



PNNL-36037

2023 Hydrological/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana

May 2024

CJ Thompson
E Koehler
PJ Chamberlain

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <<https://www.ntis.gov/about>>
Online ordering: <http://www.ntis.gov>

2023 Hydrological/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana

May 2024

CJ Thompson¹
E Koehler²
PJ Chamberlain³

Prepared for
Talen Montana, LLC, Colstrip, Montana, and Genie Land Company,
Colstrip, Montana, under an Agreement for Commercializing Technology
with the U.S. Department of Energy, Contract DE-AC06-76RLO1830

Pacific Northwest National Laboratory
Richland, Washington 99354

¹ Pacific Northwest National Laboratory, Richland, Washington

² KC Harvey Environmental, LLC, Bozeman, Montana

³ Enviro-Sci Consulting, Big Lake, Alaska

Executive Summary

Pacific Northwest National Laboratory/Battelle–Pacific Northwest Division and KC Harvey Environmental, LLC personnel have monitored and collected water-quality data in the drainage basins of Cow Creek and Pony Creek in Rosebud County, Montana, since 1984. This work is performed for the signatories of Stipulation 12(d) under contract with Talen Montana (formerly PPL Montana LLC), which is in Colstrip, Montana. This report presents the collection, analysis, and interpretation of the water-quality and hydrologic data obtained during 2023.

The objective of the 2023 water monitoring study was to determine whether there were changes in water-quality and hydrological parameters downgradient from the effluent holding pond (EHP) east of Colstrip Power Plant Units 3 and 4 on Genie Land Company property since the initial 1984–1986 baseline study was completed. A further objective was to establish the current characteristics and conditions of water-quality and hydrological parameters so the effects of any ongoing or future events associated with the EHP and related operations could be evaluated. Specific conductance and boron and sulfate concentrations have been monitored since 1984 as indicators of leakage from the EHP. Since 2009, bromide has been included in the list of monitored constituents as an additional, unique indicator for detecting potential impacts from EHP operations.

There were no reportable releases from the EHP during 2023.

Groundwater levels measured in the alluvial wells in 2023 averaged about 0.75 ft higher than those in 2022, consistent with the higher levels of precipitation in the latter half of 2023. Between the July and October sampling campaigns, water levels rose in most monitored wells. The 2023 groundwater levels do not indicate any major changes in groundwater flow patterns for Cow Creek and South Fork Cow Creek.

Statistical analyses of conductivity, sulfate, and boron data identified five site-level trends where water quality had changed from previous years; no drainage-level trends were detected. Cow Creek had an increasing trend for boron at PW 735, which taps a deeper aquifer. Pony Creek had increasing trends for conductivity at Genie old well (GOW) 4 and Genie spring (GSP) 4, a decreasing trend for boron at GOW 4, and an increasing trend for boron at GSP 6. No abrupt increases were detected at any of the sites identified in Stipulation 12(d). These findings do not indicate any apparent impacts on groundwater quality from the operations of the EHP.

Using data from the entire observation period (1984–1987 and 1990–2023) and based on the selected sampling sites, the Pony Creek drainage was determined to have the lowest mean levels of conductivity, sulfate, and boron. Water from the Cow Creek drainage was distinctly highest in conductivity and sulfate, and the South Fork Cow Creek drainage means fell between the values from the other drainages except for boron, for which the mean was identical to that of Cow Creek.

Bromide was formally added as a monitoring constituent in 2010 in response to Talen Montana's recent addition of bromide to the flue gas scrubber solution. Bromide data from the 2023 sampling events provided no evidence of impingement of water from the EHP in the study area.

Continuous-monitoring probe measurements in private monitoring well 736 showed relatively stable conductivity except for a sharp decrease followed by an increase over a 1-week period in September. The cause of the changes is unknown, but it may be related to soluble minerals that previously accumulated in the unsaturated zone and were mobilized during heavy spring and summer precipitation. Groundwater elevation declined during most of the year. The probe data do not indicate any releases from the EHP during 2023.

Acronyms and Abbreviations

DTW	depth to water
EHP	effluent holding pond
ERA	Environmental Resources Associates
ft	foot (feet)
GAS	Genie alternative supply
GEL	General Engineering Laboratories
GKW	quality control sample
GNW	Genie new well
GOW	Genie old well
gpd/ft ²	gallons per day per square foot
GSP	Genie spring
GSW	Genie surface water
in.	inch(es)
μm	micron(s)
MDL	method detection limit
mg/L	milligram(s) per liter
mL	milliliter(s)
MPV	most probable value
NSI	NSI Lab Solutions
OD	outside diameter
PLR	Predictive Likelihoods Ratio
PNNL	Pacific Northwest National Laboratory
PNWD	Battelle–Pacific Northwest Division
ppm	parts per million
PVC	polyvinyl chloride
PW	private monitoring well
QA	quality assurance
QC	quality control
RPD	relative percent difference
μS/cm	microsiemens per centimeter
TDS	total dissolved solids
TOC	top of well casing
W-1	Montana State University-installed well

Contents

Executive Summary	iii
Acronyms and Abbreviations	v
1.0 Introduction.....	1.1
1.1 Study Area	1.2
1.2 Study Purpose and Scope.....	1.3
1.3 Report Contents and Organization.....	1.4
2.0 Hydrological Evaluation	2.1
2.1 Groundwater Flow System	2.3
2.2 Cow and Pony Creek Monitoring Network	2.10
2.3 Results and Discussion	2.11
3.0 Water Quality in the Cow Creek and Pony Creek Drainages	3.1
3.1 Water Sampling	3.1
3.2 Sampling Protocols	3.1
3.2.1 Alluvial Wells.....	3.1
3.2.2 Spring and Surface Water.....	3.3
3.2.3 Sample Handling, Quality Control Samples, and Field Measurements	3.4
3.2.4 Sample Custody.....	3.5
3.3 Sample Analysis	3.6
3.4 Quality Control for Water Samples	3.6
3.5 Statistical Analysis of Chemical Data.....	3.10
3.5.1 Objective 1: Identify Any Time Trends at the Site or Drainage Level in the Parameters Conductivity, Sulfate, and Boron.....	3.10
3.5.2 Objective 2: Estimate and Compare Overall (Time-Averaged, Site-Averaged) Drainage-Mean Values for Cow Creek, South Fork Cow Creek, and Pony Creek	3.11
3.5.3 Sites and Data Used in the Statistical Analyses.....	3.11
3.5.4 Statistical Testing Methods	3.13
3.5.5 Results of Statistical Analysis	3.14
3.5.6 Overall Conclusions of the Statistical Analysis.....	3.16
3.6 Results and Discussion	3.16
3.6.1 Sampling Data from the Genie Alternative Supply Wells.....	3.17
3.6.2 Sampling Data from Pony Creek Drainage	3.18
3.6.3 Sampling Data from South Fork Cow Creek.....	3.25
3.6.4 Sampling Data from Cow Creek	3.29
3.6.5 Comparison of Cow Creek, South Fork Cow Creek, and Pony Creek Drainages	3.32
3.6.6 Analysis of Bromide.....	3.34
3.6.7 Continuous-Monitoring Probe Results	3.35

4.0	Conclusions.....	4.1
5.0	References.....	5.1
	Appendix A Comments Provided by Signatories of Stipulation 12(d).....	A.1
	Appendix B Field Record Sheet and Chain-of-Custody Form Used for Sample Control	B.1
	Appendix C Analytical Methods Used for Water Analysis	C.1
	Appendix D Photos of Spring and Surface-Water Sites	D.1
	Appendix E Site and Drainage-Mean Values for Data Collected in 2023	E.1
	Appendix F Analytical Results for Water Samples Collected in 2023	F.1
	Appendix G Statistical Analysis – Methods and Results	G.1

Figures

Figure 1.1.	Colstrip Units 1–4 (from right to left).....	1.1
Figure 1.2.	Location of study area (modified from Montana Department of State Lands and U.S. Office of Surface Mining 1983).....	1.3
Figure 1.3.	Map showing the location of the Colstrip Units 3 and 4 EHP near the headwaters of Cow Creek.....	1.5
Figure 2.1.	Annual precipitation at Colstrip, Montana, 1984–2023. Data were compiled from the National Oceanic and Atmospheric Administration. ¹	2.2
Figure 2.2.	Monthly distribution of annual precipitation near Colstrip, Montana, 1984–2023. Data were compiled from the National Oceanic and Atmospheric Administration. ¹ Data for June–August 2023 were from the Brandenburg, MT climate station.	2.3
Figure 2.3.	Measuring DTW with a water-level meter.....	2.5
Figure 2.4.	Monitoring locations in the study area.....	2.8
Figure 2.5.	Water table in the alluvial aquifers of Pony Creek, Cow Creek, and South Fork Cow Creek.....	2.9
Figure 3.1.	Well sampling: (a) Lowering the submersible pump into well GNW 6. (b) Filling sample bottles with flow from the submersible pump. Water from the well is pumped through the sampling manifold, a 3-ft section of ¼-in. OD poly tubing, and an in-line filter before flowing into the sample containers. (c) Lowering a bailer into well GNW 7. (d) Sampling the pump outfall at well GOW 12 using a peristaltic pump.....	3.3
Figure 3.2.	Measuring the flow rate at a spring.....	3.4
Figure 3.3.	Preparing an equipment blank sample. The orange-colored peristaltic pump was used to pump deionized water from the 1-gallon container on the right through the sample filter and into the sample bottle.	3.5
Figure 3.4.	Plots and data for conductivity, sulfate, and boron in the Pony Creek Drainage during 2023. Values are single measurements from the July sampling event.....	3.19
Figure 3.5.	Trend plot for conductivity at GSP 4. Outliers identified in Appendix G, Section G1.2, were omitted from the graph.	3.21
Figure 3.6.	Trend plot for conductivity at GOW 4.....	3.21
Figure 3.7.	Trend plot for boron at GOW 4.....	3.22
Figure 3.8.	Trend plot for boron at GSP 6. Outliers identified in Appendix G, Section G1.2, were omitted from the graph.	3.22
Figure 3.9.	Trend plots for conductivity, sulfate, and boron in the Pony Creek Drainage, 1985–2023.....	3.24
Figure 3.10.	Plots and data for conductivity, sulfate, and boron in South Fork Cow Creek Drainage during 2023. Values are averages from the July and October sampling events.	3.26
Figure 3.11.	Trend plots for conductivity, sulfate, and boron in South Fork Cow Creek, 1984–2023.....	3.27
Figure 3.12.	Plots and data for conductivity, sulfate, and boron values for alluvial wells and springs in Cow Creek Drainage during 2023. Values shown are averages from the July and October sampling events.....	3.30

Figure 3.13.	Trend plot for boron at PW 735.	3.31
Figure 3.14.	Trend plots for conductivity, sulfate, and boron in Cow Creek Drainage, 1984–2023.	3.32
Figure 3.15.	2023 Bromide concentrations from sampling locations within the Cow Creek, South Cow Creek, and Pony Creek Drainages. The sites are ordered according to their position in each drainage (upgradient to downgradient). ND = non-detected result; Dry = site was dry (no sample was collected).	3.34
Figure 3.16.	Continuous-monitoring probe.	3.35
Figure 3.17.	Well PW 736 with companion datalogger enclosure, solar panel, and fenced enclosure.	3.36
Figure 3.18.	Groundwater temperature trend measured at PW 736 in (a) 2023 and (b) 2010–2023. Data collected in 2023 are shown in red.	3.37
Figure 3.19.	Groundwater elevation trend at PW 736 during (a) 2023 with monthly total precipitation plotted as columns associated with a secondary y-axis and (b) 2010–2023. Data collected in 2023 are shown in green.	3.38
Figure 3.20.	In situ PW 736 specific conductance trend measured by the continuous monitoring probe during (a) 2023 with field and laboratory values from the July and October sampling events and (b) 2010–2023. Probe data collected in 2023 are shown in magenta.	3.39

Tables

Table 2.1.	2023 groundwater-level data (feet).	2.6
Table 2.2.	2023 status of springs and surface-water sites.	2.7
Table 3.1.	Sites sampled during 2023 and sample acquisition methods. ^(a)	3.2
Table 3.2.	Analytical detection and reporting limits.	3.6
Table 3.3.	Analytical results for QC samples from the July 2023 sampling trip.	3.8
Table 3.4.	Analytical results for QC samples from the October 2023 sampling trip.	3.9
Table 3.5.	Wells and springs used in the statistical analysis.	3.12
Table 3.6.	Time trends detected in 2023. ^(a,b)	3.15
Table 3.7.	95% confidence limits on drainage-mean values using all data. ^(a)	3.15
Table 3.8.	Composition of EHP solution (based on a sample collected in May 2015).	3.16
Table 3.9.	Values for conductivity, sulfate, and boron in samples collected from GAS wells in 2023.	3.17
Table 3.10.	Drainage-mean values for conductivity, sulfate, and boron in Pony Creek Drainage, 1985-2023.	3.25
Table 3.11.	Drainage-mean values for conductivity, sulfate, and boron in South Fork Cow Creek, 1984–2023.	3.28
Table 3.12.	2023 analysis results from the PW wells.	3.31
Table 3.13.	Drainage-mean values for conductivity, sulfate, and boron in Cow Creek drainage, 1984-2023.	3.33

1.0 Introduction

The Colstrip coal-fired electric power generation complex in southeastern Montana consists of four units that previously had a combined generating capacity of 2,276 megawatts (Figure 1.1). Units 1 and 2 were retired in January 2020. Active Units 3 and 4 are currently capable of producing up to 1,480 megawatts of electricity. Solid waste from the scrubber system of Units 3 and 4 is slurried and piped to a nearby effluent holding pond (EHP) that was placed in service in 1983. In 2004, a paste plant was placed in service to receive and dewater the scrubber slurry. This material is then pumped to a dry disposal system that was placed into service in October 2022, and dry scrubber solids are placed into J-1 Cell of the EHP.



Figure 1.1. Colstrip Units 1–4 (from right to left).

This report describes the 2023 monitoring and analysis activities for the groundwater and surface-water downgradient from the EHP on Talen Montana (formerly PPL Montana LLC) and Genie Land Company properties. Assessment of the hydrologic conditions and water quality in the study area have been performed since 1984.

During the first quarter of 1984, Stipulation 12(d) was signed by the Montana Board of Natural Resources and Conservation, the Department of Natural Resources and Conservation, the Department of Health and Environmental Sciences, the Northern Plains Resource Council, the Rosebud Protective Association, the Genie Land Company, G.M. Garfield, the Rosebud County Board of Commissioners, and the Montana Power Company (now Talen Montana). This action provides for a water monitoring program downgradient (east) from Colstrip Power Plant Units 3 and 4 EHP in the Cow Creek and Pony Creek drainage basins in Rosebud County, Montana. After reorganization of the Montana State government in 1995, the roles and responsibilities of the Board of Natural Resources and Conservation, the Department of Natural Resources and Conservation, and the Department of Health and Environmental Sciences relative to Stipulation 12(d) were transferred to the Department of Environmental Quality and the Board of Environmental Review. Appendix A of this report provides comments about the report by the signatories of Stipulation 12(d).

Stipulation 12(d) provided for a water monitoring study that began in 1984, was suspended from May 1987 through August 1989, and was resumed in September 1989. At that time, Battelle–Pacific Northwest Division (PNWD), operator of the U.S. Department of Energy’s Pacific Northwest National Laboratory, assumed responsibility for project management, data analysis, and data interpretation. KC Harvey Environmental, LLC, based in Bozeman, Montana, is currently responsible for sample collection and handling and site maintenance. Enviro-Sci Consulting, through a contract with PNWD, performs the statistical estimation and testing of constituent trends and drainage-level comparisons of indicator parameter mean values.

Talen Montana also has an extensive groundwater monitoring program in the area upgradient from this study (i.e., in the vicinity of the Units 3 and 4 EHP). The latest site characterization reports are available from the Montana Department of Environmental Quality website.¹

1.1 Study Area

The study area is located near Colstrip, Montana, which is about 30 mi south of Forsyth and about 140 mi east of Billings (Figure 1.2). This region lies in the Great Plains Physiographic Province (Shimer 1972) in the northwestern part of the Powder River Basin. The study area is immediately east of Colstrip and east of the Western Energy Company Rosebud Mine (Figure 1.3; T 2 N, R 42 E and 43 E). The area includes three creeks with intermittent flows: Pony Creek, Cow Creek, and South Fork Cow Creek. The creeks are tributaries to Rosebud Creek, which drains north to the Yellowstone River near the unincorporated town of Rosebud, Montana. The drainages are divided by well-dissected uplands and buttes. Elevation differences between the creeks and adjacent uplands range from 20 to 30 ft near the headwaters to approximately 300 ft in the central part of the study area. The Colstrip Units 3 and 4 EHP complex is near the headwaters of Cow Creek (Figure 1.3). The pond occupies a small drainage tributary to Cow Creek. The EHP and clear well have a combined design surface area of approximately 324 acres and a usable surface area of approximately 180 acres (Hydrometrics 1990).

Throughout this study, Pony Creek, the northernmost drainage (Figure 1.3), is included in comparisons with the other drainages because Pony Creek represents background conditions. The drainage is hydrologically and geochemically similar to both the Cow Creek and South Fork Cow Creek drainages, but it is presumed to be unaffected by EHP activities. However, local variations in geology may differentially affect water quality. In addition, mining reclamation has progressed down the Pony Creek–Spring Creek and Pony Creek–Cow Creek divides to a point near Genie spring (GSP) 4 in Pony Creek, but the alluvial stream channel has not been directly disturbed. This mining reclamation could influence water-quality and hydrological parameters within the drainage.

¹ <http://deq.mt.gov/cleanupandrec/Programs/colstrip>

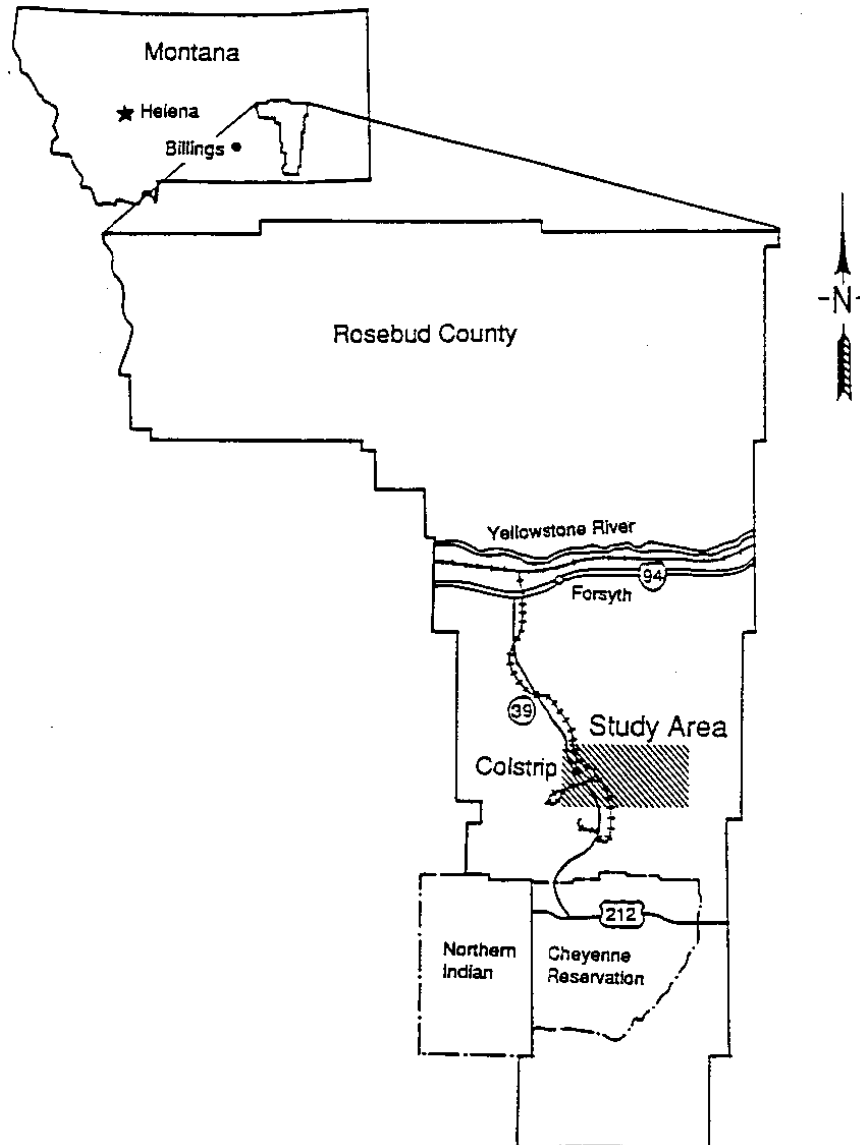


Figure 1.2. Location of study area (modified from Montana Department of State Lands and U.S. Office of Surface Mining 1983).

1.2 Study Purpose and Scope

The objective of the 2023 water monitoring study was to include and compare data collected during 2023 with data obtained in previous years to determine whether any changes had occurred in the water-quality constituents or hydrological parameters (water levels and groundwater flow directions) east of Colstrip Units 3 and 4 EHP. A further objective was to establish the current characteristics and conditions of water-quality and hydrological parameters to evaluate any effects of ongoing or future events associated with the EHP (such as accidental releases) and related operations. For example, in August 2020, Hydrometrics conducted a hydraulic pumping test on the Units 3 and 4 EHP underdrain to collect data that can support incorporating the underdrain into the full-scale implementation of the DEQ-approved remedy for the EHP, better estimate volumes within the ash, and evaluate groundwater conditions directly below the EHP. Details of the pumping test may be obtained from the Montana Department of

Environmental Quality website. Such pumping could reduce the fluxes of water and constituents of interest into the drainages of the study area.

To meet the study objectives, water samples were collected from surface-water sites, springs, and wells during July and October 2023. Within those water sources are six categories of sample sites: 1) Genie new well (GNW), 2) Genie old well (GOW), 3) Genie alternative supply (GAS) well, 4) Genie spring (GSP), 5) Genie surface water (GSW), and 6) a group of three private monitoring wells (PW)² that were installed in June 2009. The PW wells are located along Cow Creek near the western border of the study area (i.e., just northeast of the EHP). A continuous-monitoring probe that measures specific conductivity, temperature, and depth to water (DTW) was installed in well PW 736 in December 2010. Data from that system were evaluated to help understand the short-term behavior of the shallow groundwater system in the upper portion of Cow Creek near the EHP.

This report presents the compilation, statistical analysis, and interpretation of hydrologic and water-quality data collected during the July and October 2023 sampling trips. Chemical analysis of the water quality samples was subcontracted to General Engineering Laboratories, LLC, in Charleston, South Carolina.

1.3 Report Contents and Organization

Chapter 2 of this report presents a hydrological evaluation of the study area for 2023. Chapter 3 presents the sampling methods and quality assurance/quality control (QA/QC) results for 2023, statistical detection of water-quality trends at selected sites and for each drainage as a whole, and statistical comparisons of indicator parameter and constituent concentrations at the drainage level. Chapter 3 also describes and provides an interpretation of the water-quality data and the statistical testing results in the Cow Creek, South Fork Cow Creek, and Pony Creek drainages. Conclusions are presented in Chapter 4, and cited references are listed in Chapter 5. Appendix A contains comments about this report by the signatories of the Stipulation 12(d). Appendices B through G contain supporting information for the report, including example sampling documentation, a summary of analytical methods, photos of sampling sites, sample data (chemical analysis results), and statistical analysis methods and results.

² Private monitoring wells were installed by Hydrometrics, Inc.

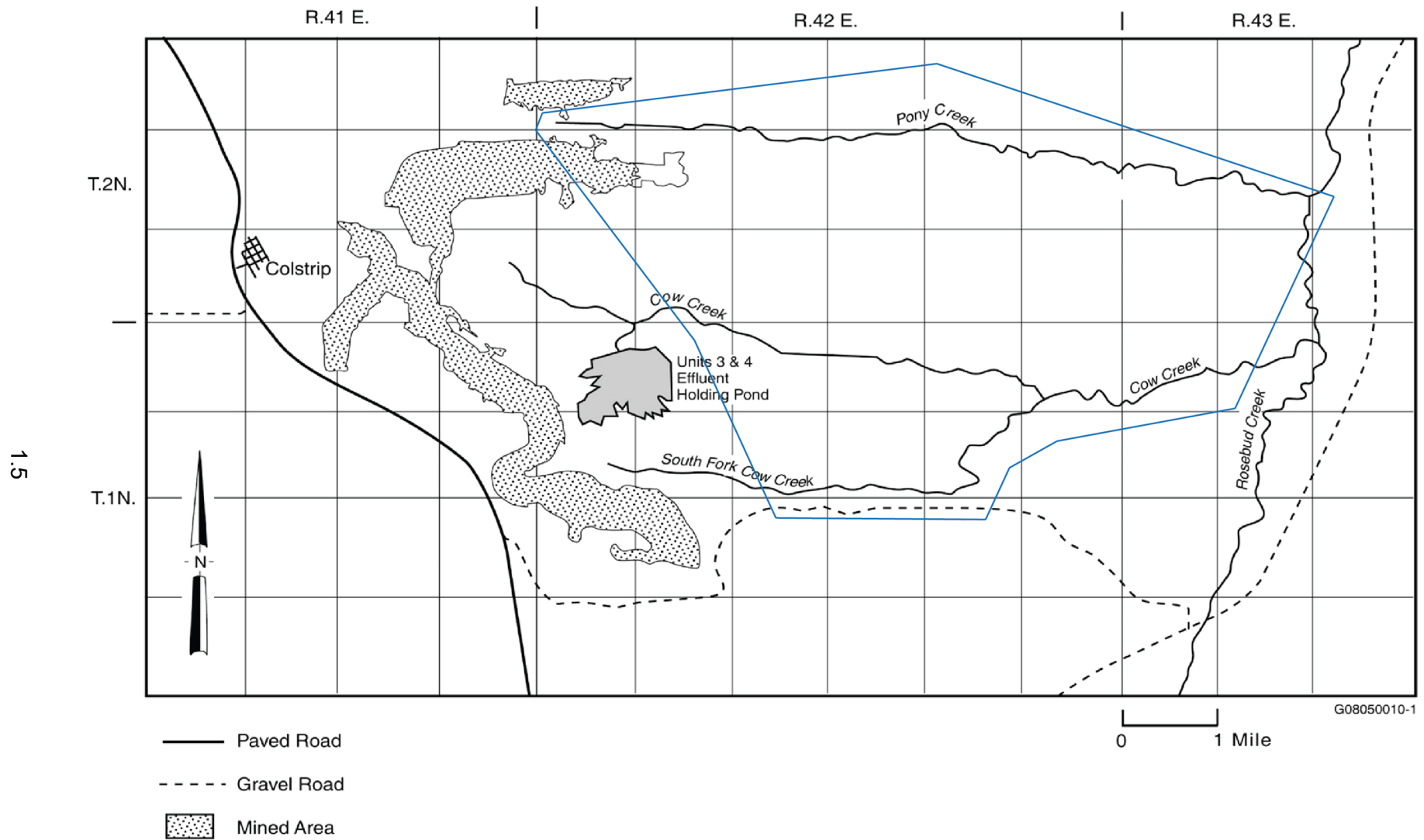


Figure 1.3. Map showing the location of the Colstrip Units 3 and 4 EHP near the headwaters of Cow Creek. The study area is within the perimeter of the blue polygon.

2.0 Hydrological Evaluation

This study focuses on monitoring surface water and groundwater in the Cow Creek and South Fork Cow Creek drainages to evaluate effects from the Colstrip Units 3 and 4 EHP. Pony Creek also is monitored as a control drainage that is presumed to be unaffected by the EHP impoundment. An essential part of this monitoring is the ongoing evaluation of hydrological data, including recent precipitation, because hydrological conditions could influence chemical and contaminant indicator levels in these water sources.

All three streams are characterized as intermittent with little or no flow during most of the year. Most surface-water flow occurs during snowmelt or precipitation events in the spring. Groundwater springs (i.e., locations where an aquifer reaches the ground surface; referred to as springs in this report) result in longer-duration flows of relatively small volume downstream from the source locations. Flows from springs vary seasonally, and some are dry during parts of the year.

Temporal patterns of precipitation may influence concentrations of analyzed parameters and water level/elevation by increasing or decreasing water flux through the hydrologic system. Also, based on evidence from the continuous monitoring probe in well PW 736 (Section 3.6.7), heavy influx of water, from rain or snow melt events, may mobilize soluble minerals that were previously stored in the partially saturated zone. This could trigger migration downward into the aquifer beneath and affect constituent levels there (Thompson et al., 2023). Therefore, in this study, it is important to monitor annual precipitation amounts and distribution trends throughout the year because the amount of precipitation likely influences indicator parameter levels. In some cases, fluctuations in precipitation may be the cause of detected changes in water quality.

Precipitation data for this report were obtained from the National Oceanic and Atmospheric Administration³. However, 2023 data from the Colstrip climate station were not available for the months of June, July, and August. Therefore, data from the Brandenburg, MT climate station was substituted for those months. The Brandenburg station is located approximately 17 miles east of the Colstrip station. While the Colstrip and Brandenburg data were likely not identical during June–August, we estimate that the Brandenburg data provide a good indication of monthly trends for Colstrip. This is supported by a very high degree of correlation (correlation coefficient = 0.96) between the two stations' precipitation data for those months when data was available at both locations (January–May and September–December).

Annual precipitation in the vicinity of Colstrip in 2023 was 20.11 in., which is approximately 30% higher than long-term, historical averages from before 1983 (15.5 in.; Montana Department of State Lands and U.S. Office of Surface Mining 1983) and from 1984–2022 (15.45 in., based on monthly data from the National Oceanic and Atmospheric Administration³). The higher total for 2023 was significantly greater than the amounts received in the previous 3 years (16.45, 8.89, and 11.78 in., respectively). Since this monitoring study began in 1984, the annual precipitation data have exhibited below-average values in the 1980s, near- and below-average values in the 1990s and early 2000s, and large 1–2-year fluctuations since 2004 (Figure 2.1). The five driest years between 1984 and 2023 were 2012 (8.04 in.), 1988 (8.58 in.), 2021 (8.89 in.), 2004 (9.04 in.), and 1984 (11.06 in.), respectively. In contrast, precipitation was above average in 2005, 2007, 2008, 2011, 2013, 2016, 2018, 2019, and 2023. The largest fluctuation recorded since 1984 was a decrease of over 16 in. from 2011 to 2012, which was just slightly larger than the increase in precipitation from 2012 to 2013 of approximately 15.5 in.

³ Data are available at https://www.weather.gov/byz/local_climate.

The monthly distribution of precipitation at Colstrip for 1984–2022 and Colstrip/Brandenberg for 2023 is shown in Figure 2.2. In 2023, approximately 73% of the annual precipitation occurred in the spring and summer months (including September), which is similar to the normal seasonal distribution of about 75% (Montana Department of State Lands and U.S. Office of Surface Mining 1983). May, June, September, and October were the wettest months in 2023, accounting for 24%, 22%, 13%, and 14% of the total annual precipitation, respectively. The winter months were relatively dry. November was the driest month with only about 0.2% of the annual precipitation. Because of limited winter precipitation, frozen (impermeable) surface soil, and insignificant upgradient catchment area, snowmelt and rainfall during winter are negligible sources of water influx for the subsurface drainages. Hence, the typically elevated groundwater levels observed during the spring monitoring campaigns are mostly due to antecedent springtime rainfall. Also, when precipitation in late summer is high, measured groundwater levels also tend to be high some weeks later. The above-average precipitation in October of 2023 (climatologically not a winter month) is expected to have undergone recharge in the same proportion as in warmer months. Precipitation during October is primarily in the form of rainfall, and any freezing of soil is quickly melted by rainwater before excessive runoff or evaporation (albeit reduced) commences. Therefore, groundwater recharge of the alluvial aquifers, which are described in the next section, is presumably rapid in this hydrogeologic system.

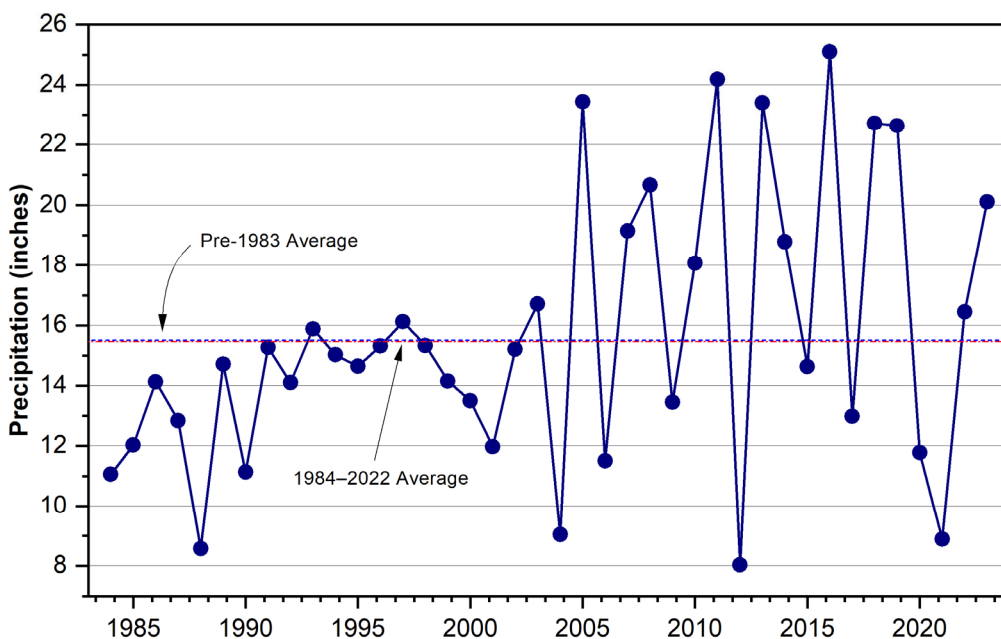


Figure 2.1. Annual precipitation at Colstrip, Montana, 1984–2023. Data were compiled from the National Oceanic and Atmospheric Administration.¹

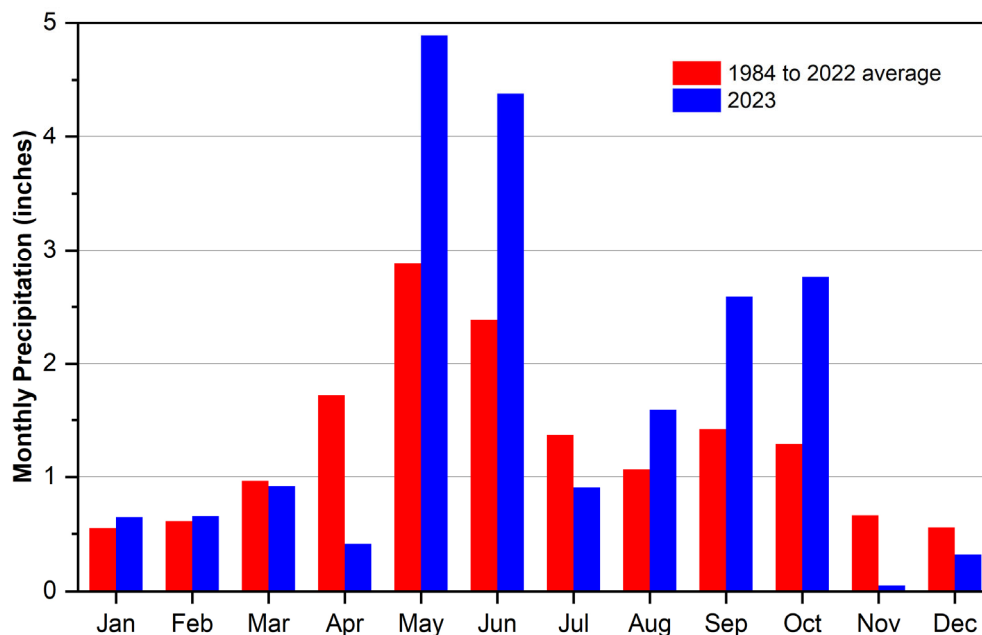


Figure 2.2. Monthly distribution of annual precipitation near Colstrip, Montana, 1984–2023. Data were compiled from the National Oceanic and Atmospheric Administration.¹ Data for June–August 2023 were from the Brandenburg, MT climate station.

2.1 Groundwater Flow System

Groundwater flow in the mine-permitted areas near Colstrip is well documented. The Montana Department of State Lands and U.S. Office of Surface Mining (1983) assessed the hydrology and geology of the area. Van Voast et al. (1977), Van Voast and Reiten (1988), Erbes (2000), and Metesh (1994) have discussed groundwater hydrology with respect to mining impacts.

The study area is in the northwestern portion of the Powder River Basin. Rock strata near the study area generally slope several degrees downward toward the southeast. The Paleocene Fort Union Formation is exposed at the surface throughout the study area except in the creek valleys, where Quaternary alluvium forms the valley fill. The Rosebud and McKay coal beds are situated in the middle portion of the Tongue River Member of the Fort Union Formation. The Rosebud coal bed is strip-mined in the Colstrip area. Much of the middle Tongue River Member and younger rocks have been eroded in the study area; consequently, sub-McKay Tongue River Member siltstones, sandstones, and coal beds form the surface in the western portion, and the underlying Lebo Shale Member is exposed in the extreme eastern portion. In some places in the past, the coal outcrops have burned, leaving formations of erosion-resistant clinker that caps most of the ridges in the study area.

More detailed stratigraphy information for the immediate study area was obtained from wells drilled in the area. The GNW series of wells was drilled into or through the alluvium in the Pony Creek, Cow Creek, and South Fork Cow Creek drainages. The wells generally penetrate inter-bedded sands, silts, and clays with occasional gravelly zones. Several GAS wells in the upland portions of the study area penetrate the sub-McKay strata, whose stratigraphy consists of alternating layers of sand, silt, siltstone, shale, silty sandstone, sandstone, and thin layers of coal. Lee (1980) noted the formation of these sediments in a deltaic to estuarine environment. Lee further noted the complex depositional processes

existing in such systems that result in sedimentary deposits characterized by lenticular beds, heterogeneous lithology, truncated units, and abrupt facies changes (Lee 1980).

In general, the Rosebud and McKay coal seams of the Tongue River Member form the major aquifers in the Colstrip area. In the EHP area, the Rosebud coal seam is burned in most places, leaving the clinker dry. The McKay coal seam has variable amounts of water. The shallow alluvial sediments and sandstone layers between the coal seams also are used occasionally in the area for groundwater supplies, depending on local conditions (Van Voast et al. 1977).

Hydrometrics (1987) and Metesh (1994) described the local groundwater flow in the sub-McKay sandstone, the McKay coal seam, and the overlying alluvium in the upper Cow Creek and South Cow Creek drainages. The reported mean hydraulic conductivity for the coal seams is about 14 gallons per day per square foot (gpd/ft^2), while the hydraulic conductivity for the overlying surficial alluvial aquifer is reported to average 900 gpd/ft^2 for the Cow Creek drainage (Metesh 1994). Hydraulic conductivity in the Pony Creek alluvium is also expected to be approximately 900 gpd/ft^2 ; however, as discussed below, it is likely much higher in the reclaimed backfill in the upper regions where mining has occurred.

Olsen et al. (1987) mapped the water table in the alluvial aquifer in the Cow Creek and South Fork Cow Creek drainages. Both surface water and groundwater generally flow to the east, parallel to the creek drainages and sub-parallel to the regional southeastward dip of the underlying sediments. Hydraulic head contours indicate that groundwater flows down the stream drainages in both the sub-McKay and alluvial sediments. Groundwater appears to discharge from the sub-McKay sediments into the alluvium. The drainage basin boundary near the headwaters of Cow Creek is on Western Energy Company property in Area E of the Rosebud Mine.

Waren and McDannel (2003) evaluated groundwater conditions in mined lands in the Colstrip area; they noted groundwater flow tends to be re-established in backfill materials that replace coal aquifers when mined lands are reclaimed. However, they also noted local perturbations of the potentiometric surface can occur that are persistent in the reclaimed environment. This finding indicates portions of the backfill materials that at least have different hydraulic properties (e.g., higher hydraulic conductivities) from those of the coal bed and adjacent sediments before they were mined. Consequently, these phenomena could affect portions of the Cow Creek and Pony Creek drainages adjacent to mined lands where hydraulic head distributions have changed in the reclaimed flow field. Therefore, as with impacts on flow resulting from fluctuations in precipitation, any induced changes in the alluvial flow field in these drainages (due to local changes in hydraulic conductivity) may affect concentrations of the sampled parameters there (e.g., increases in flow rate could have a diluting effect on constituent concentrations, although movement of any plume to sites downgradient would be hastened).

Hydrological measurements for this study were conducted during late July and early October 2023 (Figure 2.3). Water-level data are shown in Table 2.1 for all wells. DTW measurements in wells screened in the alluvium ranged from approximately 5 ft to 23 ft in 2023.



Figure 2.3. Measuring DTW with a water-level meter.

Overall, groundwater elevations in monitoring wells have fluctuated within a few feet over the course of the study, and changes in the alluvial groundwater levels generally have reflected changes in both annual and seasonal precipitation. During 2023, the overall average groundwater-level measured in all wells was 0.75 ft higher than their corresponding values from the summer and fall sampling trips in the previous year. Individual wells with the greatest year-over-year changes from 2022 to 2023 include GNW 2 (1.0 ft higher in the summer and 2.4 ft higher in the fall), GNW 3 (1.7 ft higher in the summer and 1.4 ft higher in the fall), GNW 6 (2.6 ft higher in the summer and 2.9 ft higher in the fall), and GOW 11 (2.3 ft higher in the summer and 2.5 ft higher in the fall). The higher water levels in 2023 at most wells are consistent with the increased precipitation in 2023, the rapidity of groundwater recharge, and the assumption that precipitation in October contributes to recharge in the same proportion as it does in warmer months.

The GOW well data are listed in Table 2.1. Many of these wells were drilled in the 1950s, and their construction and lithology is not well documented. GOW 5 is deeper and may penetrate through the Lebo Shale Member into the Tullock Member. Geochemical evidence for discerning the source of water to the GOW wells was documented by Olsen et al. (1991) and updated by Thompson et al. (2011). Only two of the GOW well levels are currently measured; GOW 1's average water level was 0.5 ft higher in 2023 than in 2022, while GOW 11's average water level was 2.4 ft higher in 2023.

Table 2.1. 2023 groundwater-level data (feet).

Site	TOC ^(a)	Total Depth	Casing Height (above ground surface)	Measured DTW ^(b) July	July Elev. ^(c) (TOC-DTW)	Measured DTW ^(b) October	Oct. Elev. ^(c) (TOC-DTW)
GNW 1	3047.84	42.5	--	5.49	3042.35	5.01	3042.83
GNW 2	3015.99	33	3.17	13.61	3002.38	15.39	3000.60
GNW 3	2991.86	32.5	2.88	17.81	2974.05	18.29	2973.57
GNW 4	2957.19	30	2.25	8.54	2948.65	8.59	2948.60
GNW 5	2950.64	35.5	3	9.58	2941.06	8.92	2941.72
GNW 6	3059.84	26	--	17.75	3042.09	18.22	3041.62
GNW 7	2989.56	17	2.88	15.46	2974.1	15.38	2974.18
GNW 8	2947.56	32	2.75	8.75	2938.81	8.81	2938.75
GNW 9	3065	--	--	-	-	--	--
GNW 10	--	~20	1.86	15.86	--	--	--
GNW 11	3210	--	1.79	--	--	--	--
GOW 1	2972.38	--	--	21.58	2950.8	22.01	2950.37
GOW 3	3124.29	--	--	--	--	--	--
GOW 4	3024.77	--	--	--	--	--	--
GOW 5	2910.26	--	--	--	--	--	--
GOW 6	3261.66	--	--	--	--	--	--
GOW 11	3065	--	--	20.17	3044.83	20.92	3044.08
GOW 12	2940	105	--	--	--	--	--
PW 734	3083.65	13	1.8	14.34	3069.31	14.36	3069.29
PW 735	3086.21	80	1.7	32.77	3053.44	32.89	3053.32
PW 736	3075.5	37	1.6	22.67	3052.83	22.93	3052.57
GAS 1	3075	100	2.25	40.42	3034.58	40.56	3034.44
GAS 2	2985	200	--	--	--	--	--
GAS 3	3015	195	1.17	--	--	--	--
GAS 4	3050	280	--	--	--	--	--
GAS 6	3080	280	--	--	--	--	--
GAS 7	2925	200	--	--	--	--	--
W-1	2880	--	2.50	11.58	2868.42	11.78	2868.22

(a) TOC = top of well casing. Values are elevations above mean sea level.

(b) DTW = depth to water.

(c) Elevation above mean sea level.

GAS = Genie alternative supply well; GNW = Genie new well; GOW = Genie old well; PW = private monitoring well; W-1 = Montana State University-installed well.

Three PW wells were installed at the upper part of the Cow Creek drainage in June 2009. These closely spaced wells monitor three different vertical intervals. PW 734 is screened in and monitors a tributary to Cow Creek. The well's screen interval is from 6 to 11 ft. This depth coincides with the uppermost aquifer in this locale, which is generally between 8 and 10 ft in depth and is often dry. Well PW 736 also monitors the uppermost aquifer with a screened depth of 16 to 36 ft. Well PW 735 monitors a deeper aquifer with a screened depth interval between 35 and 80 ft. Well PW 734 did not have sufficient water

for sampling in July 2023, but a reduced-volume sample was collected in October. This well was also dry during several previous sampling trips (2010 spring and fall, 2014 spring, 2015 fall, 2016 spring and fall, 2017 spring and fall, 2021 fall, and 2022 fall). Water levels in the shallower PW wells may be affected by the timing of monitoring in relation to precipitation events as well as operations associated with the EHP, which include pumping a network of wells installed between the EHP and the PW wells. As discussed in Section 3.6.7, continuous water level measurements are recorded in PW-736. Those data showed 0.3 and 0.6 ft water-level increases during February and March that were attributed to snowmelt events. Interestingly, the 2023 water level data generally decreased through the remainder of the year and was not correlated with high levels of rainfall in May, June, September, and October.

The measuring point elevations (i.e., top of the casing) were resurveyed before the June 2010 measurements for GNW 1, GNW 2, GNW 3, GNW 4, GNW 5, GOW 1, and PW 734. The largest change in measuring point elevation occurred at GNW 5, where the new survey revealed a 33.62 ft lower elevation from the top of the casing. The other resurveyed measuring point elevations increased between 0.01 ft and 2.64 ft. Table 2.1 shows the most recent top-of-casing elevations for all wells. The new measuring point surveys should be accounted for when comparing water-level elevations in this report with earlier reported elevations that were based on inaccurate elevation surveys.

The hydrological conditions of springs and surface-water sites are listed in Table 2.2 (Figure 2.4 shows the site locations). Appendix D contains photos of these sites from the July and October sampling campaigns.

Table 2.2. 2023 status of springs and surface-water sites.

