



United States Department of the Interior

U.S. GEOLOGICAL SURVEY
National Water Quality Laboratory
P.O. Box 25585
Denver, CO 80225-0585

NATIONAL WATER QUALITY LABORATORY TECHNICAL MEMORANDUM 2019.01

February 19, 2019

Subject: Characterization of bias and variability in dissolved trace elements determined by inductively coupled plasma–mass spectrometry (ICP-MS) attributable to a deviation in the acid concentration of calibration standards during a period starting no earlier than October 2005 and as late as 2008, and ending August 8, 2015

Effective: August 8, 2015

Authors: Sarah J. Stetson, Method Research and Development Program, NWQL, sstetson@usgs.gov, 303-236-3277
Daniel Mendez, Analytical Services, NWQL, dmendez@usgs.gov, 303-236-3289
Carl Harris, Analytical Services, NWQL, Retired
Christopher Kanagy, Method Research and Development Program, NWQL, ckanagy@usgs.gov, 303-236-3262

Revision: Not applicable

1 PURPOSE

This technical memorandum alerts data users to potential shifts in measured concentration and variability due to the correction of a deviation in the standard operating procedure (SOP) used in the determination of trace elements in filtered acidified water (FA) by inductively coupled–plasma mass spectrometry (ICP-MS). The SOP deviation occurred from an unknown start date of no earlier than October 2005 (possibly as late as 2008) and ending August 8, 2015. The deviation was a change in the concentration of acid used in calibration standards, which could affect concentrations measured in environmental samples. The potential impact of this deviation on the data produced was investigated experimentally and is discussed; however, the results for specific environmental samples are also affected by additional factors that could not be tested. Data users are encouraged to consider the experimental results in this memorandum along with additional quality-control data, including results from field spikes and replicates and from the USGS Quality Systems Branch Inorganic Blind Sample Program (<https://qsb.usgs.gov/ibsp/>), when interpreting environmental sample results from the affected period.

2 SUMMARY OF KEY FINDINGS

- A deviation was identified in the SOP for the preparation of calibration standards for the analysis of filtered water samples for dissolved trace elements by ICP-MS by the USGS National Water Quality Laboratory (NWQL).
- The SOP deviation occurred from an unknown start date of no earlier than October 2005 (possibly as late as 2008) and ending August 8, 2015. The consistency with which the deviation was applied during this period is unknown.
- For an unknown period, calibration standards and QC samples were formulated in 1.0 percent nitric acid rather than the 0.4 percent nitric acid specified in the SOP.
- The largest effects observed in our experiments were for beryllium, selenium, and zinc, with data affected by the SOP deviation being biased high by a median of 5.7 percent (Be), 8.0 percent (Se), and 3.1 percent (Zn), respectively, compared to data from adherence to the SOP.
- Variability was not appreciably affected by the SOP deviation.
- Laboratory users are encouraged to evaluate method performance in their own sample matrices through the use of field spikes as described by [Paul and others \(2016\)](#) and Office of Water Quality Technical Memorandum [2017.04](#).

3 BACKGROUND

Shortly prior to August 8, 2015, a deviation was identified in the SOP for the preparation of calibration standards for the analysis of filtered water samples for dissolved trace elements by ICP-MS by the USGS National Water Quality Laboratory (NWQL). The NWQL SOP for determination of dissolved trace elements by ICP-MS specifies that calibration standards and laboratory quality-control (QC) samples are made using reagent water acidified to 0.4 percent with nitric acid. For an unknown period, calibration standards and QC samples were formulated in 1.0 percent nitric acid rather than the 0.4 percent nitric acid specified in the SOP. The formulation of calibration standards and QC samples in 1.0 percent nitric acid began sometime after USGS method I-2020-05 (Garbarino and others, 2006) was implemented at the NWQL in October of 2005, and possibly did not begin until 2008. The consistency with which the deviation was applied during this period also is unknown. On August 8, 2015, the SOP deviation was corrected, and calibration standards and QC samples were again formulated in a 0.4 percent nitric acid matrix.

Filtered environmental water samples to be analyzed for dissolved trace elements by ICP-MS are preserved in the field by acidifying to 0.4 percent with nitric acid, and are referred to as “filtered acidified samples” or “FA.” The potential effects of a mismatch in the acid concentration used in the calibration standards compared to the concentration used in the environmental samples were unknown. There is wide agreement in the ICP-MS community that standards should be matrix matched to samples wherever possible, including matching the acidity of the calibration standards to that of the environmental samples. Methods used at the NWQL for the determination of trace elements in FA samples by ICP-MS specify the use of calibration standards made in an acidic matrix similar to that of preserved samples (Garbarino, 1999; Garbarino and others, 2006). The effects of acid concentration and other matrix components on analysis in ICP instrumentation has been extensively studied, although very few studies address newer ICP-MS instruments that utilize lenses or a quadrupole for mass filtering and a collision cell for interference removal (Agatemor and Beauchemin, 2011; Aguirre and others, 2014; Allain and others, 1991; Canals and others, 2002; Fansworth and Spencer, 2017; Maestre and other, 1999; Olesik and Jiao,

2017; Pickford and Brown, 1986; Stewart and Olesik, 1998; Todoli and Mermet, 1998, US EPA, 2004). No studies to date have investigated small (less than one percent) differences in acid concentration.

