	
	Record any toluene detection events	
	Record any ethylbenzene detection events	
	Record any m-xylene detection events	
	Record any o-xylene detection events	
	Record any p-xylene detection events	
	Record any SO_2 detection events	

Public Website Review (5-min data)		
TDLAS	Record any instances of invalid data	
	Record any instances of missing data	
	Record any H_2S detection events	
FTIR	Record any instances of invalid data	
	Record any instances of missing data	
	Record any ammonia detection events	
	Record any 1,3-butadiene detection events	
	Record any hexane detection events	
All Compounds	Record any instances of missing data	