Site Name	Site Elevation (ft)	July Status	October Status
GSW 1	3047.80	OK(s) ^(a)	OK(s)
GSW 2	3026.51	OK(s)	OK(s)
GSW 3	3005.14	Dry	OK(s)
GSW 5	2951.85	OK (s)	OK(s)
GSW 6	2905.43	OK(s)	OK(s)
GSP 1	3036.03	OK(s)	OK(s)
GSP 2 ^(b)	2949.87	OK(s)	OK(s)
GSP 3	3193.42	OK(s)	OK(s)
GSP 4	3275.67	OK(s)	NS ^(c)
GSP 5	3233.67	OK(s)	NS
GSP 6	3180.00	OK(s)	NS
GSP 7	2952.92	OK(s)	NS
GSP 8	2828.29	Dry	NS
GSP 9	3251.31	OK(s)	NS
GSP 10	3120	OK(s)	NS

(a) OK(s) = site okay for sampling (i.e., adequate water was available)

(b) GSP 2 is also known as Stinking Spring.

(c) NS = site not scheduled for sampling.

GSP = Genie spring; GSW = Genie surface water.

Several of the springs provide enough water for livestock watering. GSW 3 and GSP 8 were dry during the July 2023 sampling trip. The remaining spring and surface-water sites were sampled as scheduled. Elevations for springs and surface-water sites were determined at the beginning of the study and have been considered accurate without resurveys. The estimated potentiometric surface is shown in Figure 2.5.

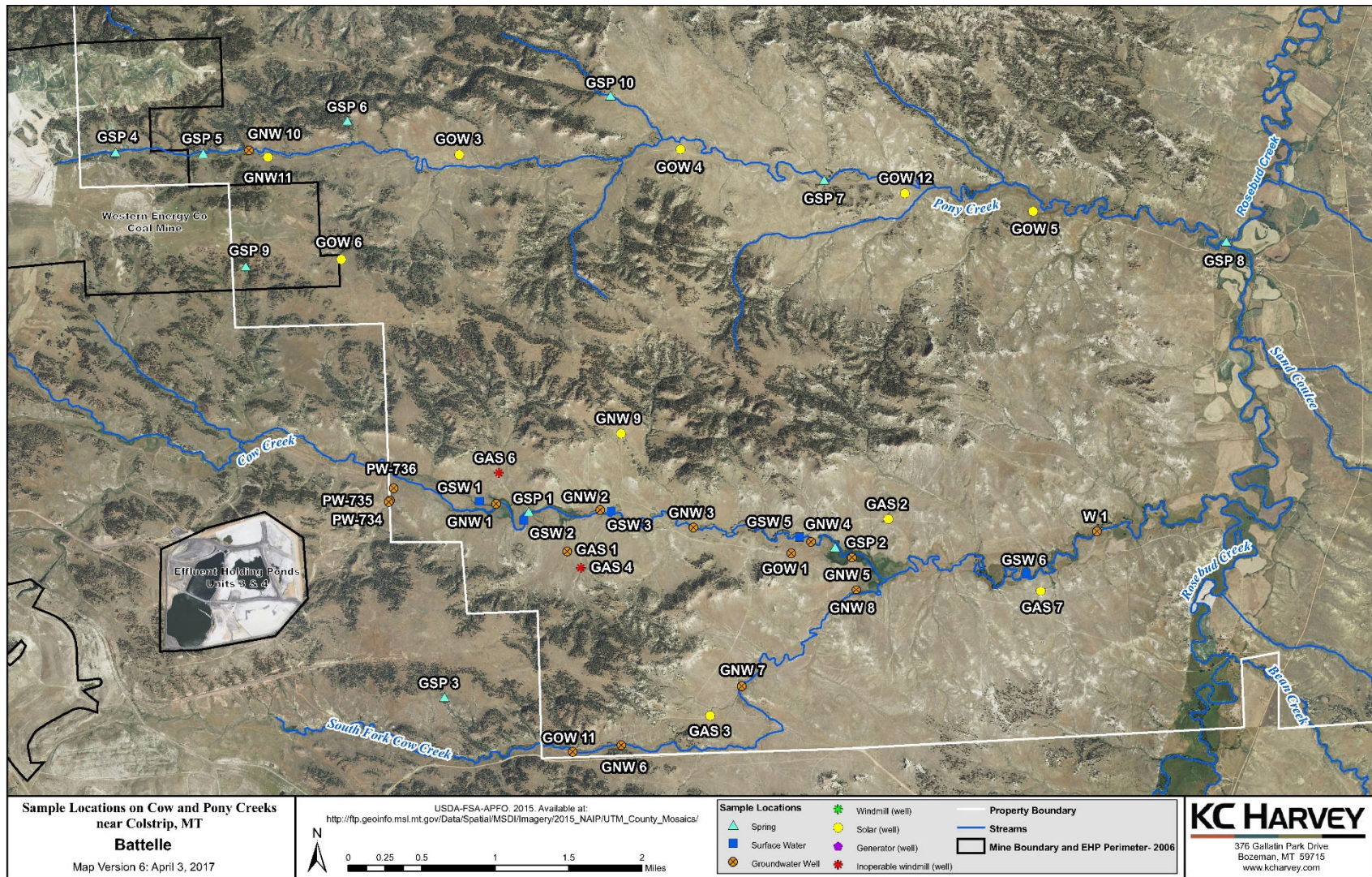


Figure 2.4. Monitoring locations in the study area.

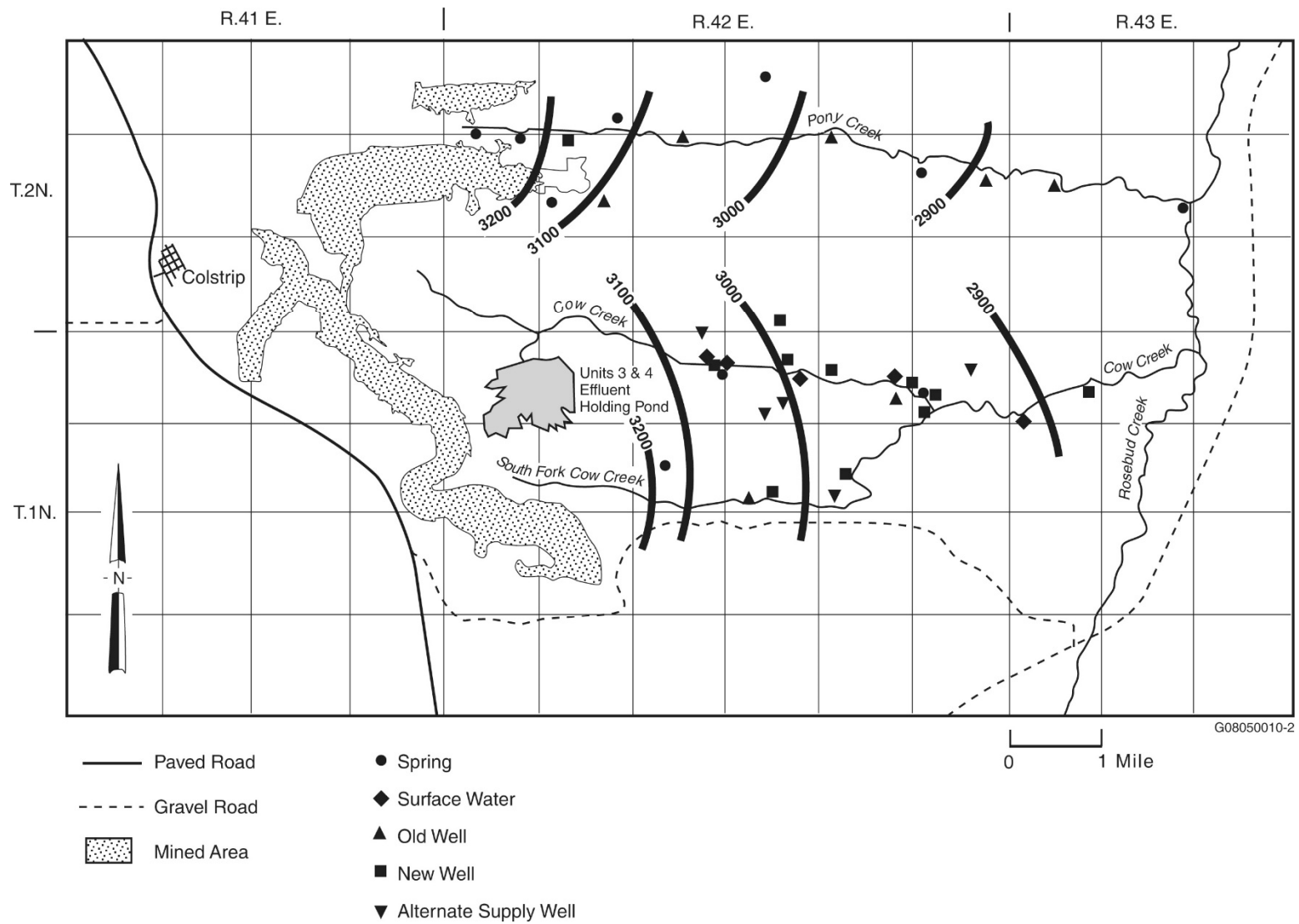


Figure 2.5. Water table in the alluvial aquifers of Pony Creek, Cow Creek, and South Fork Cow Creek.

Relatively minor differences have been observed since the early 1990s in the Cow Creek and South Fork Cow Creek drainages. Water levels in October 2023 generally were within several feet of levels observed in the early 1990s. The fall measurement was mapped because it tends to be more representative of base flow conditions; the effects of spring runoff and winter recharge are minimized. Most of the measured water levels are from wells that penetrate the alluvial aquifer. The hydraulic head contours are based on the assumption that the sub-McKay strata are hydraulically connected to the alluvial aquifers in each creek valley (Hydrometrics 1987); thus, a hydraulic connection is assumed for all wells. This assumption is consistent with the hydro-stratigraphic location of these units above the Lebo Shale, a major aquitard in the system. Although the aquifers are hydraulically connected, because of their parallel flow directions and the large distances between monitoring locations in different drainages, it is assumed that any water-quality changes at sites in one drainage do not directly affect the sites in the other drainages.

Groundwater gradients were calculated based on water-level measurements taken in the alluvia for Cow Creek and South Fork Cow Creek in October 1990. The October 2023 measurements do not indicate any significant changes in these calculations. The groundwater gradient in Cow Creek is about 0.008 ft/ft between GSW-1 and GNW 5. The new measuring point survey results do not substantially affect this calculation of the gradient, and the recalculated groundwater elevation at GNW 5 is more consistent with nearby wells GNW 4 and GNW 8. The groundwater gradient in Pony Creek is about 0.01 ft/ft between GSP 4 and GSP 7. Because the hydraulic conductivities in these drainages are essentially unchanged from previous years, the groundwater velocities, which are proportional to the product of gradient and conductivity, are also expected to be unchanged. Therefore, any detected site trends would not appear to be attributable to changes in velocities (i.e., not related to precipitation).

2.2 Cow and Pony Creek Monitoring Network

The sampling network includes a combination of wells, springs, and surface-water locations (Figure 2.4). See Section 3.2 for the locations where water samples were collected at least once during 2023. Specific well construction details are included in Table 2.1.

The GNW wells and PW 736 are generally less than 50 ft deep and were drilled to monitor the valley fill alluvium. Wells PW 735, GNW 9, and GNW 11 are deeper and penetrate through the valley fill alluvium into the sub-McKay part of the Tongue River Member. Well PW 734 monitors the uppermost aquifer in a relatively shallow zone at about 8–12 ft below ground surface. All GNW and PW wells are accessible for measuring water levels except for GNW 9 and GNW 11, which have dedicated, solar-powered pump configurations that prohibit measurements.

The GAS wells generally penetrate into the sub-McKay portion of the Tongue River Member and deeper units of the Fort Union Formation. Only GAS 1 is accessible for water-level measurements. The water level has remained fairly constant in this well, typically varying less than 1 ft over the entire long-term study. This behavior is consistent with that of a confined aquifer. Obtaining access to the remainder of the GAS wells for water-level measurements would provide more insight into the regional groundwater flow characteristics of the sub-McKay flow regime. The geochemical characteristics of water from these deeper wells are compared with those of the alluvium in the geochemical discussion (Section 3.6.1).

The configuration of several of the GOW and GAS wells makes it difficult to obtain water-level measurements at those sites since most of these wells were outfitted with windmills. In 1992, access ports were installed on some of the wells, but because of limited space between the riser pipe and the well casing, obtaining water-level measurements was impractical. Several of the older GOW wells have fallen into disrepair and are no longer used to provide stock water. Solar-powered stock-watering systems have been installed at five locations (Figure 2.4). Water levels cannot be measured in the solar-powered systems, but water-quality samples have been collected at these sites (GAS 7, GNW 9, GNW 11, GOW 4, and GOW 12). As discussed in Section 3 and Appendix G of this report, well GOW 12 is suspected to be

tapping a geochemically different water supply than it accessed before the solar-powered pump installation.

2.3 Results and Discussion

Water-level data collected during 2023 confirm previous observations of hydrological conditions and groundwater movement in the Pony Creek, Cow Creek, and South Fork Cow Creek drainages. Water-level measurements for wells withdrawing water from the alluvial materials that fill the valleys of Pony Creek, Cow Creek, and South Fork Cow Creek continue to indicate that groundwater flows in an easterly, down-valley direction toward Rosebud Creek.

Groundwater levels measured in the alluvial wells during 2023 averaged about 0.75 ft higher than those of the previous year. Approximately half of the spring/summer water levels were higher in 2023, while all fall water levels were higher except for well PW 736. These trends are generally consistent with the higher levels of precipitation in the latter half of 2023. Water levels rose in most wells between the July and October 2023 sampling campaigns. The largest increases occurred at wells GNW 2 (1.78 ft), GNW 6 (0.47 ft), GOW 1 (0.43 ft), and GOW 11 (0.75 ft).

Groundwater levels measured during 2023 do not indicate any major changes in groundwater flow patterns for Cow Creek and South Fork Cow Creek. The EHP for Colstrip Units 3 and 4 is situated within the drainage near the headwaters of Cow Creek. The headwaters of all three drainages are within areas previously disturbed by mining.

Further studies could be considered to assess the aquifer properties and better quantify the potential for groundwater movement and potential contaminant migration through groundwater in the study area. Groundwater recharge over much of the alluvial aquifers is suspected to be quite rapid because the majority of recharge occurs during spring. As noted above, this is due to little available wintertime precipitation and seasonally frozen surface soil; therefore, effective groundwater recharge is negligible. Another important aspect of the hydrology in the drainage is the role of creek flow, particularly in combination with recharge potential. When present, surface-water flow in the creek is likely the fastest path of potential contaminant migration through the flow system. No creek flow was observed within the South Fork Cow Creek, Cow Creek, and Pony Creek drainages during the July and October 2023 sampling trips with the exception of minor flows in GSP 1 (Cow Creek), GSP 6 (Pony Creek), and GSP 10 (Pony Creek) in June.

Water levels in the alluvial aquifer generally are 5 ft to 20 ft below the creek bed in the Cow Creek drainage, indicating the creek would tend to lose water to the aquifer. Such losses would depend on the hydrogeology of any given creek segment. Consequently, seasonally flowing surface water is a potential pathway for quick movement and uneven distribution of constituents within a drainage creek. Additional study (e.g., surface-water modeling) would be required to understand the potential for contaminant migration through creek flow and for creek flow to act as recharge to the alluvial aquifer throughout the drainage where flow occurs.

3.0 Water Quality in the Cow Creek and Pony Creek Drainages

Water-quality data gathered from the 2023 sampling trips are presented in Sections 3.1 through 3.3. A statistical analysis of time trends and a comparison of drainages are provided in Section 3.5, followed by interpretation of the results in Section 3.6. Throughout this chapter, concentrations of chemical parameters are listed in units of parts per million (ppm), which is equivalent to milligrams per liter (mg/L). Specific conductance (also referred to as conductivity in this report) is reported in units of microsiemens per centimeter ($\mu\text{S}/\text{cm}$). The naming convention for samples is the site name followed by a dash and the sampling trip number. For 2023, the July sampling campaign was trip 78, and the October sampling campaign was trip 79.

3.1 Water Sampling

To determine the current water-quality status of the Cow Creek and Pony Creek drainages, three water source categories (surface, springs, and wells) were sampled and analyzed. Samples were taken from the six site categories noted in Chapter 1.0: GAS, GNW, GOW, GSP, GSW, and PW. Samples were collected from 11 wells and springs in the Pony Creek drainage; a spring and a well on the Cow Creek/Pony Creek divide; 21 wells, springs, and surface-water sites in the Cow Creek drainage; and 6 wells and springs in the South Fork Cow Creek drainage. Table 3.1 lists the sites that were sampled in 2023, and Figure 2.4 shows the sampling locations. Sampling in the Pony Creek drainage and the Cow Creek/Pony Creek divide was conducted during July only. Pony Creek spring GSP 8 was dry during the July campaign and was not sampled. Wells in the upper portion of the alluvial aquifer, springs, and surface-water sites in Cow Creek and South Fork Cow Creek were predominantly sampled in both July and October. Cow Creek surface-water site GSW 3 was not sampled in July because that site was dry. Data gathered from the sampling trips are discussed in Sections 3.3 through 3.6.

3.2 Sampling Protocols

The procedures described in this section are the sample collection techniques used for this hydrological/water-quality study. The protocols were developed to 1) minimize the possible contamination of samples, 2) ensure that samples closely represent the water quality of the sampling site, and 3) ensure accurate identification of samples collected during this study. As each sample site was visited, observations including general weather conditions, sample site conditions, and purging data for wells sampled with a submersible pump or bailer were recorded in a field notebook.

Water samples were collected by the methods most appropriate for the sample source. Table 3.1 summarizes the sites sampled during 2023 and the primary equipment used at each location. Springs, wind-driven and solar-powered stock-watering wells, and surface-water sites were sampled with a battery-powered peristaltic pump with an in-line, disposable 0.45 μm filter. Twelve wells were sampled with a submersible pump, and three wells were sampled with a bailer due to low yield or high levels of fine sediments that preclude efficient pumping.

3.2.1 Alluvial Wells

Water levels were recorded at most GNW- and PW-labeled wells and at GOW 1, GOW 11, GAS 1, and W-1 before they were pumped or bailed. Solar-powered pumps have been installed in wells GNW 9 and GNW 11, and it was impractical to obtain water-level measurements at those sites. The reference point for the water-level measurements was the top of the well casing (marked location or lowest point on the top of the casing). Water levels were recorded to the nearest 0.01 ft using a water-level meter.

Table 3.1. Sites sampled during 2023 and sample acquisition methods.^(a)

South Fork Cow Creek		Cow Creek		Cow/Pony Creek Divide		Pony Creek	
GSP 3	P	PW 734	P	GSP 9	P	GSP 4	P
GOW 11	S	PW 735	S	GOW 6 ^(b)	P	GSP 5	P
GNW 6	S	PW 736	S	--		GNW 10	B
GAS 3	P	GSW 1	P	--		GNW 11 ^(b)	P
GNW 7	B	GNW 1	S	--		GSP 6	P
GNW 8	S	GSW 2	P	--		GOW 3	P
--		GSP 1	P	--		GSP 10	P
--		GAS 1	S	--		GOW 4 ^(b)	P
--		GNW 2	S	--		GSP 7	P
--		GSW 3	P	--		GOW 12 ^(b)	P
--		GNW 9 ^(b)	P	--		GOW 5 ^(b)	P
--		GNW 3	S	--		--	
--		GOW 1	S	--		--	
--		GSW 5	P	--		--	
--		GNW 4	S	--		--	
--		GSP 2	P	--		--	
--		GNW 5	S	--		--	
--		GAS 2	P	--		--	
--		GSW 6	P	--		--	
--		GAS 7 ^(b)	P	--		--	
--		W-1	B	--		--	

(a) Sample acquisition methods: P = pumped with portable peristaltic pump, S = pumped with 4-in. submersible pump, B = bailed with 3.5-in. outside diameter (OD) PVC bailer, then bailer was sampled with peristaltic pump.

(b) Well is pumped by a dedicated, solar-powered pump. Outflow from the pump was sampled in-line (i.e., before flowing into the nearby stock-watering tank).

GAS = Genie spring; GSW = Genie surface water; PVC = polyvinyl chloride; PW = private monitoring well installed by Hydrometrics, Inc.; W-1 = Montana State University-installed well. GSP 2 is also known as Stinking Spring.

Most well samples were acquired with a 4-in. submersible pump (Figure 3.1a and b) or a 3.5-in. OD polyvinyl chloride bailer (Figure 3.1c). A battery-powered peristaltic pump was used to filter water obtained with the bailer. When a well was sampled with the submersible pump, a minimum of 100 gal (i.e., greater than three well volumes) of water were pumped before sampling. This purging process helps ensure that the collected water is representative of the groundwater in the vicinity of the well. The total discharge was recorded on a totalizing flow meter and written in the field notebook to the nearest 0.1 gal.

Well GNW 7, a low-yield well, was bailed dry during both the July and October 2023 sampling events and allowed to recharge for 24 hours or more before sampling. The bailer (Figure 3.1c) is a hollow PVC cylinder (approximately 3.5 feet long and 4 inches in diameter) with a loose ball inside and a round hole in the bottom. When the bailer is lowered into a well, water fills the sampler through the bottom hole. Raising the bailer causes the loose ball to settle into the hole, creating a seal that prevents the water from draining out as the sampler is retrieved from the well.



Figure 3.1. Well sampling: (a) Lowering the submersible pump into well GNW 6. (b) Filling sample bottles with flow from the submersible pump. Water from the well is pumped through the sampling manifold, a 3-ft section of ¼-in. OD poly tubing, and an in-line filter before flowing into the sample containers. (c) Lowering a bailer into well GNW 7. (d) Sampling the pump outfall at well GOW 12 using a peristaltic pump.

Wells GAS 2, GAS 3, GAS 7, GOW 3, GOW 4, GOW 5, GOW 6, and GOW 12 previously were equipped with windmill-driven pumps but now have solar-powered pumps, which eliminates the need for wind during sampling. Relatively deep wells GNW 9 and GNW 11 are also equipped with solar-powered pumps. All wells with solar-powered pumps are sampled at the pump outfall using a battery-powered peristaltic pump (Figure 3.1d). These wells are generally not purged before sampling because the solar-powered pumps are usually running continuously during daylight hours. In cases where a well pump was not running steadily (e.g., due to overcast conditions), the well was purged by operating the pump for several minutes before sampling. For July 2023, GNW 11 and GOW 3 were sampled from the respective stock ponds because the pump at GNW 11 was inoperable and an active hornet nest was present near the pump outfall at GOW 3. The windmill at GAS 4 has not been functional since 2003; therefore, no samples were collected from that location.

3.2.2 Spring and Surface Water

Spring and surface-water samples were obtained using a peristaltic pump and an in-line filter assembly to fill the sample bottles directly at the sampling site. Flow measurements or estimates were made at the

time of sample collection for all surface-water sites and springs, when possible (Figure 3.2). Appendix D contains photos of the spring and surface-water sites from both sampling campaigns.



Figure 3.2. Measuring the flow rate at a spring.

3.2.3 Sample Handling, Quality Control Samples, and Field Measurements

Field samples were preserved based on the recommendations in standard procedures used by the analytical laboratory. Each water sample was divided into two aliquots: 1) 250 mL of a 0.45 μm filtered sample with 1 mL of 1/1 concentrated HNO_3 /water preservative (for inductively coupled argon plasma analyses for magnesium, calcium, sodium, potassium, and boron) and 2) 1 L of a 0.45 μm filtered sample with no added preservative (for pH, conductivity, alkalinity, total dissolved solids [TDS], bromide, chloride, and sulfate analyses).

Sample fluid from submersible pumped wells was flowed through a stainless-steel sampling manifold equipped with a disposable in-line 0.45 μm filter assembly. After well purging, the filter was flushed with approximately 500 mL of sample water before the sample was collected. Sample water was filtered directly into the appropriate sample container. Spring and surface-water samples were obtained similarly by attaching a 0.45 μm filter to the peristaltic pump tubing, flushing the filter, and delivering a field sample into the appropriate container.

To improve the consistency of results (i.e., ion charge balance) between dissolved metals and anions at surface-water and spring sites, we pumped water from each of those sites through a filter and into a pre-rinsed 1 gallon container. The sample was then thoroughly mixed before pouring the water into sample containers. As mentioned in previous reports, this protocol was adopted in 2016 and has been shown to improve the charge balances at sites that tend to have little water available, such as GSP 1, GSP 2, GSP 3, GSP 4, GSP 5, and GSW 1. Another benefit of this mixing protocol is that it improves the sampling precision at these sites. The approach is similar to composite soil sampling, but it has a different goal of

reducing the total sampling/analysis “error” (as defined in Sections 3.5 and G.2) within samples from springs and surface water.

At each sampling site, an unfiltered sample aliquot was dispensed into a test tube for field parameter analysis. Measurements of conductivity and pH were made within 8 hours of sampling. These field measurements provide initial screening information and help verify laboratory results. Water temperature was determined in situ or in the pump-flow discharge stream with a dial or digital thermometer. These instruments were checked for accuracy with an ice bath and a laboratory mercury thermometer and, in all cases, were found to be accurate within $\pm 1^{\circ}\text{C}$. Sample pH was determined using a pH meter calibrated using standard buffer solutions ranging from pH 4 to pH 10. Sample conductivity was measured with a conductivity meter calibrated using various calibration standards ranging from 447 to 12,880 $\mu\text{S}/\text{cm}$. All calibration data were recorded in the project field notebook.

For both sampling campaigns, field blank and duplicate samples were collected and sent to the analytical laboratory. Equipment blanks were prepared by running distilled water through the peristaltic pump and in-line filter assembly (Figure 3.3). Bottle blanks were prepared by filling sample containers with distilled water. Two duplicate samples were collected during each campaign: one from a submersible pumped well and the other from a location where the peristaltic pump and in-line filter assembly were used. All samples were placed immediately in coolers with ice.



Figure 3.3. Preparing an equipment blank sample. The orange-colored peristaltic pump was used to pump deionized water from the 1-gallon container on the right through the sample filter and into the sample bottle.

3.2.4 Sample Custody

Sample custody was managed by Pacific Northwest National Laboratory (PNNL) personnel. After collecting and processing the samples, KC Harvey Environmental, LLC staff transferred the samples to

the on-site PNNL representative for the appropriate custodial procedures (i.e., sealing, sample packing, storing, and shipping). An example of a chain-of-custody form is provided in Appendix B.

3.3 Sample Analysis

Samples were analyzed by General Engineering Laboratories (GEL) at its main laboratory facility in Charleston, South Carolina. Appendix C lists the analytical methods used by GEL, and Appendix F contains tables of analytical results from both sampling campaigns in 2023. GEL's detection and reporting limits for all measured parameters are listed in Table 3.2.

Table 3.2. Analytical detection and reporting limits.

Parameter	Method ^(a)	Detection Limit ^(b) (ppm)	Reporting Limit ^(c) (ppm)
Alkalinity	2320B	0.725	2.00
Conductivity	120.1	1.00 $\mu\text{S}/\text{cm}$	1.00 $\mu\text{S}/\text{cm}$
pH	9040	0.01 pH units	0.100 pH units
TDS	2540C	2.38	10.0
Bromide	9056	0.067	0.200
Chloride	9056	0.067	0.200
Sulfate	9056	0.133	0.400
Boron	6010	0.015	0.050
Calcium	6010	0.050	0.200
Magnesium	6010	0.110	0.300
Potassium	6010	0.050	0.250
Sodium	6010	0.100	0.300

(a) Method references are listed in Appendix C.

(b) The detection limit is the lowest concentration that can be measured with 99% confidence. When samples are diluted, the detection limit scales with the dilution factor (e.g., a 10-fold dilution results in a detection limit that is 10 times higher).

(c) The reporting limit is an estimate of the lowest concentration that can be reliably quantified.

3.4 Quality Control for Water Samples

Three types of QC samples were submitted to GEL during the two 2023 sampling campaigns to assess sampling and analysis performance. Duplicates, blanks, and blind standards provide measures of reproducibility, contamination, and accuracy, respectively. In all cases, the QC samples were submitted to the analytical laboratory in double-blind fashion (i.e., they were disguised as regular monitoring samples). At the laboratory, an additional level of QC was added by randomly selecting samples to be analyzed in duplicate and by spiking samples to calculate recovery values for measured parameters. Table 3.3 and Table 3.4 list all QC samples sent to the laboratory for analysis and present the analytical results. Appendix C lists the analytical methods used for water analysis.

Analytical results received from GEL were evaluated using Microsoft Excel[®] and Access.⁴ Ion balances were calculated on the spreadsheets from the analytical results reported by the laboratory. If an ion balance deviated by more than $\pm 10\%$, the analytical laboratory attempted to identify the problem. PNNL scientists reviewed the data for questionable results and typographical errors. GEL was contacted to

⁴ Access and Excel are registered trademarks of Microsoft Corporation.

verify any questionable results. The data were then compared with analytical results from previous sampling trips (Olsen et al. 1987, 1991–2006; McDonald et al. 2007; Thompson et al. 2008–2023). These data were also compared with their means, and a statistical outlier identification procedure was applied (Appendix G, p. G.2). Any questionable data that could not be reconciled with the analytical laboratory were still entered into the database but flagged as questionable and omitted from all statistical tests, as discussed in Section 3.5. Extremely low or high data identified as outliers are listed in Appendix G, Table G.2. *Outliers* are errantly extreme data whose causes are unknown but are often observed to occur randomly. They are statistically identifiable and are concluded to have resulted from a different process than that of other data; therefore, outliers are excluded from subsequent data analyses.

Outliers may also occur systematically but are less common. For example, recent data from samples at GOW 12 were omitted from the statistical analysis because they were determined to be systematically low and are not representative of the water in Pony Creek’s alluvium. The causes of the extreme values are not considered to be laboratory or sampling related, rather the data are thought to be representative of the water being sampled. These extreme values may be attributable to changes in that well’s water influx source.

The results of the QC samples from the July 2023 sampling trip are summarized in Table 3.3. Two field blanks (GKW 3-78 and GKW 6-78) were collected to evaluate the potential for sample contamination from the bottles and sampling equipment. The bottles used for the blanks were the same type as those used for routine monitoring samples. Results from analysis of the blanks did not identify any significant contamination from bottles or equipment for the constituents measured.

Duplicate samples were used to assess the variation in laboratory analysis results (i.e., precision). The values for each constituent or parameter were compared by computing the relative percent difference (RPD), which is the absolute value of the difference between the results divided by their average. RPDs not exceeding 20% are generally considered acceptable if the analytical results are at least five times larger than the laboratory analysis method detection limit (MDL). At levels below five times the detection limit, higher RPDs may occur and be acceptable. The results for duplicate samples (GKW-1-78 and PW-735-78; GKW-2-78 and GSP-4-78) demonstrated good reproducibility for all measured parameters (RPDs were <10% with the exception of 12% for TDS for the first duplicate pair), indicating low sample-to-sample variation and acceptable analytical precision.

Laboratory accuracy was evaluated using blind standards, which are samples that contain known levels of constituents. For the July 2023 sampling event, two blind standards were submitted: a sample consisting of a NSI Lab Solutions standard, QCI-136 (Minerals QC CRM; sample GKW-4-78), and two Environmental Resources Associates (ERA) standards, WasteWatr™ Trace Metals and Minerals (sample GKW-5-78). To assess precision, the laboratory results were compared with the most probable value (MPV) for each constituent by computing the percent error, which is the measured value minus the MPV expressed as a percentage of the MPV. Results within 20% are generally considered acceptable if the analytical results are at least five times the detection limit. GEL demonstrated strong performance on the blind standards—all results were acceptable, and the error percentages were less than ~12%.

Table 3.3. Analytical results for QC samples from the July 2023 sampling trip.

Sample	Date Sampled/ Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br ⁻ (ppm)	Cl ⁻ (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GKW-1-78 (Field Duplicate)	25-Jul-23	367	7.45	3250	<0.067	16.8	1800	2120	1.13	236	254	12.7	202	41.78	45.29	-8.06
PW-736-78	25-Jul-23	365	7.43	3260	<0.067	17.3	1800	2380	1.17	246	273	13.3	212	44.29	45.26	-2.16
RPD	—	0.5	NC	0.3	NC	2.9	0.0	11.6	3.5	4.1	7.2	4.6	4.8	—	—	—
GKW-2-78 (Field Duplicate)	26-Jul-23	377	7.30	2290	<0.067	9.11	1030	1840	0.858	181	225	21.9	48.5	30.21	29.24	3.26
GSP-4-78	26-Jul-23	377	7.41	2280	<0.067	8.94	1010	1710	0.871	184	228	22.2	49.4	30.65	28.82	6.17
RPD	—	0.0	NC	0.4	NC	1.9	2.0	7.3	1.5	1.6	1.3	1.4	1.8	—	—	—
GKW-3-78	Bottle Blank	<0.725	5.88	1.41	<0.067	0.170	<0.133	<2.38	<0.015	<0.05	<0.11	<0.05	0.165	NC	NC	NC
GKW-6-78	Equip. Blank	<0.725	6.21	1.23	<0.067	<0.067	<0.133	<2.38	0.0246	0.142	<0.11	<0.05	0.110	NC	NC	NC
GKW-4-78	27-Jul-23	81.5	9.21	404	<0.067	46.9	22.2	299	35.9	<0.05	<0.11	18.8	73.4	NC	NC	NC
NSI Standards	MPV	89.8	NR	397	—	51.5	22.8	300	—	—	—	—	—	NR	NR	NR
Percent Error	—	9.2	NC	1.8	NC	8.9	2.6	0.3	NC	NC	NC	NC	NC	—	—	—
GKW-5-78	27-Jul-23	79.0	9.23	333	<0.067	38.8	10.2	239	0.921	<0.05	<0.11	<0.05	0.214	NC	NC	NC
ERA Standards	MPV	78.1	NR	325	NR	41.7	10.6	272	0.881	NR	NR	NR	NR	NR	NR	NR
Percent Error	—	1.2	NC	2.5	NC	7.0	3.8	12.1	4.5	NC	NC	NC	NC	—	—	—

ERA = Environmental Resources Associates; GKW = QC sample; GNW = Genie new well; GSP = Genie spring; GSW = Genie surface water; MPV = most probable value; NC = not calculated; NR = not reported; STD = standard; TDS = total dissolved solids.

Table 3.4. Analytical results for QC samples from the October 2023 sampling trip.

Sample	Date Sampled/ Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (µS/cm)	Br ⁻ (ppm)	Cl ⁻ (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GKW-1-79 (Field Duplicate)	10-Oct-23	424	7.38	4010	0.143	18.1	2290	4140	1.92	234	311	17.8	422	56.07	56.67	-1.06
PW-735-79	10-Oct-23	429	7.37	3970	<0.067	23.0	2330	3840	2.00	241	342	18.7	421	58.95	57.74	2.08
RPD	—	1.2	NC	1.0	NC	23.8	1.7	7.5	4.1	2.9	9.5	4.9	0.2	—	—	—
GKW-2-79 (Field Duplicate)	10-Oct-23	413	7.58	3960	0.164	20.6	2390	3800	1.73	252	360	18.5	302	55.80	58.60	-4.90
GSP-2-79	10-Oct-23	413	7.48	4040	<0.067	20.4	2420	3760	1.68	260	391	19.7	311	59.17	59.22	-0.08
RPD	—	0.0	NC	2.0	NC	1.0	1.2	1.1	2.9	3.1	8.3	6.3	2.9	—	—	—
GKW-3-79	Bottle Blank	<0.725	5.67	6.43	<0.067	<0.067	<0.133	<2.38	<0.015	<0.05	<0.11	<0.05	0.387	NC	NC	NC
GKW-6-79	Equip. Blank	1.00	6.21	1.49	<0.067	<0.067	<0.133	<2.38	<0.015	0.162	<0.11	<0.05	<0.1	NC	NC	NC
GKW-4-79	11-Oct-23	82.5	9.12	356	<0.067	35.0	40.5	256	29.3	<0.05	<0.11	31.7	55.2	NC	NC	NC
NSI Standards	MPV	72.7	NR	370	NR	35.9	40.9	281	NR	NR	NR	33.3	58.2	NR	NR	NR
Percent Error	—	13.5	NC	3.8	NC	2.5	1.0	8.9	NC	NC	NC	4.8	5.2	—	—	—
GKW-5-79	11-Oct-23	142	9.28	449	<0.067	46.8	28.7	376	0.960	<0.05	<0.11	0.131	0.424	NC	NC	NC
ERA Standards	MPV	123	NR	472	NR	49.9	29.5	409	0.881	NR	NR	NR	NR	NR	NR	NR
Percent Error	—	15.4	NC	4.9	NC	6.2	2.7	8.1	9.0	NC	NC	NC	NC	—	—	—

ERA = Environmental Resources Associates; GKW = QC sample; GNW = Genie new well; GOW = Genie old well; MPV = most probable value; NC = not calculated; NR = not reported; PW = private monitoring well; STD = standard; TDS = total dissolved solids.

A summary of the results of QC samples from the October 2023 sampling trip is presented in Table 3.4. No constituents were detected at significant levels in the bottle or equipment blanks (GKW-3-79 and GKW-6-79, respectively). The duplicate samples collected at PW 735 and GSP 2 demonstrated good sampling and analysis reproducibility with all RPDs <10% except for chloride (24%) at PW 735. The chloride concentration in sample PW-735-79 was slightly high compared to previous trends, so the sample was reanalyzed. Unexpectedly, the reanalysis result for chloride was also elevated; thus, the reason for the discrepancy is unknown. Two blind standards similar to those from July were submitted to the laboratory with the October 2023 samples. The first was a standard from NSI Lab Solutions (GKW-4-79) and the second consisted of two ERA WasteWatR™ standards (GKW-5-79). All results for the blind standards were acceptable.

Five samples from July had conductivity values that were ~10-20% lower than the results from the same sites in 2022. The samples were from wells GNW 2 (Cow Creek), GNW 3 (Cow Creek), GNW 6 (South Fork Cow Creek), GNW 10 (Pony Creek), and GOW 11 (South Fork Cow Creek). The samples were reanalyzed, and the reanalysis results were consistently higher than the original values. However, the new results still appeared to be anomalously low. The analytical laboratory investigated the results in question, but they were unable to identify any problems with the analyses. None of the suspicious results were identified as outliers during the statistical analysis (Section G.1.2).

Four samples collected in October from GNW wells 2 through 5 had minor charge imbalances and metals results that were slightly low compared to historical data. The samples were reanalyzed, and in all cases, the reanalysis results were more consistent with previous data. No systematic concerns were identified for metals results at other sites for samples collected in July or October.

Agreement between field and laboratory pH and conductivity readings was reasonable for both sampling campaigns this year. On average, the field pH readings were approximately 0.3 pH units lower than the laboratory values. For conductivity, the field readings also tended to be lower than the lab values. Some of the discrepancies may be due to chemical changes in the samples that occur between the time the measurements are made in the field and at the laboratory (e.g., outgassing of samples from deeper wells and redistribution of inorganic carbon). Most of the conductivity measurements made in the field and at the laboratory agreed within 10%. One of the surface water samples from October, GSW-6-79, had a notably lower field value (6,550 $\mu\text{S}/\text{cm}$) compared to the lab result. The reason for the poor agreement at this site is unknown, but the lab value is more consistent with the data from this site from the past two years. Field data are not used for statistical trend detection.

3.5 Statistical Analysis of Chemical Data

The primary objectives of the 2023 hydrologic/water-quality study were to determine whether there were changes in water quality in the Cow Creek, South Fork Cow Creek, and Pony Creek drainages from the previous year, and to establish the current characteristics and conditions of water quality within the three drainages studied. Stipulation 12(d) identifies two parameters that indicate leakage: conductivity level and boron concentration. In addition, high sulfate values are considered a signature of leakage from the EHP. To help meet the study objectives, a statistical analysis was performed on the water-quality data, focusing on these three indicators. The objectives of the statistical analysis are described in the next two sections.

3.5.1 Objective 1: Identify Any Time Trends at the Site or Drainage Level in the Parameters Conductivity, Sulfate, and Boron

The *site-level* analysis is performed to detect any consistently increasing trends at individual sites; this is of primary importance for decision-making. Stipulation 12(d) mandates that if an abrupt increase occurs

in conductivity or boron at any well between the Talen Montana/Genie Land Company property line and Stinking Spring (GSP 2) or at any point where an adequate interception system could be constructed, Talen Montana would intercept the reduced-quality water. Therefore, the site-level trend analyses are primarily for detecting and quantifying any increasing trends at the Cow Creek or South Cow Creek sites. The trend analyses for sites at the Pony Creek drainage are useful for comparing and verifying any detected trends at South Cow Creek or Cow Creek. For example, the Pony Creek analysis can help verify whether any increasing trends at the other drainages are due to changes in regional hydrogeochemical properties or weather-related events. Pony Creek is hydrologically and geochemically similar to both Cow Creek drainages and is presumed to be unaffected by EHP activities. Therefore, any time trends detected at Pony Creek sites are expected to result from natural changes in drainage flow conditions. However, mining reclamation (Rosebud Mine Area D) is under way on the north and south sides of Pony Creek approximately 0.5 mi upgradient of GSP 4, and local hydrogeological conditions may differ somewhat between the Cow and Pony Creek drainages because of ongoing activities or natural differences. Alternatively, if transport of EHP effluent by storm-related creek flow events occurred at Cow Creek and South Fork Cow Creek, nearly simultaneous trends at multiple sites could ensue in both drainages. The possibility of contamination by this pathway should therefore be ruled out before concluding that all increasing trends are weather related. Spatial waterfall plots (introduced in last year's study) can provide a visual indication of this mode of contamination, as discussed in Section G.3.

Our site-level analysis also tests for decreasing trends that may indicate plume movement through the drainage as exhibited by the reduction of key parameters at certain sites. Under accidental-release conditions that produce a distinct plume, the concentrations of selected parameters at an affected site will eventually decrease as the plume moves further downgradient. Therefore, identification of recent decreases can be useful for characterizing the extent of contamination and for providing insights about the rate of contaminant migration. Decreasing trends may also be caused by naturally changing aquifer characteristics or climatological factors. To supplement these tests and the tests for increasing trends, spatial waterfall plots (Section G.3) should qualitatively reveal any movement of one or more plumes through a drainage and also assist in our distinguishing between contaminant- and climatological-induced changes.

In addition to site-level time-trend analyses, *drainage-level* time trends for each of the three indicator parameters are assessed using data from several sites in each drainage. Because drainage-trend estimates possess considerable uncertainty and reflect spatial averaging, they are more suitable for qualitatively depicting mass-balance changes within a drainage. Further discussion of drainage trends is provided in Appendix G, Section G.4.

3.5.2 Objective 2: Estimate and Compare Overall (Time-Averaged, Site-Averaged) Drainage-Mean Values for Cow Creek, South Fork Cow Creek, and Pony Creek

Average constituent levels at the drainages are evaluated as part of an ongoing characterization effort. Pony Creek represents a control drainage, although hydrogeologic conditions in Pony Creek are similar but not identical to those in the Cow Creek and South Fork Cow Creek drainages.

3.5.3 Sites and Data Used in the Statistical Analyses

The sites used for the statistical analysis are identified in Table 3.5.

Table 3.5. Wells and springs used in the statistical analysis.

Cow Creek	Pony Creek	South Fork Cow Creek
PW 735 ^(a)	GSP 4	GSP 3
PW 736	GSP 6	GOW 11
GNW 1	GOW 3	GNW 6
GSP 1	GOW 4	GNW 7
GNW 2	GSP 7	GNW 8
GNW 3	GOW 12 ^(b)	--
GOW 1	GSP 8	--
GNW 4	--	--
GSP 2	--	--
GNW 5	--	--

(a) Used for site-level trend analysis only.

(b) Only past data from GOW 12 were used, and only at the drainage-level as discussed in Section 3.6.2.

GNW = Genie new well; GOW = Genie old well; GSP = Genie spring; PW = private monitoring well installed by Hydrometrics, Inc.; GSP 2 is also known as Stinking Spring.

In general, these sites are the same as those examined by Olsen et al. (1987, 1991–2006), McDonald et al. (2007), and Thompson et al. (2008–2023). Sites added since earlier studies are wells PW 735 and 736, which have been included since the early 2010s. Data from site GOW 12 have been excluded after sampling event 60. A new well was installed at GOW 12 in 2014, and the water chemistry of the new well is distinct (e.g., much lower sulfate) from that of the former well. The statistical site analysis will resume inclusion of GOW 12 if these anomalies resolve and after sufficient data for reliable testing have been collected for the new well.

The ordering in Table 3.5 corresponds to the distance of each site from the headwaters (closest first). These wells and springs for sampling groundwater were selected as the ones most likely to yield results representative of each aquifer. Site W-1 was not included because it is downstream from the confluence of Cow Creek and South Fork Cow Creek; therefore, any effects at site W-1 are not uniquely attributable to either of these drainages. This precludes its usefulness in identifying differences in drainage characteristics, either temporal or between drainages. Also, any site trend at W-1 is also of no value for site-based inferences of drainage conditions. However, W-1 data could be useful for other investigations that are currently outside the scope of this report. For example, W-1 could be used to represent a “composite” of Cow and South Fork Cow drainages, in a comparison with selected downgradient sites at Pony Creek. Therefore, W-1 continues to be included in the sampling campaigns.

Data from sampling years 1985–1987 and 1990–2023 were examined. Statistical analyses were performed for conductivity ($\mu\text{S}/\text{cm}$), boron (B, ppm), and sulfate (SO_4^{2-} , ppm). Data from samples obtained in 2023 are listed along with other parameters in Appendix F, and the data from previous years are provided in earlier reports (Olsen et al. 1992–2006; McDonald et al. 2007; Thompson et al. 2008–2023).

Outliers in the data (defined in Section 3.4) were identified and excluded from all statistical analyses as described in Appendix G, Section G.1.2. In this year’s outlier investigation, conductivity at GNW 5, event 1 was removed because it was inordinately low and inconsistent with the adjacent years’ data. Recent results from GSP 7 (with the exception of the low conductivity value from event 70) continue to be retained in the analysis because those results appear to be more representative than data from 2017–2018

based on conditions observed at that site during the sampling trips. The data from GOW 12 after sampling event 60 continue to be omitted from the drainage time trend and drainage comparisons because their values are unrepresentative of the Pony Creek aquifer.

All outlying data were kept in the database but were omitted from the site- and drainage-level time-trend analyses and drainage comparisons. Outliers excluded as a result of previous outlier analyses are provided in earlier reports (e.g., Thompson et al. 2008–2023), and all cumulatively excluded data are identified in Appendix G, Table G.2.

3.5.4 Statistical Testing Methods

Two statistical tests were used to objectively evaluate whether changes in the indicator parameters occurred at individual sites during the most recent four years. A cursory description of both methods is provided below; Appendix G includes more detailed information. When testing for Stipulation 12(d) parameters boron and conductivity, the two tests are evaluated as a composite test. A recent increase is concluded if *either* test determines that an increase occurred. The conventional test only was applied to the remaining time trend tests— i.e., site-level tests for increases in sulfate or decreases in any of the three parameters, and all drainage level tests (all drainages, all parameters, increases, and decreases).

All conclusions regarding temporal changes and in the comparisons of drainages are based on classical statistical testing, which *assumes a default hypothesis and only abandons that hypothesis if the observed data are highly incompatible with that condition*. The default hypotheses in our analyses are 1) there is no increasing or high-valued trend (and no decreasing trend when testing for decreases), and 2) in the drainage-comparison tests, the drainage constituent means are all equal.

