Technical Memorandum

Date: July 1, 2023

To: Demian Ebert, PacifiCorp

From: Jennifer Vaughn, Watercourse Engineering, Inc.
Mike Deas, Watercourse Engineering, Inc.

Re: Quality Assurance Review of KHSA 2022 dataset

Suggested Citation: Vaughn, J., M. Deas. 2023. Technical Memorandum Re: QA Review of KHSA 2022 dataset. July 1. 4 pp.

1. Introduction

A quality assurance (QA) review of the Klamath Hydroelectric Settlement Agreement (KHSA) 2022 dataset was conducted by Watercourse Engineering, Inc. (Watercourse) by comparing the dataset values associated with randomly selected Sample IDs with the values of those Sample ID results in the original lab reports. Sample IDs were selected from the Baseline Monitoring Program (Baseline) general dataset, the Baseline algae species dataset, and the Public Health Monitoring Program (Public Health) dataset. The review was accomplished with the cooperation of field crews and stakeholders involved in the KHSA sampling programs, including PacifiCorp, Karuk Tribe, Yurok Tribe, and E&S Environmental Chemistry, Inc.

2. Review Methods

The 2022 KHSA data review followed similar methods to the 2009-2021 KHSA dataset reviews. For each program, one percent of the Sample ID values in the 2022 KHSA dataset were compared with values in the original lab reports provided to stakeholders by their respective labs during the sampling programs. To select samples for review, the number of Sample IDs in 2022 was counted and each Sample ID was assigned an integer as a reference number. One percent of the total number of Sample IDs in 2022 was then calculated and rounded to the nearest integer. Using the Random function in MS Excel (RANDBETWEEN), a random value was generated between one and the total number of Sample IDs for 2022. A random value was generated again until the total number of random integers equaled one percent of the total number of Sample IDs in 2022. The random values were then used to choose random Sample IDs, selecting those Sample IDs with integer reference numbers matching the randomly generated values.

Once the Baseline general dataset, Baseline algae species dataset, and Public Health dataset had a sufficient number of randomly selected Sample IDs to total one percent of the total number of Sample IDs for each dataset, the appropriate stakeholders and field crews were contacted, and all original lab reports associated with the randomly selected Sample IDs were requested. Original lab reports included portable document format (pdf) versions of paper or digital lab reports and MS Excel files of paper or digital lab data. Sample IDs and their associated lab reports were documented by Watercourse.

When the original lab reports were obtained, a comparison of each possible result for each Sample ID was conducted. In the Baseline general dataset, there were a total of 19 constituents possible for each Sample ID, and in the Public Health dataset, there were a total of 30 possible results. Though there was no algae species data collected in 2022 for the Public Health sampling program, the dataset still contained algae species fields to be consistent with previous datasets; therefore, those fields still existed and needed to contain null values. Each of these possible null values were counted as possible results. This is consistent with the method of including the null values for Biological Oxygen Demand (BOD) and Volatile Suspended Solids (VSS) in the Baseline general dataset, as those constituents were no longer collected, but those fields were still in place to be consistent with historical datasets. In the Baseline algae species dataset, the number of possible comparisons was calculated as the number of fields in the spreadsheet original lab reports (15) multiplied by the total number of species identified for all selected Sample IDs. Each possible result was examined to determine if the KHSA dataset value matched the original lab report value. Regardless of dataset, if a Sample ID did not have a value for a specific constituent within the dataset, Watercourse confirmed that constituent had not been analyzed by any lab associated with that Sample ID using lab reports obtained from the appropriate stakeholders or field crew.

If the result in the KHSA dataset did not match the result in the original lab reports, that result was flagged. If the values did not match, the error was labeled a “true” error. If the value was rounded from the original lab report result, the error was labeled as a “significant figure issue.” Error rates were calculated as the percentage of results that did not match the original lab report values. Error rates were calculated for: (a) all non-matches; (b) only true errors; and (c) only significant figure issues. The accuracy of the datasets for the Baseline general dataset, Baseline algae species dataset, and Public Health programs were calculated as the percentage of results that matched the original lab report values, ignoring significant figure issues (i.e., only true errors).