Acid concentrations can influence sample matrices and can subsequently influence results for environmental samples calculated using a calibration curve derived from standards composed of a different matrix. Identifying the effects of a deviation in acid concentration is made difficult by additional factors that contribute to the bias and variability in analyses of trace elements by ICP-MS. Day-to-day tuning of the ICP-MS, variability in the performance of the instruments and analysts over time, and daily and yearly maintenance can contribute to minor variations in bias and variability. Paul and others (2016) investigated other sources of bias and variability. The use of internal standards helps adjust for some of these fluctuations. Control limits are used to identify large biases and variability. For example, after tuning, the instrument must meet minimum criteria, including minimum sensitivity, maximum oxide interference, and maximum doubly charged ion interferences, which are checked daily. In addition, continuing calibration verifications are analyzed every 10-20 samples and must be within acceptance criteria in order to accept the sample data.

Although the SOP deviation was not found to be responsible for results outside of the control limits, the NWQL found it prudent to further investigate for potential effects not identified using the control limits. An experiment was undertaken to characterize potential effects to environmental samples, including any bias, shift in measured concentration, or variability introduced into the data.

4 METHODS

4.1 Experimental Design

An experiment to characterize the potential bias and variability attributable to the SOP deviation was designed to incorporate many of the factors that are included in a typical validation of method changes at the NWQL. The experiment was conducted under conditions of normal method operation, including multiple analysts, multiple analysis events at various times of the year, and a variety of environmental sample matrices. Because a single instrument had been used to produce most of the data for dissolved trace elements, that instrument was used exclusively for this experiment. Order of analysis was not varied: the calibration standards in 0.4% nitric acid matrix were always run on the first day and the calibration standards in the 1.0% nitric acid matrix were always run on the second day of each analysis event. By not randomizing this variable, systematic bias may have been introduced to the data. Minimum instrument performance criteria were achieved each day including minimum sensitivity, maximum oxide interference, and maximum doubly charged ion interferences, and QC results were acceptable each day such that the instrument was performing under typical conditions each day of the study. Analytes and detection limits used for the experiment are listed in Table 1.

By including variables such as multiple analysts and days of analysis, acid concentration was not isolated as a single variable in this study. Significant differences in bias and variability identified from the population of experimental samples are likely chiefly attributable to the difference in acid concentration in the calibration standards; however, the results of this study cannot be used to identify specific effects on specific environmental samples. One objective of the study design was to characterize the magnitude and direction of potential shifts in measured concentration of environmental sample data that may have occurred when the acid concentration was changed from 1.0 percent nitric acid back to 0.4 percent nitric acid consistent with the intended SOP on August 8, 2015.

Table 1. Analytes, method references, lab codes, method and parameter codes, and detection limits used for the experiment.

[NEMI, National Environmental Method Index; NWIS, National Water Information System; µg/L, micrograms per liter]

Analyte (element symbol)	Analyte name	Method reference in NEMI ¹	Lab code	Method code in NWIS	Parameter code in NWIS	Detection limit (µg/L) ²
Li	lithium	I-2477-92	2505	PLM40	01130	0.22
Be	beryllium	I-2477-92	1787	PLM43	01010	0.02
B	boron	I-2477-92	2504	PLM40	01020	5.0
Al	aluminum	I-2477-92	1784	PLM43	01106	2.2
V	vanadium	I-2020-05	3134	PLM10	01085	0.08
Cr	chromium	I-2020-05	3126	PLM10	01030	0.3
Mn	manganese	I-2477-92	1793	PLM43	01056	0.40
Co	cobalt	I-2020-05	3124	PLM10	01035	0.05
Ni	nickel	I-2020-05	3130	PLM10	01065	0.20
Cu	copper	I-2020-05	3128	PLM10	01040	0.8
Zn	zinc	I-2020-05	3138	PLM10	01090	2.0
As	arsenic	I-2020-05	3122	PLM10	01000	0.10
Se	selenium	I-2020-05	3132	PLM10	01145	0.05
Sr	strontium	I-2477-92	2507	PLM40	01080	0.8
Mo	molybdenum	I-2477-92	1794	PLM43	01060	0.05
Ag	silver	I-2477-92	1796	PLM43	01075	0.02
Cd	cadmium	I-2477-92	1788	PLM43	01025	0.03
Sb	antimony	I-2477-92	1785	PLM43	01095	0.027
Ba	barium	I-2477-92	1786	PLM43	01005	0.25
W	tungsten	I-2477-92	3228	PLM43	01155	0.03
Tl	thallium	I-2477-92	2508	PLM40	01057	0.03
Pb	lead	I-2477-92	1792	PLM43	01049	0.04
U	uranium	I-2477-92	1797	PLM43	22703	0.014

1. I-2477-92 is Garbarino (1999) and Faires (1993); I-2020-05 is Garbarino and others (2006).

2. Detection limits were determined using the US EPA MDL procedure (U.S. EPA, 2014) and are those in place during this study.