Testing for increases involves assessing how likely it would be to observe our measured data values if the underlying trend curve is actually not increasing. To evaluate this likelihood, we use a *test statistic* that is based on the data. To test our hypothesis, we generate a large number of simulated datasets under that assumption (e.g., each having a non-increasing trend of the type corresponding to each test as described below) and then tally the percentage of test statistics that are similar to the one we have observed (i.e., based on the actual monitoring data). If this percentage is very low, then we conclude that an underlying increasing trend exists (i.e., the data are inconsistent with the no-trend hypothesis).

3.5.4.1 Conventional Test

Our “conventional” test has been used for most of the duration of this monitoring program. The test is used to detect both increasing and decreasing trends for all three indicator parameters at individual sites and at each drainage. In brief, the conventional test uses a locally flexible curve-fitting algorithm to determine whether a steady increase or decrease occurred over the most recent four years (or longer). If this curve exhibits a continual recent change, then the same curve-fitting algorithm is applied to a large population (100,000) of simulated datasets (each based on the assumption of a flat recent trend curve with randomized sampling and analysis errors added). If fewer than 0.07% (0.16% for composite testing) of the curves obtained from the simulated datasets exhibit the same type of change (increasing or decreasing) as the curve based on original monitoring data, then a continual trend is concluded. For example, in applying this algorithm to testing for recent increases, by adding randomly selected observed (realistic) errors to an assumed recently flat trendform, plausible data series under this assumption are repeatedly generated, and we are evaluating how likely (how frequently among 100,000 iterations) it would be to observe a recent increase among curves fitted to each of those data series. If fitted recent increases are very rare, then we reject the flat-curve assumption and accept that there is actually a recent increase in the underlying trend. The underlying trend is defined as the curve that we would observe if there were no sampling and analysis errors.

3.5.4.2 PLR Test

The Predictive Likelihoods Ratio (PLR) test has been used to assess trends in this program since 2017. This test is applied to detect recent increases in conductivity and boron at the site level only. In the PLR test, a large population (1,000,000) of datasets is generated based on nonincreasing trends with random measurement errors added. These datasets represent scenarios that could occur if the true trend was not increasing—each dataset consists of a randomly selected, non-increasing trend with random measurement (sampling plus analysis) errors added. For each dataset, a PLR statistic is calculated that is the probability that the data represent an increasing trend divided by the probability that the data represent no increasing trend. This statistic tends to be larger when there actually is an increasing trend. Consequently, if fewer than 0.16% of the PLR values exceed the PLR computed from the original data, we conclude that there is an increasing recent trend (i.e., it is unlikely, under a non-increasing situation, to observe a PLR as large as the one computed from the actual monitoring data).

Performance-wise, the PLR test is far superior to the conventional test because the PLR test has greater detection ability when actual trends exist, and the PLR test's false-detect rate is exactly controllable. The conventional test is nevertheless retained in our analysis as part of the composite test, for all other site trend testing and for drainage-trend tests. The conventional test may perform at its best for longer detected trends, and its results often may be visually verified when the data vary only slightly around an increasing estimated trend.

3.5.4.3 Evaluation of Drainage Means

Drainage averages are evaluated by comparing box plots. Medians and 95% confidence intervals are used for estimating and comparing overall drainage means. Appendix G, Section G.4.5 provides more information about the drainage comparisons.

3.5.5 Results of Statistical Analysis

This discussion of results is based on the time-trend and drainage-comparison analyses described in Appendix G of this report. More technical foundations for the methods are provided by Thompson et al. (2014) and Chamberlain (2018).

3.5.5.1 Assessment of Drainage Data over Time

For each constituent, the average of non-outlying data among all sites within each drainage for 1985-1987 and 1990-2023 was examined for strict trends over the most recent four or more years (see Appendix G, Figure G.8). Drainage averages from each campaign were used as individual observations. As indicated in Table 3.6, no continual recent drainage-level trends—either increasing or decreasing—were found at any of the drainages.

3.5.5.2 Assessment of Sites over Time

We examined sampling data for each site for 1985-1987 and 1990-2023 for 1) increasing or extremely high-valued time trends in conductivity or boron (using the composite test discussed earlier and in Appendix G, Section G.3.3), 2) decreasing trends at any site, and 3) increasing trends in sulfate at any site. All trends were defined as continual (steadily increasing or decreasing) over the last four or more years. Test results are given in both Figure G.9 and in the overlaid time plots of the sites in Appendix G, Figure G.10. Summarized results of these trends also are listed below in Table 3.6.

Increasing trends in conductivity were found at Pony Creek sites GSP 4 and GOW 4. Increasing trends for boron were concluded at Pony Creek site GSP 6 and Cow Creek site PW 735, and a decreasing boron trend was found at Pony Creek site GOW 4. No sulfate trends, either increasing or decreasing, were found at any site.

This is the 4th consecutive year in which an increasing boron trend was detected at PW 735. However, in contrast to the previous 3 years, no increasing conductivity trend was found at that site. The boron trend appears to be tapering off (successively smaller over the past 3 years) based on its posterior distribution (Figure G.7) which indicates that the trend does not exceed 15% (over the recent 4 years); this upper limit is considerably lower than the maximum increases of 25% and 53% (as indicated in the 2022 and 2021 reports, respectively). Well PW 735 taps a deeper aquifer than that supplying the nearby alluvial wells.

Table 3.6. Time trends detected in 2023.^(a,b)

Site	Conductivity (μS/cm)	B (ppm)	Sulfate (ppm)
Cow Creek PW 735	No site trends	No decreasing site trends +(a,c)	No site trends
Drainage level	No drainage trend	No drainage trend	No drainage trend
South Fork Cow Creek	No site trends	No site trends	No site trends
Drainage level	No drainage trend	No drainage trend	No drainage trend
Pony Creek	No decreasing site trends		No site trends
GSP 4	+		
GOW 4	+	-(d)	
GSP 6		+	
Drainage level	No drainage trend	No drainage trend	No drainage trend

(a) Blue text indicates composite test results (conductivity and boron only), and black text conveys standard test results (all three parameters).

(b) No increases were detected at any sites identified in Stipulation 12(d).

(c) “+” composite test concluded an increasing time trend (total 5% level of false detection error among all composite tests in the drainage).

(d) “-” standard trend test found a decreasing time trend (total 5% level of false-detection error among all non-composite trend tests, including drainage-level tests).

GNW = Genie new well; GOW = Genie old well; GSP = Genie spring; PW = Private monitoring well.

3.5.5.3 Comparisons between Drainages

Data for all sites within a drainage were pooled across time so the drainage means could be compared (Appendix G, Figure G.11). The estimated drainage-mean values and their 95% confidence intervals are provided in Table 3.7. For the overall observation period (1985–1987 and 1990–2023), the mean values for conductivity and sulfate in the Pony Creek drainage are less than the values for the South Fork Cow Creek drainage, both of which are less than those in the Cow Creek drainage (Table 3.7). For boron, the mean values in both the Cow Creek and South Fork Cow Creek drainages are not statistically different, but each value is greater than the mean in the Pony Creek drainage.

Table 3.7. 95% confidence limits on drainage-mean values using all data.^(a)

Parameter	Cow Creek	Pony Creek	South Fork Cow Creek
Conductivity (μS/cm)	3,495 ±64	2,010 ±104	2,600 ±115
Boron (ppm)	1.40 ±0.03	0.60 ±0.05	1.39 ±0.06
Sulfate (ppm)	2,000 ±40	770 ± 62	1,240 ±86

(a) 1985–1987 and 1990–2023 data

3.5.6 Overall Conclusions of the Statistical Analysis

The *drainage time-trend test results* are as follows:

- No recent increasing or decreasing drainage trends for conductivity, boron, or sulfate were found at any of the drainages.

The *site-level time-trend test results* are as follows:

- No increases were detected at any of the sites identified in Stipulation 12(d).
- An increasing trend was found for boron at Cow Creek site PW 735, but this trend is likely tapering off.
- Based on the inclusion of the 2023 monitoring data, there is no longer an increase in conductivity at Cow Creek site PW 735 (a recently increasing trend was reported last year [Thompson et al. 2023]).
- Increasing recent trends in conductivity were detected at Pony Creek sites GSP 4 and GOW 4.
- Two boron trends were detected at Pony Creek: a recent increase at site GSP 6 and recent decrease at site GOW 4.

Drainage comparisons indicated that the Pony Creek drainage had the lowest values for conductivity, sulfate, and boron. The Cow Creek drainage was distinctly highest in conductivity and sulfate, and the South Fork Cow Creek drainage means fell between the values of the other drainages except for boron, whose mean did not differ from that of Cow Creek.

3.6 Results and Discussion

This section summarizes the monitoring data collected during 2023 and is organized into subsections for monitoring locations (alternate supply wells and the three creek drainages), bromide measurements, and continuous-monitoring data at well PW 736. Recent trends identified in the preceding section are interpreted in terms of their significance, based on the magnitude of change relative to previous observations.

Drainage profiles of conductivity, sulfate, and boron are plotted in Figure 3.4, Figure 3.10, and Figure 3.12, respectively, in ensuing sections. These parameters were selected based on the requirements in Stipulation 12(d) and because they are the most likely indicators of leakage from the Colstrip Units 3 and 4 EHP. For reference, Table 3.8 lists composition data from a sample that was collected from the EHP in 2015. There were no reportable releases from the EHP area during 2023.

Drainage-mean values of conductivity, sulfate, and boron are plotted for each drainage in Figure 3.9, Figure 3.11, and Figure 3.14 (in ensuing sections) as a function of time from 1985 through 2023. Piper diagrams illustrating the geochemical water types of the GAS wells and sites along the three creek drainages were presented in the 2010 report (Thompson et al. 2011). Because the water chemistry has not changed appreciably in the past 13 years, those plots were not updated for this report.

Table 3.8. Composition of EHP solution (based on a sample collected in May 2015).

Sample Site	Alkalinity (ppm CaCO ₃)	pH	Cond (μS/cm)	Br ⁻ (ppm)	Cl ⁻ (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)
3+4 EHP B-CELL	<4	3.0	32,200	1940	1,160	35,600	34,400	278	560	8,070	132	2,660

The temporal profiles that follow are used qualitatively to support the discussion of the analytical data. For quantitative estimates of drainage time trends and their corresponding tests of statistical significance, refer to Appendix G, Figure G.8. Except where noted, outlying data (as identified in Appendix G, Section G.1.2) were typically not removed in the calculations of these profiles.

The site-specific average values were used to generate temporal drainage-mean values (see Appendix E) in each of the tables for Cow Creek, South Fork Cow Creek, and Pony Creek. The individual analytical results for each sample collected during 2023 are listed in Appendix F. All less-than (<) values were ignored when an average value was calculated for a site. When only a single value was present, no mean value was calculated.

3.6.1 Sampling Data from the Genie Alternative Supply Wells

The GAS wells are located in the Cow Creek and South Fork Cow Creek drainages. These wells were installed to provide alternative supplies of water for stock in lieu of surface and upper unconfined groundwater that has elevated sulfate concentrations. Stipulation 12(d) does not apply to the GAS wells, nor are they currently useful for water quality comparisons at other sites; therefore, statistical analyses were not applied to the GAS wells. These wells are being monitored to ensure that their water quality is not being negatively affected by potentially poor-quality alluvial groundwater within the drainages. Because of their depth, they have the potential to tap deeper aquifers than the GNWs or GOWs in the drainages. Chemical data were used to classify the GAS wells according to their possible water sources. Initial geochemical description of the GAS wells was presented by Olsen et al. (1991) and was updated in the 2010 report (Thompson et al. 2011). Briefly, GAS 1 taps what is suspected to be a shallow alluvial aquifer that contains a magnesium/calcium sulfate and sodium/potassium sulfate mixed-water system very similar to that of GOW 1 and the GNW wells within the Cow Creek drainage. GAS 1 is not being used as an alternate supply well because its water chemistry is similar to that of the GNW wells within the Cow Creek drainage (it may be tapping the same alluvial aquifer). The geochemical compositions of water from GAS 2 and GAS 3 are somewhat similar. GAS 2 taps an aquifer containing a sodium sulfate/sodium bicarbonate mixed-water system, and GAS 3 draws water from a sodium sulfate aquifer system.

Wells GAS 1, GAS 3, and GAS 7 were sampled in July and October, while GAS 2 was sampled only in July. The mean analytical values for conductivity, sulfate, and boron from water samples collected in the GAS wells are provided in Table 3.9.

Table 3.9. Values for conductivity, sulfate, and boron in samples collected from GAS wells in 2023.

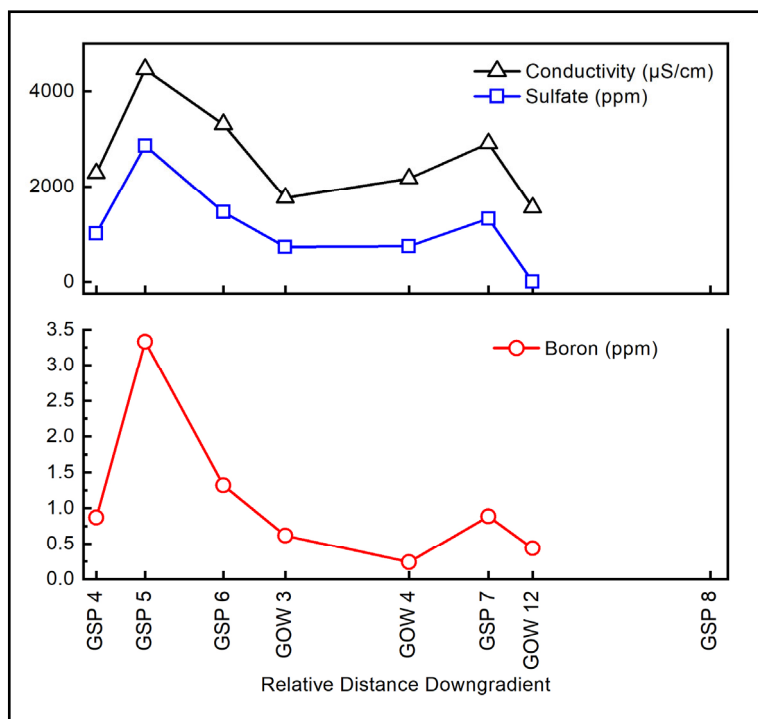
Site	Comment	Conductivity (μS/cm)	Sulfate (ppm)	Boron (ppm)
GAS 1	Mean value	2,775	1,205	0.942
GAS 2	Single value	1,690	413	0.325
GAS 3	Mean value	2,780	996	0.299
GAS 7	Mean value	1,990	578	0.298

Among the GAS wells, GAS 3 had the highest value for conductivity, while GAS 1 had the highest concentrations for sulfate and boron. GAS 1 and GAS 3 had similar values for conductivity and sulfate, but boron concentrations were significantly higher at GAS 1. These relative patterns were essentially unchanged from those of the past 12 years. The overall stability of these parameters and their relatively low values indicate that the GAS wells continue to be unaffected by anthropogenic activities. All chemical data from the GAS wells can be found in Appendices E and F.

3.6.2 Sampling Data from Pony Creek Drainage

The Pony Creek drainage contains 12 sampling points: 6 springs, 4 GOW stock-watering wells, and 2 GNW wells. Because the Pony Creek drainage is used to assess background hydrologic and geochemical conditions, its sampling sites are monitored only once per year. All the sites except GSP 8 were sampled during July 2023 (GSP 8 was dry during the July sampling campaign). Data from GNW 11 and GSP 10 are not representative of the upper alluvial aquifer in the Pony Creek drainage, so they were not included in the statistical analysis (Olsen et al. 1991). Additionally, the values from GOW 12 were omitted for samples after June 2014, because a new well was installed at that location that summer, and subsequent data suggest that the new well taps a deeper source with a different water type that is not representative of the upper alluvial aquifer. If this anomaly resolves, future site-level trend analyses will include results from GOW 12 after sufficient data have been collected from the new well for reliable trend analyses.

Analytical results for the indicator parameters measured at sample sites in the Pony Creek drainage during 2023 are presented in Figure 3.4. Conductivity values ranged from 1,550 $\mu\text{S}/\text{cm}$ at GOW 12 to 4,460 $\mu\text{S}/\text{cm}$ at GSP 5, with a drainage-mean value of 2,815 $\mu\text{S}/\text{cm}$. Sulfate concentrations ranged from 0.418 mg/L at GOW 12 to 2,860 ppm at GSP 5, with a drainage-mean value of 1,353 ppm; and boron ranged from 0.242 ppm at GOW 4 to 3.33 ppm at GSP 5, with a drainage-mean value of 1.21 ppm.



[†]Values for GOW 12 were not included in the drainage mean.

^{*}GSP 8 was dry during the July sampling campaign.

Figure 3.4. Plots and data for conductivity, sulfate, and boron in the Pony Creek Drainage during 2023. Values are single measurements from the July sampling event.

For 2016 and 2017, data from GSP 7 were not included in the trend analyses and drainage-mean calculations because the values were unusually high compared to historical averages and were formally confirmed to be outliers. This appeared to be caused by dry conditions and apparent evaporation at that site (Thompson et al. 2018). Since the July 2019 sampling campaign, more water has been present, and the concentrations have been more consistent with historical data. Trend testing for GSP 7 resumed in 2020.

Four site-specific time trends were detected in the Pony Creek drainage: an increase in conductivity at GSP 4, an increase in conductivity and a decrease in boron at GOW 4, and an increase in boron at GSP 6. These site trends have been observed in previous years. Figure 3.5, Figure 3.8, Figure 3.7, and Figure 3.8 show the cumulative trend plots for these sites and constituents.

The increasing trend in conductivity at GSP 4 has been detected every year since 2008 except 2016, 2018, and 2020. The cumulative trend plot (Figure 3.5) shows strong 1–3 year fluctuations over the past 17 years that are superimposed on an overall increase with time. The fluctuations may be related to the highly variable precipitation since about 2001 (Figure 2.1). The July 2023 result of 2,280 $\mu\text{S}/\text{cm}$ is slightly lower than last year's value of 2,402 $\mu\text{S}/\text{cm}$; these recent data points are the highest conductivity values that have been observed at this site. We suspect the increase is associated with mining reclamation near the upper portions of Pony Creek, although specific causes have not been identified. Mining in the Pony Creek drainage ceased in 2013.

The increasing trend for conductivity at GOW 4 reflects the slightly higher values since 2017 (Figure 3.6). However, the 2022 and 2023 results (2,120 and 2,170 $\mu\text{S}/\text{cm}$, respectively) are lower than many of the values dating back to September 1985. The recent changes are minor compared to previous fluctuations and probably do not indicate impingement of a contaminant plume at this site. Increasing trends for conductivity at GOW 4 were also determined in 2010, 2012, and 2021 with the same interpretation—no obvious plume movement could be concluded based on the relatively small magnitude of the associated 4-year trends and the long-term variability in the entire data set.

A decreasing boron trend was also detected at GOW 4. The trend plot (Figure 3.7) shows relatively large swings between 1985 and 1995 and smaller fluctuations (up to ± 0.1 ppm) during the past 20 years. The estimated trend at this site (Figure G.9) exhibits a gradual decline over the entire study period and was concluded to be correctly decreasing. A decreasing boron trend was also detected at GOW 4 each year from 2015–2021. However, the recent decreasing trend is not indicative of plume passage at this site, because a similar decline throughout the study period is also exhibited at site GOW 12, possibly suggesting insignificant influx of contaminated runoff in the vicinity of these sites prior to 1984 followed by gradual attenuation. Regardless, the boron levels at these sites are all very low. The 2023 boron concentration (0.242 ppm) is similar to the values from 2020–2021 (0.233–0.237 ppm).

Consistent with last year, the increasing boron trend identified at GSP 6 is negligible in magnitude. Boron levels at this site slowly increased between 1989 and 2013 but have been relatively stable since 2014 (Figure 3.8). The latest value from July (1.32 ppm), is nearly identical to the concentration from June 2022 (1.33 ppm) and is also very similar to the values from 2018 and 2019 (1.35 and 1.37 ppm, respectively).

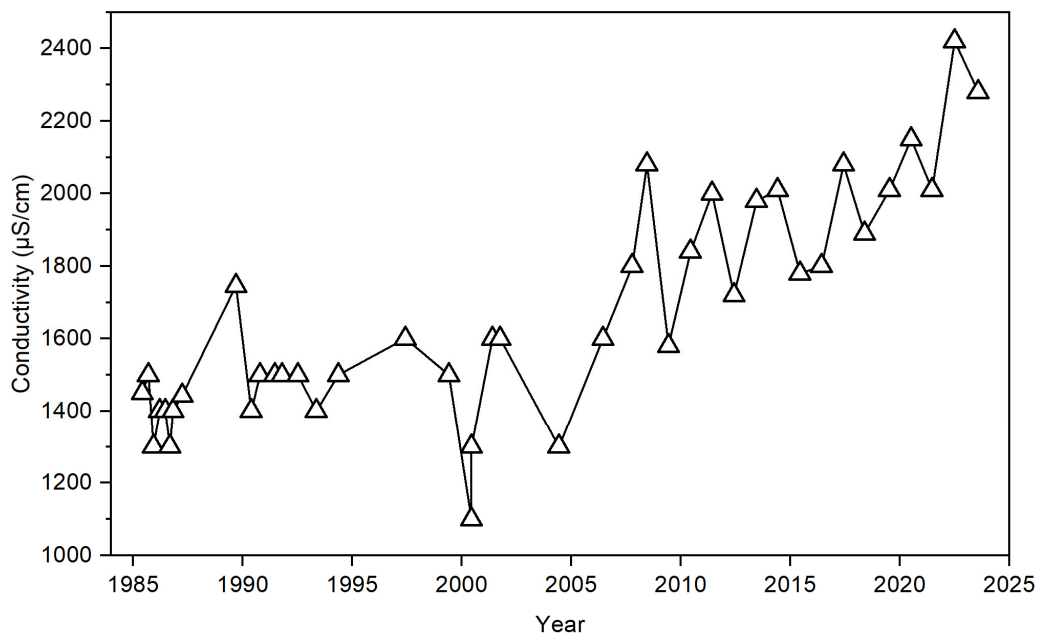


Figure 3.5. Trend plot for conductivity at GSP 4. Outliers identified in Appendix G, Section G1.2, were omitted from the graph.

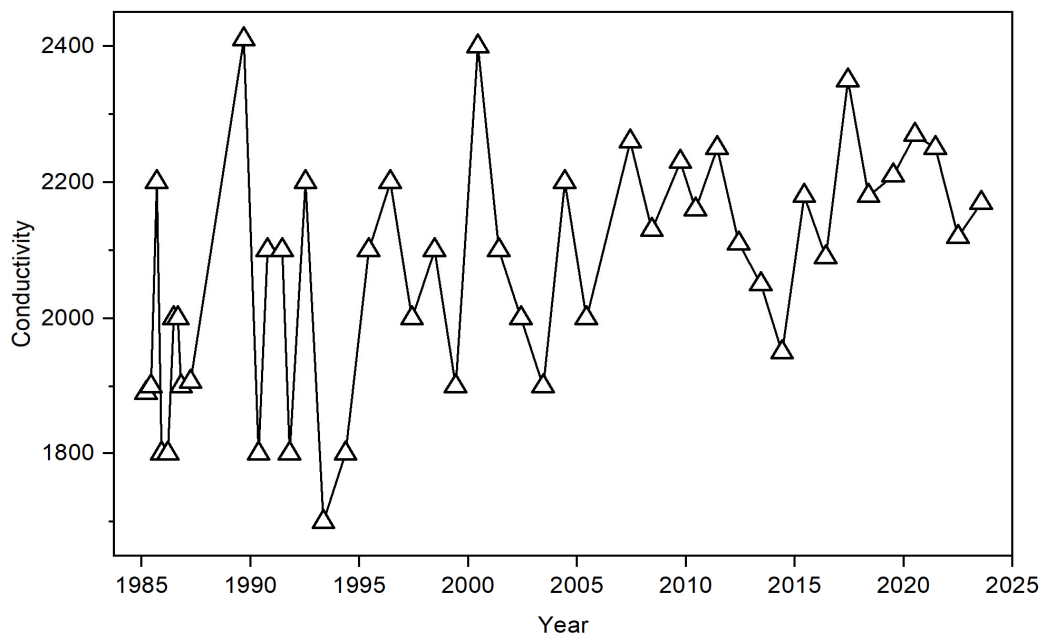


Figure 3.6. Trend plot for conductivity at GOW 4.

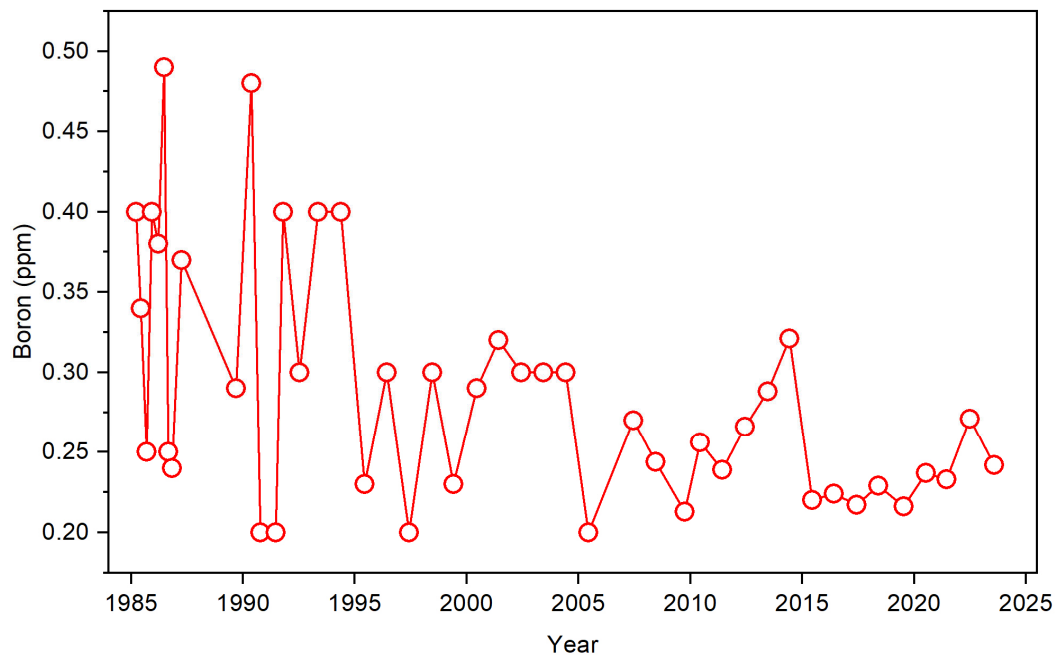


Figure 3.7. Trend plot for boron at GOW 4.

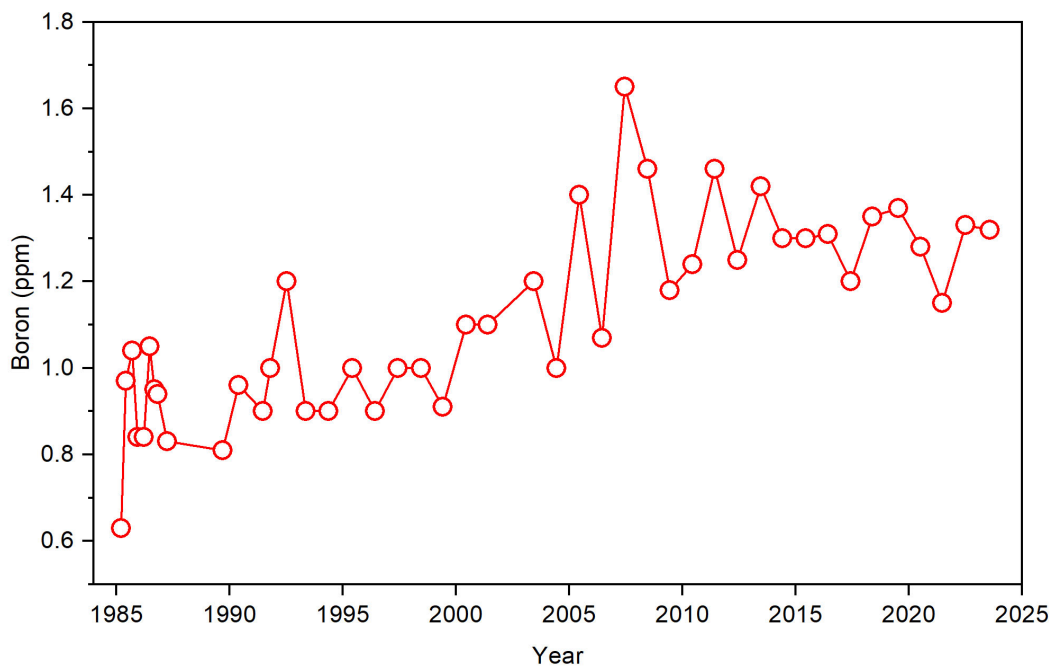


Figure 3.8. Trend plot for boron at GSP 6. Outliers identified in Appendix G, Section G1.2, were omitted from the graph.

Comparing the drainage means over time (1985–1986 and 1990–2023; Figure 3.9 and Table 3.10) show that conductivity values and sulfate and boron concentrations are correlated. Each parameter shows minor fluctuations superimposed on a gradually increasing tendency since about 1990. No drainage-level trends were detected for any of the parameters in 2023. Taken together, the estimated continual increases in

conductivity since 1995 at the 5 uppermost sites in Pony Creek (GSP 4, GSP 6, GOW 3, GOW 4, and [until 2020] GSP 7) seem to indicate long-term, continual contamination via surface flow directly to the springs, and indirectly to the two GOW wells due to recharge at upgradient locations. A contradiction to this explanation is that the downgradient locations among site types have often exhibited higher concentrations, but this can be rationalized if the downgradient sites have naturally different background levels of TDS and other constituents that influence conductivity. As noted earlier, mining activity is underway near GSP 4; the reclamation process involves redistribution of the coal seam's original overburden back over the mined area according to the land's original topography, then adding top soil, and finally reseeding and possibly fertilizing. Previous reports have shown or predicted increased concentrations of dissolved constituents in groundwater from mining spoils near the study area (Metesh 1994, Van Voast and Reiten 1988). Therefore, the above apparent increases at the five uppermost sites at Pony Creek may be attributable to local application of mining spoils and topsoil, and our ability to formally detect increasing trends at only two of these sites could be due to higher sampling variability at the other sites.

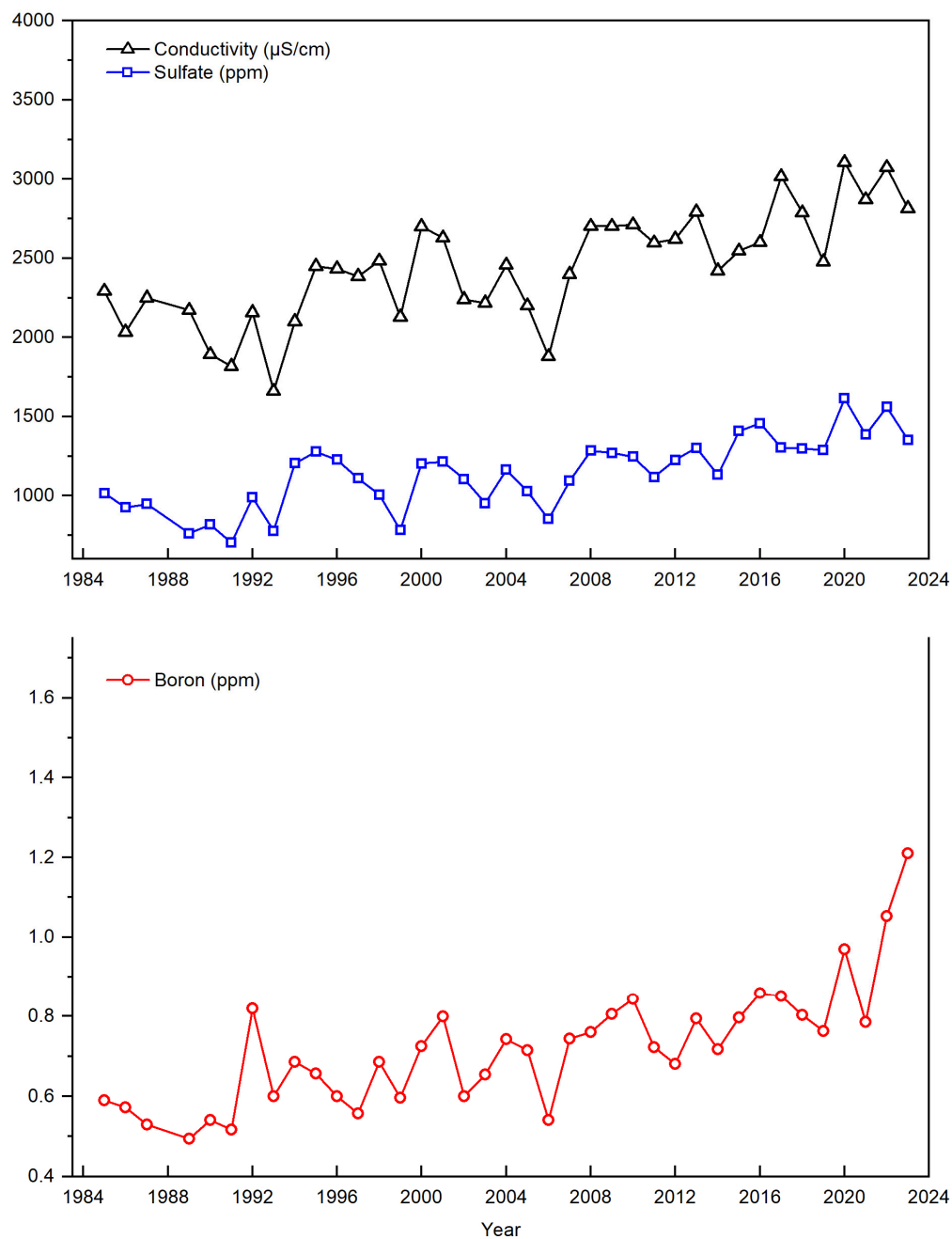


Figure 3.9. Trend plots for conductivity, sulfate, and boron in the Pony Creek Drainage, 1985–2023.

Table 3.10. Drainage-mean values for conductivity, sulfate, and boron in Pony Creek Drainage, 1985-2023.

Year	Drainage Mean		
	Conductivity ($\mu\text{S}/\text{cm}$)	SO_4^{2-} (ppm)	B (ppm)
1985	2,292	1,015	0.59
1986	2,034	926	0.57
1987	2,250	948	0.53
1989	2,171	762	0.49
1990	1,892	818	0.54
1991	1,817	703	0.52
1992	2,157	988	0.82
1993	1,660	778	0.60
1994	2,100	1,205	0.69
1995	2,450	1,278	0.66
1996	2,433	1,227	0.60
1997	2,386	1,110	0.56
1998	2,483	1,006	0.69
1999	2,129	7,83	0.60
2000	2,700	1,203	0.73
2001	2,629	1,215	0.80
2002	2,240	1,105	0.60
2003	2,217	952	0.65
2004	2,457	1,163	0.74
2005	2,200	1,027	0.71
2006	1,880	853	0.54
2007	2,398	1,093	0.74
2008	2,701	1,284	0.76
2009	2,701	1,269	0.81
2010	2,712	1,246	0.84
2011	2,597	1,117	0.72
2012	2,620	1,225	0.68
2013	2,791	1,302	0.79
2014	2,421	1,132	0.72
2015	2,546	1,409	0.80
2016	2,600	1,457	0.86
2017	3,016	1,302	0.85
2018	2,788	1,299	0.80
2019	2,478	1,288	0.76
2020	3,104	1,614	0.97
2021	2,870	1,385	0.79
2022	3,073	1,561	1.05
2023	2,815	1,353	1.21

3.6.3 Sampling Data from South Fork Cow Creek

There are five primary sampling sites within the South Fork Cow Creek drainage (GSP 3, GOW 11, GNW 6, GNW 7, and GNW 8) and one alternative supply well (GAS 3). GSP 3 is situated on a small tributary that feeds into the main flow of South Fork Cow Creek; the remaining four sites are situated along the main stem of South Fork Cow Creek. All six sites within the South Fork Cow Creek drainage were sampled during both the July and October 2023 sampling trips. Although GAS 3 is included in the

South Fork Cow Creek drainage, water from that well is not representative of the upper alluvial aquifer. Therefore, GAS 3 was not included in the statistical analyses.

The 2023 analytical results for indicator parameters at sample sites in the South Fork Cow Creek drainage are presented in Figure 3.10. Values shown are the averages from the July and October sampling events. Conductivity averages ranged from 1,815 $\mu\text{S}/\text{cm}$ at GNW 7 to 3,820 $\mu\text{S}/\text{cm}$ at GOW 11, with a drainage-mean value of 2,779 $\mu\text{S}/\text{cm}$. Mean sulfate concentrations ranged from 721 ppm at GNW 7 to 2,250 ppm at GOW 11, with a drainage-mean value of 1,400 ppm. Mean boron concentrations ranged from 0.826 ppm at GNW 8 to 1.58 ppm at GOW 11, with a drainage-mean value of 1.23 ppm.

Figure 3.10 shows spatial trends for conductivity, sulfate, and boron in the South Fork Cow Creek drainage. Conductivity and sulfate were spatially correlated, with maxima near the upper end of the drainage (GOW 11), decreasing levels from GNW 6 to GNW 7, and increasing values from GNW 7 to GNW 8. Boron concentrations followed a similar profile, with the highest concentrations at GOW 11 and steadily decreasing concentrations down the drainage. No site-specific trends were detected in the South Fork Cow Creek drainage.

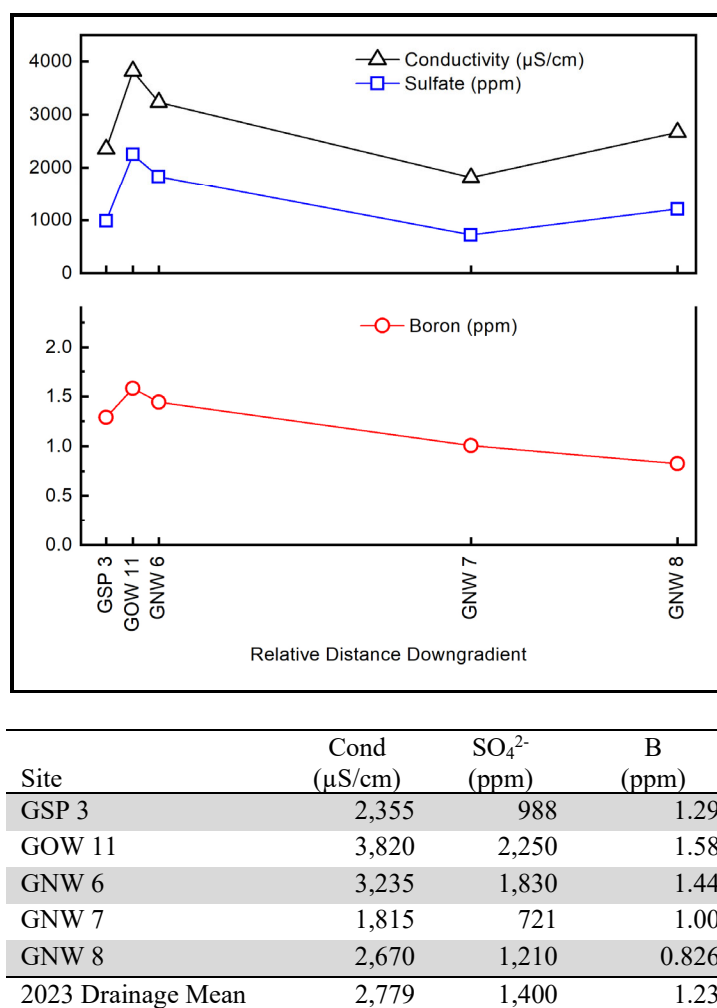


Figure 3.10. Plots and data for conductivity, sulfate, and boron in South Fork Cow Creek Drainage during 2023. Values are averages from the July and October sampling events.

Figure 3.11 and Table 3.11 show the annual drainage means since monitoring began in 1985. No drainage-level trends were detected for South Fork Cow Creek this year. In 2022, an increasing conductivity trend was concluded, but the change was negligible in magnitude and did not coincide with any site-specific trends. Mean values for all three indicator parameters have not exhibited significant changes over the past decade. The 2023 means decreased slightly (10% or less) relative to the averages from last year.

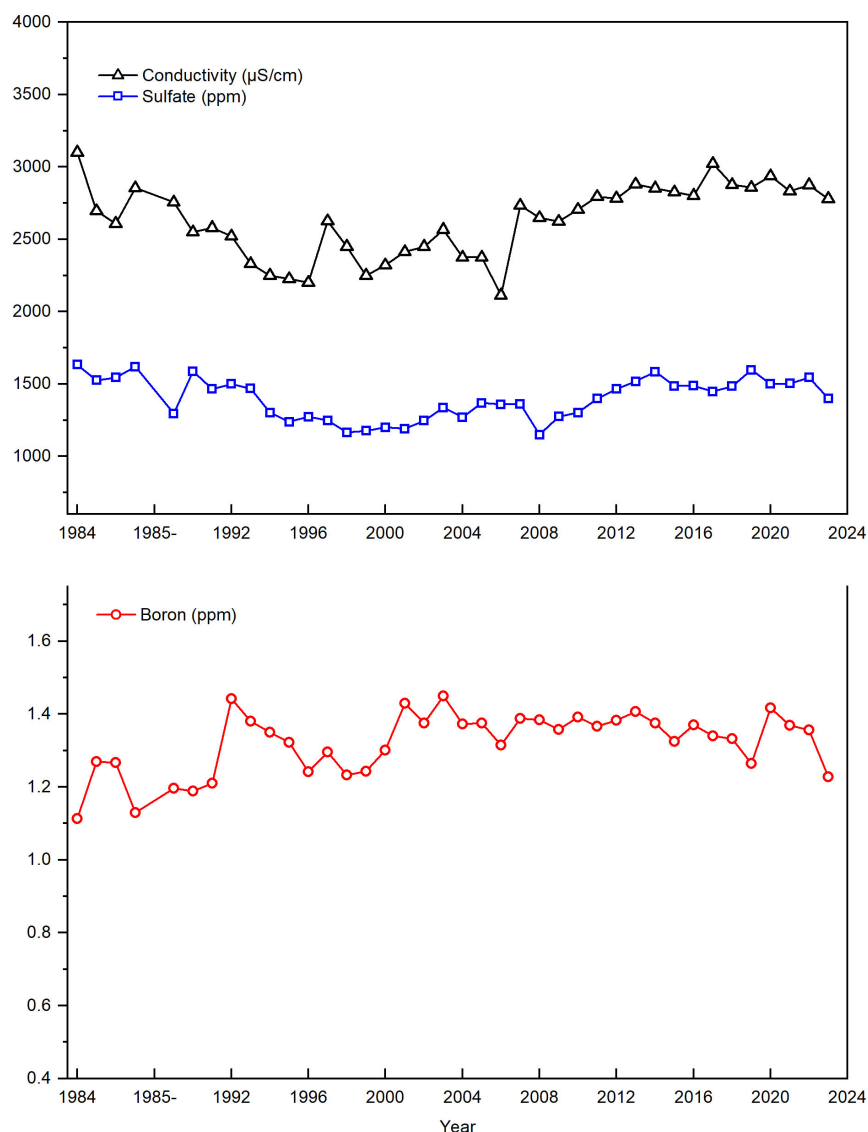


Figure 3.11. Trend plots for conductivity, sulfate, and boron in South Fork Cow Creek, 1984–2023.

Table 3.11. Drainage-mean values for conductivity, sulfate, and boron in South Fork Cow Creek, 1984–2023.

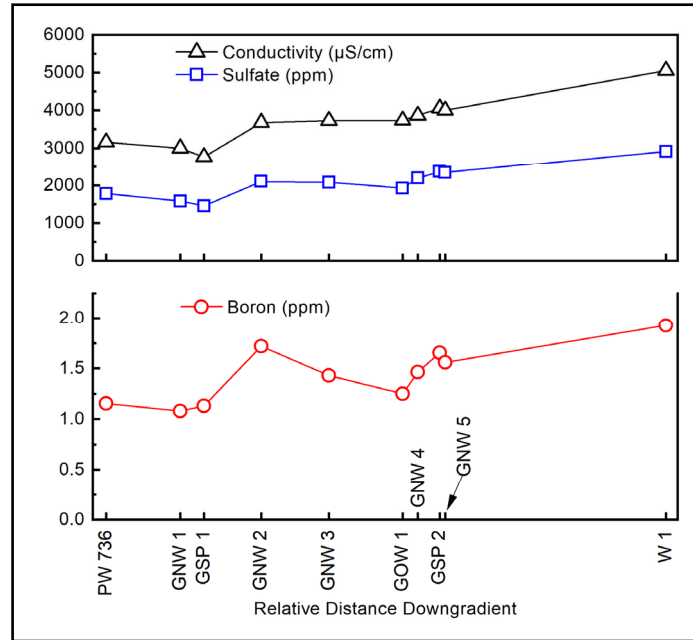
Year	Drainage Mean		
	Conductivity ($\mu\text{S}/\text{cm}$)	SO_4^{2-} (ppm)	B (ppm)
1984	3,100	1,633	1.11
1985	2,695	1,525	1.27
1986	2,608	1,544	1.27
1987	2,853	1,620	1.13
1989	2,755	1,296	1.20
1990	2,550	1,587	1.19
1991	2,580	1,465	1.21
1992	2,520	1,500	1.44
1993	2,330	1,468	1.38
1994	2,250	1,301	1.35
1995	2,225	1,239	1.32
1996	2,200	1,273	1.24
1997	2,625	1,248	1.30
1998	2,450	1,165	1.23
1999	2,250	1,176	1.24
2000	2,321	1,200	1.30
2001	2,413	1,191	1.43
2002	2,450	1,246	1.38
2003	2,567	1,337	1.45
2004	2,375	1,270	1.37
2005	2,375	1,368	1.38
2006	2,113	1,359	1.32
2007	2,734	1,362	1.39
2008	2,648	1,149	1.38
2009	2,623	1,275	1.36
2010	2,705	1,302	1.39
2011	2,793	1,399	1.37
2012	2,782	1,466	1.38
2013	2,880	1,517	1.41
2014	2,853	1,585	1.38
2015	2,826	1,486	1.32
2016	2,802	1,487	1.37
2017	3,023	1,448	1.34
2018	2,876	1,484	1.33
2019	2,857	1,597	1.26
2020	2,936	1,500	1.42
2021	2,831	1,503	1.37
2022	2,875	1,544	1.36
2023	2,779	1,400	1.23

3.6.4 Sampling Data from Cow Creek

Along Cow Creek, 21 sites were sampled in 2023. This included GAS wells 1, 2, and 7 and the three PW monitoring wells near the western boundary of the study area (PW 734, PW 735, and PW 736; Figure 2.4). However, PW 734 had very limited water during both sampling trips. No sample was collected in July, and a reduced-volume sample was collected in October. Surface-water site GSW 3 was dry in July (no sample could be collected) but had sufficient water for sampling in October. Well GNW 9, the GAS wells, and PW 735 were not included in the drainage-mean calculations because most of them do not monitor the upper alluvial aquifer in Cow Creek. However, PW 735 is nevertheless formally evaluated for time trends in target parameters. Well PW 736 taps the alluvial aquifer as do wells GNW 1, 2, 3, 4, and 5, and could be considered a sentinel well that would provide a warning if EHP-contaminated water escaped past the upgradient capture system. Well PW 735 is included in the statistical site-specific trend analyses to determine if there are significant changes in water quality at that deeper location. The 21 monitoring sites in the drainage are divided into three categories: wells, springs, and surface waters. Surface-water sites (GSW) were segregated from the alluvial wells and springs because concentrations of constituents in surface waters are affected by evaporation and dilution, depending on season and weather. Therefore, GSW data also were not included here nor in any of the statistical analyses.