3. Types of Errors and Issues

There are several types of true errors that can occur within a large dataset, but the most common two are transcription errors and omission errors. Transcription errors occur when a value entered into a dataset clearly did not match the value in the original lab report. Omission errors occur when a value was not included in the dataset that should have been. Other errors in the dataset can include a result added to the dataset that should have been omitted, or a value assigned to a wrong Sample ID.

The issue of significant figures was also included in the QA review of the KHSA dataset. At this time, there has been no decision on the number of significant figures that should be included in the KHSA dataset for each constituent. For this review, if the dataset value was rounded from the original lab-reported value, that value was flagged.

4. Possible Sources of Errors

As a further investigation into the dataset accuracy, Watercourse investigated the possible source of error or significant figure issue for each flagged value. This was done by using the compilation spreadsheets and other files provided to Watercourse by stakeholders that had been used to create the annual KHSA datasets and reports. Because of this, original lab reports could be compared to the KHSA dataset files to determine when an error or significant figure issue most likely had been introduced into the dataset. The number and percentages of each of those possible sources of each error was calculated and documented.

5. Review Results

For the 2022 Baseline general dataset of 295 Sample IDs, three Sample IDs were randomly selected. However, the two Sample IDs selected from the Baseline algae species dataset also had general data associated with them, and therefore a total of five Sample IDs were reviewed for the Baseline general dataset. Each value of the 19 possible constituents for the four Sample IDs were reviewed, resulting in the assessment of 95 values (Table 1~~Table 1~~). The 152 sample IDs included in the Baseline algae species dataset required two randomly selected Sample ID for review. However, two of the Sample IDs selected from the Baseline general dataset also had algae data associated with them, and therefore a total of four Sample IDs were assessed for the Baseline algae species dataset. Each of the 15 possible constituents for each of the species identified for each sample were reviewed, resulting in the assessment of 1,335 values for the Baseline algae species dataset (Table 1~~Table 1~~). The 2022 Public Health dataset of 202 Sample IDs required two randomly selected Sample IDs. Including the null fields for algae species that were included to be consistent with historical datasets, a total of 60 values were assessed for the Public Health dataset, as the dataset consisted of 30 possible constituents for each of the two randomly selected Sample IDs (Table 1~~Table 1~~).

The QA review results found no errors in the Baseline general dataset, Baseline algae species dataset, or the Public Health dataset, resulting in an accuracy of 100 percent for all three 2022 datasets assessed. However, there were sixteen significant figure issues found in this review of the 2022 datasets. All sixteen significant figure issues were in the Baseline general dataset and were introduced by the sampling entity.

Table 1. KHSA 2022 Possible Results Reviewed by Program.

Program Part	Baseline – General	Baseline – Algae Species	Public Health
Number of Constituents	19	15	30
Number of Sample IDs	295	152	202
Number of Reviewed Sample IDs	5	4	2
Number of possible results examined during QA review	95	1335	60

Table 2. KHSA 2022 Accuracy Estimates.

	Baseline – General		Baseline - Algae Species		Public Health	
Number of possible results	95		1335		60	
Number of results that were not exact matches	16	(17%)	0	(0%)	0	(0%)
Number of result non-matches that were significant figure issues	16	(17%)	0	(0%)	0	(0%)
Number of results that were true errors	0	(0%)	0	(0%)	0	(0%)
Estimated Accuracy of dataset	**100%**		**100%**		**100%**	

Table 3. KHSA 2022 Sources of Significant Figure Issues.

	Baseline - General		Baseline - Algae Species		Public Health	
Number of significant figure issues	16		0			
Significant figure issue, unknown source	0	(0%)	0	(0%)	0	(0%)
Significant figure issue, introduced by sampling entity	16	(100%)	0	(0%)	0	(0%)
Significant figure issues, introduced by Watercourse	0	(0%)	0	(0%)	0	(0%)
Significant figure issues, other sources	0	(0%)	0	(0%)	0	(0%)