The experiment was conducted using (1) spiked water samples composed of blank water and matrix waters that had been previously tested, (2) standard reference samples (SRSs) supplied by the USGS Quality Systems Branch, and (3) environmental water samples, as follows:

1. *Spikes:* Blank water used for spiked samples (called “blank spikes”) was ASTM Type I water. Groundwater matrix came from a domestic well near Elizabeth, CO. Surface-water matrix was provided by Denver Water and was collected from the raw-water inlet to the Foothills Water Treatment Plant, which comes from the South Platte River at the outlet of Strontia Springs Reservoir. Aliquots of groundwater and surface-water were filtered using a 0.45-micrometer (µm) capsule filter before spiking. All matrices were acidified to 0.4 percent ultrapure nitric acid and then spiked to the intended concentrations using a stock solution containing certified amounts of

all analytes listed in table 1. Spiked samples included concentrations near the estimated detection limit (DL) when possible, and in the lower and upper thirds of the calibration range or range of concentrations previously observed in samples received at the NWQL.

2. *SRSs*: USGS Quality Systems Branch standard reference samples T-215 and T-219 were supplied acidified to 0.4 percent nitric acid and were used as is.
3. *Environmental samples*: 124 environmental samples from throughout the United States were selected to span a wide range of conductivities (ionic strength) and hydrologic conditions. Approximately 40 percent of the selected samples were groundwater and approximately 60 percent were surface-water.

The spikes, SRSs and environmental samples were distributed throughout four batches. The sample sequence within each batch was randomized and each batch was analyzed on two consecutive days using the Agilent Technologies 7500ce ICP–MS instrument that was typically used in NWQL fee-for-service analysis for filtered samples (referred to as “instrument 7” in Paul and others (2016)). On the first day of each analysis event, the sample results were quantitated using calibration standards made with 0.4 percent nitric acid. On the second day, the same sample results were quantitated using calibration standards made with 1.0 percent nitric acid. Each spike and SRS was injected one or more times during at least three separate analysis events and the repeated injections are referred to as *replicate analyses*. Three analysts performed the analyses between May 26, 2016 and October 17, 2016. A single analyst performed both days of analysis for a given batch.

All data were reviewed using routine procedures. Where QC did not pass for an analyte (which occurred rarely, for reasons unrelated to the acid concentrations), the data point was eliminated from the analysis. Results with concentrations greater than the highest calibrator also were eliminated from the data analysis (rather than diluting and re-analyzing those samples on a different day) because the experiment was designed to minimize the number of variables. Results below the detection limit were eliminated from some of the paired data analysis (detailed in section 4.2). These practices resulted in the number of data points (n) used for each analyte varying.

4.2 Statistical Methods and Calculations

Percent change (%) in sample concentration, C , resulting from using calibration standards in 1.0 percent nitric acid ($C_{1\%}$, deviating from the SOP) compared to calibration standards in 0.4 percent nitric acid ($C_{0.4\%}$, adhering to the SOP) was calculated for each sample using equation 1:

$$\text{Percent change (\%)} = \frac{(C_{1\%} - C_{0.4\%})}{C_{0.4\%}} \times 100\% \quad [\text{Equation 1}]$$

Percent change was positive for samples with $C_{1\%}$ greater than $C_{0.4\%}$. Percent change was negative for samples with $C_{1\%}$ less than $C_{0.4\%}$. Percent change was calculated relative to $C_{0.4\%}$ because $C_{0.4\%}$ is the accepted practice and thus serves as the basis for comparison.

Concentration differences are expressed using equation 2:

$$\text{Concentration difference} = C_{1\%} - C_{0.4\%} \quad [\text{Equation 2}]$$

Where most probable values (MPV) or expected concentrations were available for spikes and SRSs *percent recoveries* were calculated using equation 3:

$$\text{Percent recovery (\%)} = \frac{C_x}{C_e} \times 100\% \quad [\text{Equation 3}]$$

Where C_x is the measured concentration and C_e is the MPV or expected concentration.

Differences between percent recoveries (equation 3) were calculated as *absolute difference in percent recovery* using equation 4:

$$\text{Absolute difference in percent recovery (percentage points)} = \text{Percent Recovery}_{1.0\%} - \text{Percent Recovery}_{0.4\%} \quad [\text{Equation 4}]$$

Mean percent recoveries of replicates were chiefly used with equation 4 and the result of equation 4 called *absolute difference in mean percent recovery*.

Where replicate analyses of a spike or SRS were performed standard deviation (SD) and relative standard deviation (RSD) were used to express variability. *Standard deviation* and *RSD* were calculated using equations 5-6:

Standard deviation:

$$SD = \sqrt{\frac{\sum_{i=1}^N (C_i - \bar{C})^2}{N-1}} \quad [\text{Equation 5}]$$

where N is the number of replicates, C_i is the concentration of replicate i , and \bar{C} is the mean of all replicates.

Relative standard deviation:

$$RSD (\%) = \frac{SD}{\bar{C}} \times 100 \text{ percent} \quad [\text{Equation 6}]$$

where SD is the standard deviation from equation 5, and \bar{C} is the mean of all replicates.