Analytical results for the indicator parameters for springs and monitoring wells within the Cow Creek drainage for 2023 are graphed in Figure 3.12. Site mean conductivity values ranged from 2,765 $\mu\text{S}/\text{cm}$ at GSP 1 to 5,050 $\mu\text{S}/\text{cm}$ at W-1, with a drainage mean of 3,554 $\mu\text{S}/\text{cm}$. Average sulfate concentrations ranged from 1,450 ppm at GSP 1 to 2,910 ppm at W-1, with a 1,978 ppm drainage mean. Mean boron concentrations ranged from 1.08 ppm at GNW 1 to 1.92 ppm at W-1, with a 1.38 ppm drainage mean. The spatial profiles for conductivity, sulfate, and boron in the alluvial wells and springs in Cow Creek drainage (Figure 3.12) show that all three parameters were correlated. Values generally increased down the drainage with localized peaks at GNW 2 and GSP 2. The overall maximum target parameters were found at W-1, which is consistent with tendencies from most previous years.

One site-specific time trend was identified in the Cow Creek drainage: PW 735 had an increasing trend for boron. This was the fourth consecutive year in which an increasing boron trend was observed at PW 735. However, in contrast to the previous four years, an increasing trend for conductivity was not detected in 2024. Figure 3.13 shows the boron concentrations at PW 735 as a function of time. The plot exhibits a sharp decrease between 2010 and 2013, a moderate increase from 2013-2014, and generally increasing values from 2017-2020. Since 2020, boron levels have been relatively stable, fluctuating from approximately 1.9 to 2.0 ppm. As discussed in Appendix G and Section 3.5.5.2, the posterior distributions from the last three years suggest that the increases are tapering off. PW 735 taps a deeper aquifer than most of the alluvial wells along Cow Creek and is monitored for water-quality changes in this deeper zone. We are unable to conclude whether the increasing trends are related to contaminant plume migration; additional monitoring data and a better understanding of the natural background variation are needed. Moreover, it is unclear whether the aquifer tapped by PW 735 is confined (i.e., hydraulically isolated) in the vicinity of the EHP capture system. If this deeper aquifer is confined, it is still possible that there is some exchange with water from the shallower aquifer at an upgradient location. However, such a scenario is speculative and is not confirmed by our monitoring data. The laboratory-measured and continuous-monitoring probe results for conductivity at nearby well PW 736, which monitors the shallower alluvial aquifer, exhibited a slight, continuous increase in conductivity during 2023. There was a brief pulse in September (discussed in Section 3.6.7.2) and a positive offset in the continuous readings following the October sampling event. However, there is no obvious indication that the recent increases in boron (and conductivity from 2018-2022) at PW 735 are impacting the alluvial aquifer.



Site	Cond ($\mu\text{S}/\text{cm}$)	SO_4^{2-} (ppm)	B (ppm)
PW 736	3,155	1,780	1.16
GNW 1	3,000	1,580	1.08
GSP 1	2,765	1,450	1.13
GNW 2	3,680	2,100	1.72
GNW 3	3,730	2,075	1.43
GOW 1	3,735	1,925	1.25
GNW 4	3,865	2,190	1.46
GSP 2	4,055	2,365	1.66
GNW 5	4,000	2,340	1.56
W-1 ^a	5,050	2,910	1.92
2023 Drainage Mean	3,554	1,978	1.38

^a Values for W-1 were not included in the drainage mean.

Figure 3.12. Plots and data for conductivity, sulfate, and boron values for alluvial wells and springs in Cow Creek Drainage during 2023. Values shown are averages from the July and October sampling events.

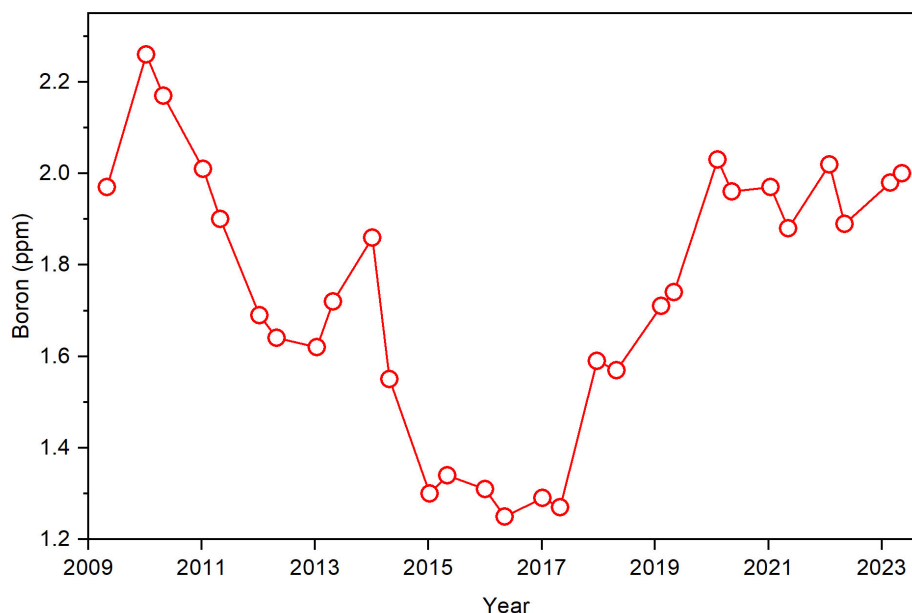


Figure 3.13. Trend plot for boron at PW 735.

Table 3.12 summarizes the 2023 analytical results from the PW wells located near the western boundary of the study area. Limited data was obtained for well PW 734 due to dry conditions. There was insufficient water in July for sampling, but a limited-volume sample of ~200 mL was obtained in October. Due to the reduced volume, the sample was not filtered in the field. The other wells were sampled in July and October. Of the three PWs, PW 736 is believed to be most representative of the alluvial aquifer; therefore, PW 736 is included in the drainage-mean calculations. Well PW 734 taps a region where the uppermost aquifer is relatively shallow (between about 8 and 10 ft), while PW 735 monitors the deeper saturated zone with a screened depth between 35 and 80 ft, and its data are included in the site trend-detection analysis as noted previously. With the exception of pH, all of the parameters in Table 3.12 had their highest average values at PW 735. The highest pH was found at PW 734, followed by PW 736 and PW 735, consistent with the data from last year. Average bromide concentrations were highest at PW 735; PW 736 and PW 734 had non-detected results. Bromide data are discussed in more detail in Section 3.6.6. Levels of the indicator parameters conductivity, sulfate, and boron for all three PW wells were within the corresponding ranges of the alluvial wells in the drainage.

Table 3.12. 2023 analysis results from the PW wells.

Site	Comment	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br ⁻ (ppm)	Cl ⁻ (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)
PW 734	Single Value	210	8.14	477	<0.067	2.48	40.8	244	0.107	81.0	14.2	2.15	9.13	5.66	5.12
PW 735	Mean Value	426	7.38	4215	0.158	21.3	2330	4060	1.99	242	342	18.2	418	58.87	57.64
PW 736	Mean Value	364	7.43	3155	<0.067	17.4	1780	2485	1.16	240	265	13.3	209	43.21	44.83

Figure 3.14 and Table 3.13 (in Section 3.6.5) show the drainage-mean values from 1984–2023. The latest conductivity and sulfate values decreased relative to the previous year, while boron increased slightly. None of these changes is significant relative to previous fluctuations. No drainage trends were identified for 2023.

3.6.5 Comparison of Cow Creek, South Fork Cow Creek, and Pony Creek Drainages

For the overall observation period (1985–1987 and 1990–2023), the mean values for conductivity and sulfate in the Pony Creek drainage are less than those in the South Fork Cow Creek drainage, which are less than those in the Cow Creek drainage. For boron, the means in both the Cow Creek and the South Fork Cow Creek drainages are not statistically different, but both are greater than the mean in the Pony Creek drainage.

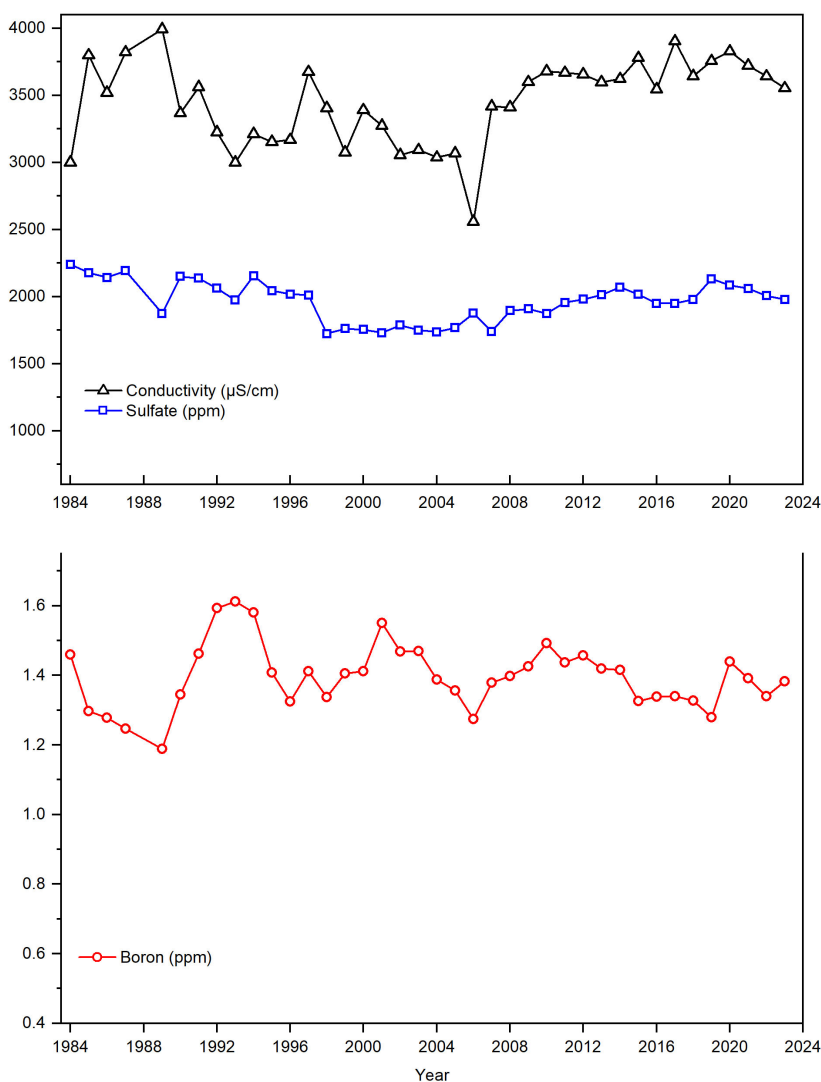


Figure 3.14. Trend plots for conductivity, sulfate, and boron in Cow Creek Drainage, 1984–2023.

Table 3.13. Drainage-mean values for conductivity, sulfate, and boron in Cow Creek drainage, 1984-2023.

Year	Drainage Mean		
	Conductivity ($\mu\text{S}/\text{cm}$)	SO_4^{2-} (ppm)	Boron (ppm)
1984	3,000	2,240	1.46
1985	3,799	2,178	1.30
1986	3,519	2,142	1.28
1987	3,823	2,193	1.25
1989	3,994	1,875	1.19
1990	3,369	2,150	1.35
1991	3,563	2,138	1.46
1992	3,225	2,063	1.59
1993	3,000	1,975	1.61
1994	3,213	2,156	1.58
1995	3,154	2,044	1.41
1996	3,169	2,019	1.33
1997	3,675	2,013	1.41
1998	3,406	1,725	1.34
1999	3,075	1,763	1.41
2000	3,391	1,755	1.41
2001	3,275	1,731	1.55
2002	3,056	1,788	1.47
2003	3,094	1,750	1.47
2004	3,038	1,738	1.39
2005	3,069	1,769	1.36
2006	2,559	1,876	1.27
2007	3,419	1,740	1.38
2008	3,413	1,896	1.40
2009	3,599	1,909	1.43
2010	3,680	1,875	1.49
2011	3,669	1,956	1.44
2012	3,657	1,981	1.46
2013	3,599	2,014	1.42
2014	3,623	2,070	1.42
2015	3,779	2,016	1.33
2016	3,546	1,949	1.34
2017	3,904	1,949	1.34
2018	3,644	1,979	1.33
2019	3,756	2,133	1.28
2020	3,828	2,085	1.44
2021	3,723	2,061	1.39
2022	3,642	2,007	1.34
2023	3,554	1,978	1.38

3.6.6 Analysis of Bromide

In December 2009, driven by a regulatory requirement of the U.S. Environmental Protection Agency, bromide (Br^-) was added to the scrubber solution to enhance the removal of mercury from the flue gas produced by the Colstrip power plant. An EHP sample collected in 2015 contained a bromide concentration of 1,940 ppm (Table 3.8). Within groundwater systems, the bromide ion is regarded as a conservative tracer, and its concentration in the EHP is sufficiently high to be useful as a leakage indicator. Therefore, in response to the addition of bromide to the EHP, bromide was added to the formal list of the chemical parameters of interest in 2010.

In previous annual reports, the first 13 years of monitoring results for bromide were reported along with data from September 2009 (before bromide was added to the EHP) for comparison (Thompson et al. 2011–2023). Several sites have had bromide concentrations that were higher than the 2009 levels, but because the data were highly variable, and almost all of the results were low (i.e., within a factor of three of the MDL), statistical trend testing to address possible EHP contamination has not been performed. Results at these low concentrations are not considered to be reliable for quantitative evaluation.

Monitoring of bromide continued in 2023. Figure 3.15 shows the concentrations for sites along the Cow Creek, South Fork Cow Creek, and Pony Creek drainages. In this plot, the ordering of sites within each drainage is based on their distance from the headwaters to help reveal any spatial patterns that might be occurring. For sites that were sampled in July and October, the average is shown. In cases for which bromide was not detected, the plot is annotated with ND to indicate that an analysis was performed. Also, in situations where the 2023 data for a site consisted of non-detected and detected values, the detected concentration was plotted rather than the average. The MDL was 0.067 ppm as indicated by the horizontal red line on the plot.

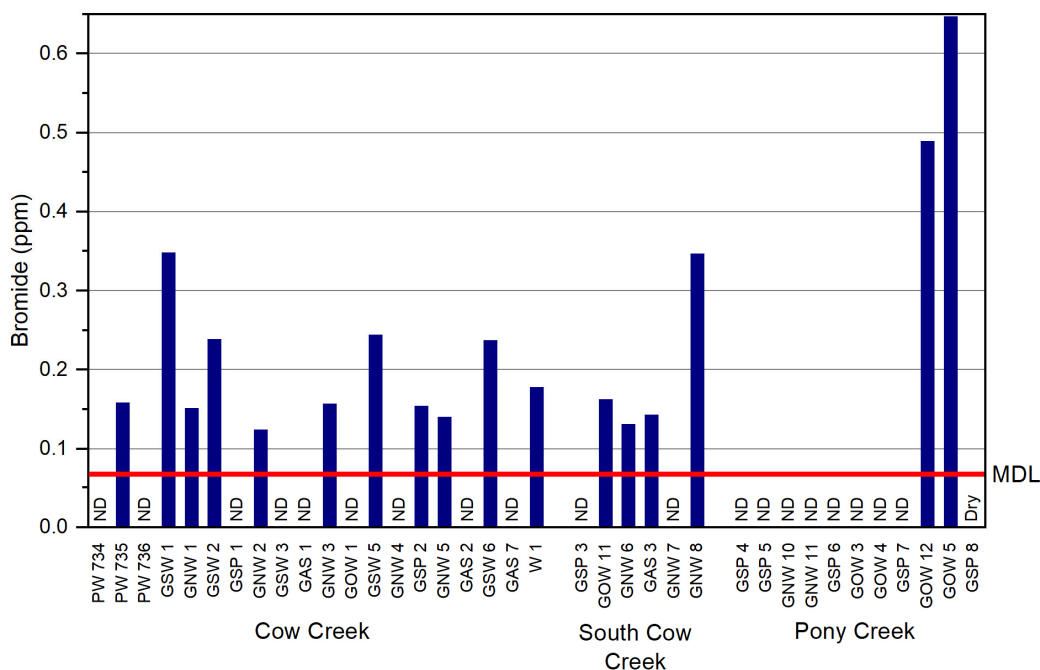


Figure 3.15. 2023 Bromide concentrations from sampling locations within the Cow Creek, South Cow Creek, and Pony Creek Drainages. The sites are ordered according to their position in each drainage (upgradient to downgradient). ND = non-detected result; Dry = site was dry (no sample was collected).

Consistent with previous data, the 2023 bromide results provide no clear evidence of contamination from the EHP in the study area. Of the 36 sites with data in Figure 3.15, fifteen had bromide concentrations that exceeded two times the MDL, and seven had levels greater than three times the MDL. The highest bromide concentrations were at Pony Creek wells GOW 5 and GOW 12. These wells have had values ranging from approximately 0.3–0.7 ppm in previous years. Most of the other sites with concentrations greater than three times the MDL were surface water sites; evaporation may account for their higher concentrations. Interestingly, South Fork Cow Creek well GNW 8's result of 0.347 ppm is the highest concentration that has been observed at that site since bromide monitoring commenced in 2010. However, this slightly elevated result does not suggest EHP leakage based on the non-detected and low concentrations observed at all upgradient sites in the drainage.

In many ways, bromide is uniquely valuable in our study. Because bromide levels are extremely low at nearly all sites in the study area, any marked increases in concentration might indicate contamination originating from the EHP. Also, because significant amounts of bromide have only been present in the EHP since around 2010, any contamination could not have reached many downgradient sites via groundwater transport. Therefore, increases in bromide at those sites would specifically identify runoff of rain/snowmelt as the contaminating pathway, because that would be a faster transport mechanism. Finally, because bromide was not detected at upgradient sites at Pony Creek nor is it associated with mining, substantial increases in bromide at those sites may not only indicate the onset of EHP contamination but also establish that there is a contaminant-sensitive hydraulic connection—either groundwater or intermittent runoff—between the EHP and Pony Creek. In the future, this analyte will become increasingly useful as a tracer for potential groundwater contamination from the EHP and downgradient sites.

3.6.7 Continuous-Monitoring Probe Results

An in situ monitoring probe was installed in well PW 736 on December 2, 2010. The probe measures groundwater specific conductance, temperature, and DTW at 4-hour intervals. The data are electronically recorded using an on-site datalogger and periodically are uploaded via satellite to an online database.

Well PW 736 is in the uppermost reaches of Cow Creek, downgradient from the Units 3 and 4 EHP. The monitoring probe (Figure 3.16) was installed in the well casing several feet below the groundwater surface and secured in place using a cable and fixed anchor point. The probe is part of a data station that is located at the ground surface above the well (Figure 3.17). The station includes a datalogger, battery, solar panel, and satellite connection, all of which work to power the probe, record, and relay probe measurements. Probe measurements for the first full year were reported in 2012 (Thompson et al. 2012). The following section presents the data collected in 2023 by the submersible continuous-monitoring probe.



Figure 3.16. Continuous-monitoring probe.



Figure 3.17. Well PW 736 with companion datalogger enclosure, solar panel, and fenced enclosure.

3.6.7.1 Data Record

This section summarizes data collected every four hours from January 1, 2023 through December 31, 2023 and includes discussions of groundwater specific conductance, elevation, and temperature. Specific conductance, or electrical conductivity, is the same parameter measured by the analytical laboratory and is an index of dissolved solutes. In addition, the datalogger records data-station battery voltage. The performance of the solar array and data-station battery voltage are recorded to monitor the station's status and maintenance needs. During 2023, the solar array and battery functioned properly.

Since its installation in December 2010, the monitoring station has recorded approximately 34,060 readings. The 2023 data includes 2,190 readings. During the year, several data records failed to be recorded and were encountered as blank cells in the downloaded file. These were infrequent and constituted approximately 1% of the total data records. Also, false records were recorded during field sampling in July and October when the probe was removed from the well for inspection and when the probe re-equilibrated after being redeployed in the well.

During the July 2023 sampling campaign, well PW 736 was sampled after inspecting the probe. As part of the inspection, the probe calibration was evaluated by immersing the probe in a standard solution of known conductivity ($1,413 \mu\text{S}/\text{cm}$) and comparing the readings with the standard's value. The probe read approximately $1,406 \mu\text{S}/\text{cm}$, and the field team concurred that recalibration was appropriate. Following recalibration, inspection of the probe's condition, and sampling, the probe was replaced in the well. False readings (three readings) of the probe were recorded following replacement of the probe and were removed from the data set.

During the October 2023 sampling event, the probe was again placed in a standard solution of known conductivity ($1,413 \mu\text{S}/\text{cm}$). The probe read $1,362 \mu\text{S}/\text{cm}$, and the field team concluded that recalibration was appropriate. Following recalibration, the probe read $1,408 \mu\text{S}/\text{cm}$. Well PW-736 was then sampled,

and the probe was replaced. False readings (three readings) of the probe were recorded following replacement of the probe and were removed from the data set. Once the probe stabilized, conductivity readings were approximately 200 $\mu\text{S}/\text{cm}$ greater than pre-calibration conditions. The likely causes of the offset were recalibration and removing and reinstalling the probe.

3.6.7.2 Results and Discussion

Groundwater temperatures during 2023 ranged from 10.30 to 10.82 $^{\circ}\text{C}$ with a mean temperature of 10.56 $^{\circ}\text{C}$ (Figure 3.18a). The 13-year (period of record) groundwater temperature fluctuation ranges from 10.28 to 11.02 $^{\circ}\text{C}$ with a mean temperature of 10.63 $^{\circ}\text{C}$ (Figure 3.18b). Seasonal variation in aquifer temperature is evident; maximum temperatures occur in February/early-March and minimum temperatures occur in late July/August, indicating, for this site, approximately a 6-month lag in heat balance relative to that of ambient air temperature. Consistent with data records from previous years, the daily groundwater temperature variation was greatest between early spring and mid-summer and is potentially related to aquifer recharge events. Smaller daily temperature variation from August to February is likely associated with less recharge during this period.

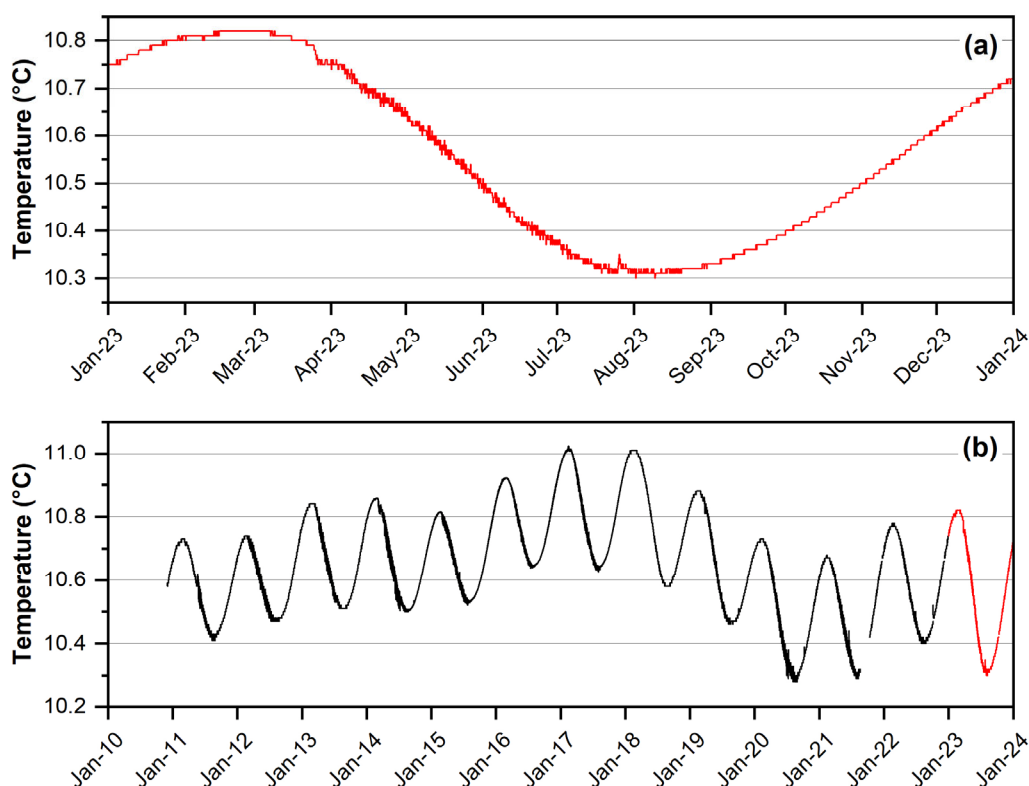


Figure 3.18. Groundwater temperature trend measured at PW 736 in (a) 2023 and (b) 2010–2023. Data collected in 2023 are shown in red.

The groundwater elevation in PW 736 fluctuated 0.82 ft during 2023, ranging from 3,052.46 to 3,053.28 ft above mean sea level (Figure 3.19a). Groundwater elevation peaked in March and decreased through the remainder of the year. This correlation between precipitation and groundwater recharge has been observed during previous years and is typical of a shallow, unconfined aquifer. The groundwater elevations presented in Figure 3.19 are referenced to the surveyed elevation of the top of the well casing which is 3075.5 ft above sea level.

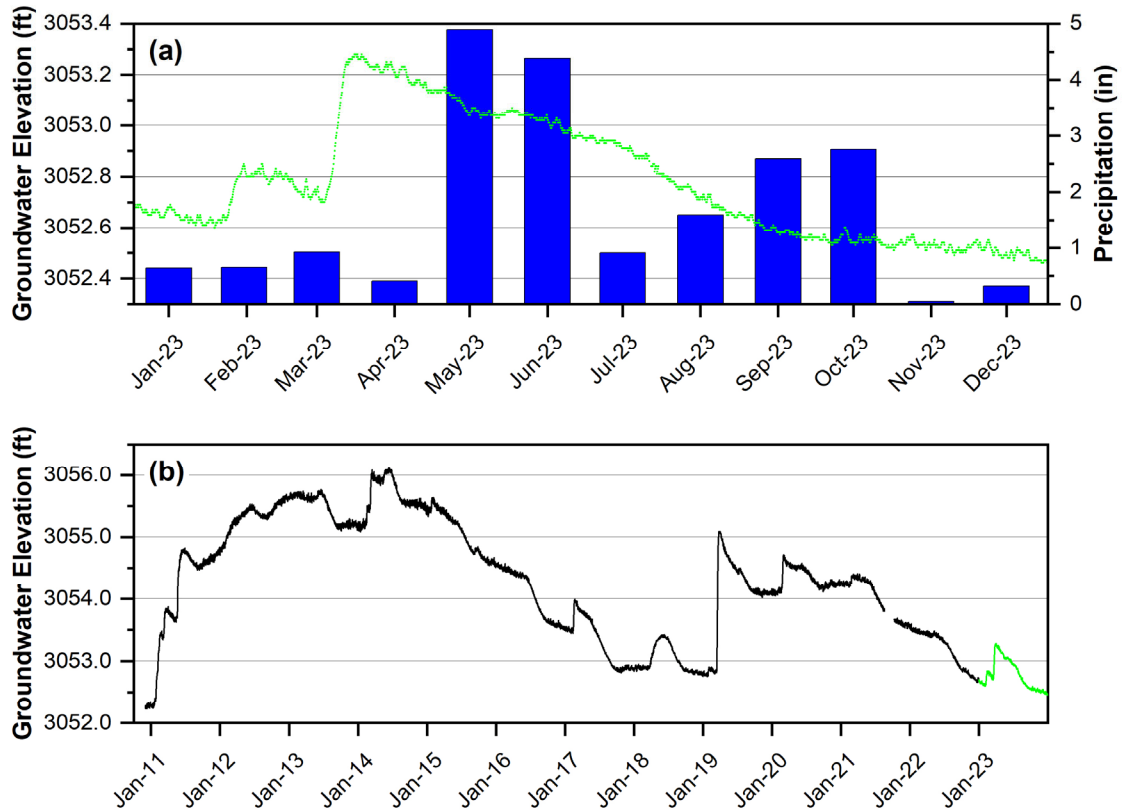


Figure 3.19. Groundwater elevation trend at PW 736 during (a) 2023 with monthly total precipitation plotted as columns associated with a secondary y-axis and (b) 2010-2023. Data collected in 2023 are shown in green.

The 13-year fluctuation in groundwater elevation is 3.89 ft, ranging from 3,052.23 and 3,056.12 ft (Figure 3.19b). These data suggest that the aquifer is seasonally recharged by spring precipitation and snowmelt during years such as 2011, 2017, 2018, 2019, 2020, 2021, and 2023; and to a lesser degree in years such as 2015, 2016, and 2022. Rapid increases in groundwater elevation occurred during May 2011, February 2017, March 2019, and March 2020. Most notably in May 2011, 10 in. of precipitation was recorded due to a heavy rain event, resulting in flooding and a rapid increase in groundwater elevation. Rapid increases also occurred from snowmelt events during February 2017, March 2019, March 2020, March 2021, February 2023, and March 2023.

During 2023, the groundwater elevation rapidly increased approximately 0.3 ft during February and 0.6 ft in March due to snowmelt events. Curiously, the water level continually decreased after March (Figure 3.19a) despite there being inordinately high rainfall in spring and summer, which appears to have flattened the declining trend instead of increasing the water level. As noted in Section 2.1, the minimal recharge from rainfall may be due to a combination of surface topography and soil composition upgradient of PW 736, which tends to preferentially induce runoff during rainfall events. Recharge via snowpack melt is not as dependent on surface topography and is less influenced by imbibition rate.

Regardless of the mode of natural recharge, punctuated and significant rises in water level (whenever a recharge pathway is established) are consistent with a shallow, unconfined alluvial aquifer with a vadose zone capable of appreciable water storage capacity. If precipitation input exceeds the storage capacity of the unsaturated zone, then water table recharge occurs. The daily variation observed in groundwater temperature during the late winter to early summer period suggests some vadose zone input, but a change

in water table surface elevation generally is not observed despite daily temperature variations, which suggests these inputs are typically minimal. For 2023, the influence of water inputs from the EHP, if any, on groundwater elevation is unknown. However, nothing in the 2023 probe data suggests EHP pulse inputs affecting groundwater elevation.

Groundwater specific conductance measured at PW 736 during 2023 ranged from 1,518 to 3,105 $\mu\text{S}/\text{cm}$ (Figure 3.20a). The 13-year trend in specific conductance is shown in Figure 3.20b.

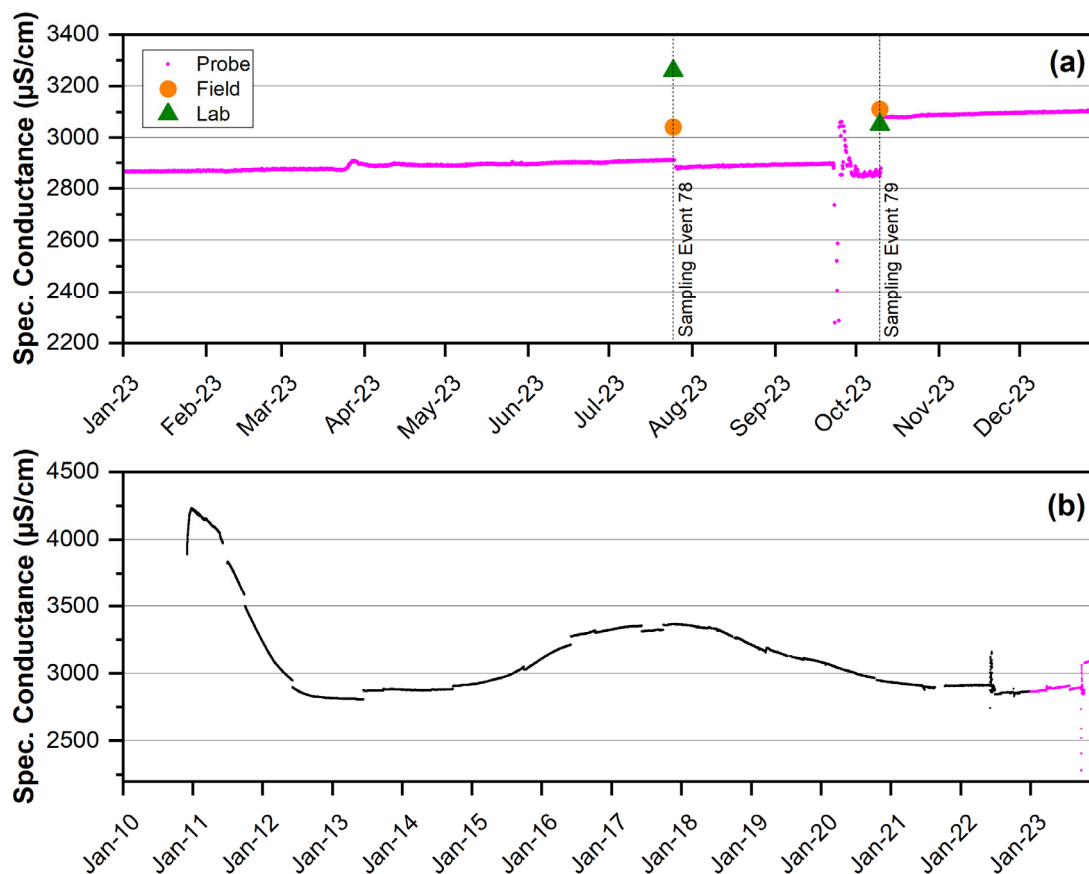


Figure 3.20. In situ PW 736 specific conductance trend measured by the continuous monitoring probe during (a) 2023 with field and laboratory values from the July and October sampling events and (b) 2010-2023. Probe data collected in 2023 are shown in magenta.

Three measurements of specific conductance were collected during each sampling event. As mentioned above, the datalogger provides a continuous measurement at 4-hour intervals. In addition, a field sample was collected and measured within a few hours of collection. The third conductivity value is a laboratory result measured within the analytical holding time. The resulting variance in the sample results indicates generally good agreement between the measurements by different instruments at different times. For example, during the July 2023 sampling event, the datalogger reported 2,913 $\mu\text{S}/\text{cm}$, the field result was 3,040 $\mu\text{S}/\text{cm}$, and the corresponding lab value was 3,260 $\mu\text{S}/\text{cm}$. The maximum difference between these measurements was 347 $\mu\text{S}/\text{cm}$. During the October 2023 sampling event, the datalogger reported 2,895 $\mu\text{S}/\text{cm}$ while the field result was 3,110 $\mu\text{S}/\text{cm}$, and the corresponding laboratory value was 3,050 $\mu\text{S}/\text{cm}$. The maximum difference between these measurements was 215 $\mu\text{S}/\text{cm}$. The RPDs between the datalogger and lab analysis measurements were approximately 11% and 5% for the July and October 2023 sampling events, respectively. Thus, the differences between datalogger and lab analysis

measurements of specific conductance measurements during 2023 are consistent with past years. These results confirm that the continuous-monitoring probe provides measurements of specific conductance that are reasonably close to the laboratory values.

The long-term trend in conductivity shown in Figure 3.20b suggests that electrical conductivity values are stable over long periods of time with only slight changes between 2012 and 2021. Abrupt changes have occurred in 2022 and 2023. During 2022 a sudden increase of approximately 250 $\mu\text{S}/\text{cm}$ occurred prior to the June event, and more recently in 2023, a sudden decrease of 600 $\mu\text{S}/\text{cm}$ and then a subsequent increase of 800 $\mu\text{S}/\text{cm}$ occurred over a 7-day period in September prior to the October 2023 sampling event.

The first short, wavelike phenomenon observed since installation of the probe occurred in June 2022; this aberration was speculated to be caused by the sudden dissolution of precipitated salts in the vadose zone (or the diffusion of dissolved salts out of capillary water there) associated with high-volume rainfall events that triggered recharge (Thompson et al, 2023). However, a statistical analysis was conducted in September of 2023, to address whether the “wave” may instead have resulted from disfunction in the probe or its powering and charging system. Recorded values of specific conductance, temperature, groundwater level, and voltage supplied to the probe were examined for: all of 2023, June only, and a period after June. For the two short periods in 2023, these variables were examined for statistical dependence upon their own recently preceding values and upon those of the other variables. Of especial interest was whether conductivity was being affected by the other measurements. Statistical testing concluded that this causality, and no other causal influences, were present during the wave period. Afterward, the system's daily charging and discharging cycles were scrutinized, and the voltage system was concluded to have been functioning properly during the wave period.

A supersmooth fit of the probe's specific conductance data was then performed, which revealed a doubly-peaked wave trend, with unusually high error variation (much higher than the variation around the flat trends characteristic of most of the 2023 data). The highly variable errors during the wave period were tested for causal influences by temperature, probe voltage, and water level and were concluded to be associated with concurrent temperature readings. The probe actually measures raw conductivity, which is then converted to specific conductance using the temperature measurement. Thus, unusually high error variation in the temperature readings may have influenced the specific conductance values' marked errors around the smoothed waves.

In summary, this special analysis found no causal disfunction in the monitoring system that would explain the wave in specific conductance during 2022, but the conductance measurements did possess significantly elevated uncertainty during the wave period. Any trendform testing of those waves (currently outside of the scope of our study) might require a formal statistical test such as our conventional or plr tests. *Nevertheless, the tentative conclusions proffered in last year's report were not refuted.* Although, no similar special investigation was applied to the period in 2023 which again exhibited a wave in specific conductance, we again conclude that this second wave was also likely due to the influx of mobilized salts.

The degree to which regional aquifer characteristics vary over time with respect to salinity is unknown. However, spring and early summer recharge events and other climatological factors (e.g., amount of snowmelt) are likely the primary causes of the fluctuations. The profile in Figure 3.20b is similar in shape to the estimated conductivity trend for PW 736 (Figure G.9).

3.6.7.3 Data Summary

The continuous-monitoring probe in well PW 736 continues to record important groundwater data downgradient from the Units 3 and 4 EHP. The specific conductance, groundwater elevation, and temperature data provide unique and useful supplements to biannual analytical sampling. Continued monitoring will allow long-term groundwater table fluctuations and specific conductance trends to be observed. No unusual temperature or elevation measurements were observed during 2023; however, a sudden decrease then a spike in specific conductance occurred in September suggesting that May and June precipitation (9.3 in.) may have influenced the release of dissolved minerals.

Probe/datalogger installations in other monitoring wells and springs or other surface water in the project area would improve our understanding of downgradient groundwater and surface-water temporal behavior in Cow Creek and the South Fork of Cow Creek. Additionally, such installations would facilitate rapid detection, characterization, and mitigation response in the event of significant leakage or a contaminated stormwater event.

4.0 Conclusions

Since 1984, staff from PNNL/PNWD and KC Harvey Environmental, LLC, have collected hydrologic and water-quality data in the drainage basins of Cow Creek and Pony Creek in Rosebud County, Montana. Information from this monitoring program is used to determine whether water-quality and hydrological parameters have changed from previous years and to provide a basis for evaluating the effects of any ongoing or future events associated with Colstrip Power Plant Units 3 and 4 EHP, which is located upgradient from Cow Creek and Pony Creek.

There were no reportable releases from the EHP during 2023.

Total estimated precipitation in the Colstrip vicinity during 2023 was 20.11 in., which is 30% higher than the average of 15.4 in. measured since 1984. May, June, September, and October were the wettest months, accounting for 24%, 22%, 13%, and 14% of the total annual precipitation, respectively. The winter months were relatively dry, consistent with historical precipitation patterns.

Groundwater levels measured in the alluvial wells during 2023 averaged about 0.75 ft higher than those of the previous year. Approximately half of the spring/summer water levels were higher in 2023, while all fall water levels were higher except for well PW 736. These trends are generally consistent with the higher levels of precipitation in the latter half of 2023. Water levels rose in most wells between the July and October 2023 sampling campaigns. The largest increases occurred at wells GNW 2 (1.78 ft), GNW 6 (0.47 ft), GOW 1 (0.43 ft), and GOW 11 (0.75 ft).

Statistical testing identified four sites (GOW 4, GSP 4, GSP 6, and PW 735) where water quality exhibited continual trends from the past four or more years. Four site-specific trends were identified for Pony Creek: GOW 4 had an increase for conductivity and a decrease for boron, GSP 4 had an increase for conductivity, and GSP 6 had an increase for boron. Among Cow Creek sites, relatively deep well PW 735 had an increase in boron. No abrupt increases were detected at any of the sites identified in Stipulation 12(d). Additionally, no recent increasing or decreasing drainage-trends were found at Pony Creek, South Fork Cow Creek, or Cow Creek. Overall, these findings do not indicate any definitive impacts on groundwater quality from the operations of the EHP.

The relatively new PW wells on Cow Creek near the western border of the study area have been monitored for water quality since 2009. Similar to previous years, the levels of conductivity, sulfate, and boron in well PW 736 were comparable to those in most of the alluvial wells in the drainage, while the nearby deeper well PW 735 had markedly higher values for those parameters. PW 735 had a recurring increasing trend for boron, but the increase may be tapering off based on a posterior probability analysis. It is unclear whether the aquifer tapped by PW 735 is confined (i.e., hydraulically isolated) in the vicinity of the capture system. Limited monitoring data were obtained for the neighboring shallower well PW 734 due to dry conditions: no sample could be collected in July, while a reduced-volume sample was collected in October. The October levels of alkalinity, conductivity, bromide, chloride, and sulfate for PW 734 were significantly lower than those in PW 735 and PW 736.

Statistical tests based on data from the entire observation period (1984–1987 and 1990–2023) and selected sampling sites concluded that water from the Pony Creek drainage had the lowest levels of conductivity, sulfate, and boron. Water from the Cow Creek drainage was distinctly highest in conductivity and sulfate, and mean values for water from the South Fork Cow Creek drainage fell between the values of the other drainages except for boron, for which the mean was the same as that from Cow Creek.

Monitoring bromide as a potential indicator of contamination from the Units 3 and 4 EHP continued in 2023. Two sites in the lower portion of Pony Creek, GOW 5 and GOW 12, had the highest bromide concentrations of approximately 0.6 and 0.5 ppm, respectively, in July. Previous data indicate that these wells tap groundwater sources with higher concentrations of bromide than other sites in the study area. The furthest downgradient site along South Fork Cow Creek, well GNW 8, also had a slightly elevated concentration of approximately 0.3 ppm. At most other sites, bromide levels ranged from non-detected (<0.067 ppm) to within a factor of three of the MDL. *Consequently, the most recent results provide no apparent evidence of encroachment of water from the EHP in the study area either via groundwater transport or by overland flow of rainwater/snowmelt.*

The continuous-monitoring probe installed in well PW 736 in December 2010 collects frequent temperature, DTW, and conductivity data to better understand the behavior of the alluvial aquifer near the EHP and provide insight into the local alluvial hydrology in the study area in general. Similar to data from previous years, the 2023 probe measurements show fluctuations in temperature that along with probe groundwater-level measurements, appear to be related to aquifer recharge. Specific conductance was relatively stable during 2023, although there was a rapid decrease of approximately $600\ \mu\text{S}/\text{cm}$ followed by an increase of $800\ \mu\text{S}/\text{cm}$ over a 7-day period in September. The cause of the pulse is unknown, but it may be related to soluble minerals in the overlying unsaturated zone that were mobilized during especially heavy precipitation. Groundwater elevation declined during most of the year. This year's continuous-monitoring data do not indicate any releases from the EHP.

5.0 References

- Chamberlain, P. J. 2018. "The Bayes Factor Reinterpreted and Renamed as the Predictive Likelihoods Ratio - It's a UMP Composite vs. Composite Test on Multiple Parameters." Technical Bulletin No. 3, Enviro-Sci Consulting.
- Erbes D. 2000. *Geochemical Transformations and Hydrologic Recovery Observed in a Coal-Strip-Mine Spoils Aquifer*. Billings Land Reclamation Symposium 2000, Striving for Restoration, Fostering Technology and Policy for Reestablishing Ecological Function, Reclamation Research Unit Publication 00-01, Montana State University, Bozeman, Montana.
- Hydrometrics. 1987. Quantity Management of Groundwater in the Upper Cow Creek Drainage, Colstrip, Montana. Hydrometrics, Inc., Helena, Montana.
- Hydrometrics. 1990. Evaluation of 1989 Hydrologic Monitoring Data from the Montana Power Company Process Pond System at Colstrip, Montana. Hydrometrics, Inc., Helena, Montana.
- Lee RW. 1980. *Geochemistry of Water in the Fort Union Formation of the Northern Powder River Basin, Southeastern Montana*. Water Resources Investigations Open-File Report 80-336, U.S. Geological Survey, Helena, Montana.
- McDonald JP, KB Olsen, JL Smoot, PJ Chamberlain, SR Jennings, and JD Goering. 2007. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2006*. Battelle – Pacific Northwest Division, Richland, Washington.
- Metesh J. 1994. *Cumulative Hydrologic Impacts on Rosebud Creek from Development Near Colstrip*. Montana Bureau of Mines and Geology Open-File Report No. 273, Butte, Montana.
- Montana Department of State Lands and U.S. Office of Surface Mining. 1983. *Comprehensive Environmental Impact Study: Western Energy's Rosebud Mine*. Helena, Montana.
- Olsen KB, DA Myers, MR Toland, and FF Munshower. 1987. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. Battelle–Pacific Northwest Division, Richland, Washington.
- Olsen KB, JL Smoot, and JW Johnston. 1991. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1990*. Battelle–Pacific Northwest Division, Richland, Washington.
- Olsen KB, JL Smoot, and PJ Chamberlain. 1992. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1991*. Battelle – Pacific Northwest Division, Richland, Washington.
- Olsen KB, JL Smoot, and PJ Chamberlain. 1993. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1992*. Battelle – Pacific Northwest Division, Richland, Washington.
- Olsen KB, JL Smoot, and PJ Chamberlain. 1994. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1993*. Battelle – Pacific Northwest Division, Richland, Washington.
- Olsen KB, JL Smoot, and PJ Chamberlain. 1995. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1994*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1996. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1995*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1997. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1996*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1998. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1997*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1999. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1998*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2000. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1999*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2001. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2000*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2002. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2001*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2003. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2002*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2004. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2003*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2005. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2004*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2006. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2005*. Battelle – Pacific Northwest Division, Richland, Washington.

Shimer JA. 1972. *Field Guide to Landforms in the United States*. The Macmillan Company, New York.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2008. *2007 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2009. *2008 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4063, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2010. *2009 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4182, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2011. *2010 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4261, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2012. *2011 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4352, Battelle–Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2013. *2012 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-22498, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2014. *2013 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10010, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2015. *2014 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10027, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2016. *2015 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10043, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2017. *2016 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10056, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, PD Thorne, PJ Chamberlain, SR Jennings, and C Lucy. 2018. *2017 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10066, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, PJ Chamberlain, and C Lucy. 2019. *2018 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10080, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, PJ Chamberlain, and E Koehler. 2020. *2019 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10098, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, E Koehler, M Larson, and PJ Chamberlain. 2021. *2020 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10108, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, E Koehler, and PJ Chamberlain. 2022. *2021 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10120, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, E Koehler, and PJ Chamberlain. 2023. *2022 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10132, Pacific Northwest National Laboratory, Richland, Washington.

Van Voast WA and JC Reiten. 1988. *Hydrogeologic Responses: Twenty Years of Surface Coal Mining in Southeastern Montana*. Memoir 62, Montana Bureau of Mines and Geology, Butte, Montana.

Van Voast WA, RB Hedges, and JJ McDermott. 1977. *Hydrogeologic Conditions and Projections Related to Mining near Colstrip, Southeastern Montana*. Bulletin 102, Montana Bureau of Mines and Geology, Butte, Montana.

Waren K and A McDannel. 2003. "An Evaluation of Ground Water Conditions in the Colstrip Coal Deposit, Montana." *Abstracts of the Twentieth Annual Meeting American Society for Surface Mining and Reclamation*, Billings, Montana, June 3–6, 2003.

Appendix A

Comments Provided by Signatories of Stipulation 12(d)

Appendix A

Comments Provided by Signatories of Stipulation 12(d)

No formal comments were received for this year's report.

Appendix B

Field Record Sheet and Chain-of-Custody Form Used for Sample Control

COLSTRIP GROUND WATER SAMPLE FIELD RECORD							
BATTELLE Pacific Northwest Division Richland, WA 99354							
Well Name _____				Project: <u>Colstrip</u>			
Recommended Purge Volume (gal) _____				Date _____			
Purge Flow Rate (gal/min) _____				Calculations: _____			
Pump Type _____							
SAMPLES COLLECTED							
SAMPLE NUMBER(S)	BOTTLE ID	PUMP TYPE	COLLECTOR				
Filtered Anions 1L P Filtered Metals 250 ml P w/HNO3 TOTAL NUMBER OF BOTTLES: 2							
FIELD MEASUREMENTS							
Previous pH	Prv. Depth	Water Levels (ft)		Time			
Steel Tape# _____	Held			pH			
pH/Cond Meter# _____	Cut			Teml (DEG.C)			
Thermometer# _____	Depth below TC			Cond (um/cm)			
Turbidimeter# _____				Turb (NTU)			
FIELD OBSERVATIONS							
Weather: _____							
General Problems/ Unusual Events: _____							
Equipment Irregularities: _____							
Container Irregularities: _____							
Comments: _____							
Well capped and locked? Yes No							
Data Recorded by: _____				Data Checked by: _____			
Chain-of-Custody form # _____				FRF Logbook and Page No: _____			

Appendix C

Analytical Methods Used for Water Analysis

Appendix C

Analytical Methods Used for Water Analysis

Listed below are the methods for chemical analysis that were used to measure parameters in the Cow Creek and Pony Creek hydrogeochemical study. Analyses were conducted by GEL Laboratories, LLC, in Charleston, South Carolina.

Bicarbonate alkalinity	<i>Standard Methods for the Examination of Water and Wastewater</i> , 19th Edition, 1995, Method 2320 B
Carbonate alkalinity	<i>Standard Methods for the Examination of Water and Wastewater</i> , 19th Edition, 1995, Method 2320 B
Conductivity	<i>Methods for Chemical Analysis of Water and Wastes</i> , EPA-600/4-79-020, March 1983 and subsequent revisions, Method 120.1
pH	<i>Test Methods for Evaluating Solid Waste, Physical/Chemical Methods</i> , EPA SW-846 Third Edition, November 1986, and its updates, Method 9040
Total dissolved solids	<i>Standard Methods for the Examination of Water and Wastewater</i> , 19th Edition, 1995, Method 2540 C
Trace inductively coupled plasma metals	<i>Test Methods for Evaluating Solid Waste, Physical/Chemical Methods</i> , EPA SW-846 Third Edition, November 1986, and its updates, Method 6010B
Anion analysis (Br ⁻ , Cl ⁻ , and SO ₄ ⁼)	<i>Test Methods for Evaluating Solid Waste, Physical/Chemical Methods</i> , EPA SW-846 Third Edition, November 1986, and its updates, Method 9056A
Ion balances	<i>Standard Methods for the Examination of Water and Wastewater</i> , 19th Edition, 1995, Method 1030 F

Appendix D

Photos of Spring and Surface-Water Sites

Appendix D

Photos of Spring and Surface-Water Sites

This appendix contains spring and surface-water site photographs that were taken during the July and October 2023 sampling campaigns. The images convey site conditions at the time of sampling and can be useful when interpreting geochemical results (e.g., surface-water sites can be subject to evaporation, which could result in higher concentrations of measured parameters). Refer to Figure 2.4 for the locations of the sampling sites.

D.1 Spring Sites

Pony Creek sites GSP 4 through GSP 10 were sampled in July only as noted in Section 3.1.

June	October
GSP 1	
	
GSP 2	
	
GSP 3	
	

June	October
GSP 4 	Site was not visited in October.
GSP 5 	Site was not visited in October.
GSP 6 	Site was not visited in October.

June	October
GSP 7 	Site was not visited in October.
GSP 8  Site was dry; no sample was collected.	Site was not visited in October.
GSP 9  	Site was not visited in October.

June

October

GSP 10



Site was not visited in October.



This site's sample was collected from the water collection basin shown in the middle photo. The lower photo shows the nearby stock-watering tank that is filled by the spring.

D.2 Surface-Water Sites

June	October
GSW 1	
	
GSW 2	
	
GSW 3	
	
Site was dry; no sample was collected.	

June	October
<p data-bbox="769 247 852 281">GSW 5</p> <div data-bbox="375 310 794 627" data-kind="parent" data-rs="2">  </div> <div data-bbox="828 310 1247 627" data-kind="parent" data-rs="2">  </div>	
<p data-bbox="769 674 852 707">GSW 6</p> <div data-bbox="375 737 794 1054" data-kind="parent" data-rs="2">  </div> <div data-bbox="828 737 1247 1054" data-kind="parent" data-rs="2">  </div>	

Appendix E

Site and Drainage-Mean Values for Data Collected in 2023

Appendix E

Site and Drainage-Mean Values for Data Collected in 2023

Table E.1. 2023 reported values for all parameters from GAS wells.

Site	Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)
GAS 1	Mean Value	336	7.52	2775	<0.067	11.8	1205	2145	0.942	111	110	8.68	384
GAS 2	Single Sample	365	8.70	1690	<0.067	10.2	413	1030	0.325	3.04	0.860	1.96	374
GAS 3	Mean Value	330	8.42	2780	0.143	13.5	996	1755	0.299	10.3	2.94	3.54	624
GAS 7	Mean Value	348	8.59	1990	<0.067	10.6	578	1255	0.298	4.12	1.24	2.34	435

Table E.2. 2023 reported values for all parameters from South Fork Cow Creek Drainage.

Site	Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)
GSP 3	Mean Value	222	7.14	1945	<0.067	7.58	969	1605	2.16	166	151	14.0	98.6
GOW 11	Mean Value	347	7.56	3820	0.162	27.0	2250	3220	1.58	324	370	12.1	210
GNW 6	Mean Value	328	7.56	3235	0.131	19.4	1830	2610	1.44	254	296	11.0	170
GNW 7	Mean Value	394	7.66	1815	<0.067	6.77	721	1465	1.00	159	119	8.64	120
GNW 8	Mean Value	338	7.64	2650	0.347	13.0	1210	1820	0.826	131	148	8.74	324
Drainage Mean		326	7.51	2693	0.214	14.7	1396	2144	1.40	207	217	10.9	184

Table E.3. 2023 reported values for all parameters from Pony Creek Drainage.

Site	Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)
GSP 4	Single Sample	377	7.41	2280	<0.067	8.94	1010	1710	0.871	184	228	22.2	49.4
GSP 5	Single Sample	442	7.72	4460	<0.067	20.6	2860	3880	3.33	319	523	31.1	180
GSP 6	Single Sample	641	7.42	3310	<0.067	11.3	1460	3070	1.32	304	312	16.0	123
GOW 3	Single Sample	324	8.24	1760	<0.067	7.01	725	1410	0.616	150	133	9.56	83.6
GOW 4	Single Sample	287	8.43	2170	<0.067	10.8	744	1290	0.242	14.1	8.41	3.16	467
GSP 7	Single Sample	413	8.38	2910	<0.067	22.5	1320	2560	0.885	176	234	11.6	201
GOW 12 ^(a)	Single Sample	680	8.64	1550	<0.455	117	0.418	906	0.436	2.43	0.884	1.75	375
GOW 5	Single Sample	685	8.52	1680	0.677	150	<0.133	942	0.457	2.68	0.843	2.01	396
Drainage Mean		414	7.93	2815	<0.067	13.5	1353	2320	1.21	191	240	15.6	184

(a) Values for GOW 12 were not included in the drainage mean.

Table E.4. 2023 reported values for all parameters from Cow Creek Drainage.

Site	Comments	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)
PW 736	Mean Value	364	7.43	3155	<0.067	17.4	1780	2485	1.16	240	265	13.3	209
GNW 1	Mean Value	336	7.53	3000	0.151	17.6	1580	2530	1.08	250	238	16.2	168
GSP 1	Mean Value	396	7.6.0	2765	<0.067	11.8	1450	2295	1.13	218	221	12.8	182
GNW 2	Mean Value	412	7.50	3680	0.124	16.4	2100	2820	1.72	244	322	21.2	256
GNW 3	Mean Value	402	7.48	3730	0.157	17.9	2075	2945	1.43	258	322	18.2	279
GOW 1	Mean Value	389	7.58	3735	<0.067	22.2	1925	3325	1.25	194	267	11.4	391
GNW 4	Mean Value	410	7.45	3865	<0.067	18.9	2190	3145	1.46	266	341	18.3	280
GSP 2	Mean Value	406	7.52	4055	0.154	21.3	2365	3670	1.66	259	397	20.8	314
GNW 5	Mean Value	412	7.56	4000	0.140	19.3	2340	3295	1.56	262	361	18.9	323
W 1 ^(a)	Mean Value	434	7.54	5050	0.178	22.8	2910	4770	1.92	299	425	20.0	496
Drainage Mean		392	7.52	3554	0.145	18.1	1978	2946	1.38	243	304	16.8	267

(a) Values for W 1 were not included in the drainage mean.

Appendix F

Analytical Results for Water Samples Collected in 2023

Appendix F

Analytical Results for Water Samples Collected in 2023

Table F.1. Sample Site: GAS 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GAS-1-78	25-Jul-23	335	7.53	2830	<0.067	12.2	1210	1970	0.939	113	116	9.01	395	27.77	28.26	-1.74
GAS-1-79	10-Oct-23	337	7.51	2720	<0.067	11.5	1200	2320	0.946	109	105	8.34	372	30.47	32.05	-5.04
MEAN		336	7.52	2775	<0.067	11.8	1205	2145	0.942	111	110	8.68	384	29.12	30.15	

Table F.2. Sample Site: GAS 2

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GAS-2-78	25-Jul-23	365	8.70	1690	<0.067	10.2	413	1030	0.325	3.04	0.860	1.96	374	16.54	16.19	2.17

Table F.3. Sample Site: GAS 3

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GAS-3-78	24-Jul-23	328	8.51	2810	<0.067	13.5	983	1840	0.301	9.93	2.87	3.48	625	28.01	27.41	2.17
GAS-3-79	09-Oct-23	331	8.34	2750	0.143	13.5	1010	1670	0.297	10.7	3.01	3.61	624	28.02	28.03	-0.04
MEAN		330	8.42	2780	0.143	13.5	996	1755	0.299	10.3	2.94	3.54	624	28.01	27.72	

Table F.4. Sample Site: GAS 7

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GAS-7-78	25-Jul-23	347	8.59	2030	<0.067	10.4	577	1250	0.298	4.16	1.23	2.28	436	19.33	19.25	0.45
GAS-7-79	10-Oct-23	349	8.59	1950	<0.067	10.8	580	1260	0.299	4.07	1.26	2.40	434	19.25	19.36	-0.58
MEAN		348	8.59	1990	<0.067	10.6	578	1255	0.298	4.12	1.24	2.34	435	19.29	19.30	

Table F.5. Sample Site: GNW 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-1-78	25-Jul-23	335	7.40	3000	0.150	17.9	1620	2910	1.10	256	250	17.0	173	41.30	40.93	0.89
GNW-1-79	10-Oct-23	337	7.66	3000	0.152	17.3	1540	2150	1.06	243	225	15.5	164	38.16	39.29	-2.91
MEAN		336	7.53	3000	0.151	17.6	1580	2530	1.08	250	238	16.2	168	39.73	40.11	

Table F.6. Sample Site: GNW 2

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-2-78	25-Jul-23	404	7.47	3600	0.124	15.9	2210	2540	1.71	239	327	20.4	255	50.44	54.54	-7.81
GNW-2-79	10-Oct-23	420	7.54	3760	<0.067	17.0	1990	3100	1.73	248	318	22.1	256	50.23	50.31	-0.15
MEAN		412	7.50	3680	0.124	16.4	2100	2820	1.72	244	322	21.2	256	50.34	52.43	

Table F.7. Sample Site: GNW 3

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-3-78	25-Jul-23	401	7.46	3680	0.157	18.4	2120	2550	1.47	256	338	18.0	276	53.04	52.68	0.69
GNW-3-79	10-Oct-23	404	7.51	3780	<0.067	17.4	2030	3340	1.39	260	305	18.5	282	50.80	50.83	-0.06
MEAN		402	7.48	3730	0.157	17.9	2075	2945	1.43	258	322	18.2	279	51.92	51.76	

Table F.8. Sample Site: GNW 4

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-4-78	25-Jul-23	409	7.40	3820	<0.067	19.2	2240	3040	1.43	272	361	17.9	284	56.08	55.36	1.30
GNW-4-79	10-Oct-23	410	7.50	3910	<0.067	18.6	2140	3250	1.50	260	321	18.7	275	51.82	53.28	-2.78
MEAN		410	7.45	3865	<0.067	18.9	2190	3145	1.46	266	341	18.3	280	53.95	54.32	

Table F.9. Sample Site: GNW 5

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-5-78	25-Jul-23	411	7.44	4000	0.140	20.3	2430	2860	1.60	273	371	18.2	322	58.61	59.38	-1.31
GNW-5-79	10-Oct-23	412	7.67	4000	<0.067	18.3	2250	3730	1.52	252	351	19.6	324	56.04	55.60	0.79
MEAN		412	7.56	4000	0.140	19.3	2340	3295	1.56	262	361	18.9	323	57.33	57.49	

Table F.10. Sample Site: GNW 6

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-6-78	24-Jul-23	323	7.57	3190	<0.067	19.7	1800	2230	1.43	256	303	11.1	170	45.38	44.49	1.97
GNW-6-79	09-Oct-23	332	7.55	3280	0.131	19.1	1860	2990	1.45	253	288	11.0	171	44.04	45.90	-4.16
MEAN		328	7.56	3235	0.131	19.4	1830	2610	1.44	254	296	11.0	170	44.71	45.20	

Table F.11. Sample Site: GNW 7

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-7-78	26-Jul-23	400	7.69	1800	<0.067	6.84	723	1450	0.999	160	123	8.76	121	23.59	23.25	1.47
GNW-7-79	10-Oct-23	387	7.64	1830	<0.067	6.70	719	1480	1.01	158	115	8.51	119	22.74	22.90	-0.70
MEAN		394	7.66	1815	<0.067	6.77	721	1465	1.004	159	119	8.64	120	23.16	23.07	

Table F.12. Sample Site: GNW 8

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-8-78	25-Jul-23	331	7.62	2670	<0.067	14.5	1210	1610	0.824	135	156	9.06	330	34.16	32.22	5.83
GNW-8-79	10-Oct-23	345	7.66	2630	0.347	11.5	1210	2030	0.828	127	141	8.42	317	31.94	32.42	-1.49
MEAN		338	7.64	2650	0.347	13.0	1210	1820	0.826	131	148	8.74	324	33.05	32.32	

Table F.13. Sample Site: GNW 9

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-9-78	25-Jul-23	330	8.47	2410	<0.067	12.9	837	1610	0.293	8.19	2.42	3.13	560	25.05	24.39	2.66
GNW-9-79	10-Oct-23	330	8.47	2390	0.138	12.5	817	1330	0.269	7.54	2.24	2.61	555	24.77	23.96	3.31
MEAN		330	8.47	2400	0.138	12.7	827	1470	0.281	7.86	2.33	2.87	558	24.91	24.18	

Table F.14. Sample Site: GNW 10

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-10-78	26-Jul-23	406	7.60	3110	<0.067	9.95	1770	2590	2.94	258	332	28.9	73.8	44.13	45.25	-2.50

Table F.15. Sample Site: GNW 11

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GNW-11-78	26-Jul-23	355	8.62	2910	<0.067	16.0	1220	2180	0.385	63.4	58.4	7.56	590	33.83	32.95	2.62

Table F.16. Sample Site: GOW 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-1-78	25-Jul-23	391	7.57	3780	<0.067	21.5	1910	3360	1.26	197	272	11.4	393	49.59	48.19	2.86
GOW-1-79	10-Oct-23	387	7.58	3690	<0.067	22.9	1940	3290	1.24	191	262	11.5	389	48.30	48.78	-0.98
MEAN		389	7.58	3735	<0.067	22.2	1925	3325	1.25	194	267	11.4	391	48.95	48.48	

Table F.17. Sample Site: GOW 3

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-3-78	26-Jul-23	324	8.24	1760	<0.067	7.01	725	1410	0.616	150	133	9.56	83.6	22.31	21.77	2.43

Table F.18. Sample Site: GOW 4

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-4-78	26-Jul-23	287	8.43	2170	<0.067	10.8	744	1290	0.242	14.1	8.41	3.16	467	21.79	21.53	1.18

Table F.19. Sample Site: GOW 5

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-5-78	26-Jul-23	685	8.52	1680	0.677	150	<0.133	942	0.457	2.68	0.843	2.01	396	17.48	17.93	-2.54

Table F.20. Sample Site: GOW 6

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-6-78	26-Jul-23	270	7.62	1010	<0.067	3.08	283	685	0.466	91.3	69.4	5.08	31.7	11.77	11.38	3.41

Table F.21. Sample Site: GOW 11

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-11-78	24-Jul-23	343	7.55	3810	0.155	26.4	2180	2810	1.58	324	375	11.9	207	56.32	52.99	6.10
GOW-11-79	09-Oct-23	351	7.56	3830	0.17	27.5	2320	3630	1.58	323	365	12.3	212	55.68	56.10	-0.75
MEAN		347	7.56	3820	0.162	27.0	2250	3220	1.58	324	370	12.1	210	56.00	54.55	

Table F.22. Sample Site: GOW 12

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GOW-12-78	26-Jul-23	680	8.64	1550	0.455	117	0.418	906	0.436	2.43	0.884	1.75	375	16.55	16.91	-2.13

Table F.23. Sample Site: GSP 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-1-78	25-Jul-23	385	7.51	2750	<0.067	12.1	1440	2350	1.09	213	225	12.3	180	37.28	38.02	-1.96
GSP-1-79	10-Oct-23	407	7.7	2780	<0.067	11.6	1460	2240	1.17	222	217	13.4	184	37.27	38.86	-4.17
MEAN		396	7.60	2765	<0.067	11.8	1450	2295	1.13	218	221	12.8	182	37.28	38.44	

Table F.24. Sample Site: GSP 2

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-2-78	25-Jul-23	399	7.55	4070	0.154	22.2	2310	3580	1.63	258	403	21.9	317	60.37	56.70	6.28
GSP-2-79	10-Oct-23	413	7.48	4040	<0.067	20.4	2420	3760	1.68	260	391	19.7	311	59.17	59.22	-0.08
MEAN		406	7.52	4055	0.154	21.3	2365	3670	1.66	259	397	20.8	314	59.77	57.96	

Table F.25. Sample Site: GSP 3

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-3-78	24-Jul-23	218	7.24	1960	<0.067	7.79	965	1630	2.28	166	150	13.9	94.3	25.08	24.67	1.65
GSP-3-79	09-Oct-23	225	7.04	1930	<0.067	7.36	973	1580	2.03	165	152	14.0	103	25.58	24.97	2.42
MEAN		222	7.14	1945	<0.067	7.58	969	1605	2.16	166	151	14.0	98.6	25.33	24.82	

Table F.26. Sample Site: GSP 4

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-4-78	26-Jul-23	377	7.41	2280	<0.067	8.94	1010	1710	0.871	184	228	22.2	49.4	30.65	28.82	6.17

Table F.27. Sample Site: GSP 5

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-5-78	26-Jul-23	442	7.72	4460	<0.067	20.6	2860	3880	3.33	319	523	31.1	180	67.57	68.97	-2.05

Table F.28. Sample Site: GSP 6

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-6-78	26-Jul-23	641	7.42	3310	<0.067	11.3	1460	3070	1.32	304	312	16.0	123	46.59	43.53	6.79

Table F.29. Sample Site: GSP 7

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-7-78	26-Jul-23	413	8.38	2910	<0.067	22.5	1320	2560	0.885	176	234	11.6	201	37.07	36.38	1.89

Table F.30. Sample Site: GSP 9

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-9-78	26-Jul-23	279	7.55	1570	<0.067	9.48	596	1210	0.755	128	140	19.8	23.9	19.45	18.26	6.33

Table F.31. Sample Site: GSP 10

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSP-10-78	26-Jul-23	343	7.58	1200	<0.067	9.06	319	818	0.559	97.6	87.7	6.31	43.8	14.15	13.76	2.83

Table F.32. Sample Site: GSW 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSW-1-78	25-Jul-23	2620	8.18	4990	0.431	62.1	791	2970	2.87	380	437	73	346	71.83	70.62	1.70
GSW-1-79	10-Oct-23	778	7.78	5200	0.266	34.4	3300	5360	2.23	451	555	40.6	374	85.47	85.24	0.27
MEAN		1699	7.98	5095	0.348	48.2	2046	4165	2.55	416	496	56.8	360	78.65	77.93	

Table F.33. Sample Site: GSW 2

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSW-2-78	25-Jul-23	470	7.69	4230	0.203	26.9	2350	2830	1.94	272	375	25.6	297	57.99	59.08	-1.86
GSW-2-79	10-Oct-23	644	8.24	4410	0.274	34.8	2510	3930	1.91	270	450	36.2	344	66.38	66.12	0.39
MEAN		557	7.96	4320	0.238	30.8	2430	3380	1.92	271	412	30.9	320	62.19	62.60	

Table F.34. Sample Site: GSW 3

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSW-3-79	10-Oct-23	387	8.32	3530	<0.067	28.1	2110	3590	1.58	208	349	30.8	295	52.71	52.46	0.47

Table F.35. Sample Site: GSW 5

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSW-5-78	25-Jul-23	1020	7.23	7480	0.244	45.3	4680	6530	3.22	492	721	37.2	590	110.48	119.11	-7.52
GSW-5-79	10-Oct-23	692	7.85	8770	<0.067	80.5	6830	10300	3.94	518	1090	62.0	961	158.90	158.31	0.37
MEAN		856	7.54	8125	0.244	62.9	5755	8415	3.58	505	906	49.6	776	134.69	138.71	

Table F.36. Sample Site: GSW 6

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K(ppm)	Na(ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
GSW-6-78	25-Jul-23	783	7.71	8220	0.305	41.1	4360	7850	3.35	336	638	33.0	992	113.24	107.60	5.12
GSW-6-79	10-Oct-23	1010	7.51	9280	0.169	58.0	5980	11800	3.79	381	808	33.0	1250	140.70	146.34	-3.93
MEAN		896.5	7.61	8750	0.237	49.6	5170	9825	3.57	358	723	33.0	1121	126.97	126.97	

Table F.37. Sample Site: W 1

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
W-1-78	25-Jul-23	434	7.51	5200	0.178	23.7	2930	4860	1.92	313	465	20.5	508	76.49	70.35	8.36
W-1-79	10-Oct-23	435	7.58	4900	<0.067	22.0	2890	4680	1.93	285	385	19.6	485	67.49	69.49	-2.92
MEAN		434	7.54	5050	0.178	22.8	2910	4770	1.92	299	425	20.0	496	71.99	69.92	

Table F.38. Sample Site: PW 734

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
PW-734-79	10-Oct-23	210	8.14	477	<0.067	2.48	40.8	244	0.107	81.0	14.2	2.15	9.13	5.66	5.12	10.08

Table F.39. Sample Site: PW 735

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
PW-735-78	25-Jul-23	424	7.40	4300	0.158	19.6	2330	4280	1.98	242	343	17.6	415	58.79	57.54	2.15
PW-735-79	10-Oct-23	429	7.37	4130	<0.067	23.0	2330	3840	2.00	241	342	18.7	421	58.95	57.74	2.08
MEAN		426	7.38	4215	0.158	21.3	2330	4060	1.99	242	342	18.2	418	58.87	57.64	

Table F.40. Sample Site: PW-736

Sample	Date Sampled	Alkalinity (as ppm CaCO ₃)	pH	Cond (μS/cm)	Br- (ppm)	Cl- (ppm)	SO ₄ ²⁻ (ppm)	TDS (ppm)	B (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Na (ppm)	Cation Sum (meq)	Anion Sum (meq)	% Error
PW-736-78	25-Jul-23	365	7.43	3260	<0.067	17.3	1800	2380	1.17	246	273	13.3	212	44.29	45.26	-2.16
PW-736-79	10-Oct-23	363	7.43	3050	<0.067	17.5	1760	2590	1.14	234	257	13.3	206	42.12	44.40	-5.27
MEAN		364	7.43	3155	<0.067	17.4	1780	2485	1.16	240	265	13.3	209	43.21	44.83	

Appendix G

Statistical Analysis – Methods and Results

Appendix G

Statistical Analysis – Methods and Results

This appendix contains the results of the statistical analyses and an overview of the methods used to conduct them. For more technical documentation of these methods, readers with a deeper statistical background are referred to Appendix F in Colstrip reports of years prior to 2015 (the most complete of these being that in Thompson et al. [2014] in which random field assumptions and expected mean square error of the sampling regime are addressed) and Chamberlain (2018). There is no required statistical background for this appendix, but there are a few terms and concepts whose coverage is outside of the scope of this report—to understand them readers may consult an introductory statistics textbook such as McClave and Sincich (2016).

We performed statistical analysis to 1) determine if data exhibit differences among drainages or manifest recent changes over time and 2) characterize the water quality for each drainage. Specific objectives were as follows:

- Identify any recent time trends in individual site data and for each drainage overall.
- Detect any statistically significant differences between overall parameter means in the Cow Creek, South Fork Cow Creek, and Pony Creek drainages. Pony Creek is used as a “control” drainage, presumed to be unaffected by EHP activities, and is hydrologically and geochemically similar to both Cow Creek drainages. However, water quality in a drainage may change as a result of mining activity taking place therein.
- Estimate overall drainage means for the parameters of interest.

G.1 Data Used

G.1.1 Sample Sites

The parameters conductivity ($\mu\text{mhos/cm}$), boron (ppm), and sulfate (ppm) were examined in the statistical analysis. The sampling sites, listed in Table G.1, were groundwater wells and springs that were selected as being representative of the aquifers of interest.

The sites are ranked in terms of their distance from the headwaters of their respective drainages, with the shortest distance listed first. Wells PW 735 and PW 736, located between the EHP and Cow Creek, were installed in 2009, and the analysis results from sampling there are reported by Thompson et al. (2010–2023) and in Appendix F of this report. An additional site that is downgradient from the confluence of Cow Creek and South Fork Cow Creek drainages (W-1) was not included in the analysis because its location does not support drainage comparisons, and in the presence of time-drainage interactions (Olsen et al. 1992), time comparisons inclusive of this well are not statistically proper. Thus, its inclusion was not appropriate in the drainage comparisons or time-trend analyses. However, because of its location downstream of the confluence of the Cow Creek and South Cow Creek drainages, the cumulative site W-1 data could be useful as a composite representation of selected sites in the lower Cow and South Fork Cow Creek drainages, in a comparison with downstream Pony Creek sites (currently outside the scope of our study).

Table G.1. Sampling sites.

Cow Creek	Pony Creek	South Fork Cow Creek
PW 735	GSP 4	GSP 3
PW 736	GSP 6	GOW 11
GNW 1	GOW 3	GNW 6
GSP 1	GOW 4	GNW 7
GNW 2	GSP 7	GNW 8
GNW 3	GOW 12	--
GOW 1	GSP 8	--
GNW 4	--	--
GSP 2	--	--
GNW 5	--	--

GNW = Genie new well; GOW = Genie old well; GSP = Genie spring;
PW = private monitoring well.

G.1.2 Identification and Removal of Outliers

This section addresses the removal of inordinately extreme data (high or low) from the time-trend analyses and drainage comparisons and describes the outlier identification method used to identify these values. In the following discussion, we first identify the outlying data that were excluded, then we describe the method used to detect these anomalous data.

Data used in the analysis are presented in Appendix F (2023 sampling data), Olsen et al. (1992–2006), Thompson et al. (2008–2023), and McDonald et al. (2007). Data plots at the site level are given later in Section G.3.4, along with the results of the time-trend analysis. Extremely high or low data, as detected in the analysis of outliers or as identified by PNNL staff, were then examined for validity by scrutinizing sampling records and laboratory QA/QC information. Conclusively anomalous data were then removed from all statistical analyses as being “outliers”- i.e., data values that are extreme (high or low) and are highly likely to be artificial (such outcomes may occasionally occur even within a very stringently controlled sampling and analysis program, hence our implementation of an outlier identification step subsequent to receipt of the final and [any] re-analyzed data).

Extremely large or small data values were excluded from the analysis according to the methodology and procedure discussed below. However, data from GOW 12 since sampling event 60 were again omitted from the drainage time-trend and drainage comparisons, along with the 2017-2018 sampling results for GSP 7 because all of these values are unrepresentative of the Pony Creek aquifer, and therefore their site trend testing was suspended pending the acquisition of additional data and information (Thompson et al. 2018–2023).

The decidedly outlying data (and directly omitted data – e.g., recent GOW 12) omitted from this year’s laboratory results (Appendix F) and those removed in previous years (as discussed in earlier reports) are identified in Table G.2, where -36, for example, indicates sampling event 36. All of these data were excluded from the statistical analyses on the basis of their being classified as spuriously extreme data resulting from unknown causes and not in any way related to the (unknown) actual values.

In this year’s outlier investigation conductivity at GNW 5 (event 1) was declared to be extremely low – it was found to be inconsistent with the adjacent years’ data as well as lying far into the data rejection region (described in next paragraph) . Recent results from GSP 7, with the exception of the low conductivity value from event 70, continue to be retained in the analysis because those results appear to

be more representative than data from 2017–2018, based on conditions observed at that site during the sampling trips.

Table G.2. Data omitted from the analysis.^(a)

Cow Creek	Pony Creek	South Fork Cow Creek
GNW 1 -9 (cond) -19 (boron) -36 (boron)	GSP 4 -2 (cond.) -24 (cond., boron, SO ₄ ⁼) -28 (cond.) -36 (cond.) -38 (cond.) -42 (cond.) -48 (boron)	GSP 3 -4 (cond.) -16 (boron) -36 (boron) -48 (cond., SO ₄ ⁼) -54 (cond., SO ₄ ⁼) -58 (cond., SO ₄ ⁼)
GSP 1 -20 (boron) -21 (cond., SO ₄ ⁼) -23 (cond., boron, SO ₄ ⁼) -76 (cond.)	GSP 6 -13 (boron) -36 (boron)	GNW 6 -15 (boron) -42 (boron) -70 (SO ₄ ⁼)
	GOW 3 -2 (SO ₄ ⁼) -10 (cond., boron, SO ₄ ⁼) -12 (boron) -18 (boron) -20 (SO ₄ ⁼)	GNW 7 -1 (cond) -3 (SO ₄ ⁼)
GNW 2 -44 (SO ₄ ⁼)	GOW 4 None omitted	GNW 8 -11 (boron) -13 (boron) -36 (boron) -41 (SO ₄ ⁼) -77 (cond.)
GNW 3 -13 (boron) -20 (SO ₄ ⁼) -36 (boron) -54 (SO ₄ ⁼)	GSP 7 -2 (cond., boron, SO ₄ ⁼) -16 (boron) -66 through -69 (cond., boron, SO ₄ ⁼) -70 (cond.)	GOW 11 -11 (SO ₄ ⁼) -15 (boron) -44 (cond) -60 (cond.)
GOW 1 -13 (boron) -71 (SO ₄ ⁼)		
GNW 4 -44 (cond.)	GOW 12 -5 (cond., SO ₄ ⁼) -18 (cond.) -10 (cond., SO ₄ ⁼) -13 (cond., boron, SO ₄ ⁼) -18 (SO ₄ ⁼) -61 through -69 (cond., boron, SO ₄ ⁼)	
GSP 2 -13 (cond., boron, SO ₄ ⁼) -15 (boron) -16 (boron) -24 (boron) -60 (cond., boron, SO ₄ ⁼)	GSP 8 -12 (cond., boron, SO ₄ ⁼) -14 (cond., boron, SO ₄ ⁼) -16 (SO ₄ ⁼) -18 (cond., boron, SO ₄ ⁼) -30 (SO ₄ ⁼)	
GNW 5 -1 (cond)		

(a) Specific sampling events are indicated by a hyphen and the sampling event number.

The statistical detection of outliers was accomplished by direct inspection of data, comparing data with smoothed values using the LOWESS model fit (discussed below) in the “R” data analysis system (R Development Core Team 2024), examining regression residuals from the smoothed fit (i.e., data values minus fitted values, for each sampling event), and using a formal (objective) outlier rejection rule. The rule stipulates that all data whose residuals from a robust trend fit that are farther than 5.2 median absolute deviations (MADs) of the residuals from the median residual (approximately zero), should be rejected (X84 outlier rejection rule of Hampel et al. 1986).

A robust estimator of the mean or some other feature of the data’s probability distribution is one that is not influenced when a certain proportion of extreme-valued data points is encountered – i.e., that estimator is not sensitive to spuriously large or small values. Therefore, it remains useful as a reference measure of the remaining “clean data,” even when the outlying data are extreme.

In our example, a robust fit over time was applied to the cumulative data from each site (using the locally weighted scatterplot smoothing [LOWESS] model [Cleveland 1981]). The fitted lines and data were plotted together in order to visually inspect for outlying values. Then 1) the regression residuals (i.e., the differences between the data values and the fitted lines evaluated at the data sampling times) were obtained, 2) the median (MED) and MAD of the residuals were computed, and 3) data whose corresponding residuals were outside of the interval defined by $\text{MED} \pm 5.2 \text{ MAD}$ were considered probable outliers.

The procedure, originally described by Thompson et al. (2015), is again depicted in Figure G.1 for a hypothetical example over the years 1990–2015. The data are plotted in Figure G.1(a) and, based on visual observation, suspiciously spurious data—outliers—are labeled “C”. The robust LOWESS fit is overlaid with the data in Figure G.1(b), and its corresponding plot of residuals is given in Figure G.1(c), along with the outlier identification lines. Data whose residuals are farther from zero than $\pm 5.2 \text{ MAD}$ (flat lines)—in this case only, one data point, labeled “C” in Figure G.1(c)—is an outlier according to the above X84 rule. After the data are “cleaned” by removing outliers and otherwise confirmed non-data, in the time trend analysis described in the next section, a more optimal, data-sensitive smoothing method may be reliably applied in lieu of the LOWESS method. Such a preferred model is displayed in light blue in Figure G.1(d) and described there as the “final estimate.”

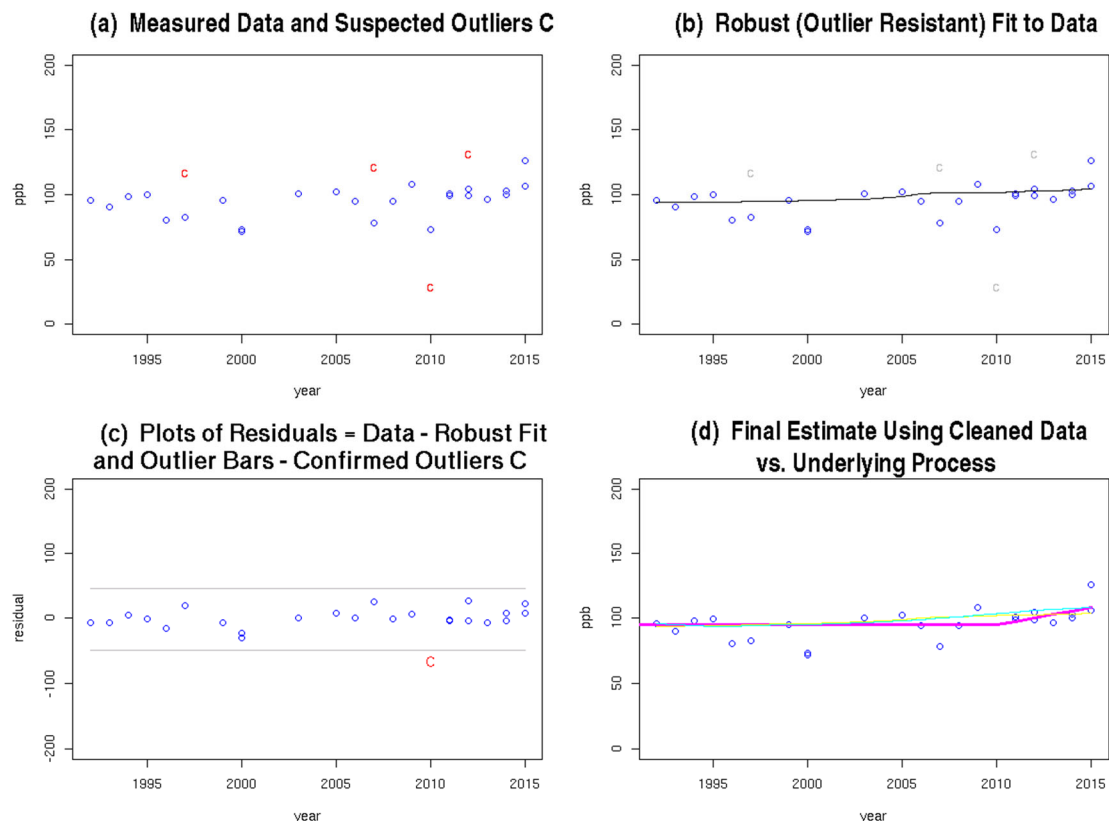


Figure G.1. Procedure for outlier analysis – hypothetical example (Thompson et al. 2015).

Data identified as outliers were then examined for validity by scrutinizing sampling records and laboratory QA/QC information. In recent years, in this study this rule has been used to determine a definitive acceptance region, rather than a rejection region; i.e., values within the 5.2 MADs from the median are definitely retained in the analysis, and those that are moderately more outlying are usually retained as well, instead of strictly rejecting all data outside the region. Also, data were retained when there was evidence of their validity. For example, there are instances in which the robust LOWESS smoother is too insensitive, not following the actual data, as at several springs (e.g., GSP 7 data from sampling campaigns in the 1990s), where it failed to adequately represent historically confirmed peaks—this resulted in some residuals being (artificially) outside of the ± 5.2 MAD envelope. In such cases, the data were typically retained in all analyses.

It should be emphasized that the analysis and detection of outliers is focused on finding "false" data that are spuriously high or low, whose theoretical probability distributions are different than the probability distribution of the actual data, and there is insufficient information about those other distributions to adjust the outlying data so that they may be included with the "well-behaved" data, in our subsequent analyses. But the outlier analysis is not useful for the more difficult task of identifying nonextreme data that derive from a probability distribution that has the same range as the actual data but is skewed or otherwise different as a result of some artificial causal mechanism. There are methods for this type of identification, however they are beyond the scope of our study. Had there been a notable bias/non-centrality of the results (i.e., having an error distribution that departs from the "bell shape" that is anticipated and confirmed for this type of data), which is due to either field sampling or laboratory analysis, such an inquiry might be justifiable and in such cases it may be possible to transform the skewed data to represent typical values, even in the presence of time trends, rather than remove those data from the analyses. This potential to recover the actual data would also apply to systematic outliers, so in

Thompson et al. (2016), a cursory examination of the distributions of outliers (as identified by “o” [instead of “c”] in Figure G.9) was performed for each applicable site and constituent; there was no conclusive evidence of any systematic causes, largely due to a (fortuitously) small number of outliers. Also, the QC results consistently indicate the nonexistence of any bias at the laboratory level. Based on the discussion in Section 3.6.2, the 2017 through 2019 GSP 7 data and GOW 12 data after sampling campaign 60, were qualitatively judged to be systematically outlying: therefore, no application of analysis was necessary to confirm their status as (systematic) outliers.

As indicated earlier, after the outlying data were removed, another model—SUPSMU—was fitted to the outlier-cleaned data as in Figure G.1(d). The supersmooth (i.e., SUPSMU) estimator reported by Friedman (1984), is a curve-fitting approach in which, for each given point (e.g., time) a straight line is fit using the data from nearby points only, and the final predictions are obtained from (the midpoints on) a moving series of shorter lines that steepen or flatten to accommodate the behavior of the data (e.g., conductivity). The window widths for the data to be included may change from one time point to the next and are selected so that the local lines are sufficiently sensitive to changes in the data, but not so much as to “chase” them too closely, causing unduly high uncertainty in the predictions. One significant advantage of the SUPSMU approach is that no particular functional form is globally preselected or imposed on all of the data—the method largely maintains any temporal changes in smoothed historical data while flexibly accommodating new changes, which may be either more or less dramatic than historical ones and may also tend in a different direction (increasing or decreasing). For representing “cleaned” data, the SUPSMU model is preferred over the LOWESS model, because SUPSMU has an appropriately higher data sensitivity, applicable when the data contain no outliers.

G.2 Statistical Testing Posture: Assume No Impact Unless Refuted by Data

All conclusions regarding temporal changes and in the comparisons of drainages are based on classical statistical testing, which *assumes a default condition or hypothesis, and only abandons the belief that it holds if the observed data are highly incompatible with that condition*. The default hypotheses in this report are 1) there is no increasing or high-valued trend (no decreasing one when testing for decreases), and 2) in the drainage-comparison tests, the drainage constituent means—based on all historical data pooled together—are all equal.

G.3 Time-Trend Testing and Drainage Comparisons

In this report, trend means time-varying, true value. Our use of the term “trend” here instead means a time-representation of a constituent level, be it cyclic, increasing, decreasing, etc. This trend is the true, underlying *level*; it is the curve that we would observe if there were no sampling or analysis errors. Sampling data always contain some “error” component, due to uncontrollable variation in both sampling and analysis, but it is desirable for the data to be close to the underlying trend values. This time trend aka “temporal trend” is not necessarily always-increasing or -decreasing over time, or even time-varying, but merely represents the true value at any point in time. Similarly, the term drainage trend does not imply a specific direction of change, only that the true drainage mean over time is being represented.

Recently Increasing or Decreasing Trends. Recently increasing or decreasing trends are, in this report, defined as those occurring continuously over at least the most recent 4 years (or eight sampling campaigns). This requirement reflects the expected nature of a contaminant plume while excluding spurious changes over only a few years that are weather-related. The 4-year requirement applies to both site- and drainage-level trend direction testing.

G.3.1 Testing for Recent Time Trends Using Our Conventional Test

This subsection describes our “conventional test”, applied in Colstrip Reports since the mid 1990s. As discussed later, the PLR test, first used in Colstrip trend testing in 2018, was also applied to the Stipulation 12(d) constituents, and the results of both tests were examined and a composited conclusion was made. Firstly, this section describes the conventional test, which addresses the following questions: “... are any recent, increasing trends at the site and drainage level valid? Or did we estimate increases ‘by chance’ while their true level is instead constant over time?” Because the data have some degree of uncertainty, fitted trends are also uncertain; so conclusions, based on visual inspection, that trend directions are truly representative could be highly unreliable (especially when the data vary greatly around a fitted trend line). So we instead examine *the probability of observing an increase in the fitted trend*. We answer the first question above by assuming (in our statistical distribution) that the second one holds true—the true level is actually constant in recent years—we abandon this assumption and conclude that there is indeed an increasing trend if the probability of observing, under the flat-trend assumption, the originally fitted increasing trend “by chance” is small. (If instead the originally fitted trendline was continually *decreasing* we would again test under the same flat-trend assumption but conclude that a *decreasing trend actually exists* if the probability was extremely low, under the flat-trend assumption, of fitting the originally decreasing continual trend.)

The testing in this section involves 1) computing the above probability, and 2) concluding that a fitted increase, at either the site or drainage level, is valid only if that probability of a “by-chance” occurrence is less than the customary threshold value of 5%. Otherwise, the observed increase is concluded to be not valid, having occurred merely by chance—a chance event having a probability greater than 5%. In the same manner we also address the case of decreasing trends, because actual decreases give us insight into a contaminant plume's movement and intensity. Use of this customary value ensures that if our assumption is true, the error rate of incorrectly deciding that an increasing (or decreasing) trend exists is small—less than 5%.

As mentioned earlier, in this study an increasing time trend is defined as a continual increase in the fitted time trend for the most recent 4 or more years. This hypothesis was constructed to provide the earliest detection of recent trends subject to requiring sufficient data for such detection. Ideally, a change in water quality would be detected very early, but with the expectation that a contaminant plume exhibits the gradual trend that commonly occurs at a site under typical conditions, 4 years was selected as the shortest duration that would be considered for the trend to occur. Shorter periods would cause our definition to be overly sensitive to sporadic increases due to sampling/measurement variation or small hydrogeologic changes. See Thompson et al. (2014, Appendix F.2) for further discussion of the groundwater transport mechanisms that would induce varying degrees of increase. Likewise, a decreasing trend requires consecutive decreases in the fitted value for 4 or more recent years.

We compute our test probabilities using a form of Monte Carlo simulation, a computational method in which in general all uncertain factors are randomly generated *repeatedly* from their respective probability distributions, and in each repetition or *iteration*, the equation or mathematical model of interest is tabulated using the generated factors for that iteration. For a large number of iterations, the collection of these tabulated results then approximately represents the *statistical population*—all possible occurrences—of the equation being simulated. In our case the data are the uncertain elements, and for each site or drainage a data series representative of the entire constituent data history is randomly generated repeatedly—100,000 times—under the assumption of no recent trend (actually a flat trendline), and in each iteration a SUPSMU trend curve is fitted to those data. The proportion of the total population (i.e., 100,000) of fitted trends that show recent increases as defined above represents the probability, *under the no-trend assumption*, of fitting an increase “by chance.” If this probability is less than 5%, we reject the assumed “by-chance” occurrence and instead conclude that the increasing trend in the originally

fitted curve is valid; i.e., the constituent is increasing at that site or drainage. (If instead a decreasing trend, over the most recent 4 years or longer, was fitted, we would test using the same simulations under the recent flat-trend assumption, and we would conclude that a recently decreasing trend exists if fewer than 5% of the simulated trends were decreasing – i.e., the probability, under the recent flat-trend assumption, of fitting a decreasing trend was too low to accept that the decrease was merely fitted by chance.)