A significance level of 0.05 was chosen for interpreting all statistical analyses. Variability was assessed using replicate analyses of spikes and SRS samples. A two-sample independent test was used to compare variances in results calculated with 0.4% nitric acid calibration standards and results calculated with 1.0% nitric acid calibration standards. The p -values from the Brown-Forsythe test were used to determine if the two treatments came from distributions with variances that were equal.

The paired spike, SRS and environmental sample data were tested for normality using skewness and kurtosis indices. From the statistical assessment, normality could not be assumed; therefore, non-parametric tests were used for subsequent analyses. The mean and median percent change (equation 1) were calculated for each analyte for all results greater than the detection limit for (a) all analyses, including spikes, SRSs, and environmental samples, and (b) the environmental samples only. A Wilcoxon signed ranks test was performed on (a) all analyses, including spikes, SRSs, and environmental samples, and (b) the environmental samples only from each analyte to determine if the concentration data produced using 1.0 percent nitric acid calibration standards was significantly different than that produced using the 0.4 percent nitric acid calibration standards. Data below the detection limit were eliminated from this analysis. Scatterplots were used to investigate the dependence between mean percent change and concentration.

Separately, means, SDs, RSDs, and mean percent recoveries were calculated for each set of replicate spike and SRS data. A Wilcoxon signed ranks test was performed on each group of paired spike and SRS concentration data from each analyte. Data below the detection limit were included in the calculations and tests performed on each set of replicate spike or SRS data. Concentrations below the DL occurred only for

some analytes in the blank spike at the detection limit. This is expected since, by definition, a spike at the DL should be detected above the DL 50 percent of the time.

5 RESULTS AND DISCUSSION

Results of the data analysis are provided in attachments 1-5 and are discussed in the following sections. Scatterplots were used to investigate the dependence between mean percent change and concentration (attachment 1). From visual inspection, no shifts in residuals were evident over the ranges of concentrations; therefore, data over the observed concentration ranges for each analyte were pooled. Project-specific data-quality objectives must be considered when deciding which of the statistically significant biases or shifts are large enough to take into consideration when interpreting data for environmental samples analyzed during the period of affected data.

5.1 Variability

Use of the Brown-Forsythe test on spiked samples and SRSs showed that the variances for results using the 1.0% nitric acid calibration standards differed from those using the 0.4% nitric acid calibration standards, but not markedly (attachment 2). Both populations of samples showed increased variability at low concentrations near the detection limit (i.e., within 10 times the DL), which is expected. Relative standard deviations of replicate analyses of spike and SRS samples (equation 6) are sometimes high and don't agree well when quantitated against the two different acid matrices (attachment 3) when the analyte concentration is near the DL (usually within 10 times the DL), especially in the blank spike at the DL. Relative standard deviations generally drop to less than five percent in spikes with concentrations greater than 10 times the DL. Relative standard deviations obtained for blank spikes in the original method validation (Garbarino and others, 2006) are comparable to RSDs obtained in this study and reported in attachment 3, with the exception of thallium (Tl). Variability of Tl results in this study for blank spikes up to 1 µg/L is higher than in the original method validation. This increased variability was only observed in blank spikes in this study measured using both calibrator matrices, and was not observed in spiked groundwater, surface-water, or reference materials. These findings suggest that variability may have been elevated for Tl at the time of the study, but it is likely not related to the acid matrix of the calibration standards.

5.2 Bias and Change in Measured Concentration

Mean percent change (equation 1), median percent change, results from the paired Wilcoxon signed ranks test (*p*-values), and number of observations (*n*) are reported in table 2 for (a) all analyses, including spikes, SRSs, and environmental samples, and (b) the environmental samples only. Percent change was expressed as the difference between the concentration from using calibration standards in 1.0 percent nitric acid ($C_{1\%}$, deviating from the SOP) compared to using calibration standards in 0.4 percent nitric acid ($C_{0.4\%}$, adhering to the SOP), relative to the concentration from standards in 0.4 percent nitric acid, as expressed in equation 1. Attachments 4 and 5 contain detailed statistical analysis and graphs of the differences between the two concentrations ($C_{1\%} - C_{0.4\%}$) versus the mean of the two concentrations.

When all data, including spikes, SRSs, and environmental samples are pooled (table 2), paired data from the two calibrator acid matrices (equation 2) are significantly different from each other for all analytes except aluminum (Al), boron (B), lithium (Li), manganese (Mn), and molybdenum (Mo) (table 2). The median percent changes (equation 1, table 2) are less than five percent for all analytes except selenium (Se) (7.8 percent) suggesting that while statistically significant, they are possibly not of practical significance for any analytes except Se.

When only environmental sample data are considered (table 2), paired data from the two calibrator acid matrices are significantly different from each other for Al, arsenic (As), beryllium (Be), cobalt (Co), chromium (Cr), nickel (Ni), lead (Pb), Se, uranium (U), and zinc (Zn). The median percent change for each of these analytes ranged from -1.9 percent to 8.0 percent and was between -5 percent and 5 percent for Al, As, Co, Cr, Ni, Pb, U, and Zn (table 2). The median percent changes are 5.7 percent for Be and 8.0 percent for Se.

Data from replicate measurements of spikes and SRSs were compared for the two calibrator acid matrices using a Wilcoxon signed ranks test when $n > 7$ (attachment 3). The statistical power of the analysis is uncertain because of the small n for some data sets. Mean recoveries, standard deviation, percent relative standard deviation, n , and p values are tabulated in attachment 3. Median recoveries and paired test statistics of the data can be found in attachment 2.

Absolute differences in mean percent recoveries (equations 3 and 4) between the two matrices are discussed along with concentration differences (equation 2) for spike and SRS data since an expected concentration or most probable value (MPV) is available. The concentration difference (equation 2) between data collected using the two calibrator acid concentrations was statistically significant for Se in SRS T-215 and T-219 (attachment 3; 6 percentage points of difference between percent recoveries for both SRSs, equation 4). The difference in measured concentration was statistically significant for copper (Cu) in T-215 (-3 percentage points of difference) and for As (2 percentage points of difference), Cr (-13 percentage points of difference), tungsten (W; percent recoveries not calculated), and Zn (5 percentage points of difference) in T-219. The MPV for Cu and Cr are within $2\times$ the detection limit. Tungsten has no MPV but the measured concentration was within $2\times$ the detection limit. While statistically significant, the differences found for Cu, Cr, and W are not likely of practical significance. Selenium appears to be low biased in both reference materials for both calibrator acid matrices. Arsenic appears to be slightly high biased in T-219 when 1 percent nitric acid calibration standards are used (attachment 3). Zn appears to be slightly low biased in T-219 when 0.4 percent nitric acid calibration standards are used.

Mean recoveries of some analytes in the SRS samples were greater than 105 percent or lower than 95 percent (As, Cu, Ni, Se, vanadium, and Zn in T-215 and cadmium (Cd), Cr, Ni, and Se in T-219), but the MPV of all those analytes except Se in both SRSs and Cd and Ni in T-219 were less than ten times the detection limit. Recoveries of all of these analytes except Cd were consistent with recoveries reported by the NWQL through the Inorganic Blind Sample Project (IBSP; <https://qsb.usgs.gov/ibsp/>). Recovery of Cd was higher in this study (107 percent recovery with 1.0 percent acid calibrators and 103 percent recovery with 0.4 percent acid calibrators) than what is reported by the NWQL to the IBSP (101percent recovery). Recoveries of Se in both the 1.0 percent and 0.4 percent acid matrices were well below 100 percent of the MPV even though the MPV is more than 10 times the DL. Se data produced by the NWQL has historically been below the MPV when the concentration of Se in the SRS is below 1 microgram per liter (<https://qsb.usgs.gov/ibsp/>). This is either due to the NWQL data being low biased or the MPV being high biased by incomplete interference removal in measurements made by other labs that participate in the round robin studies that are used to determine the MPVs. Spike recoveries of Se in groundwater and surface-water in this study were close to 100 percent, including those spikes that were below 1 $\mu\text{g/l}$. Because of the historical low bias in NWQL Se data for some SRS samples compared to the MPV, it is difficult to tell from the SRS data whether the 0.4 percent acid matrix data was low biased or the 1 percent acid matrix data was high biased. Based on recoveries relative to the MPV, the 1 percent matrix calibration standards yielded higher recoveries relative to the MPV for Se than the 0.4 percent matrix calibration standards.

Table 2: Statistical analysis of results for spikes, SRSs, and environmental samples obtained using calibration standards made with 1.0 percent nitric acid and 0.4 percent nitric acid. Mean and median percent changes are relative to data collected using 0.4 percent nitric acid calibration standards, so percent change is positive for samples where the measured concentration with 1.0 percent nitric acid calibrators ($C_{1\%}$) is greater than the measured concentration with 0.4 percent nitric acid calibrators ($C_{0.4\%}$) and negative for samples with ($C_{1\%}$) less than ($C_{0.4\%}$). A Wilcoxon signed ranks test was used to determine whether differences in concentration data collected with calibrators of different acid concentrations were statistically significant for each analyte at a significance level of 0.05.

[m/z, mass to charge ratio; %, percent; p, test statistic from Wilcoxon signed ranks test; n, number of samples used for statistical analysis; SRS, standard reference sample]

Analyte	Symbol	m/z	Detection Limit (µg/L)	Pooled spike, SRS, and environmental sample data				Environmental sample data only			
				Mean Change (%)	Median Change (%)	p	n	Mean Change (%)	Median Change (%)	p	n
Silver, WF, ICP-MS	Ag	107	0.02	0.1	-0.8	0.0067	109	0.3	0.3	0.4205	22
Aluminum, WF, ICP-MS	Al	27	2.2	3.6	0.4	0.0707	198	5.8	1.5	0.0066	96
Arsenic, WF, cICP-MS	As	75	0.1	3.5	2.4	<0.0001	216	3.9	2.0	<0.0001	115
Boron, WF, ICP-MS	B	11	5.0	0.4	0.2	0.8773	201	0.0	0.3	0.6599	108
Barium, WF, ICP-MS	Ba	137	0.25	2.1	0.5	0.0232	233	0.6	0.0	0.8556	118
Beryllium, WF, ICP-MS	Be	9	0.02	5.2	3.5	0.0018	127	13.7	5.7	<0.0001	36
Cadmium, WF, ICP-MS	Cd	111	0.03	0.7	0.7	0.0092	136	-0.4	-0.4	0.895	44
Cobalt, WF, cICP-MS	Co	59	0.05	-0.8	-0.8	<0.0001	193	-0.6	-0.7	0.009	102
Chromium, WF, cICP-MS	Cr	52	0.3	-2.0	-1.5	<0.0001	107	-0.5	-1.4	0.0013	34
Copper, WF, cICP-MS	Cu	63	0.8	-0.6	-0.8	0.0161	142	-0.4	-0.6	0.0633	69
Lithium, WF, ICP-MS	Li	7	0.22	-0.5	0.1	0.8752	209	-0.2	0.2	0.8718	101
Manganese, WF, ICP-MS	Mn	55	0.40	0.8	-0.4	0.2182	197	0.3	-0.4	0.8207	111
Molybdenum, WF, ICP-MS	Mo	95	0.05	0.5	-0.4	0.2590	230	1.0	-0.6	0.4785	117
Nickel, WF, cICP-MS	Ni	60	0.20	-1.9	-1.3	<0.0001	198	-1.6	-1.6	0.0035	102
Lead, WF, ICP-MS	Pb	208	0.04	-1.8	-1.9	<0.0001	157	-2.3	-1.9	0.0002	65
Antimony, WF, ICP-MS	Sb	121	0.027	0.9	0.5	0.0101	204	0.0	0.3	0.3342	103
Selenium, WF, cICP-MS	Se	78	0.05	9.7	7.8	<0.0001	195	10.0	8.0	<0.0001	91
Strontium, WF, ICP-MS	Sr	88	0.8	1.4	0.7	0.0062	210	0.8	0.5	0.2489	116
Thallium, WF, ICP-MS	Tl	205	0.03	-2.5	-2.2	<0.0001	86	-1.5	-2.0	0.0885	23

Uranium, WF, ICP-MS	U	238	0.014	-0.4	-0.7	0.0282	202	-1.5	-1.3	0.0015	106
Vanadium, WF, cICP-MS	V	51	0.08	5.8	-1.1	<0.0001	155	8.8	-1.0	0.1786	114
Tungsten, WF, cICP-MS	W	182	0.03	-2.3	-1.9	<0.0001	155	-1.1	-1.6	0.2338	66
Zinc, WF, cICP-MS	Zn	66	2	3.9	3.2	<0.0001	167	3.4	3.1	<0.0001	56

Similar to the SRS data, there were several instances where the differences in the blank spikes using different calibrator acid matrices were statistically significant by the Wilcoxon signed ranks test. Generally, these occurrences were not consistent across the spike levels, and the difference in mean recovery was within -5 to +5 percent in the blank spikes of low and high concentrations. The exceptions are W and Se, where there were two instances each of the differences in the blank spikes being statistically significant and the differences in recovery being five percent or greater. The differences in recoveries for W in spikes composed of groundwater and surface-water were less than 5 percent.

Mean percent recoveries of all spikes, including blank, groundwater, and surface-water spikes generally agreed well between the two calibrator matrices. There were two exceptions to this: Tl and Se. Thallium spike recoveries in blank water varied more than 10 percent between the two calibrator matrices and the recoveries with 1 percent calibration standards were sometimes higher and sometimes lower than the recovery with 0.4 percent calibration standards. This behavior is consistent with the increased variability for Tl over that observed in the source method validation study (Garbarino and others, 2006), and Tl is therefore not considered further. Differences in percent recoveries for Se in all spike matrices ranged between +3 and +19 percentage points of difference (equation 4), with the values measured with 1.0 percent nitric acid calibration standards always being higher than those measured with 0.4 percent nitric acid calibration standards. The percentage points of difference in percent recoveries (equations 2 and 4) of Se in the spikes were comparable in magnitude and direction to the mean percent differences (equation 1) observed when comparing environmental sample data (table 2, attachment 3).

Recoveries of low and high spikes in groundwater, surface-water, and blank water with both calibrator matrices ranged between 91 percent and 111 percent (attachment 3), with the exception of one Tl spike (78 percent recovery) and two Cu spikes (86 and 87 percent recoveries). Increased variability in recoveries is expected for concentrations near the DL and was observed in spikes within ten times the DL. Recoveries of the spike at the lowest spiked concentration (within ten times the DL) for each water ranged from 70 percent to 134 percent with the exception of one V spike that had a mean recovery of 45 percent. Recoveries of Se were acceptable at all spike levels, but consistently slightly low biased when 0.4 percent nitric acid calibration standards were used and consistently slightly high biased when 1 percent nitric acid calibration standards were used. With the exception of the elements discussed above, recoveries of spikes in blank water, groundwater, and surface were comparable to those reported in Garbarino and others (2006).

5.3 Relating this study to previously reported bias in dissolved ICP-MS results

Previous concerns with regard to trace-element concentrations measured by ICP-MS being higher in dissolved samples than in whole-water samples from the same source ($C_F > C_{UF}$) led to an investigation into sources of bias and a report (Paul and others, 2016). That study was performed using two ICP-MS instruments at the NWQL, including the one (instrument 7) used for the investigation described in the present report. At the time of the study, the investigators were unaware that the calibration standards for the filtered line were formulated using 1 percent nitric acid. They were made aware during the writing of Paul and others (2016) and included the following statement:

“Currently, it is assumed that changes in strength of the HNO_3 matrix from 0.4 to 1 percent during respective calibrations of instrument 7 had little to no effect on the analytical results obtained from that instrument.”

It is now known that calibration standards for the whole water recoverable analysis were formulated in 0.4 percent nitric acid and 2 percent hydrochloric acid and all quantitation on instrument 6 was done using calibration standards in that matrix. On instrument 7, all quantitation most likely used calibration

standards formulated with 1.0 percent nitric acid. Therefore, where “instrument” was a variable in that study, “acid matrix” was also a variable.

6 SUMMARY

No consistent change in variability (precision of replicate analyses of a single sample, attachment 2) was observed for any analyte, suggesting that precision of the method was not affected by using calibration standards made in 1.0 percent nitric acid with samples preserved in 0.4 percent nitric acid. The precision observed with both calibration standard matrices was consistent with the precision observed in the source method (Garbarino and others, 2006) with the exception of Tl, where the variability was elevated in this study for both calibrator matrices.

For most analytes, the difference between sample concentrations measured with 1.0 percent acid calibration standards and 0.4 percent acid calibration standards was between -5 percent and +5 percent or was above 5 percent for one or two spikes, usually when the spike was within ten times the DL and the difference not distinguishable due to the elevated variability at low concentrations.

For Se, spike, SRS, and environmental sample data consistently show a difference in recoveries between the two acid matrices that is greater than 5 percent, and the mean or median percent change (equation 1) or absolute difference in percent recovery (equation 4) is always positive (table 2, attachment 3). This difference is sometimes statistically significant and sometimes not, possibly due to low statistical power in the smaller data sets. Spike, SRS, and environmental sample data show a difference in recoveries between the two acid matrices that is between 2 and 5 percent for As, Be, and Zn, although somewhat less consistently than for Se, and the percent change or percentage points of difference in percent recovery is usually positive (table 2, attachment 3). These differences are sometimes statistically significant.

The initial change from 0.4 percent nitric acid in calibration standards to 1.0 percent nitric acid was not documented, but occurred no earlier than October 2005 and possibly as late as 2008, therefore it is impossible to identify or attribute any initial shift in reported results. The change back to 0.4 percent nitric acid from 1.0 percent occurred on August 8, 2015 and therefore a shift of approximately 3-19 percent in environmental data for Se, approximately 0 to 4 percent for As, 0 to 11 percent for Be, and 0 to 5 percent for Zn corresponding to this time may be attributable to the change in acid matrix of the calibration standards. Within 5-10 times the detection limit, these differences may not be discernable due to increased variability near the detection limit. Data users are encouraged to consider the experimental results in this memorandum along with additional quality-control data, including results from field spikes and replicates and from the USGS Quality Systems Branch Inorganic Blind Sample Program (<https://qsb.usgs.gov/ibsp/>), when interpreting environmental sample results from the affected period.

When this SOP deviation was discovered, corrective actions were taken by the NWQL to ensure this method adhered to the approved SOP. Since 2015, during routine audits, the Quality Assurance Section (QAS) of the NWQL has increased their scrutiny of process compliance with SOPs and the source methods. The NWQL is also committed to devising appropriate procedures and training that will help to ensure that all personnel at the NWQL who work with methods understand a) what constitutes a method change and b) the process that must be completed prior to implementing a method change.

Any questions should be directed to Sarah Stetson (sstetson@usgs.gov).

7 ACKNOWLEDGEMENTS

The author gratefully acknowledges assistance from Heather Miller in the collection and handling of data for this report. Thanks to Teresa Williams, Tim Oden, Michael Rosen, John Garbarino and Tony Bednar for thoughtful reviews of the report.

8 REFERENCES

- Agatemor, C., and Beauchemin, D., 2011, Matrix effects in inductively coupled plasma mass spectrometry: A review: *Analytica Chimica Acta*, v. 706, no. 1, p. 66-83.
- Aguirre, M.A., Fialho, L.L., Nobrega, J.A., Hidalgo, M., and Canals, A., 2014, Compensation of inorganic acid interferences in ICP-OES and ICP-MS using a Flow Blurring[registered sign] multinebulizer: *Journal of Analytical Atomic Spectrometry*, v. 29, no. 7, p. 1218-1227.
- Allain, P., Jaunault, L., Mauras, Y., Mermet, J.M., and Delaporte, T., 1991, Signal enhancement of elements due to the presence of carbon-containing compounds in inductively coupled plasma mass spectrometry: *Analytical Chemistry*, v. 63, no. 14, p. 1497-1498.
- Canals, A., Gras, L., and Contreras, H., 2002, Elimination of nitric acid interference in ICP-AES by using a cyclonic spray chamber/Nafion membrane-based desolvation system: *Journal of Analytical Atomic Spectrometry*, v. 17, no. 3, p. 219-226.
- Faires, L.M., 1993, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory – Determination of metals in water by inductively coupled plasma-mass spectrometry: U.S. Geological Survey Open-File Report 92-634, 28 p.
- Farnsworth, P.B., and Spencer, R.L., 2017, Ion sampling and transport in Inductively Coupled Plasma Mass Spectrometry: *Spectrochimica acta. Part B: Atomic spectroscopy*, v. 134, p. 105-122.
- Garbarino, J.R., 1999, Methods of Analysis by the U.S. Geological Survey National Water Quality Laboratory – Determination of Dissolved Arsenic, Boron, Lithium, Selenium, Strontium, Thallium, and Vanadium Using Inductively Coupled Plasma-Mass Spectrometry: U.S. Geological Survey Open-File Report 99-093, 31 p.
- Garbarino, J.R., Kanagy, L.K., and Cree, M.E., 2006, Determination of Elements in Natural-Water, Biota, Sediment, and Soil Samples using Collision/Reaction Cell Inductively Coupled Plasma -- Mass Spectrometry: U.S. Geological Survey Techniques and Methods, Book 5, section B, chap. 1, 88 p.
- Maestre, S., Mora, J., Todoli, J.-L., and Canals, A., 1999, Evaluation of several commercially available spray chambers for use in inductively coupled plasma atomic emission spectrometry: *Journal of Analytical Atomic Spectrometry*, v. 14, no. 1, p. 61-67.
- Olesik, J.W., and Jiao, S., 2017, Matrix effects using an ICP-MS with a single positive ion lens and grounded stop: analyte mass dependent?: *Journal of Analytical Atomic Spectrometry*, v. 32, no. 5, p. 951-966.
- Paul, A.P., Garbarino, J.R., Olsen, L.D., Rosen, M.R., Mebane, C.A., and Struzeski, T.M., 2016, Potential sources of analytical bias and error in selected trace-element data quality analyses: U.S. Geological Survey Scientific Investigations Report 2016–5135, 58 p.

Pickford, C.J., and Brown, R.M., 1986, Comparison of ICP-MS with ICP-ES: detection power and interference effects experienced with complex matrices: Spectrochimica Acta Part B: Atomic Spectroscopy, v. 41, no. 1, p. 183-187.

Stewart, I.I., and Olesik, J.W., 1998, Steady state acid effects in ICP-MS: Journal of Analytical Atomic Spectrometry, v. 13, no. 12, p. 1313-1320.

Todoli, J.-L., and Mermet, J.-M., 1998, Minimization of acid effects at low consumption rates in an axially viewed inductively coupled plasma atomic emission spectrometer by using micronebulizer-based sample introduction systems: Journal of Analytical Atomic Spectrometry, v. 13, no. 8, p. 727-734.

U.S. EPA. 1994. "Method 200.8: Determination of Trace Elements in Waters and Wastes by Inductively Coupled Plasma-Mass Spectrometry," Revision 5.4. Cincinnati, OH

U.S. Environmental Protection Agency, 2014, Guidelines establishing test procedures for the analysis of pollutants (Part 136, Appendix B. Definition and procedure for the determination of the method detection limit-Revision 1.11, June 30, 1986): U.S. Code of Federal Regulations, Title 40, July 1, 2014 edition, p. 344-347. (Also available at <http://www.gpo.gov/fdsys/pkg/CFR-2014-title40-vol23/pdf/CFR-2014-title40-vol23-part136-appB.pdf>.)

9 ATTACHMENTS

Attachment 1. Bivariate fit of percent change vs concentration measured in 0.4 percent nitric acid calibration standards.

Attachment 2. Summary table of Brown-Forsythe test for variance results and statistical analyses including variance and paired tests.

Attachment 3. Mean, standard deviation, and relative standard deviation of replicate measurements of spikes and standard reference samples, along with mean percent recoveries, number of replicates measured, and p values from a Wilcoxon signed ranks test of the replicate measurements made using calibration standards made with 1.0 percent nitric acid and 0.4 percent nitric acid.

Attachment 4. Statistical analysis of results for matched pairs for pooled spikes, standard reference samples, and environmental samples analyzed using calibration standards made with 1.0 percent nitric acid and 0.4 percent nitric acid.

Attachment 5. Statistical analysis of results for matched pairs of environmental samples analyzed using calibration standards made with 1.0 percent nitric acid and 0.4 percent nitric acid.

/signed/
Jeff McCoy, Chief
National Water Quality Laboratory
Branch of Analytical Services

Supersedes: N/A

Key words: ICP–MS, inductively coupled plasma–mass spectrometer, trace elements, trace metals, acid matrix, dissolved, matrix effects

Distribution: by Rapi–Note announcement
Posted to <http://wwwnwql.cr.usgs.gov/USGS> (the NWQL USGS–visible intranet; internal USGS access only) and <http://nwql.usgs.gov/Public> (the NWQL public internet)

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.