The details of our simulation and probability testing are illustrated in Figure G.2 and were originally presented by Thompson et al. [2015]) using hypothetical data and are given in the following steps:

1. Examine the fitted trend for consistent changes (see Figure G.2a). Fit a trend to the outlier-cleaned data (using the SUPSMU regression model) and examine this trend for continual increases and continual decreases over the most recent 4 or more years. If no such changes are seen in the trend line, then do not test—the conclusion is that *no trend was found*. Alternatively, if a continual change is found, such as the increase since 2010 in Figure G.2a, then compute the probability of a chance occurrence by continuing Steps 2 through 8.
2. Compute the trend residuals (see Figure G.2b) by subtracting the fitted trend from the data value at each sampling event date.
3. Construct the assumed trend line (see Figure G.2c). Assume that no recently changing trend exists and construct this hypothesized “true” trend line as 1) the fitted trend for the pre-change period (e.g., before 2010), and 2) the average over the recent trend period, of the fitted trend; use this average for the entire change period as in Figure G.2c.
4. Generate the population of 100,000 fitted trends (see Figure G.2d) using the following procedure. In each Monte Carlo iteration, generate a new data series representing the entire sampling history—another data history that might have occurred—under the assumption that no recent continual trend has occurred, and fit a trend to this series (representing a trend that could have occurred under the same assumption). Instead of generating this data series from a probability distribution as in an ordinary Monte Carlo simulation, randomly resample with replacement, a new series of residuals from the computed residuals; that is, for each time point, select a residual “out of a hat that contains our residuals,” add that residual to the hypothesized trend (from Step 3) to obtain a simulated data value for that time point, then return the residual “back into the hat” for possible reselection at another time point. This is *sampling with replacement* so that all possible realizations—the entire distribution of residuals—are always sampled from.
5. Figure G.2d shows one series of these data, reconstructed under the flat-trend assumption.
6. Figure G.2, subplots e and i through h and l illustrate four iterations of our approach to Monte Carlo simulation. The resampled residuals for the four iterations are given in Figure G.2, subplots e through h and the corresponding trends fitted to data reconstructed from those residuals are given in Figure G.2, subplots i through l.
7. In each iteration, examine the fitted trend for an increase over the originally identified period (of increase). (If the original trend exhibited a consistent decrease, evaluate the simulated trend fits for decreases instead.)
8. Compute the probability, under the no-trend assumption, of obtaining a spuriously occurring trend increase, and make the test conclusion, as follows. After the Monte Carlo iterations are completed, compute the proportion of the 100,000 trend fits that exhibited an increase. If that proportion is less than 5%, abandon the position that no increase is occurring and conclude that the originally fitted recent increase is valid (because if the no-increase assumption were true, it is unlikely [less than a 5% chance] that an increasing trend would be fitted). Use of the 5% rule ensures that the “false-detect” error probability is under 5%.

9. As an example of this probability computation, if we tested for an increase using only the four iterations (we actually use 100,000 iterations, to obtain a better probability estimate) in Figure G.2, subplots i through l, we note that in one of those iterations we fitted an increase, so our estimated probability of such an increase occurring by chance is one in four, or 25%. So we would say that spurious increases are too likely (chance of 1 in 4, i.e., 25%)—more likely than 5%—for our concluding that our original increase is valid, so we maintain that there is actually no increasing trend. Alternatively, if we again based our conclusion on only four iterations but we instead found that no increases were fitted among them we would conclude that the true trend is an increasing one because our estimated probability of obtaining an increasing fit (if the no-trend hypothesis is true) would be 0 in 4 (i.e., 0%, which is less than 5%). While using just a few iterations has a simplistic appeal, the foregoing estimated probability will vary widely between groups of simulations of four iterations, even from 0% to 100% (i.e., it has a very high uncertainty level). So we instead use many iterations (100,000) to obtain a stable probability estimate which also closely estimates the true probability of fitting an increasing trend (when a flat trend actually exists). However, as discussed later in Section G.3.3.2, because many tests are performed, we decrease the “5% cutoff” level in each individual test so that the overall false detection rate of 5% is maintained.

The foregoing example for testing applied to an increasing trend. We actually test for either an increasing trend or a decreasing one, depending upon the direction (of the most recent 4 years) of the originally estimated trend – if this trend is decreasing for 4 or more of the most recent years, we test for a decreasing trend (the number of *decreasing* trends in the above Monte Carlo simulation is instead used in computing the p-value).

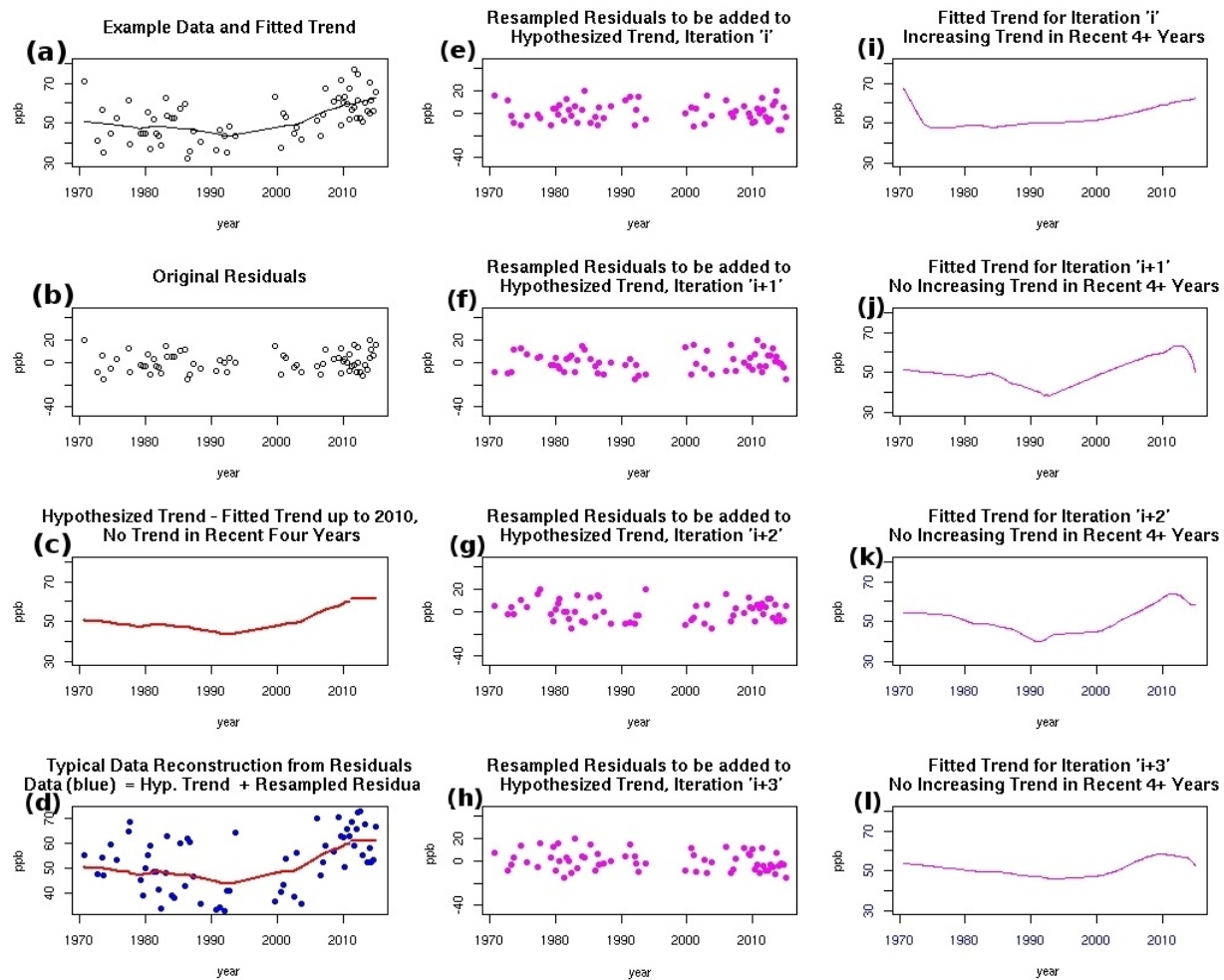


Figure G.2. Procedure for testing increasing trends – hypothetical example.

G.3.2 Model Uncertainties – Minimum and Maximum Trend Values at Each Time Point

In addition to the testing described in the previous section, minimum and maximum among “all possible” estimates were tabulated to describe the uncertainty in our trends (fitted to the original data) using Monte Carlo simulation. But in this simulation, no assumption was made concerning a recent constant mean; the original estimated trend was used as our best estimate of the mean over the entire time history. As with the testing simulation, in each iteration resampled residuals were added to a trend—in this case, the originally estimated trend—to obtain a simulated set of data, and then a trend was fitted to these data. For each time point, the minimum and maximum of the (simulated population of) fitted trends was obtained and plotted along with the data and originally fitted trends. Doing so formed an envelope around the trends that describes, for each time point, the minimum and maximum values that could have occurred from this estimation procedure.

Figure G.3 (Thompson et al. 2015) depicts simulation results for the same generic data in Figure G.2 (i.e., plots of the fitted trend from each iteration). Also, as is shown here for only year 2015, probability distributions of the model could be obtained at some or all of the other time points as well. However, for the site trend fits illustrated in the next sections, we are reporting only the extreme-most values instead of estimated probability distributions.

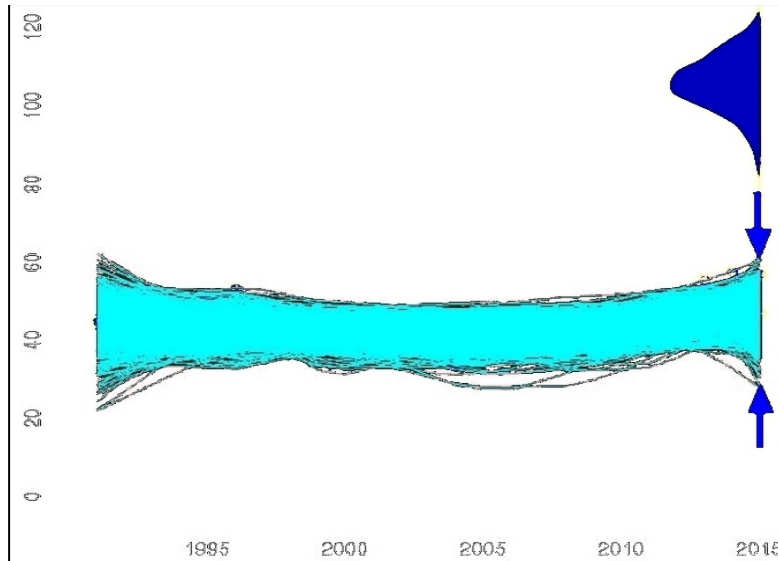


Figure G.3. Uncertainty assessment on fitted model: Overlay of all simulation results and probability distribution on 2015 predictions – hypothetical data (Thompson et al. 2015).

G.3.3 Testing for Increasing Trends in Conductivity or Boron: PLR Test Compositied with the Standard Trend Test

G.3.3.1 PLR Test for Increasing or Very High Trends

In statistical testing between the default ("null") hypothesis and an alternative hypothesis, the most fundamental test is the Likelihood Ratio test. Most hypothesis tests are based on this one, the actual forms of their *test statistic* being actually derived from an underlying Likelihood Ratio. Originally introduced in the 1930s, this ratio and its test have become the standard basis for statistical hypothesis testing, largely because when applied to two simple competing hypotheses, its ability to detect when the alternative hypothesis is true is maximum possible. This statistic is the ratio of the likelihood of the data's occurring if the alternative is true to the likelihood of their occurrence if the null hypothesis is true: it is also known as the *odds ratio* (unfortunately though this latter reference falsely suggests that we may merely "examine the odds" and make a conclusion, it will presently be shown that the actual testing involves instead determining the *probability distribution* of this ratio and where the observed ratio scores in that distribution). Also, data are often continuously distributed – they can occur over a continuum rather than taking on discrete values (e.g., whole numbers) so their probability of occurring at any exact value is infinitesimally small: analogous to being represented by a probability their distribution is instead represented as a probability curve (i.e., a "bell curve" or other-shaped one), known as a likelihood function when testing, which is evaluated at the data value. This curve will be different, in shape and/or location (ie being shifted to the right or left), for each hypothesis. Also, if each data point's curve is unaffected by previously measured data, the data are considered to be mutually independent, in which case the likelihood of the joint occurrence of the values all the data have taken is the product of their individual curve values. Therefore, the Likelihood Ratio is usually a ratio of two products of hypothesis-specific curve values.

The Likelihood Ratio is itself a statistic, having a probability curve of its own, and that curve will also be different depending upon which hypothesis is true. If the alternative hypothesis is true, the Likelihood Ratio will tend to be large because the data tend to fit the numerator likelihood better than the denominator one; that is, the curve values in the numerator's product will (on average, geometrically) be higher than those in the denominator's one. But if the null hypothesis is true, the Likelihood Ratio will

tend to be small, reflecting the better fit of the denominator's Joint Likelihood. However, the two hypothesis-specific ranges that the ratio can take often overlap somewhat, sometimes considerably.

The decision in testing the null hypothesis using the Likelihood Ratio is much the same as the testing decision described in Section G.3.1: the null hypothesis is assumed to be true and if the probability (under that assumption) of the Likelihood Ratio's being as large as (or larger than) the one observed (i.e., the ratio's p-value) is smaller than 5%, we conclude that the ratio is very unlikely to have occurred “by chance” under the null assumption, so we reject the null hypothesis and conclude that the alternative one is true. In this part of the discussion, we again preselect 5% as our acceptable “false-detect” rate: under the null hypothesis, spuriously large Likelihood Ratio values that have p-values less than 5%, and thus trigger (falsely) rejecting the null assumption, occur only 5% of the time. Finally, computing the p-value requires estimating the distribution of the Likelihood Ratio when the null hypothesis is true: in our more complex application, this estimated distribution is obtained by Monte Carlo sampling; that is, by repeatedly generating a series of data under the null hypothesis assumption and evaluating the resulting Likelihood Ratio in each iteration. The resulting “ensemble” of values approximates the population of all possible Likelihood Ratios that could occur if the null hypothesis was true, and if less than 5% of them are larger than the one computed from the original data, we reject the null hypothesis. In testing multiple hypotheses, a false reject rate lower than 5% may be specified for each individual test so that the overall error rate—of one or more false rejects—does not exceed 5%.

The Likelihood Ratio test would be useful if we were testing the null hypothesis of no trend (as in our customary test) versus a single, specific increasing trend. However, in testing the condition of Stipulation 12(d) where an abrupt increase at Cow Creek is cause for remedial intervention there, in addition to a flat trend's being acceptable, we are also not concerned about small trend oscillations and decreasing trends, whereas large “jumps” in trend and possibly continually increasing ones, trigger intervention. Therefore, the candidate trendforms under each hypothesis are numerous—actually infinite—and the simple Likelihood Ratio framework will not accommodate this realistic complexity. Hence, the PLR (Chamberlain 2018) was employed, as the ratio of Predictive Likelihoods under each hypothesis. If all possible trendforms over the four most recent years are approximated by “dot-to-dot” connections between their values at the eight sampling campaign times, similar to the plots of Figure G.4 (but over the applicable most recent years), they may then be grouped by those that increase relative to a pre-trend reference value (i.e., 50 in these plots) or exhibit large increases in the time period being tested (both shown in the lower plot, having a 4:1 relative frequency of occurrence)—all potentially triggering intervention—and by those that do not (upper plot), being of no concern with respect to Stipulation 12(d). (Only 100 trendforms were generated for these plots, though more than 10^{10} are used for our actual testing, which uses a grid of 31 possible values per time point.) Very many trendforms are considered because multiple historic accidental releases of varying durations, coupled with many different precipitation levels and seasons, could potentially result in one of almost an innumerable number of trendforms of various shapes (shapes of overlapping plumes and weather-related “phantom” plumes) at some sites, so this large of a set is expected to include at least one trendform that is close to the actual one.

Each trend in a group represents a possible state that has occurred—a *state of nature*. A *Predictive Likelihood* is a term used in Bayesian statistics when a specific characteristic that determines the shape or location of the probability curve—for example, a parameter, or here a trendform that impacts the Joint Likelihood curve—is considered to have a probability distribution itself: the Predictive Likelihood is the distribution of the joint data curve after probability-weight averaging that curve over all possible values of the characteristic. The probability distribution on the characteristic, known as the *Bayes prior distribution* (or *Bayes posterior distribution*, when historic data are combined with a prior distribution) may be limited to discrete values or be defined over a continuum of them, according to the researcher's and/or decisionmaker's belief about the characteristic's value. This probability averaging produces a joint

probability curve for the data that is no longer dependent upon only one specific value of that characteristic, but rather is shaped by all of them. The Predictive Likelihood may also be partially evaluated, over only a subset of the characteristics of interest; for example, over all trendforms in a hypothesis trendform group, producing a likelihood that is shaped only by the trendforms in that group, representing the accumulation of the combined possible occurrence of each and the plausibility of the data it induces. Therefore, in this application, the Predictive Likelihood under the alternative hypothesis that intervention is needed combines, over “all possible” undesirable trends, the data Joint Likelihoods (described earlier) evaluated at each of those trends, whereas under the null hypothesis (intervention not needed) the Predictive Likelihood combines the Joint Likelihoods applied at each of the “innocuous” trends. For each group, the predictive likelihood was tabulated as a simple average of the Joint Likelihoods and represents an equiprobability-weighted average of the occurrences of each trend in a group (innocuous or not), reflecting an “unbiased” viewpoint that each trendform in a group is equally likely (these probability-weighted likelihoods, before their accumulation, are proportional to the Bayesian *posterior distribution*, discussed later). Under the additional unbiased assumption that each trendform group is equally likely to occur, the PLR is the ratio of the two (partial) Predictive Likelihoods just described, and in similarity to the Likelihood Ratio, it will tend to be large if the data suggest more strongly that one or more increasing/high-valued trendforms best fit them, but if instead some innocuous trendforms fit the data better, this ratio will tend to be smaller thus supporting the null hypothesis that there is no concern relative to the Stipulation 12(d). As in our conventional tests, the actual testing involves repeatedly generating a data series under the null hypothesis—here, in each iteration, based on a randomly selected innocuous trendform—then evaluating the PLR at each generated series. If less than 5% of these simulated PLR values exceed the PLR computed from the actual data, we say that the observed PLR is uncommonly high if the null hypothesis was true, and so we reject that hypothesis (of an innocuous underlying trend, which includes no trend), and we conclude that either there is an increasing recent trend or there are one or more abruptly large trend values (i.e., “jumps”) in some of the recent 4 years.

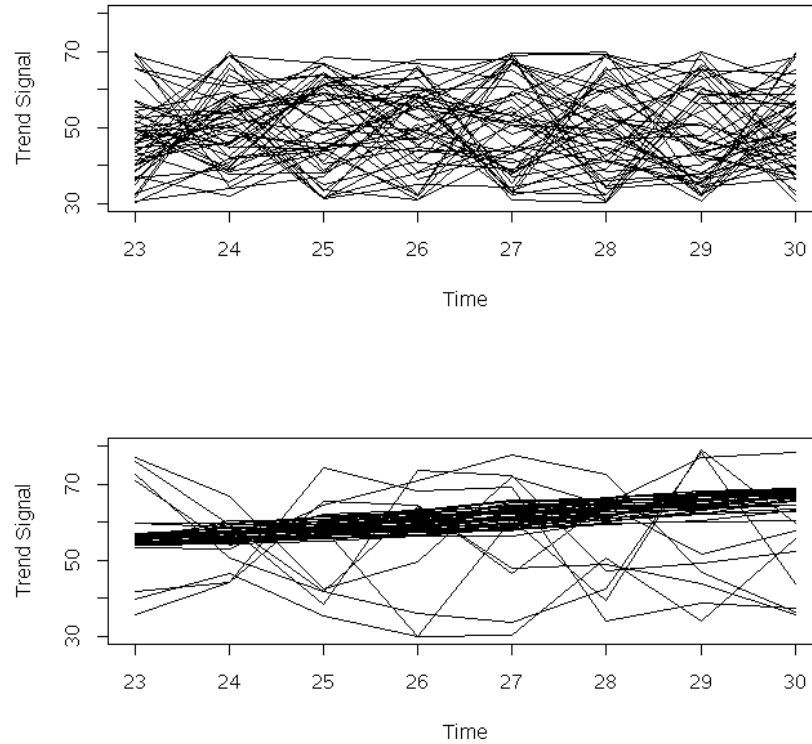


Figure G.4. Generic example of 50 possible trendforms, under non-increasing hypothesis (top) and increasing or high-valued (exceeding 70 units at least once in the timeframe considered) trendform hypothesis (bottom) (from Chamberlain, 2018).

Within the PLR testing, the generated data were obtained in a way similar to our conventional test for a trend, as described in Section G.3.1, Steps 1–4, and illustrated in Figure G.2(a)–(d) there, but in this case data were reconstructed for only the most recent eight sampling time periods and here, in each iteration a series of random measurement errors (taken from the original error estimates) was added to an innocuous trendform, which was randomly selected as the underlying trend for that series, rather than the errors' being added to a single underlying recently flat trendform as was done in the other test. The remaining part of the PLR test procedure is however different from the conventional test; whereas in that test a trend was fitted to the simulated data and, after all iterations, the p-value was the proportion of those trends that agreed, in direction, with the originally estimated trend's, here in each iteration a PLR was computed from the simulated 8-data series and the p-value was the proportion of all simulated PLR values that exceeded the PLR computed from the data.

The pre-trend reference value used to define an increasing trend (which increases continuously above this value) and, therefore, to identify the triggering group, was the most recent pre-trend value from a local (i.e., moving) polynomial fit similar to supersmooth—30 grid points were constructed using this reference value, usually symmetrically around it. The grid range's half-distance, added to and subtracted from, the reference value to respectively obtain the maximum and minimum grid values, was computed as the range of the data over the entire study period plus 2 error standard deviations, except the lower grid value was set to zero when the subtracted half-distances would otherwise result in negative grid values. The extreme, high values that define an “abrupt increase” were taken as the uppermost three grid points. Figure G.5 illustrates the grid for boron at GNW 2.

Boron: GNW2

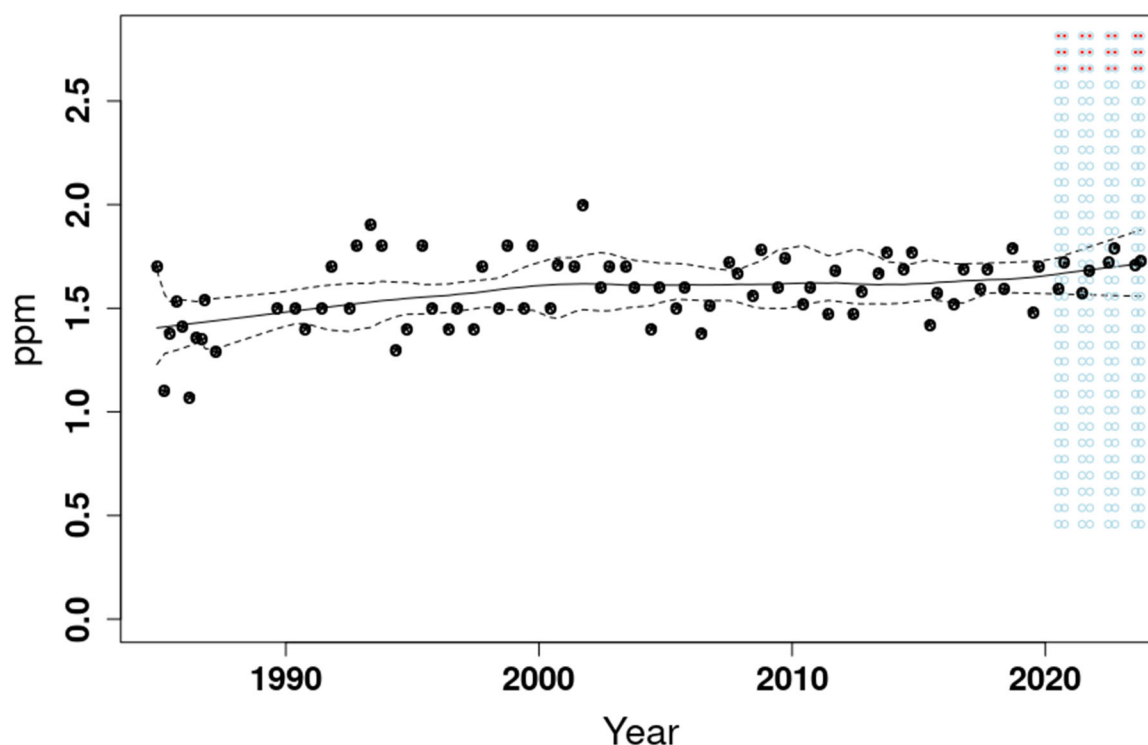


Figure G.5. Example gridpoints (blue, red, open circles) used in approximating all possible trendforms.

The occurrence (*prior*) probabilities for the subgroups, increasing and high-valued (trends having one or more threshold-exceeding values in eight sampling times), were respectively 0.4 and 0.1, and the innocuous group's occurrence had the remaining 0.5 prior probability. Returning to the PLR computation, as described earlier the PLR was obtained as the ratio of Predictive Likelihoods for the two groups and for each group, the simple average, among all trendforms in the group, of the trendform-specific Joint Likelihoods was tabulated at the generated series. Each trendform's joint likelihood was the product of the likelihood curve density function values - viz. when that function was evaluated at each of the eight data values (representing four or more most recent years) and the particular trendform's values at the respective datum times. Specifically, the probability curve used in the Joint Likelihood computations actually characterizes the distribution of original measurement errors, being estimated from them, so the curve was evaluated based on the *difference* between each generated data value (or the actual data value, when the originally observed PLR was to be computed) and the trendform's value at that time. It should be noted that in each iteration these Joint Likelihood computations were applied to all trendforms in a group—approximately 28^8 in the innocuous group and 30^8-28^8 in the triggering one—and in all but one case (when evaluating the denominator Predictive Likelihood) they were not the same trendform as the (null-based) one randomly selected and used to construct the data for that iteration. When the difference between datum and trendform was small in magnitude (representing a close match) the curve value was high (peaking around zero [recall that the observed error components—aka residuals—from the original trend fit, are the differences between the data and that fit and can be either positive or negative]). And further, if the product of these curve evaluations was also high, then that trendform closely corresponded to the generated (or observed) data and its contribution to the average—the Predictive Likelihood for that

hypothesis—was also somewhat high. Conversely, the trendforms whose values in each time point were distant from the generated/observed data resulted in differences that were too large in magnitude to be in the typical range of the errors, and therefore their resulting curve values were very small or even zero, causing the respective Joint Likelihoods to contribute very little or zero to the Predictive Likelihood for that hypothesis group, and if all contributions were small, the Predictive Likelihood for the competing hypothesis was probably largest. Therefore, because the data are generated under the null assumption, we would expect the PLR's distribution to cover small PLR values (usually fitting the null trendforms best and hence causing the denominator Predictive Likelihood to tend to be larger than the numerator one); i.e., values smaller than the resulting PLR values if the data were generated under the alternative hypothesis. So, we then compare this distribution with the observed PLR and compute the latter's p-value as the percentage of simulated PLR values that exceed the observed one. (It should be emphasized that the observed PLR is evaluated based on the actual data which have an unknown underlying trendform whose group is also unknown.) If the observed PLR is very large relative to the (null hypothesis based) simulated distribution of the PLR—for example, the observed PLR has a p-value smaller than 5%—we would reject the null hypothesis in favor of the alternative one, concluding that either an increasing trend exists, or one or more underlying means in recent years was very high.

The probability curve, technically referred to as the *probability density function*, was obtained by kernel density estimation (Silverman, 1986) using as its data the original observed errors from a (supersmooth-like) nonparametric trend fit that did not impose any trendform assumptions. Unlike a histogram for example, the kernel density estimate is a continuous function, providing a useful quantitative density value at any error value. This density becomes zero for extremely large or small (i.e., large, negative) errors. For example, within the evaluation of a given trendform's Joint Likelihood computation, if a generated/observed datum was within the Joint Likelihood computation, if a generated datum was very negative or distant from that trendform, their difference was extreme, and the density's resulting near-zero value indicated that the probability of that difference's being a measurement error was extremely low—so, virtually no errors this large or larger in magnitude could occur. This suggests quantitatively that the datum's underlying mean at that time point was different from the trendform value then (and hence dampened the Joint Likelihood for that trendform and its contribution, in the PLR computation, to the overall Predictive Likelihood for the trendform's group).

G.3.3.2 The Composite Test for Increasing or High-Valued Site Trends: Combining the Best Performance Features of the Trend and PLR Tests

In 2018, a special study using the data for conductivity and boron at Cow Creek sites was conducted to evaluate the performance of our conventional test and the PLR one. The evaluation involved Monte Carlo simulation repeatedly generating data whose underlying trendform was either flat, or in a separate set of runs, increasing, applying the tests in each iteration, and evaluating both tests' results from all the runs for that trendform. This simulation was an “outer” one, in which instead of applying the PLR test just once to laboratory data, the PLR test was repeatedly applied (each testing instance representing an “inner” computation in the overall framework) to data that were repeatedly reconstructed from each of the trendforms just mentioned by adding to them random selections from relevant site-constituent residual data pools. Study simulations were applied to the residuals of smoothed data (through 2017) for each site in Cow Creek, for both conductivity and boron, to assess the overall drainage testing error rates discussed later, and because for a given constituent, the distributions of the estimated residuals used to generate data in the outer simulation and within each test were site-specific; the residuals each represented the total measurement errors—the sum of (site-specific) sampling error and laboratory error components. Both (testing) error rates were then assessed for each test—the proportion of iterations resulting in false detect and the proportion for which it failed to detect, when the underlying trends were respectively flat, and increasing. At all sites, the PLR test was found to have a much higher rate of detecting increasing underlying trends (i.e., corrected detection rate). As expected, regarding false detection, the conventional

test had a false-detect rate lower than 5% (because testing is only done if the originally estimated trend is increasing). From an overall drainage perspective addressing whether one or more Cow Creek sites is increasing in either boron or conductivity, the false-detect results of both test results were approximately equal. They showed very high false-detect rates for the drainage overall (one or more site-level false detects when no increasing trends exist at any site there), while the drainage-level fail-to-detect rate—no detections when at least one site was increasing—were 6% for the current test and approximately zero for the PLR test.

When doing multiple hypothesis tests, the probability of at least one false detect among them is always higher than it is for an individual test, as high as the sum of the individual test's false-detect rates. So to reduce this overall testing error, a common approach is to use Bonferroni's correction (Milliken and Johnson 1984) by setting the individual error rates to the target overall false-detect rate (e.g., 5%), divided by the number of tests; that is, in an individual test, reject the null hypothesis when the observed p-value is lower than this adjusted rate and is rejecting less often. This correction ensures that the total error does not exceed the target rate, but it is often overly conservative and overstates the actual overall rate. For any given test, this correction greatly reduces the false-detection rate, but it increases the fail-to-detect rate, also known as the Type II error. Therefore, in the second study, to achieve the overall drainage false-detect level of 5%, the allowable individual Type I error rates for testing each of the eight Cow Creek sites, for boron and conductivity, were reduced to 0.003 (equaling 5% divided by 16 tests). However, because the conventional test's fail-to-detect rate was very high as a result of this correction, a composite test using both the PLR and conventional test was constructed to exploit the PLR's high correct-detect rate observed at even these more detection-reluctant testing levels. The composite test was as follows: conclude an increasing (or high) trend if either test concludes one. Also, because this test involves two tests, the target error rate for each test was further reduced by one-half to control the event of one or more false detects in that test at each site-constituent combination. The resulting performance of the composite test was remarkable, producing both a low overall false-detect rate in the drainage (under the flat-trend assumption) and a low fail-to-detect rate (when one or more increasing trends in either constituent was occurring at Cow Creek). *Therefore, since 2017, the composite test continues to be applied to the cumulative (boron and conductivity) data in testing for increasing or high site trends in the drainages.*

The nonessential recent trend tests - for decreases in conductivity or boron, increases or decreases in sulfate, and drainage-level trend testing in either direction – 94 tests in all (testing at GOW 12 is currently suspended) – each use the conventional test alone however, and the individual allowable (type I) error rate for each test was set at 0.05/94, again being based on Bonferroni's correction.

Why We Test for Increasing Trends of Any Magnitude Instead of Testing for Only Those That Are Meaningfully Large

It is important, even essential, that we test for any level of increasing trend. Here are the primary reasons:

- **Maintaining Statistical Test Sensitivity in Detecting Actual Increases – i.e., the Power of the Test.** If negligibly increasing trends - those below some minimum "cutoff" value - were moved into the innocuous set, because there is considerable overlap in the range of possible data when a trend is below a cutoff and the range when a trend exceeds that cutoff, trends on either side would frequently not be statistically distinguishable. In other words, when there is a significantly increasing trend (i.e., a severe one) just above a cutoff, the probability of a no-increase decision would be high, or at least much higher than it currently is, and so the fail-to-detect rate would be much greater. Moreover, only trends far worse than those that minimally qualify as severe would be reliably detected. *Therefore, we choose to maintain a cutoff very low - far away from highly increasing levels, so that we may detect them with high likelihood.*

- Also, when an increasing trend is detected, the supersmooth trend estimate and the posterior probability over all possible increasing trends, respectively provide accurate deterministic and probabilistic trend estimates; *testing for a wide importance range of increasing trends doesn't reduce our ability to correctly estimate any severe trends that may occur.*
- **Characterization of Extent, if Contamination Occurs.** Because most Cow Creek sites are hydrogeologically distant from the EHP, any encroaching groundwater plume should have increasingly spread its extent while becoming more dilute (largely due to typical contaminant spreading/dispersing mechanics acting in the flow direction and laterally, arising from varying groundwater velocities within each pore path and soil matrix configuration): therefore, constituent levels at these sites should initially manifest a gradual increase rather than a "jump," and so testing for small increases downgradient of a highly contaminated site enables establishing, or at least limiting, the plume's extent – e.g., the plume is just reaching the downgradient site, or, it hasn't arrived there yet.
- **Characterization of Nature of Contamination.** Further, if no negligible increase was ever detected previously at a highly contaminated site, this would tend to indicate that the source isn't contaminated groundwater at all, but instead is the result of contaminated streamflow/stormwater runoff that is sufficiently close to produce a contaminant breakthrough ("jump") at the site, facilitating our characterization of the nature of the contamination. Finally, if many sites exhibit concurrent, small increases, we also may suspect that the source is contaminated runoff because in the absence of rapid groundwater movement due to karstic or fracture flows, it is impossible for a single plume to affect all sites simultaneously with the onset of a small increase (and it is highly unlikely to observe simultaneous increases even if multiple plumes are present in the drainage). Also, detection of initially small increasing trends that are widespread in this way enables more timely remedial response.

G.3.3.3 The Posterior Distribution – For Interpreting the Observed PLR, and for Evaluating the Severity of Detected Trends

In Bayesian analysis, the true “state of nature,” here the true trendform, is assumed to have randomly occurred according to a probability distribution on it - a prior distribution – e.g., which reflects our belief about the entire set of trendforms that could possibly occur; to maintain "unbiasedness" our prior distribution assumes that each trendform is equally likely to have occurred. After data are observed, the prior distribution is reweighted – multiplied - by the (trend-specific) data distribution evaluated at each trendform; the resulting updated distribution on the trendforms - the posterior distribution - is obtained from this reweighting. As described next, *among the initial trendforms considered, those implausible ones, given the data that we've observed, are essentially filtered-out in the posterior distribution and its contribution to the observed PLR.*

The data distribution is similar to a “bell curve” and evaluating it at various trendforms, given the observed data, is similar to evaluating a single data value’s bell curve score when the curve is shifted by varying amounts and directions – when the curve’s peak is close to the datum the score will be high, but the score will decrease as the peak is moved away, in either direction, from the data value, indicating that it is incompatible with trendform values that are far away. Therefore, as discussed earlier, trendforms that are “distant” from the data at several time points will have low scores that will dampen the original prior probabilities at those trendforms, whereas the prior probabilities for trendforms that are close to the data at most or all time points will be augmented by high scores. Because this reweighting refines the prior distribution by augmenting/dampening it at each trendform according as the data are/aren’t compatible with it, the posterior values will be approximately zero at impossible trends given the observed data: in general, the posterior gives us an improved, more correct and focused set of probability estimates for the trendforms. The posterior and prior distributions for (approximately) all possible increasing trendforms for boron at PW 735 are illustrated in Figure G.6; note the posterior’s redistribution of probability, as compared to the equally likely prior distribution toward the trendforms with moderately high 8-campaign

percentage increases, away from both the more extreme ends that were admitted by the initial prior distribution. The probabilities in this posterior distribution and in the prior one, if we conclude that the actual trendform is increasing, sum to 1.0 in each of these distributions. Before making that conclusion though for this year's results, the overall posterior imputes around 99.4% probability to this “conditional” posterior on the increasing trendforms and the remaining 0.6% is imputed to the set of all innocuous trendforms (the posterior's probability for the third set, of all abruptly increasing—jumping—trendforms was nearly zero for boron at PW 735).

Posterior Distribution (blue) Re-weights the Equally-Likely Prior Distribution (red) on the Increasing Trendforms, in Accordance with their Compatibility with the Data (Each distribution sums to 1.0.)

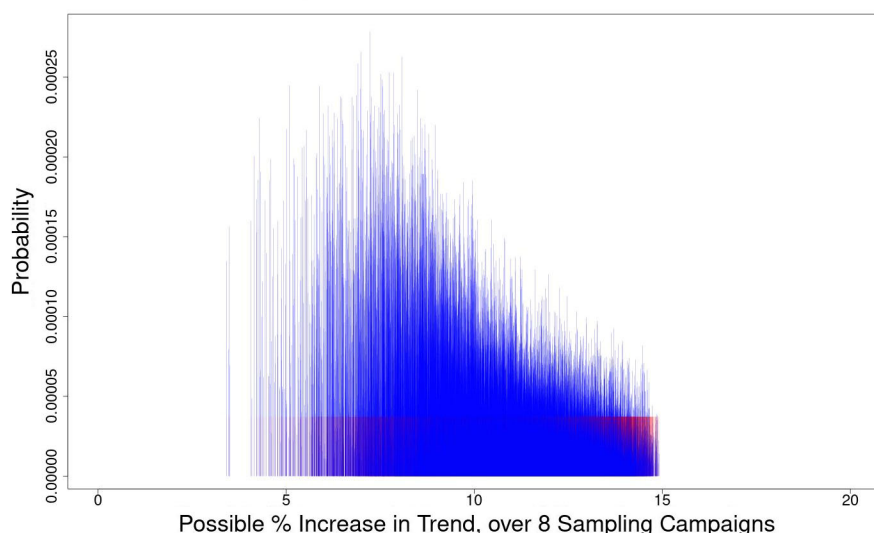


Figure G.6. Equally likely prior distribution and posterior (data-updated) distribution on approximately all possible increasing trendforms for boron at PW 735 (here individual trendform probabilities are ordered by their percentage increase over the 8 recent sampling campaigns).

Returning to the PLR, the observed PLR value (i.e., computed from the actual sample data, not a simulated value within the testing) has a simple interpretation in terms of the posterior. The observed PLR is equivalent to the probability, under the entire (including 0.6% probability for the innocuous set) posterior distribution, of the true trend's being increasing/high-valued relative to the probability that it is among the innocuous set: the observed PLR accumulates the posterior probabilities over both the increasing/high-valued and innocuous sets and is computed as their ratio. For boron at PW 735, noting from the preceding paragraph the 99.4%-0.6% probability split, the observed PLR (identical to this split) was approximately 1000-to-5.5, i.e., 182. As discussed earlier, although the PLR is an odds ratio (one could conclude based on this value alone) we only compute this value as the statistic for testing; because a large PLR value could possibly occur when the innocuous assumption applies, we do not conclude until we have done the actual (classical) testing described in the previous subsection, the PLR may not be inordinately large after all. So concluding based on merely comparing the observed PLR with an arbitrary, predetermined ratio would tend to either lead to excessively many false detects or too often failing to detect even when there actually is a contaminant plume at the site. For boron at PW 735, after testing—having repeatedly generated data series under the no-increase assumption and computed each resulting PLR value—we found that few of the generated PLR values actually exceeded the observed PLR value of 182; it was too rare to have occurred under that assumption. Therefore, as indicated in Figure G.10, boron at PW 735 was concluded to be increasing, and so we accept the posterior distribution

on only the increasing trendforms as the exclusive distribution, reassigning 0 probability to the innocuous group.

As just stated, after concluding an increase, we assume that the portion of the posterior relating to increasing trendforms accounts for all the possibilities; therefore its probabilities are rescaled to sum to 1. Cumulative probability functions (successively accumulated when going from left to right) are useful for assessing total less-than probabilities. But instead of using these accumulated probabilities (or percentiles as read from the horizontal axis) and assessing the totaled “less than” probability at every value of percentage increase, we subtract that progressive accumulation from 1.0 and plot the more useful *exceedance probability curve* shown in Figure G.7. ***From this plot, we can readily conclude that the maximum percentage increase in boron at PW 735 is approximately 15% (i.e., it is impossible [having zero probability] that the true increase is greater than 15%), and it is very likely to be much smaller than this value. Furthermore, in comparing this posterior with that obtained in Thompson et al. (2023) for this site (boron at PW 735), we find that the percentage increase is likely to be considerably smaller than it was, ending last year (i.e., the increasing trend is diminishing).***

Although the PLR test never provides an individual estimated trend, the posterior distribution completely characterizes the probabilities on all the candidate trendforms and so it may be used to obtain a “best estimate” of the true, underlying trendform. One optimal estimate is a probabilistically weighted average of all the trendforms – this is the probability-mean of them. Rather than obtaining an overall estimated trendform this way and then computing its percentage increase (of primary interest to this study) to approximate the average percentage increase among all increasing trendforms, we instead obtain the *exact* probability-average of the percentage increases, using their posterior (on these increases) already examined. Equivalently, we could accumulate the exceedance probability curve as it ranges from 0 to the maximum possible percentage (i.e., the mean of any positive-valued random variable is alternatively obtained by integrating the exceedance probability curve over its entire range). ***Based on the exact calculation just noted, the mean increase in boron (over the most recent 8 sampling campaigns) at PW 735 is 10.4%.***

G.3.4 Time-Trend Testing Results

The increasing trend concluded for boron at PW 735 was already addressed as part of our discussion of the PLR testing approach. Now we provide the complete set of test results. The results for the drainage- and site-level trend analyses are given in Figure G.8 through Figure G.10, where a plus (+) or minus (-) after the title or site, respectively, indicate that a strictly increasing or decreasing recent time trend was concluded. Figure G.9 combines the SUPSMU results with data plots for the sites in which the outlying data points that were removed from the analysis are identified with an “o,” as discussed in Section G.3. Figure G.10 overlays the smoothed site-level results within each drainage.

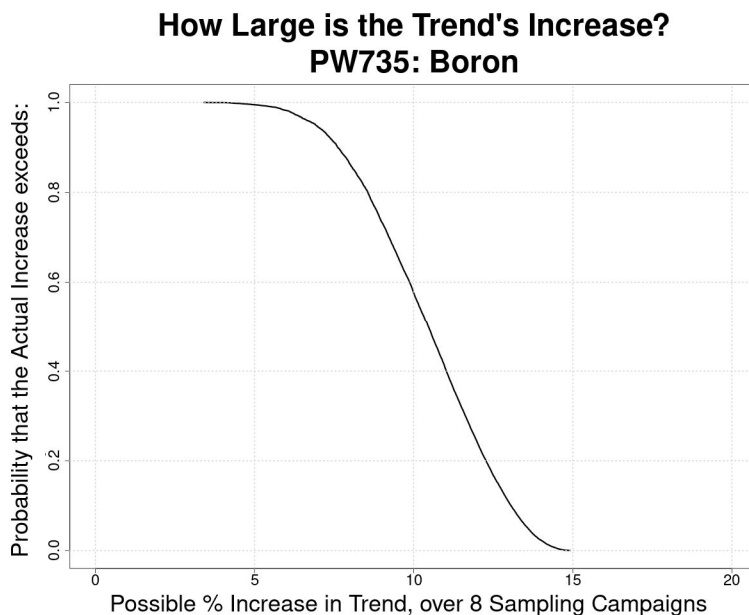


Figure G.7. Exceedance probability curve on percentage increase over recent 4 years, in boron trend at Cow Creek Site PW 735, as derived from the posterior distribution (which is depicted in blue in Figure G.6).

Site Trend Test Results. The site trend testing results are exhibited in the individual site plots of Figure G.9 and in the overlaid site plots of Figure G.10. The composite test was used to examine for increasing trends in conductivity and boron, whereas the conventional test was otherwise applied (i.e., for detection of decreases in all constituents and increases in sulfate).

An increasing trend in boron was found at private monitoring well PW 735, which taps a deeper aquifer than that supplying the alluvial wells. However, based upon the posterior distribution for boron at this site, the increasing trend is probabilistically smaller than it was in the previous two years, being consecutively smaller over the entire three years. Therefore the increasing boron trend at PW735 clearly appears to be tapering off.

At Pony Creek, increasing trends in conductivity were concluded for sites GSP 4 and GOW 4, and for boron ad increasing trend and a decreasing one were detected at sites GSP 6 and GOW 4, respectively.

No sulfate trends, either increasing or decreasing, were found at any site.

Drainage-Trend Test Results.

The results of the drainage-trend tests, for recent increases, and recent decreases are indicated in Figure G.8 and summarized as follows:

No drainage level trends were detected – neither in conductivity, boron or sulfate - at any of the drainages.

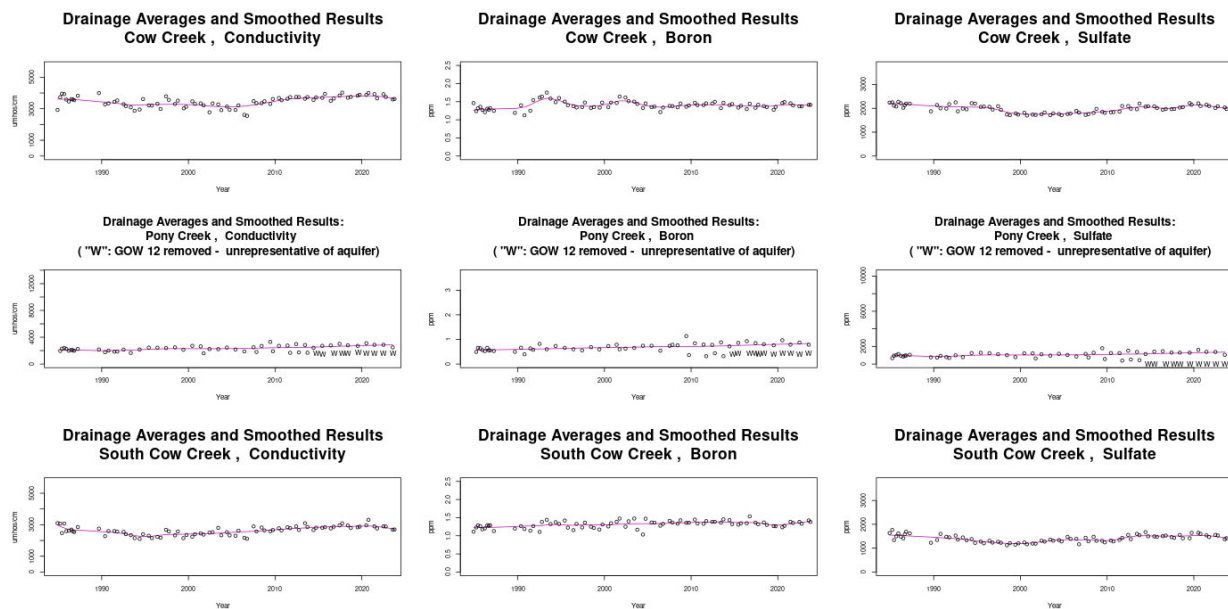


Figure G.8. Drainage-level trend analyses. In past reports, a “+” in the title denotes an increasing time trend detected over the most recent 4 or more years at the 5% level of decision error (among all non-Stipulation 12(d) trend tests, site level and drainage level), and conversely a “-” denotes a decrease over that same timeframe. No continual trends, either increasing or decreasing, were detected at any of the three drainages.

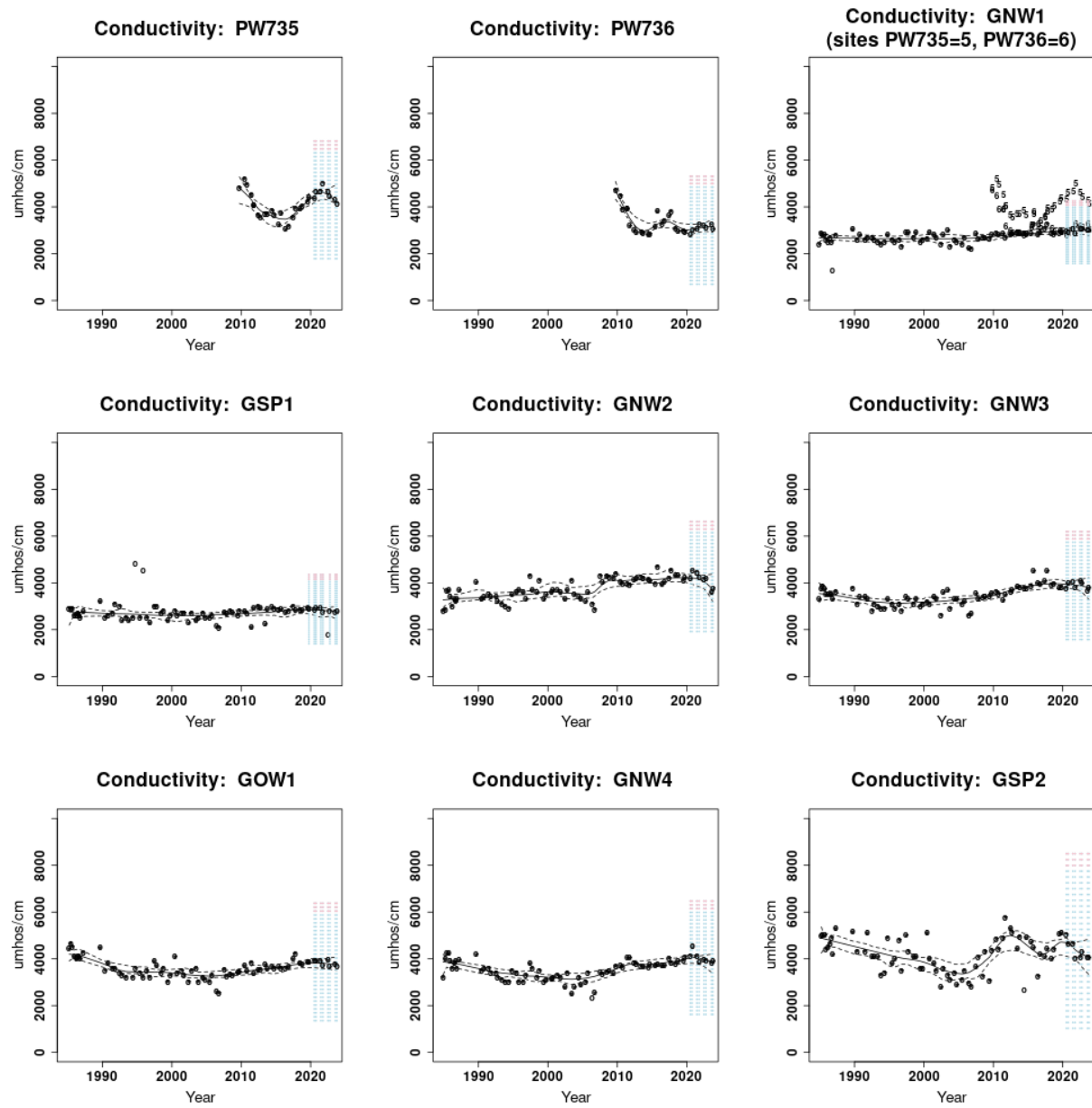


Figure G.9. Site-level trend analysis (1 of 8)

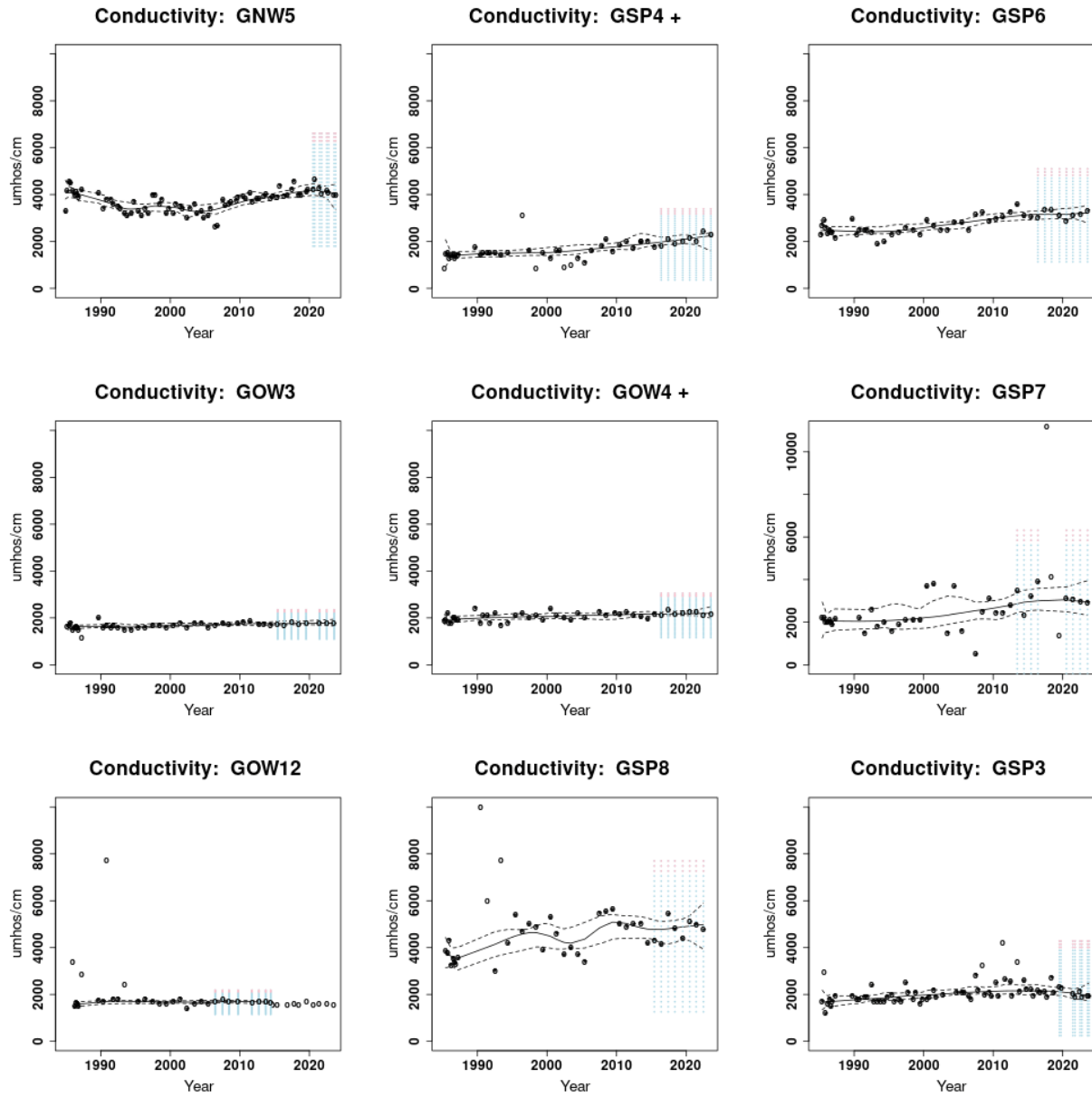


Figure G.9. (Continued, 2 of 8) A “+” denotes an increasing trend being found by the composite test at less than the 5% level of *overall decision error for all composite testing in the drainage*.

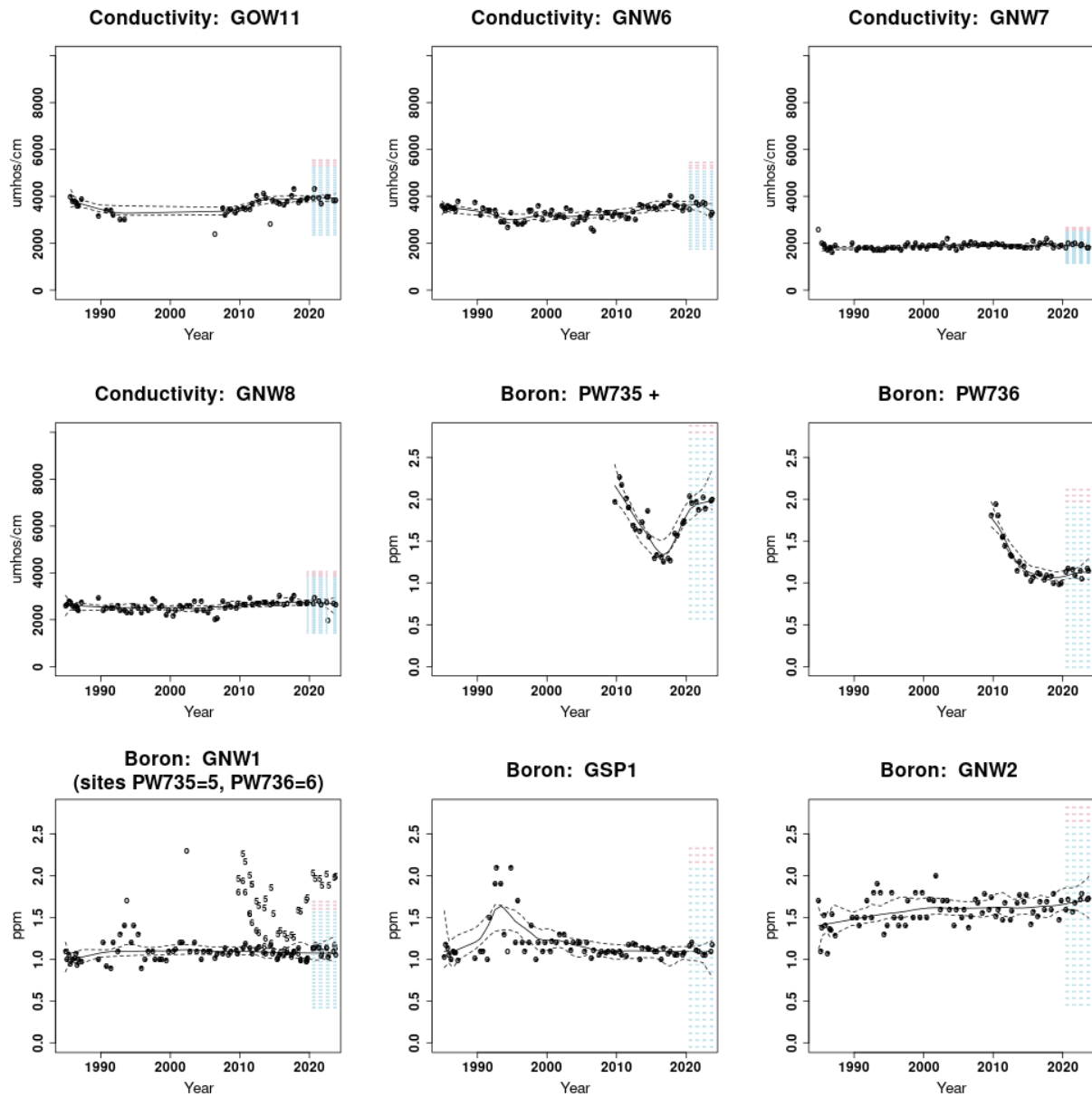


Figure G.9. (Continued, 3 of 8). A “+” denotes an increasing trend being found by the composite test at less than the 5% level of *overall decision error for all composite testing in the drainage*.

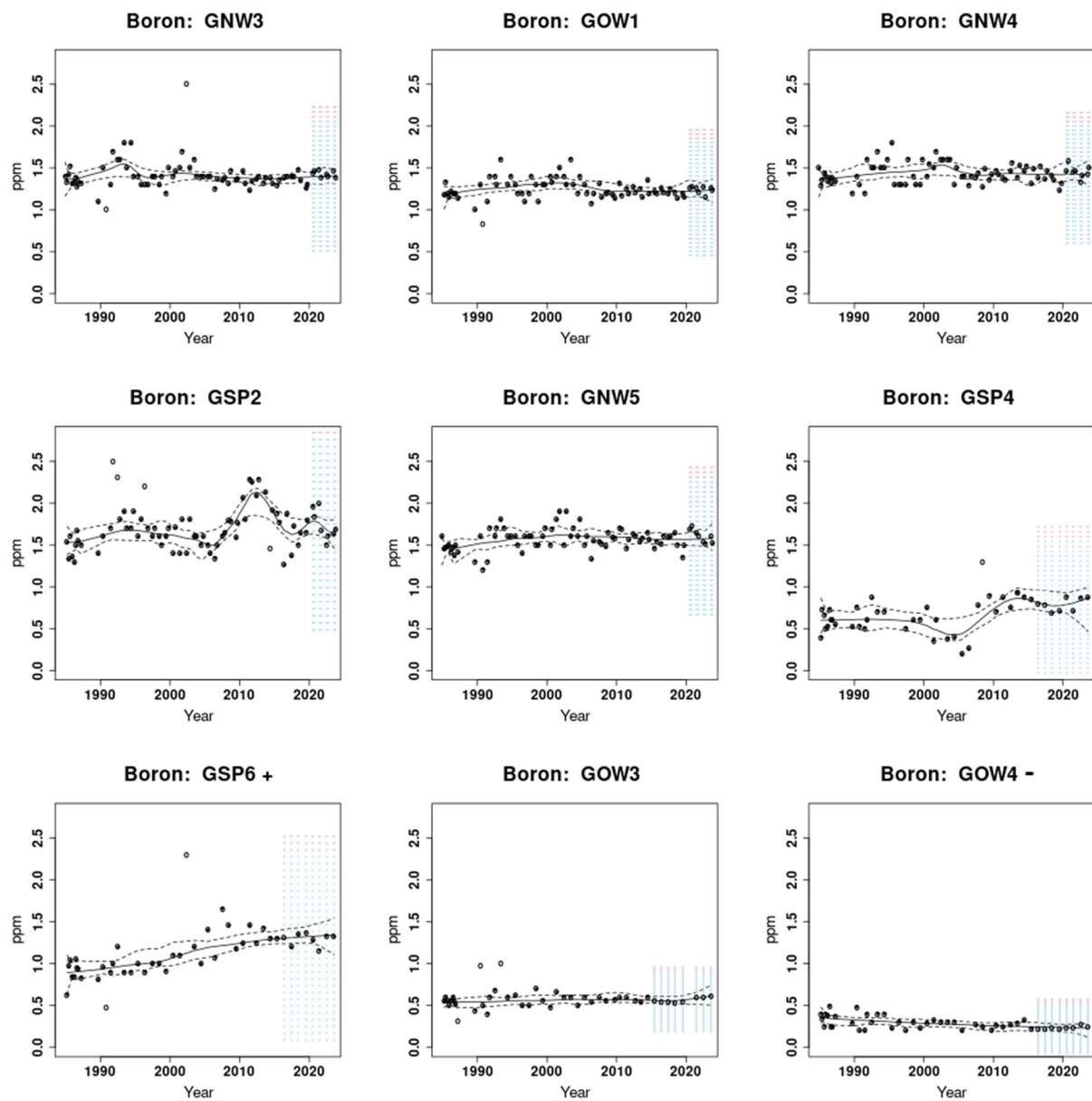


Figure G.9. (Continued, 4 of 8). A “+” denotes an increasing trend found by the composite test at less than the 5% overall decision error for all composite testing in the drainage. A “-” denotes a decreasing trend being found by the conventional test at less than a 5% overall decision error rate among all non-composite site tests and drainage tests (98 tests in all).

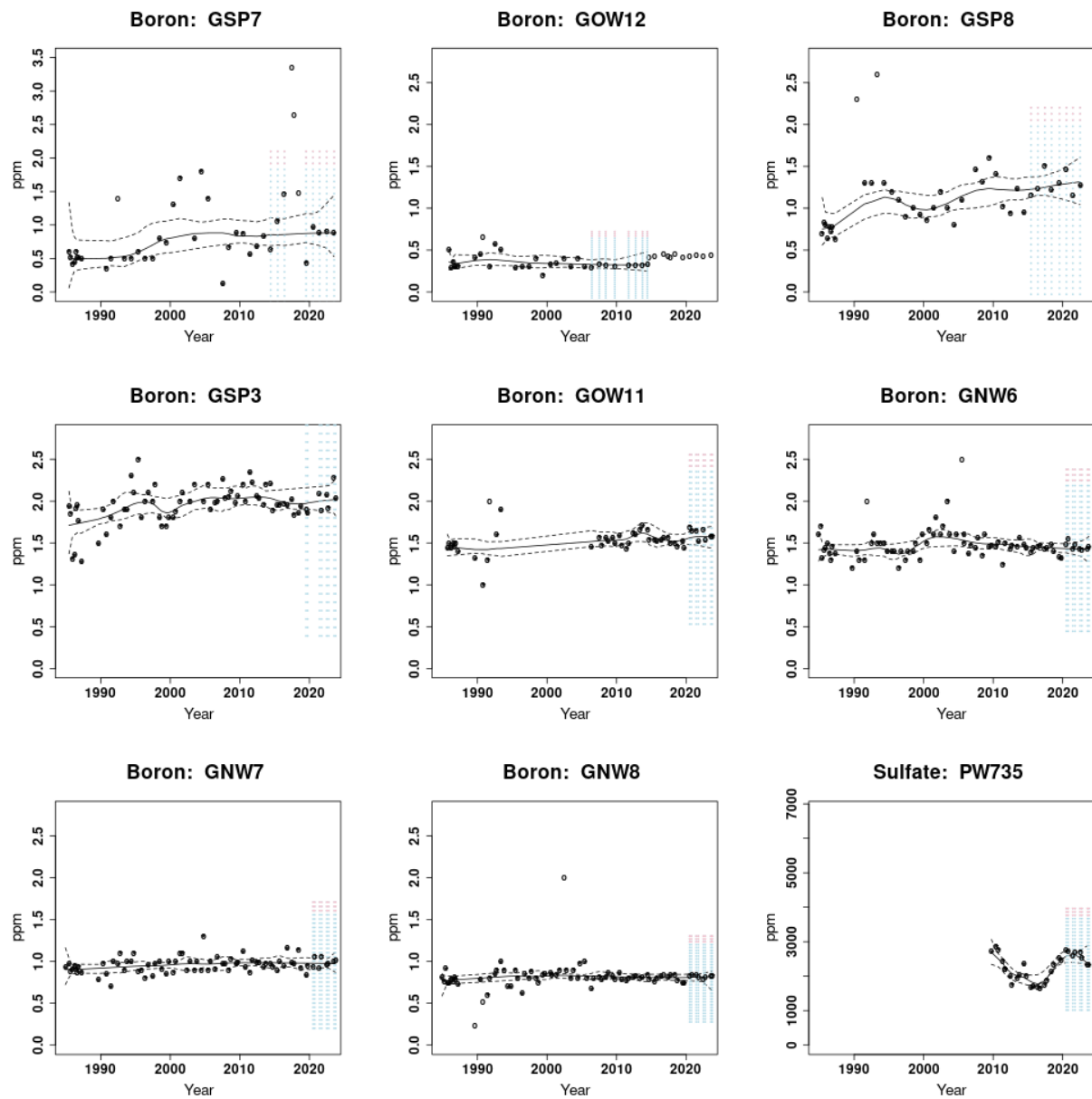


Figure G.9. (Continued, 5 of 8)

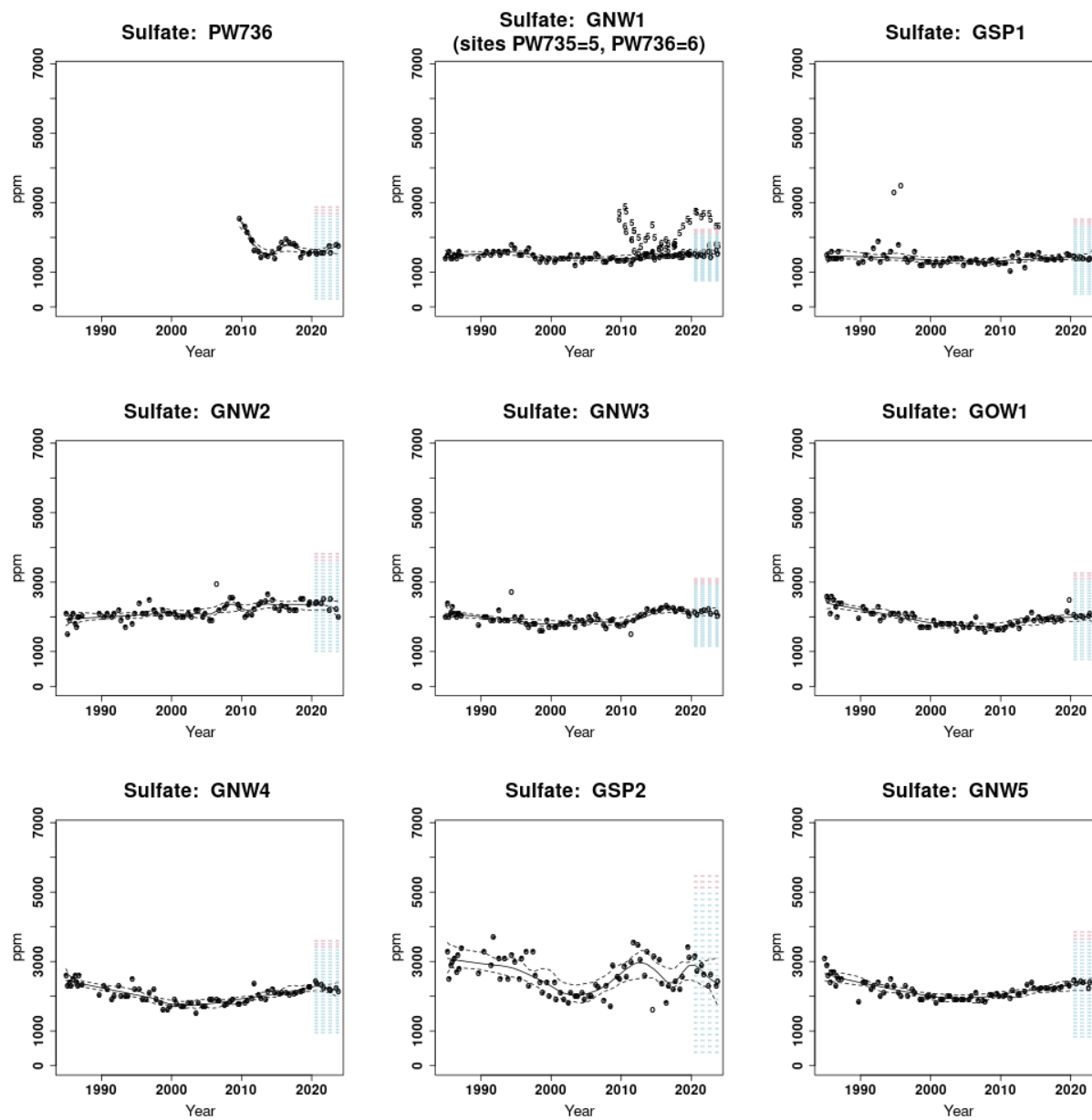


Figure G.9. (Continued, 6 of 8)

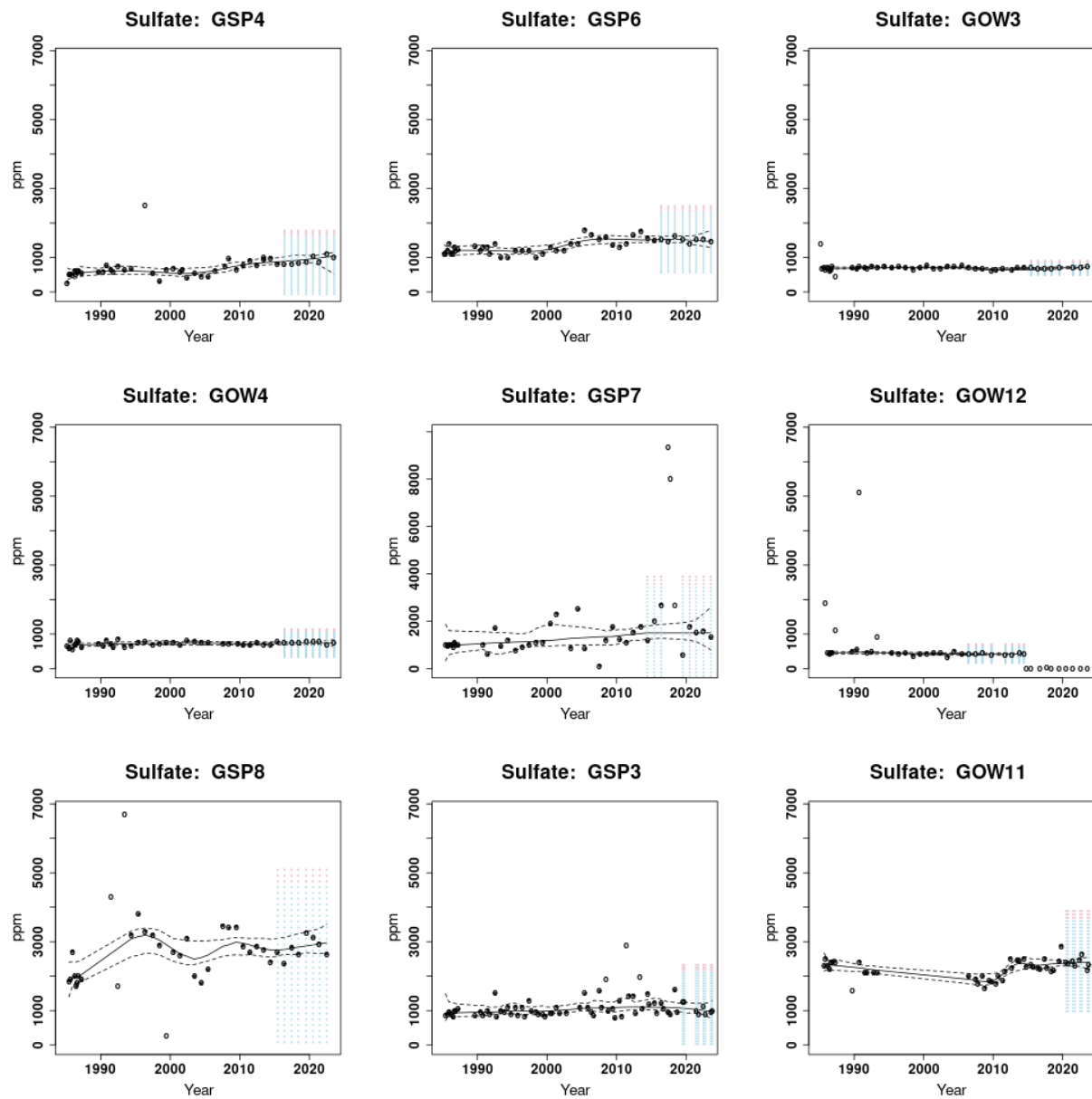


Figure G.9. (Continued, 7 of 8)

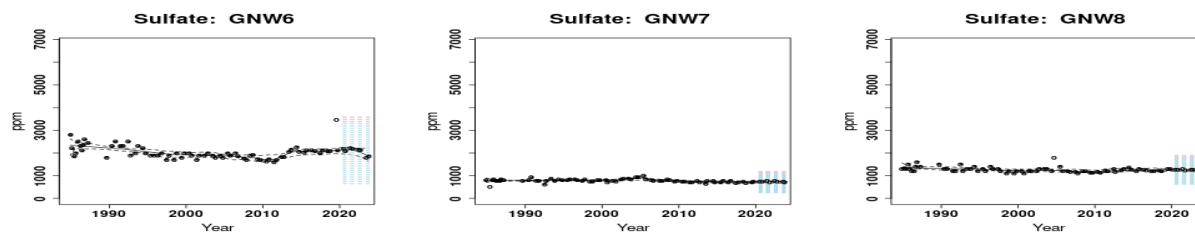


Figure G.9. (Continued, 8 of 8)

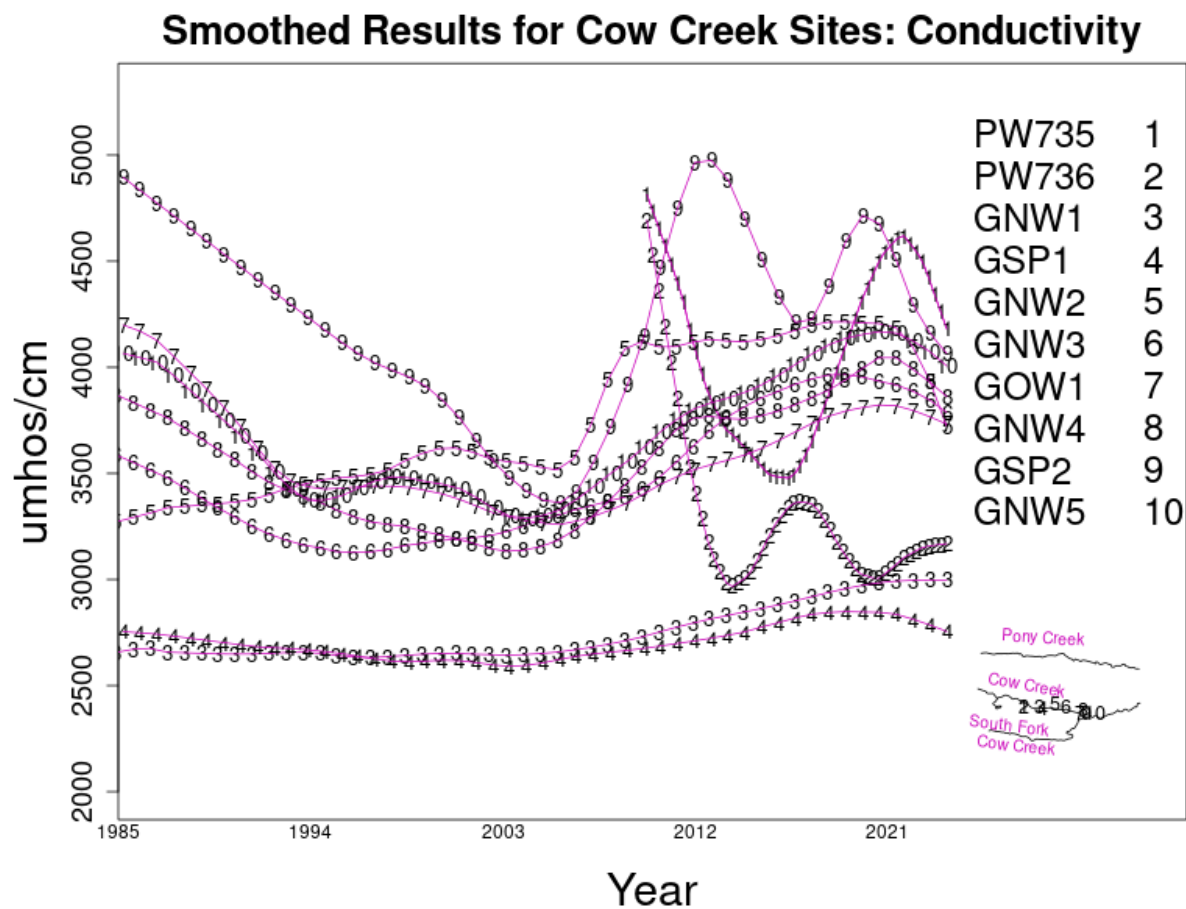


Figure G.10. Smoothed site-constituent levels over time (1 of 9)

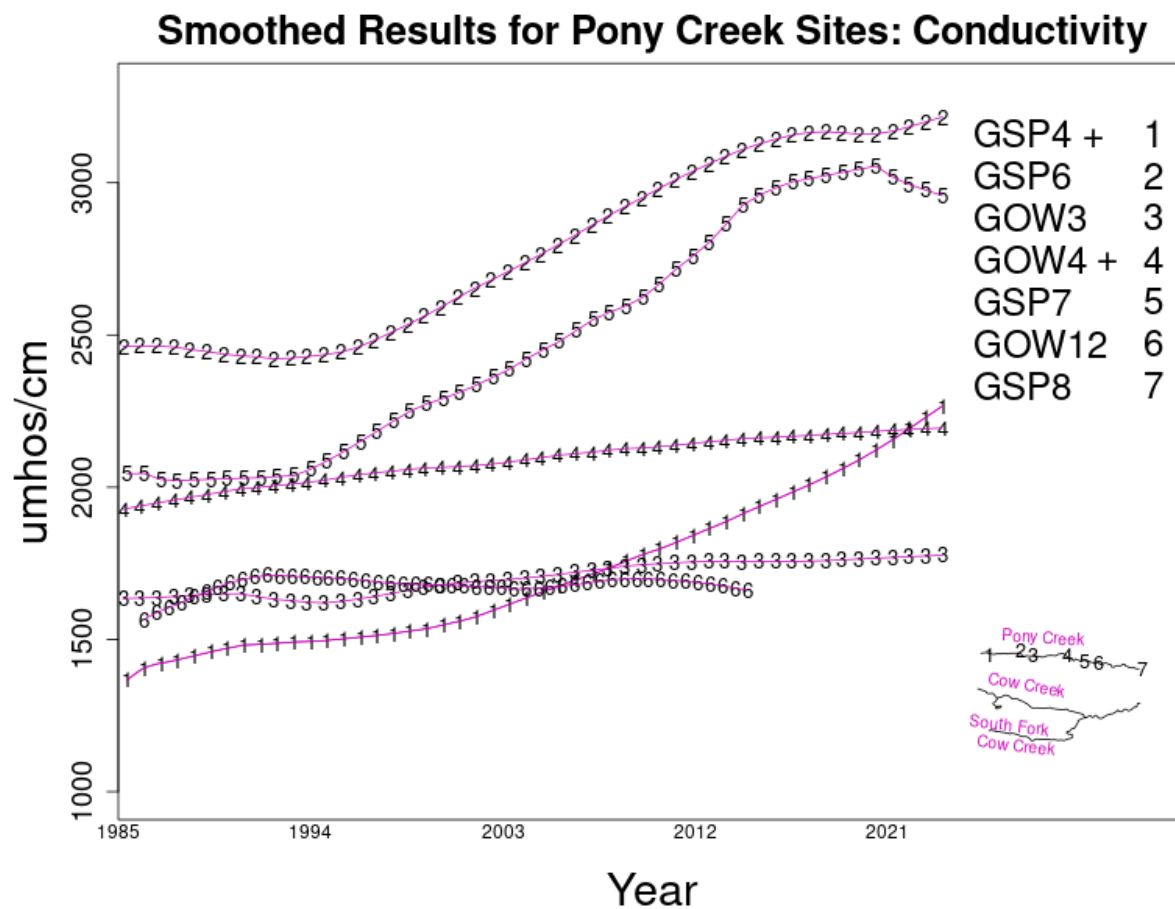


Figure G.10. (Continued, 2 of 9, trend testing at GOW12 is suspended pending further data.) A “+” denotes an increasing trend was found by the composite test at less than the 5% level of overall decision error for all composite testing in the drainage.

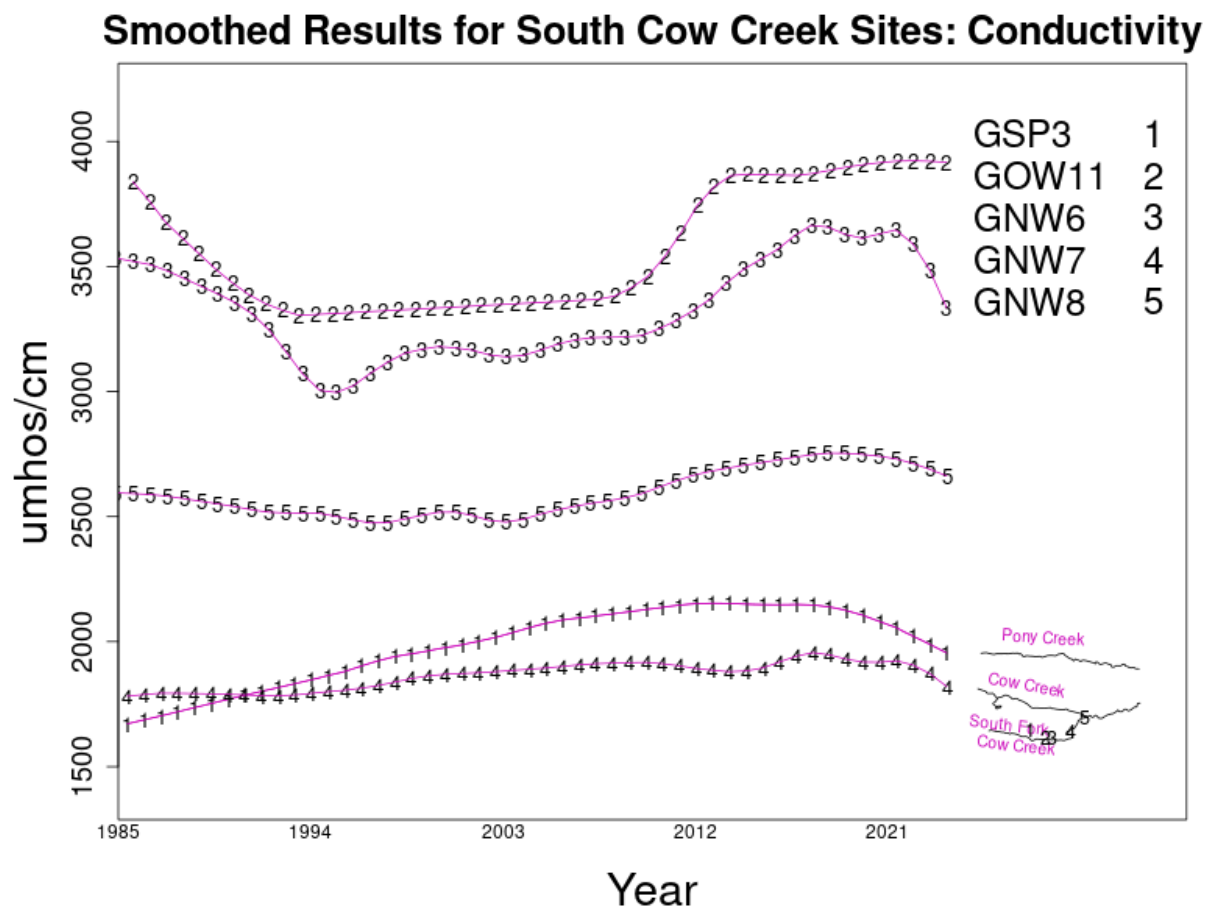


Figure G.10. (Continued, 3 of 9)

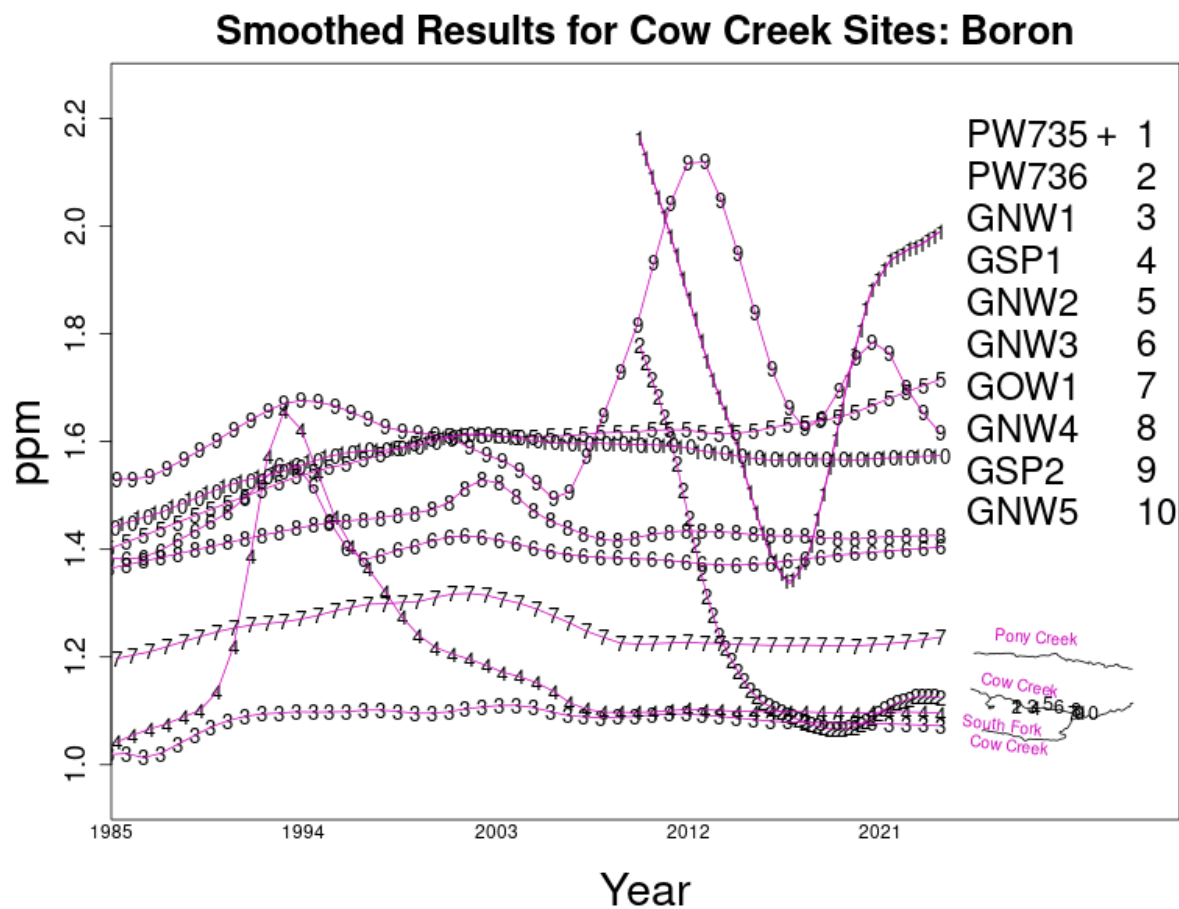


Figure G.10. (Continued, 4 of 9) A “+” denotes an increasing trend was found by the composite test at less than the 5% level of *overall decision error for all composite testing in the drainage*.

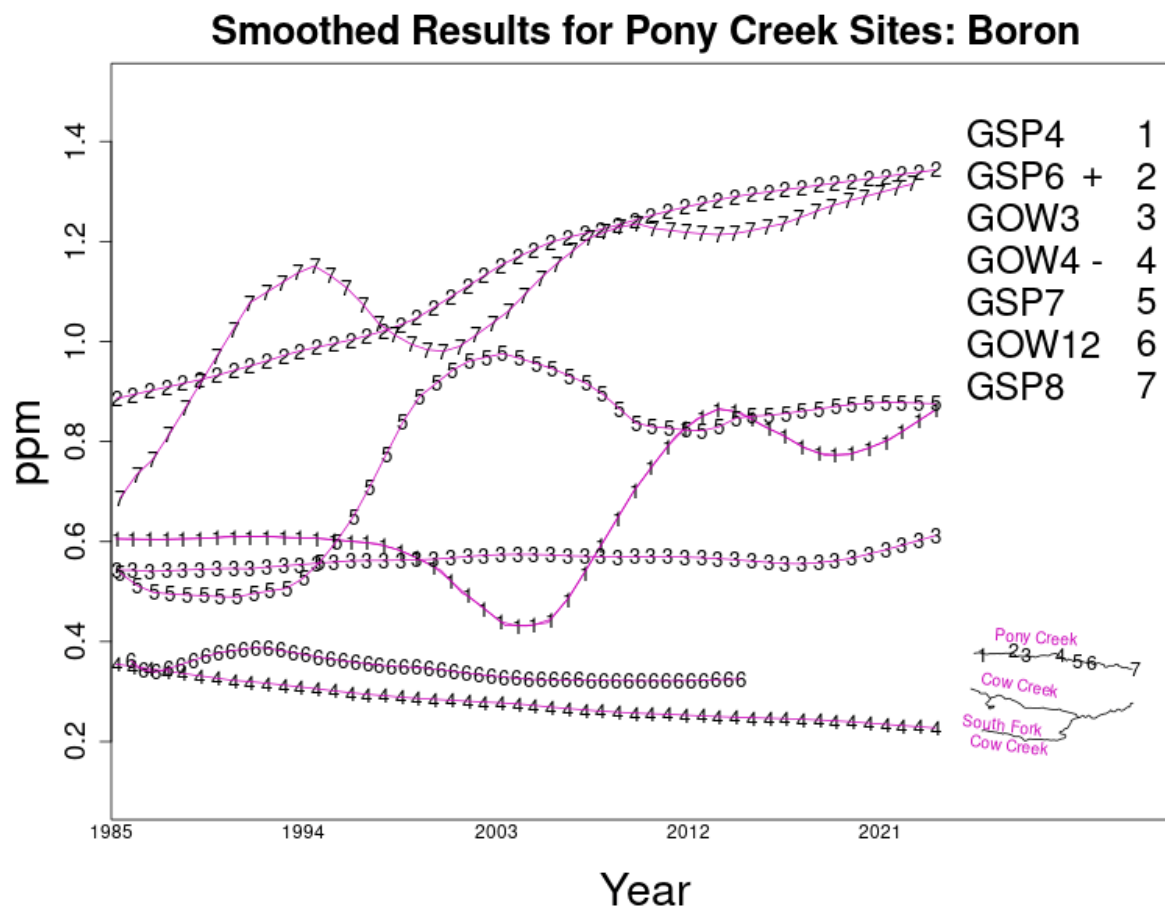


Figure G.10. (Continued, 5 of 9) A “+” denotes an increasing trend was found by the composite test at less than the 5% level of *overall decision error for all composite testing in the drainage*. A “-” denotes a decreasing trend being found by the conventional test at less than a 5% overall decision error rate among all non-composite site tests and drainage tests (98 tests in all).

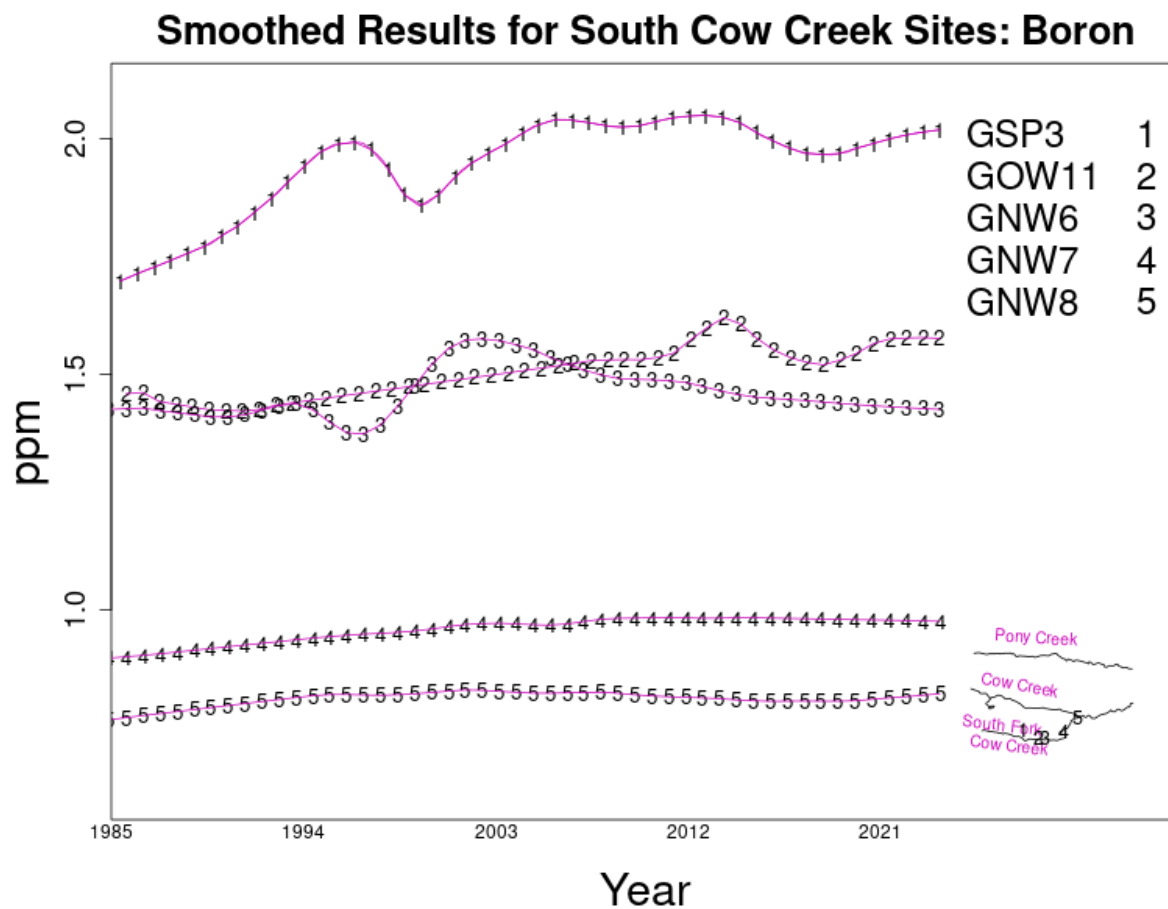


Figure G.10. (Continued, 6 of 9)

Smoothed Results for Cow Creek Sites: Sulfate

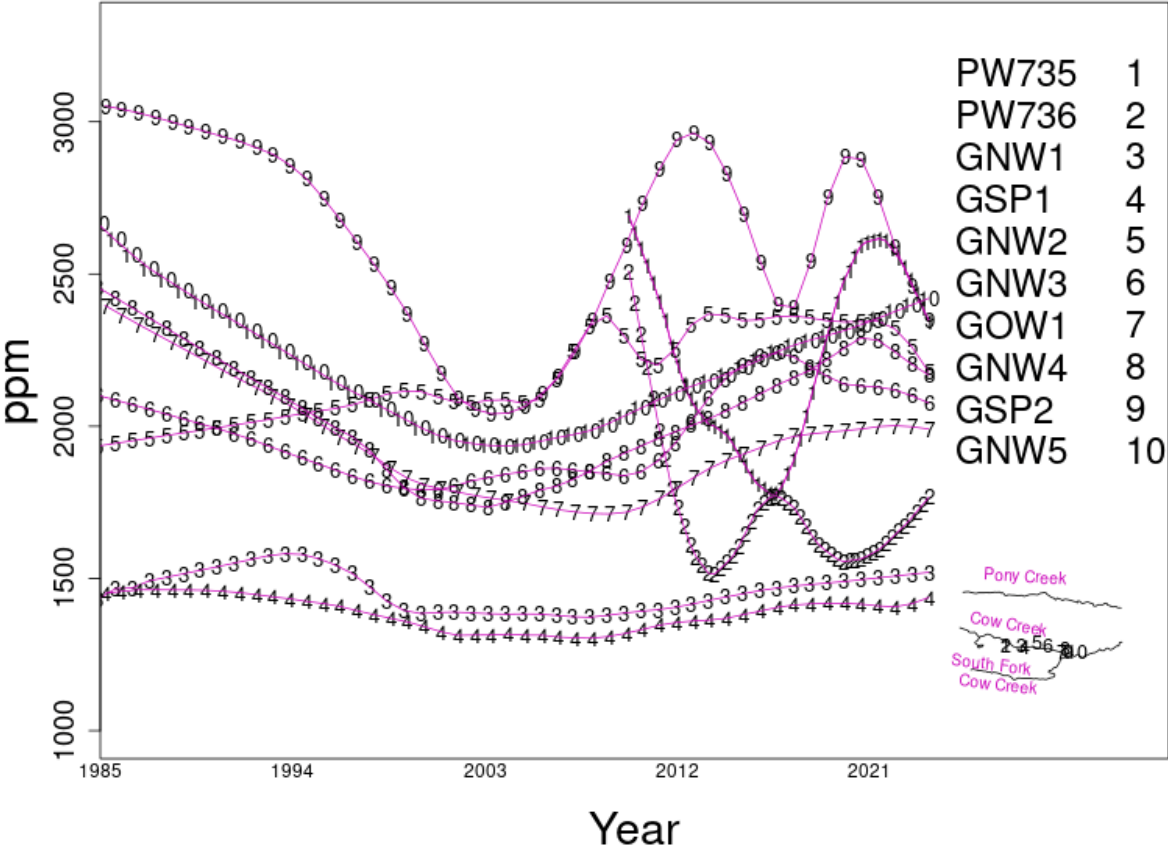


Figure G.10. (Continued, 7 of 9)

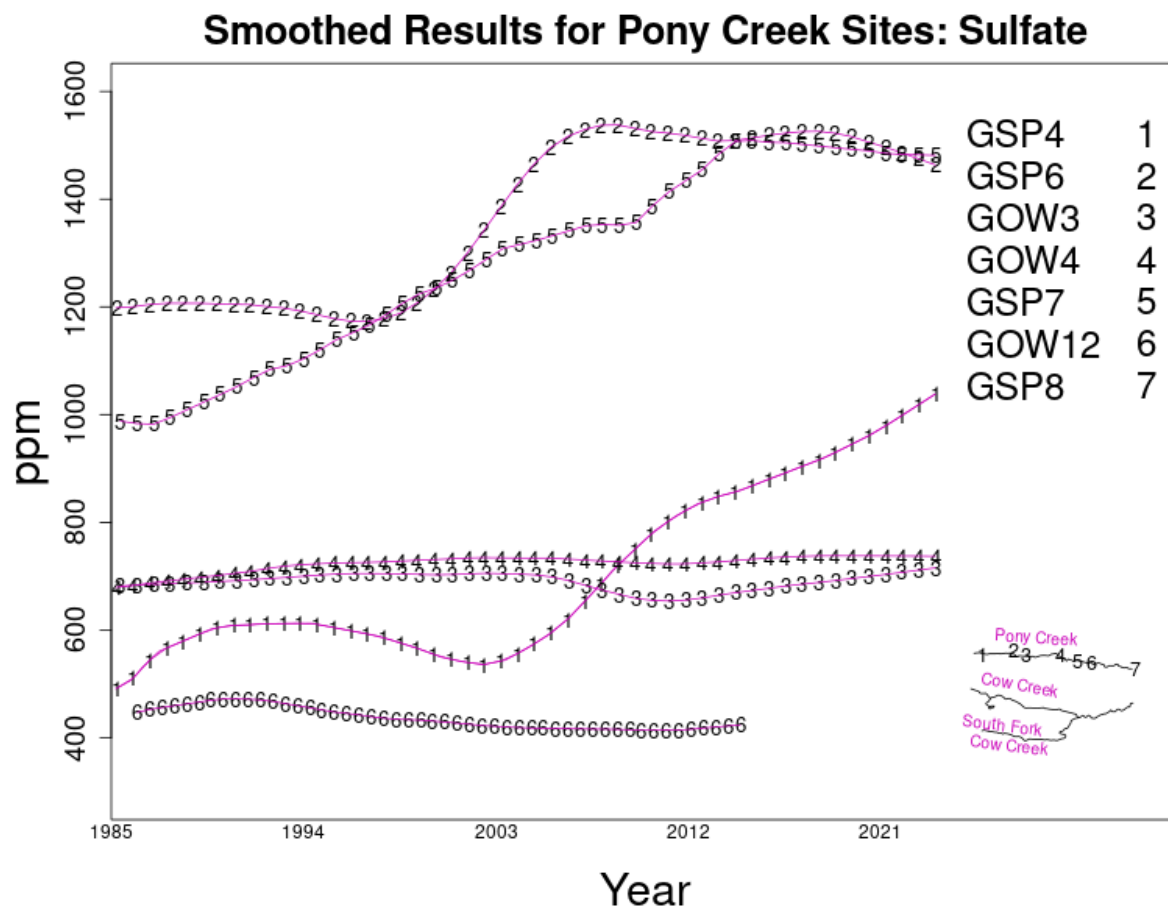


Figure G.10. (Continued, 8 of 9. Trend testing at GOW12 is suspended pending further data.)

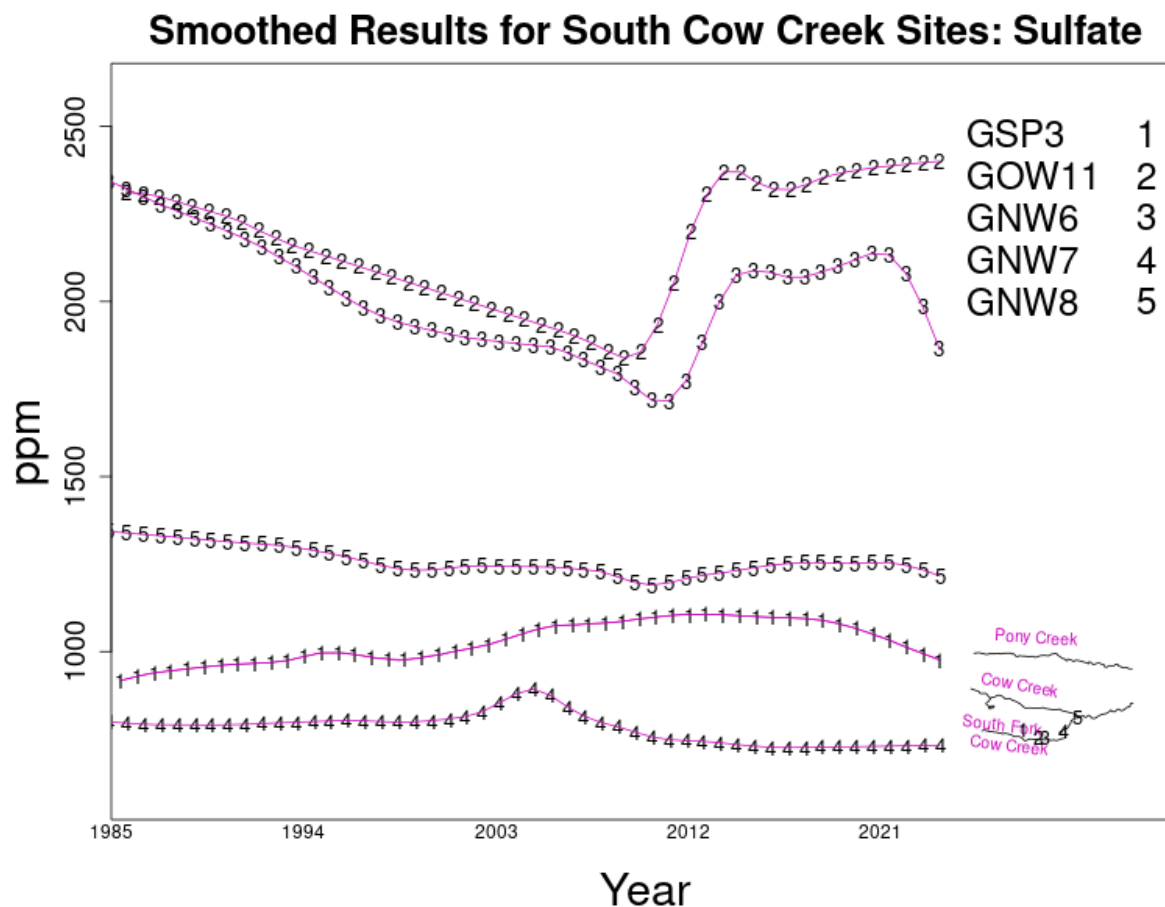


Figure G.10. (Continued, 9 of 9)

G.3.5 Overall Comparisons between Drainages Using Confidence Interval-Based Tests of Significance

To compare drainages, a graphical test was used based on the box plot (Tukey 1977) shown in Figure G.11. These plots summarize the distributional properties of the data and provide confidence intervals on the median (equaling the mean for symmetric distributions). The centerline within the box represents the sample median of the data, while the lower and upper ends denote the 25th and 75th percentiles of the data, respectively. The length of each whisker is either the distance to the next value from the box (lower or higher, related to the lower or higher whisker) or the interquartile range (the difference between the 75% percentile and the 25% percentile) multiplied by 1.5, whichever is smaller. Data points farther away from the end of the whisker are extreme and are plotted to bring attention to them. The “notch” in the box represents a confidence interval for the median (or mean, under our assumption) that is centered on the sample median. For small data sets, the notch width can be longer than the box; in this case, the box will have “ears.” The confidence interval is a range that is expected to contain the true mean with a specified level of assurance; for example, we are 95% certain that the true mean is contained in this interval. When comparing the underlying means of two sets of data, we initially assume that their means are equal, and we reject this assumption if the two confidence intervals do not overlap.

The drainage comparison tests were done as follows: the means of two or more drainages were judged to be identical (different) if their respective confidence intervals overlap (do not overlap).

These confidence intervals are approximately 95%, with a false-detect error rate of about 5% on each comparison (to reduce the overall error rate from all three pairwise comparisons, wider intervals could be used or other approaches may be adopted [Milliken and Johnson 1984]).

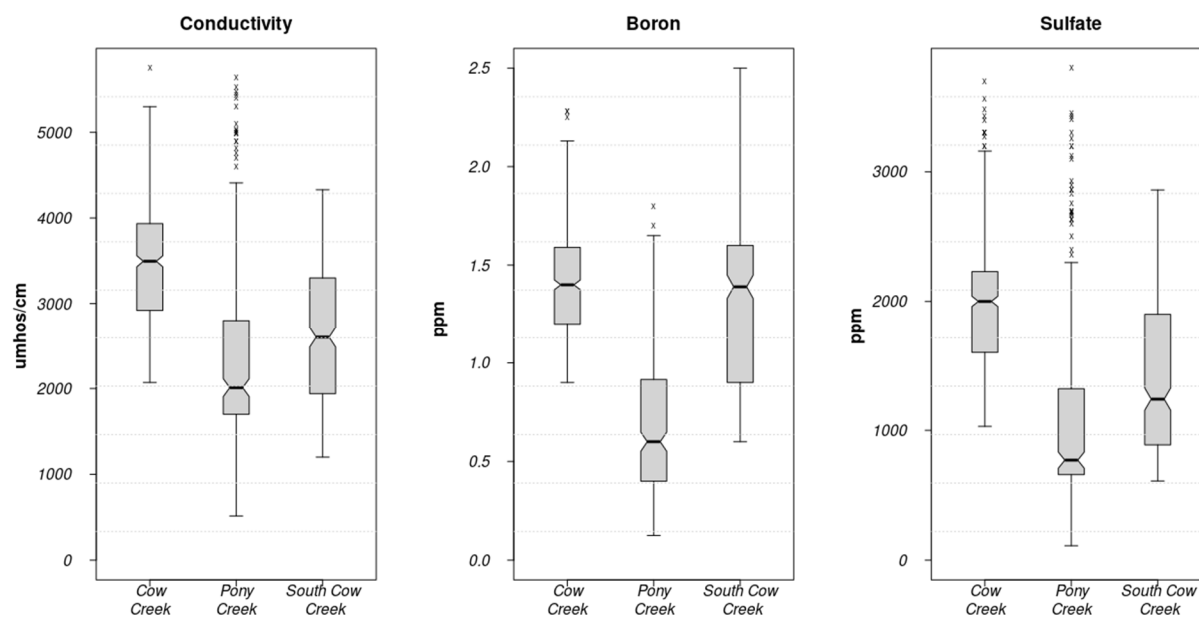


Figure G.11. Comparison of drainages using all data.

Results: Figure G.11 illustrates the box plot comparisons, based on data from all sampling years. The differences in mean constituent levels between the South Fork Cow Creek and Pony Creek drainages are conclusively nonzero (South Fork Cow Creek has higher means). The conductivity and sulfate means for Cow Creek are higher than the respective means at South Fork Cow Creek, but the mean boron levels are not statistically different in these two drainages.

G.3.6 The Spatial Waterfall Plot: An Analytical Tool for Revealing Groundwater Contamination's Nature, Extent, and Causes

This section contains a newly devised (or independently rediscovered) graphical method that will support the analysis of prospective contamination in the drainages, for verification whether there are apparent plumes at several sites in a drainage, if they are traveling down the drainage, or if they seem to be attributable to influences unrelated to holding pond activities such as land uses in the study area or weather patterns. This method is based on Spatial Waterfall plots which will be introduced next, first using exact data (no measurement error) from a hypothetical drainage similar to Cow Creek, and afterward applied to actual smoothed data, for each constituent in all three drainages.

For those unfamiliar with groundwater transport of contaminants, an introduction may be obtained from Anderson and Wang (1982), Freeze and Cherry (1979), or any introductory level text on groundwater. The subplots in figure G.12 exhibit contaminant plumes at each of 24 sites, spaced at 1,000 feet intervals along a hypothetical aquifer that is identical (hydrogeochemically) to the alluvial aquifer in the Cow Creek Drainage. These sites are increasingly distant, left-to-right then downward, from a contaminant source at the upper end of the aquifer, 1000 feet upgradient from the first site. These plumes were generated by evaluating the 1D groundwater contaminant concentration and movement solution of Agata and Banks (van Genuchten and Alves 1983, equation A.1) where in this example, initial concentration

was zero and the concentration associated with a constant in-flux (leak) of contaminated water into the upper boundary of the aquifer was set to 1.0 (any volumetric units, or it can be called a dimensionless concentration).

The results from using a depth-averaged 2D model, or evaluating a 2D or 3D model at a specific depth, and width location, with model flow only in the x direction would be similar to this model's solution. Superposition (adding together model solutions) of single-flux event results was used to evaluate the complete solution associated with multiple influx events, each occurring at a different time. The Darcy ("velocity" or specific discharge) of approximately 1 ft/day was obtained from this study for Cow Creek hydrology (discharge of 900 gallons/day [equaling 120 ft³/day] multiplied by groundwater gradient, equaling 0.008 ft/ft), which when divided by effective porosity of 0.2 (Newfields 2017) results in a pore-water velocity of approximately 5 ft/day. The hydrodynamic dispersivity (diffusion plus mechanical) was arbitrarily set to 800 ft²/day, and may actually be greater than that within the Cow Creek alluvial aquifer (and it varies slightly by contaminant). The pore water velocity is also the velocity of the contaminant or chemical property, because interaction between it and soil was assumed to be negligible. The time horizon in the plots is 2200 days.

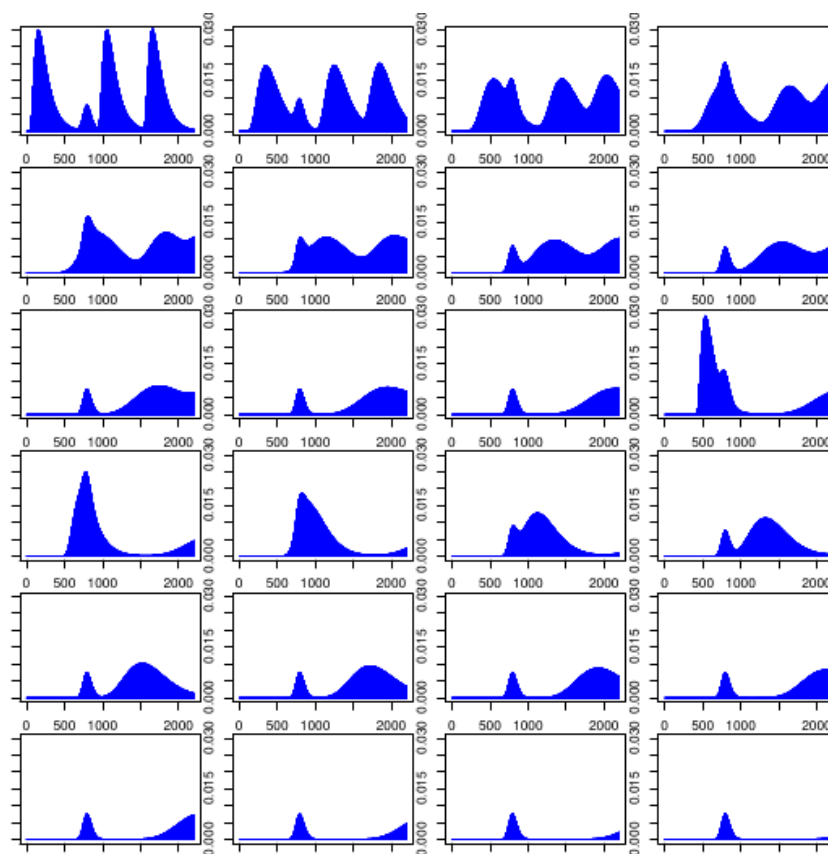


Figure G.12. Groundwater contaminant plumes at sites along a hypothetical drainage, hydrologically similar to Cow Creek (depicted plumes and this drainage are illustrative and hypothetical).

Although there is a time evolution of concentrations (per unit of effluent flux's concentration), at most sites the origins and timing plume incidences are not obvious and are even obscured; one must start with the first site's plumes and, recursively project what will occur later at the next site, 1000 feet farther down the drainage. This type of analysis requires considerable focus and thought. However, as will be

demonstrated next by examining the spatial waterfall plots among the sites, we can immediately identify and distinguish the different causes of the plumes, as well as closely estimate when they occurred.

In the spatial waterfall plot depicted in Figure G.13, the same 24 sites each have their own vertical "streak" over time (in similarity to spectral waterfalls used to assess multiple radio transmissions or other events received by an antenna, over an entire frequency spectrum): here the horizontal axis is in feet, and as indicated early the sites are separated in distance by 1000 feet, starting with the leftmost being at 1000 feet from the upper boundary of the aquifer. The intensities are *globally* normalized concentrations (brighter being higher), where the normalization is as follows: $\text{normalized value} = \{\text{concentration} - \text{minimum concentration}\} / \{\text{maximum concentration} - \text{minimum concentration}\}$, where minima and maxima are taken over all the constituent's values within the time-space domain. This normalization also removes any constant background level, so the waterfall plots can be said to include background values.

The plots represent a snapshot of a downward time-feed from the top, the most recent values at each site are at the upper edge of the plotted area, and those values at the bottom of the plot are from 2200 days past. From the plot, we first note three distinct plumes at the first site, one peaking at around 2000 days previous, the next two coming though maximally at around 900 days and 400 days. At the first site, each plume is tightly distributed. Then, after the first site, without any or much thought needed, we see that each plume is moving across the drainage with time and that (because of hydrodynamic dispersion) the plumes are occurring closer together, even overlapping. A central line drawn through each of these diagonal plumes has slope equaling $1/v$ (inverse of pore water velocity) or 0.2 days/ft: for unequally spaced sites (e.g., our monitoring sites) a straight line can still be drawn between the mass centers of the peaks—the time intervals between them would be proportional to their distances apart. Also, the accidental release duration for these plumes was 7 days.

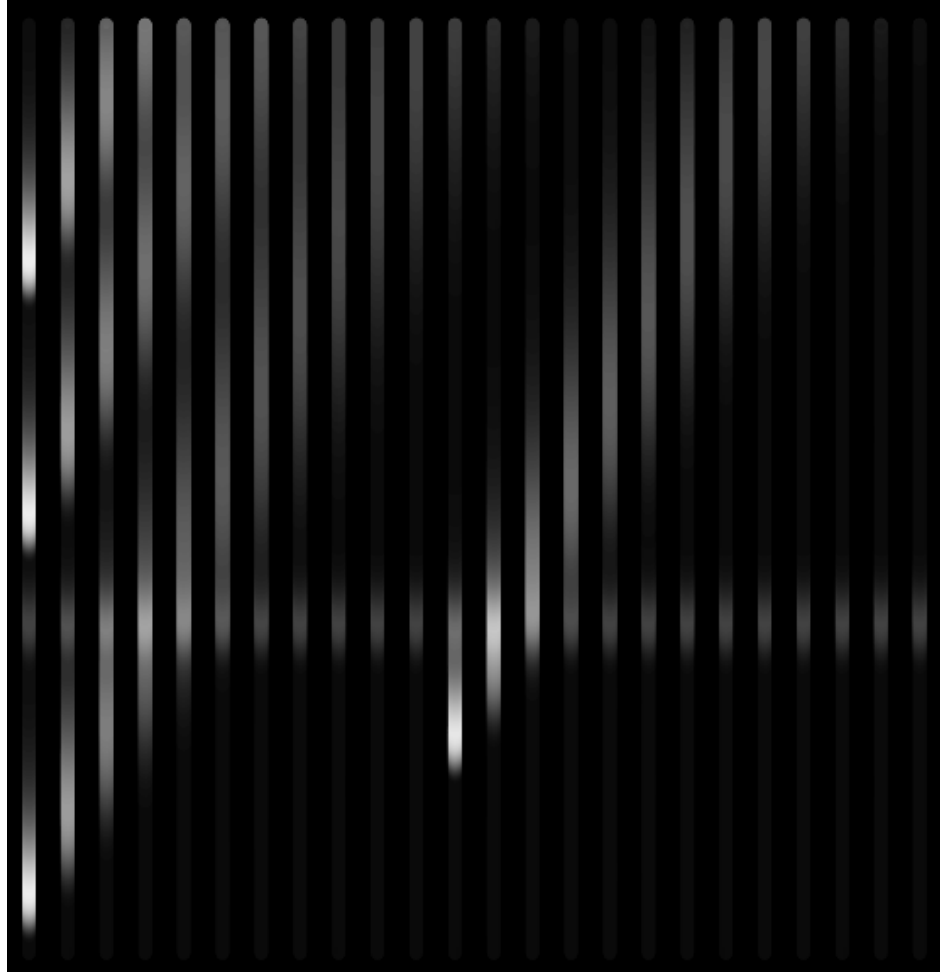


Figure G.13. Spatial waterfall plot (time vs. location) revealing modes, nature, and extent of contamination at 24 sites increasingly downgradient along a hypothetical drainage (vertical streaks exhibit temporal history of contamination [higher values are brighter] for each site, topmost ends are most recent occurrences). These plots are analogous to spectral waterfall plots that present a continuously updated feed from above the plot.

In this plot, we also see two other types of events. The first, depicted by the diagonal band starting in the right-half region three-quarters through the time history, is due to influx at site 12 (12,000 feet from hypothetical drainage's upper headwater boundary) from overland transport of contamination then infiltration into groundwater near that site only. This results in another contaminant plume, beginning at that site, which travels downgradient and disperses in the same way as do the principal plumes and its widening diagonal band reflects this. The remaining event is the horizontal band, reflecting the concentrating effects of a drought epoch, rather than a contaminant plume's occurring simultaneously at all sites. The "phantom" plume displayed was computed by arbitrarily augmenting the existing concentrations at each site over a short duration because, the one-dimensional model used for determining the other plumes, has no depth component. (Therefore, an increase, in depth-averaged concentration at each site, reflecting reduced aquifer depth due to evaporation and in the absence of offsetting groundwater recharge, could not be obtained from the model.) Visually speaking, if this horizontal band had, starting at each site, an ensuing diagonal band (similar the lower right band), collectively resulting in an overall wedge above the band, we would suspect that all sites were concurrently contaminated from overland transport. Hence the absence of any trailing plumes indicates that this horizontal band reflects a drought

epoch instead. In this example, the spatial waterfall enabled us to quickly identify globally transported plumes and to distinguish and explain all apparent plumes based on the signatures they produced in this plot.

To reiterate, spatial waterfall plots can be used to rapidly decipher a set of events that likely occurred and which are not easily identified by studying traditional temporal plots. These plots can also be used as evidence to refute the occurrence of a type of event (e.g., a horizontal band without trailing plumes or the absence of a plume trail starting at any site after the first one rules out overland transport).

The spatial waterfall plots representing the sample data are depicted in Figures G.14 through G.16; however, because the data contain error components due to sampling and analysis variations, the waterfall plots are of the smoothed data (smoothing based on a locally flexible algorithm similar to Supersmooth). It must be emphasized that smoothed values are estimated trends that may possess considerable uncertainty (and they may not be statistically different from actual underlying flat trends); although waterfall plots based on them assume that these estimates represent true underlying trends, this assumption is not necessarily correct. Therefore, waterfall plots provide supporting evidence only and are not solely reliable for rigorous conclusions. For making definitive statements about the spatial waterfall results, it is possible to also display, within this plotting framework, uncertainties of the concentrations, or uncertainties of the space-time extents for each of several fixed concentrations. Addressing these uncertainties is currently outside the scope of this study.

Also, in Figures G.14 through G.16, anticipated (time-space) directions are plotted as blue lines and are useful as guides that indicate the progression of a plume's mass-center downgradient, given that the contaminant velocity is 5 ft./day as calculated earlier (based on the studies referenced then); therefore, actual groundwater contamination's mass center should progress parallel with these lines. Likewise, interplume relief periods will also roughly align with these lines, until plumes downgradient disperse widely enough to overlap and diminish these relief "troughs." As in the previous example, aquifer contamination from contaminated overland runoff or surface water would be depicted as narrow, concentrated source events at the sites that are impacted, followed by widening plumes downgradient, each dispersing while running parallel to the red lines, similar to global contaminant progression due to contaminant influx (e.g., due to accidental release from the holding pond interception system) at the upper boundary of the aquifer. Finally, concentration due to drought, or dilution due to excessive and sustained precipitation, would respectively appear as plumes and antiplumes, concurrently arising at all or most sites, but these aberrations would resolve simultaneously without any trailing plume's being exhibited at downgradient sites later (i.e., parallel to the red lines). The plots should be examined for one or more of the various features just described.

Focusing first on conductivity, there are no clearly distinct "plume swathes" at any of the three drainages, neither starting at the uppermost sites nor possibly starting later as if due to overland flow's being a contaminant source. Although conductivity levels at downgradient sites GSP 2 (Cow Creek) and GSP 8 (Pony Creek) are among the highest in their drainages, there is no evidence for upgradient causes of those levels—single or multiple plumes—and (as reported in Thompson 2022) there is no trailing plume downgradient of GSP 2. Looking carefully and squarely at the plot, two diluting periods are evident as exhibited by dark bands at most sites within Cow and Pony Creek, prior to 1995 and after 2005 (attenuation between plumes would not be simultaneous because plumes—one or more—would offset in time); curiously though, no simultaneous attenuations occurred at South Fork Cow Creek, and no attenuations in boron or sulfate occurred during those epochs. As a reminder, Rosebud mining activities, either mining or reclamation, were underway in and around those time periods at both Area D and Area E, possibly (beneficially) affecting Pony Creek and Cow Creek alluvia, respectively, then for example via moderated levels of TDS influx. In the earlier 1990s, and possibly also related to the attenuation near the mid 1990s, were increasing trends, detected in this study, at several Cow Creek sites (e.g., Olsen et al.

1992), and many of those sites exhibited an apparent decline in trends afterward and prior to 1995. *Therefore, the simultaneous attenuation bands (prior to 1995 and after 2005) seem to be mining related; had they been precipitation-driven, the bands would have also appeared across many South Fork Cow Creek sites. Based on these plots, there is new evidence that the elevated levels detected at many sites in Cow Creek in the early 1990s were not due to contaminated runoff, because, the attenuation, discussed here, occurred instead of there being trailing plumes downgradient from each affected site.*

Since the second attenuation, starting around 2008 according to the plots (recall though that they are actually estimates with possibly excessive uncertainty that precludes making reliable statements based on them) there is a curious steady-state-like declining pattern (going downgradient) in all three parameters, occurring among Cow Creek sites GNW 2, GNW 3, GOW 1, and GNW 1. Tentatively, no other noteworthy inter-site patterns appear in these waterfall plots, especially not over recent years.

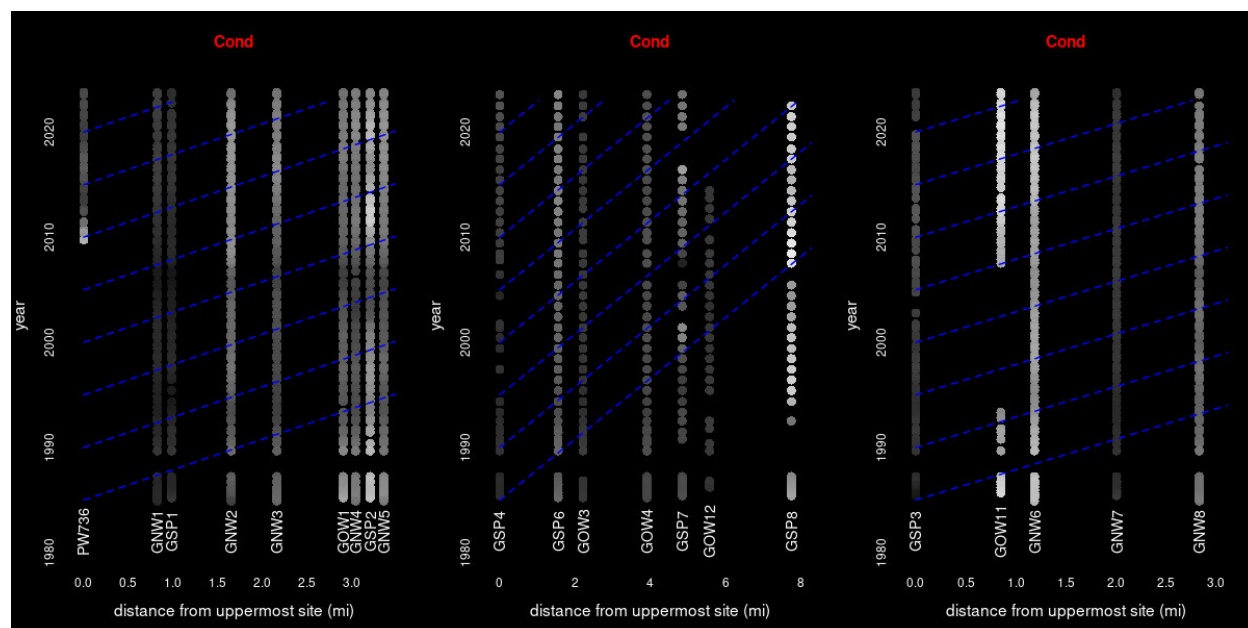


Figure G.14. Spatial waterfall plots of smoothed conductivity histories, at all sites examined in the statistical analysis.

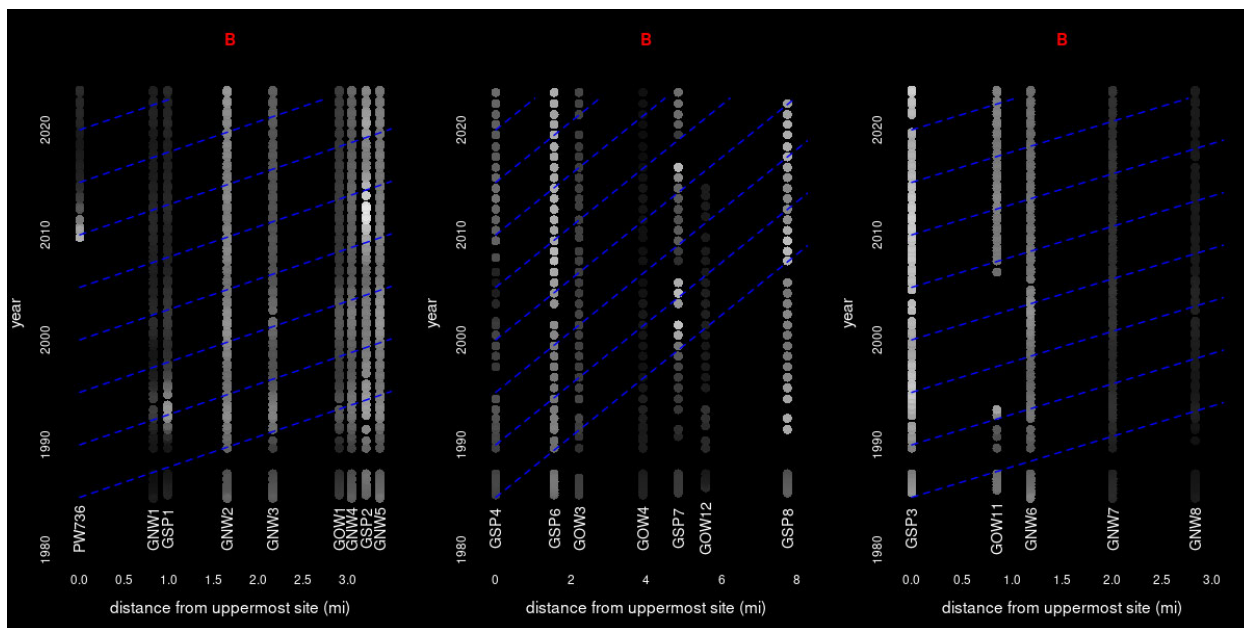


Figure G.15. Spatial waterfall plots of smoothed boron histories, at all sites examined in the statistical analysis.

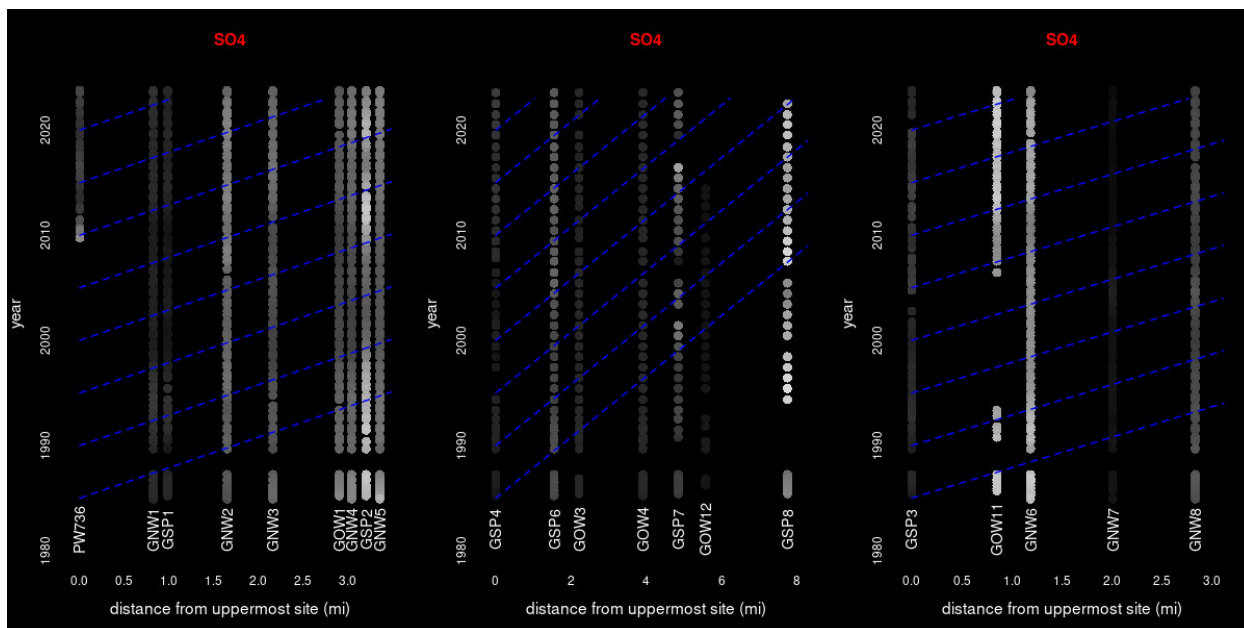


Figure G.16. Spatial waterfall plots of smoothed sulfate histories, at all sites examined in the statistical analysis.

G.3.7 Summary of the Statistical Analysis Testing

The *drainage time-trend test results* are:

- No trends were found - in conductivity, boron, or sulfate – at any of the three drainages.

The *site-level time-trend test results* are as follows:

- As in the two previous years' studies, an increasing trend in boron was found at Cow Creek site PW 735 which is clearly tapering off.
- No abrupt increases were detected, in either conductivity or boron, at any of the Cow Creek sites mentioned in Stipulation 12(d).
- Increasing conductivity trends were found at Pony Creek sites GSP 4 and GOW 4.
- Regarding boron, an increasing trend and a decreasing one were respectively found at Pony Creek sites GSP 6 and GOW 4.
- No trends in sulfate were found at any site, either decreasing or increasing.
- The Spatial Waterfall plots do not suggest any recent trends, but they appear to provide new evidence that in the mid 1990s (and after 2005) mining activity has affected many sites at Cow and Pony Creek drainages rather than those sites being influenced by contaminated overland runoff or weather patterns.
- The Spatial Waterfall plot is an extremely useful tool which will assist us in the future in our interpretation and inference, and it will continually reveal new insights as well.

Drainage comparisons indicated that the Pony Creek drainage had the lowest values for conductivity, sulfate, and boron. The Cow Creek drainage was distinctly highest in conductivity and sulfate, and the South Fork Cow Creek drainage means fell between the values of the other drainages except for boron, whose mean did not differ from that of Cow Creek.

G.4 References

- Anderson, M. and H. Wang. 1982. *Introduction to Groundwater Modeling*. San Francisco: Freeman and Co.
- Chamberlain, P. J. 2018. "The Bayes Factor Reinterpreted and Renamed as the Predictive Likelihoods Ratio - It's a UMP Composite vs. Composite Test on Multiple Parameters." Technical Bulletin No. 3, Enviro-Sci Consulting.
- Cleveland WS. 1981. "LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression." *American Statistician* 35:54.
- Freeze, R.A. and Cherry, J.A. 1979. *Groundwater* Vol. 7632, 604. Englewood Cliffs: Prentice-Hall Inc.
- Friedman JH. 1984. A variable span scatterplot smoother. Laboratory for Computational Statistics, Technical Report No. 5, Stanford University, Palo Alto, California.
- Hampel FR, EM Ronchetti, PJ Rousseau, and WA Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley and Sons.
- McClave JT and TT Sincich. 2016. *A First Course in Statistics, 12th ed*. Pearson, New York, New York.
- McDonald JP, KB Olsen, JL Smoot, PJ Chamberlain, SR Jennings, and JD Goering. 2007. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2006*. Battelle – Pacific Northwest Division, Richland, Washington.
- Milliken GA and DE Johnson. 1984. *Analysis of Messy Data, Volume 1: Designed Experiments*. Lifetime Learning Publications, Belmont, California.
- Olsen KB, JL Smoot, and PJ Chamberlain. 1992. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1991*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1993. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1992*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1994. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1993*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1995. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1994*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1996. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1995*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1997. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1996*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1998. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1997*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 1999. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1998*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2000. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 1999*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2001. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2000*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2002. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2001*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2003. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2002*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2004. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2003*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2005. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2004*. Battelle – Pacific Northwest Division, Richland, Washington.

Olsen KB, JL Smoot, and PJ Chamberlain. 2006. *Hydrogeochemical Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for CY 2005*. Battelle – Pacific Northwest Division, Richland, Washington.

R Development Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.

Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, New York.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2008. *Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana, for 2007*. Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2009. *2008 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4063, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2010. *2009 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4182, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2011. *2010 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4261, Battelle – Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and PS Blicher. 2012. *2011 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNWD-4352, Battelle–Pacific Northwest Division, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2013. *2012 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-22498, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2014. *2013 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10010, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2015. *2014 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10027, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2016. *2015 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10043, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, KB Olsen, PD Thorne, PJ Chamberlain, SR Jennings, and LMB Franklin. 2017. *2016 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10056, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, PJ Chamberlain, and C Lucy. 2019. *2018 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10080, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, PJ Chamberlain, and E Koehler. 2020. *2019 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10098, Pacific Northwest National Laboratory, Richland, Washington.

Thompson CJ, E Koehler, M Larson, PJ Chamberlain,. 2021. *2021 Hydrologic/Water Quality Study of Cow and Pony Creek Drainages, Rosebud County, Montana*. PNNL-ACT-10108, Pacific Northwest National Laboratory, Richland, Washington.

Tukey JW. 1977. *Exploratory Data Analysis*. Boston: Addison-Wesley.

van Genuchten, T.H and W.J. Alves. 1983. Analytical Solutions of the One-Dimensional Convective-Dispersive Solute Transport Equation. US Dept. of Agriculture, Technical Bulletin No. 1661.

Distribution

No. of Copies

OFF-SITE (6)

Mr. Gordon Criswell
Talen Montana, LLC
P.O. Box 33
Colstrip, MT 59323

Mr. David Davenport
63 Diamond Ranch Lane
Forsyth, MT 59327

Mr. Karson Kluver
Genie Land Company
3034 Rosebud Creek Road
Forsyth, MT 59327

Mr. Eric Koehler
KC Harvey Environmental, LLC
376 Gallatin Park Drive
Bozeman, MT 59715

Ms. Barbara Needham
Genie Land Company
3088 Rosebud Creek Road
Forsyth, MT 59327

Ms. Kordelle Stephenson
Talen Montana, LLC
P.O. Box 33
Colstrip, MT 59323

No. of Copies

ON-SITE (1)

Christopher Thompson
Pacific Northwest National Laboratory
MS P7-54

ELECTRONIC (7)

Dr. Peter Chamberlain
Enviro-Sci Consulting
envirosci@enviro-sci.com

Ms. Maggie Gordon
Northern Plains Resource Council
maggie@northernplains.org

Mr. Ed Joiner
Rosebud County Commissioners
ejoiner@rosebudcountymt.com

Mr. Svein Newman
Northern Plain Resource Council
svein@northernplains.org

Ms. Sarah Seitz
Montana Department of Environmental Quality
sarah.seitz@mt.gov

Ms. Sandy Moisey Scherer
Montana Board of Environmental Review
sandy.moiseyscherer@mt.gov

Mr. Rick Thompson
Montana Department of Environmental Quality
rithompson@mt.gov

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov