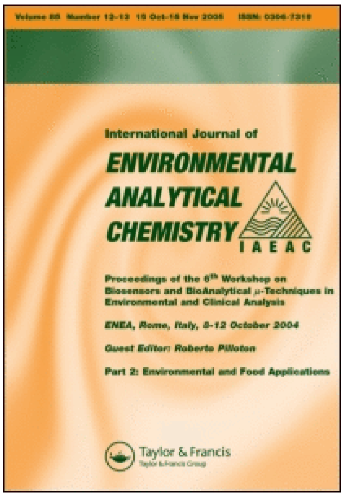

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

The role of internal standards and their interaction with soils impact accuracy of volatile organics determinations

Michael H. Hiatt[a]
[a] US Environmental Protection Agency, National Exposure Research Laboratory, Environmental Sciences Division, Las Vegas, Nevada 89193-3478, USA

Online publication date: 16 April 2010

To cite this Article Hiatt, Michael H.(2010) 'The role of internal standards and their interaction with soils impact accuracy of volatile organics determinations', International Journal of Environmental Analytical Chemistry, 90: 8, 591 — 604
To link to this Article: DOI: 10.1080/03067310903045497
URL: http://dx.doi.org/10.1080/03067310903045497

Intern. J. Environ. Anal. Chem.
Vol. 90, No. 8, 15 July 2010, 591–604

The role of internal standards and their interaction with soils impact accuracy of volatile organics determinations

Michael H. Hiatt*

US Environmental Protection Agency, National Exposure Research Laboratory, Environmental Sciences Division, PO Box 93478, Las Vegas, Nevada 89193-3478, USA

(*Received 12 March 2009; final version received 14 May 2009*)

Both US Environmental Protection Agency (EPA) SW-846 Methods 8260C/5035 and 8261A include mixing soil with water and addition of internal standards prior to analyses but the equilibration of internal standards with the soil is not required. With increasing organic carbon content and no effort to equilibrate internal standards with the matrix, results are less likely to be accurate. Adding internal standards to soils prior to diluting the sample with water gives more accurate determinations but less reliable quality control (QC). Extending times for equilibration of internal standards improves accuracy but is conducive to analyte degradation not normally observed during analyses. Soil-matrix effects on a given analyte can be greatly understated using a single internal standard as described in Method 8260C while the use of multiple internal standards as described in Method 8261A is more accurate. Method 8261A's reporting error when spiking soils before adding water provides confidence intervals with accuracy near the experimental rule (75.2, 95.7 and 99.5%) with the exception of two analytes that require overnight equilibration.

Keywords: soil; analyses; spiking; accuracy

1. Introduction

The level of uncertainty in environmental analyses is of concern for those who use analytical data to make environmental decisions [1,2]. Soil is a complex medium that can influence the behaviour of volatile organic analytes (VOAs) as general matrix effects (such as organic phase uptake), biological activity and even varying kinetics (fast and slow) adding to the difficulty in assessing the viability of VOA determinations [3,4]. Soil has been difficult to prepare as a reference material for VOAs. Researchers have used vapour addition of analytes to dry soils for developing performance soils [5]. It has also been found that volatile analytes react with the reference soil matrix after their addition [6]. Just how the soil matrix might impact the accuracy of volatile organic determinations is poorly understood.

The difficulties encountered in the preparation of soil reference materials also describe challenges for analysts in preparing soils for analyses and may differ depending on method [7]. Internal standards are commonly used in gas chromatography/mass spectrometry

*Email: hiatt.mike@epa.gov

ISSN 0306–7319 print/ISSN 1029–0397 online

DOI: 10.1080/03067310903045497
http://www.informaworld.com

(GC/MS) analyses and they are typically added to sample extracts prior to injections to normalise variations between injections. For determination of VOAs, however, internal standards are added to the sample prior to a concentration step (such as purge-and-trap concentration) exposing the internal standards to matrix effects. US Environmental Protection Agency's (EPA's) method for determining VOAs (SW-846 Method 8260C) assumes the chemical properties of analytes and their internal standards are identical and any relative change in response of an internal standard would be the same for analytes [8]. This assumption will introduce inaccurate results when the behaviour of an analyte and its internal standard are sufficiently different during the concentration from a matrix.

The equilibration of internal standards with soils is not required in EPA's methods for soil analyses. However, when analytes are at equilibrium with a soil and internal standards are not, the internal standards may be more effectively recovered than analytes. Using recovery of the internal standard for the recovery of an analyte ignores that portion of analyte bound with the matrix resulting in an understated analyte concentration. The impact of allowing equilibration time and different approaches to adding internal standards are evaluated in this work.

Method 5032 describes vacuum distillation as the means to separate volatile compounds from the soil [9] and is used as an optional concentration step for Method 8260C. Vacuum distillation had been found to be more efficient in extracting volatile analytes from soil than headspace, ambient purge-and-trap, and heated purge-and-trap techniques [10]. Therefore observations in this work as to how compounds are affected by the soil matrix are also relevant to headspace and purge-and-trap analyses and may address problems observed using non-vapour phase (methanol) extractions [7].

Method 8261A incorporates vacuum distillation and differs from Method 8260C/5032 in that Method 8261A incorporates a battery of internal standards to interpret an analyte's response and while Method 8260C specifies multiple internal standards are to be used, each analyte is related only to its nearest (in retention time) internal standard. Where the variation in response of a single internal standard translates to the same variation in its associated analytes, the battery used in Method 8261A interprets the analyte response as a function of its boiling point and relative volatility [11]. It would be expected that parsing matrix effects by chemical properties would generate more accurate results. This study processes raw experimental data by the two internal standard approaches and compares the relative accuracy of their determinations.

Method 8261A generates an error term with each analyte determination and these have been shown to be accurate in describing the analytical error for a variety of water matrices [12]. The reporting error generated as described in Method 8261A was therefore a primary tool in evaluating this study's experimental results [9]. While this study assesses the reporting errors as an analytical error it also evaluates which quality control parameters need be implemented to improve accuracy. Multiple replicate analyses provided the data for calculating the frequency that confidence intervals included the amount of analyte added to the samples. Because each analyte data set was small, a less restrictive Chebyshev's inequality was used in place of the empirical rule to identify analytes where reporting errors where inconsistent with experimental errors [13]. The analytes were also evaluated for stability in the soil samples. The matrices studied include three soils that had been used in performance studies [6], acid-washed sand and Greenwich Bay (MA) sediment.

2. Experimental

2.1 *GC/MS*

The vacuum distiller is interfaced to a GC/MS so that the vacuum distillate is transferred directly to the GC/MS for analysis after a distillation. In this study, the GC/MS was a Thermo DSQ mass spectrometer and Trace GC. The GC capillary column was a $30\,m \times 0.25\,mm$ i.d., $1.5\,\mu m$ film VOCOL column (Supelco, Bellefonte, PA). The GC operating conditions were 2.5 min at –20°C, 40°C/min ramp to 60°C, 5°C/min ramp to 120°C and held at 120°C for 1 min, 20°C/min ramp to 220°C and held for 12 min resulting in a GC run time of 34 min. The injection was split 60:1 with a constant flow rate of 1.4 ml/min. The mass spectrometer scanned between 35 and 300 amu at 1 scan/sec.

2.2 *Vacuum distiller*

A Cincinnati Analytical Instruments Model VDC1012 vacuum distiller (Indianapolis, IN) performed the distillations in the study. Samples were vacuum distilled for 7.5 min with a 2.5 min transfer to the GC/MS through a transfer line held at 200°C.

2.3 *Quantitation*

Calibration was performed as described in Method 8261A. The monitoring internal standards are listed in Table 1 (the complete list of internal standards can be found in the Supplementary Material, available online). Vinyl chloride-d_3 was added as a surrogate for gases [12]. The higher boiling point compounds, 1,2,3-trichlorobenzene-d_3, 3,5-di-tert-butyltoluene, 3,5-dibromotoluene, azulene, *a*,*a*′-dichloro-*o*-xylene were added as part of the internal standard/surrogate suite to more fully describe the boiling point effects at the upper range of the method. The software used to perform calibration and quantitation was SMCReporter 4.2 available from EPA's web pages [14].

The surrogates used to monitor method performance are presented in Table 2. These analytes were monitored as representative of three classes of compounds: volatile class was for compounds with boiling point less than 159°C, non-purgeable class as the volatile class but with relative volatility greater than 100, and the semi-volatile class representing compounds that boil at or above 159°C. A sub-class of *gases* was added to the volatile class and related to vinyl chloride-d_3.

The calibration range of analytes (listed in Table S3, Supplementary Material available online) in this study was nominally 5 to 500 ng per analyte. A review of calibration ranges was conducted to ensure the range was linear for each analyte. The lower points of some analytes were not used in generating their calibration curves when interferences at the lower point were observed. The lowest standard mass in each analyte calibration curve was used as the limit of quantitation (LOQ). Any analyte response that fell below the LOQ was considered as potentially less accurate and segregated from results that fell within the calibration range. The amount of analyte added to each sample is related to the calibration range and was at three levels, LOQ, mid-level (1/10 upper calibration amount) and high (1/2 upper calibration amount).

Samples were analysed in the same manner as was previously conducted for studying the accuracy of reporting errors for water analyses [12]. That is a multi-point calibration was run prior to analyses, and then a check standard, blank, unspiked soil, and 5, 10 and 15 g aliquots of soil spiked at the three levels (high, mid-point, and LOQ) were analysed

Table 1. Experimental relative responses of monitoring internal standards (by matrix and spike technique).

		Sand		NV		GA		OR		Sed	All[b]	
	SOM01[a]	Dry[c]	Wet[d]	Dry	Wet	Dry	Wet	Dry	Wet	wet	Dry	Wet
Volatile Class												
1,4-difluorobenzene	50–200	82–103	103–127	48–77	56–129	103–141	83–128	54–81	71–89	101–132	48–141	56–129
chlorobenzene-d_5	50–200	83–111	107–133	27–57	43–110	106–134	62–134	31–56	50–71	94–140	27–134	43–134
Volatile and Semivolatile Classes												
1,2-dichlorobenzene-d_4	50–200	65–111	105–135	7–26	21–61	78–121	26–136	11–27	24–40	67–104	7–121	21–136
Non-Purgeable Class												
tetrahydrofuran-d_8	NA[e]	46–202	85–217	87–201	(29)[f] 153–235	125–239	154–240	145–230	81–162	167–297	46–239	81–240
1,4-dioxane-d_8	NA	69–163	53–96	25–135	(8) 65–113	60–162	65–144	62–176	49–98	60–218	25–176	49–145
Semivolatile Class												
naphthalene-d_8	50–200	32–138	96–191	3–23	12–39	63–128	12–166	4–20	10–18	60–118	3–138	10–191

[a]Range of recoveries expected for internal standards from reference [15].
[b]Summary for soils not including sediment.
[c]Internal standards added to soil before diluting with water. Includes sample sizes 5, 10 and 15g.
[d]Internal standards added after diluting soil with water. Samples were 5g.
[e]This internal standard not monitored in CLP protocols.
[f]Values in parenthesis were for sample with low response due to bad seal and indicate an unacceptable recovery.

Table 2. Surrogate range of acceptable recoveries found experimentally by matrix.

Surrogates	Method[a] 8261	Sand		NV		GA		OR		SD	Summary	
		Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	wet	Dry	Wet
Gases												
vinyl chloride-*d3*	NA	(42.1)[b] 21.4–96	82–101	32–151	76–116 [2][c]	23–107	76–116	27–117	79–141	62–112	42–151	76–141
Volatile Class												
methylenechloride-d_2	75–125	83–108	77–107	76–134	62–96 [131]	68–98	78–110	75–107	80–97	49–92	68–134	62–131
benzene-d_6	75–125	90–99	99–106	94–115	93–100 [98]	89–97	93–100	92–104	85–96	91–101	89–115	85–106
1,2-dichloropropane-d_6	75–125	96–110	93–102	95–107	90–97 [96]	96–107	90–107	95–109	91–97	97–112	95–110	90–107
1,1,2-trichloroethane-d_3	50–150	99–117	96–121	83–109	92–109 [105]	97–117	94–109	85–106	93–102	113–148	83–117	92–121
4-bromofluorobenzene	75–125	94–102	93–100	66–82	70–85 [88]	96–103	70–100	69–78	69–85	87–97	66–103	69–100
Non-Purgeable Class												
nitromethane-^{13}C	65–135	83–104	72–128	64–87	77–96 [98]	67–92	77–103	74–102	83–94	85–117	64–104	72–128
ethyl acetate-^{13}C	65–135	100–111	91–124	0–1	0–7 [2]	77–102	0–104	0–5	22–80	7–90	0–111	0–124
Semivolatile Class												
decafluorobiphenyl	35–175	53–93	50–88	113–303	158–331 [149]	84–114	90–331	147–215	179–379	78–123	53–303	50–379
nitrobenzene-d_5[d]	25–150	84–133	105–135	65–286	115–180 [168]	112–148	101–161	74–121	66–144	2–98	65–286	66–180
acetophenone-d_5[d]	25–150	69–140	68–138	32–295	97–160 [205]	80–123	65–160	53–114	60–150	158–259	32–295	60–160
1,2,4-trichlorobenzene-d_3	NA	86–103	84–94	55–85	64–89 [78]	72–97	64–99	65–90	74–99	66–83	55–103	64–99
1-methylnaphthalene-d_{10}	NA	90–115	93–105	67–99	65–92 [139]	93–119	65–102	51–72	56–79	100–129	51–119	56–105
azulene	NA	18–82	70–170	14–56	13–66 [94]	0–14	13–44	0–16	58–168	56–170	0–82	0–192
3,5-di-*tert*-butyl toluene	NA	13–54	85–112	73–206	157–530 [194]	35–62	95–530	106–296	218–444	44–171	13–296	85–530
a,a-dichloro-*o*-xylene	NA	75–136	112–215	55–251	92–195 [480]	113–208	92–181	66–118	58–168	0–25	55–251	58–215

[a]Ranges posted with Method 8261A for soil [9].
[b]Surrogate windows narrowed to value in parenthesis to eliminate Chebyshev outliers.
[c]Results in brackets were those for a sample that was not well sealed during distillation.
[d]These surrogates and their ranges apply to the semivolatile analytes. Narrower limits would apply if they were used for just their analogues.

daily and replicated six times. This structure of analyses did not allow for interpretation until all analyses of a matrix were completed. However, when there were additional tests to clarify results, these were conducted using a two-point calibration (before samples and after samples) and samples were spiked with the calibration amounts. The additional tests used the high spike amount (1/2 upper point in Table 3 calibration range) and only 5 g of sample to minimise impact of background analyte concentrations.

2.4 *Samples*

Samples were prepared for analyses in the vacuum distiller sample vessel. First the necessary amount of soil was added to the vessels and the vessels closed by connecting to the vacuum distiller. Addition of a 5-uL methanol solution containing the internal standards and surrogates to the dry samples was next. This was done by touching the glass surface under the sample with the syringe before releasing the solution. The analytes were then added to the sample in the same manner. Finally, water was added to the soil. The mixing of the soil, spikes, and water varied by sample type and are discussed later.

Five different soil samples were evaluated. Three sample amounts (5, 10 and 15 g) of each soil were analysed. In addition, three different levels of analyte concentration were analysed at each sample size. The amounts of the spikes were 0.5 and 0.2 of the upper limit of the calibration range and at the limit of quantitation (0.02 of upper limit).

Acid-washed sand, three soils (NV, GA, OR), previously used to create performance sample [6] and a sediment (SD) were used for samples.

3. Results and discussion

The results are addressed in three sections and follow the sequence the study was performed. The first section discusses evaluating the Method 8261A analyses of soil and identifying any QC requirements that need to be implemented to improve accuracy. The second section addresses additional analyses that were necessary to clarify analyte behaviour or other observations. The last section compares experimental results using Method 8261A (multiple internal standards per analyte) and Method 8260C (single internal standard per analyte) quantitation procedures in order to evaluate the impact of their differing internal standard roles.

3.1 *Method 8261 accuracy*

Method 8261A includes reporting an analytical error with each analyte result. The study began with evaluating if the reporting errors were consistent with random errors of measurements. For random error measurements, the frequency that one, two, and three confidence intervals include a true value follows the experimental rule (68, 95 and 99.7%) and systematic errors would cause the average result to be biased. To evaluate consistency of Method 8261A errors with being random, the errors were analysed by analyte, matrix type and amount, and concentration. Initially only the high and mid-level spikes were investigated and those samples spike at the limit of quantitation (LOQ) were evaluated separately. This resulted in evaluating 12 results when combining the high and mid-level results or only 6 when looking at a single spike level. Therefore, due to the small number of results, the frequency that the reporting-error confidence interval includes the true value

Table 3. Summary of accuracy of reporting error confidence intervals for determinations by matrix.

					Results that meet criteria						
						Confidence intervals[c] (%)			Recovery[d] (%)		
Matrix	Results[a]	Fail mixing criteria	Fail calibration or blanks criteria	Removed by surrogate or internal standard criteria	Number[b]	1	2	3	Avg	Dev	Comments
Sand	2808	320	61	76	2351	81.5	98.2	99.8	99.6	12.7	Hexachlorobutadiene[e]
NV	1512	152	80	12	1268	80.7	97.9	99.3	103.9	26.5	
GA	2268	26	111	221	1910	65.7	91.6	99.0	98.8	15.2	Carbon disulfide and gases[e]
OR	1800	54	100	84	1562	72.7	95.1	99.7	97.5	25.0	Dibromomethane[e]
All soils	8388	558	350	389	7091	75.2	95.7	99.5	99.7	19.6	
Soils at LOQ	4194	765	1662	42	1091	61.3	88.9	95.6	110.5	22.8	
Soils at LOQ	4194	1213	0	118	1091	72.9	96.1	98.9	108.2	23.3	15% min dev
Sediment 5–15g	2232	208	84	149	1801	68.2	89.0	93.7	107.1	21.2	6 outliers
Sediment 1g	510	0	0	0	510	91.9	98.9	99.2	101.2	12.4	1 outlier (chloroethane)

[a]The sum of all experimental determinations in the study by matrix after removing analytes with more than 5% of medium spike present in matrix before spikes.
[b]Number of determinations that meet criteria.
[c]The % frequency that a result and confidence interval include the known value at 1, 2, and 3 standard deviations.
[d]The average recovery of all analytes that meet criteria and one standard deviation.
[e]Compounds not included with results.

was compared to Chebyshev's rule [13]. Analytes that did not meet Chebyshev's minimum frequency values for both two and three standard deviation in a matrix were considered outliers.

On occasion, analytes would pass Chebyshev's rule criteria but their average concentration for a matrix was significantly different from the true value (systematic error). In this study, if an average result varied more than 25% from the true value, it was considered biased.

Method 8261A reporting errors were assessed after eliminating those results compromised by background or calibration defects. Analyte results were not used if that day's continuing calibration check difference exceeded 40%. Analytes were not used for a matrix when its background content in the matrix exceeded 15% of the mid-level spike amount for 15 g aliquots. Based upon a previous investigation of water samples, a minimum relative confidence interval of 6% was applied for all analyte reporting errors [12]. Determinations that the Method 8261 reports generating software, SMCReporter [14], qualified as above calibration or below calibration range were not included in evaluations.

It was an initial assumption that if an analyte was an outlier, there would be an observable cause that would also impact one of the QC parameters. Data were reviewed by matrix so that any severe effects related to a matrix would be detected. All of the determinations of an analyte for a matrix that failed Chebyshev's rule were examined to identify those analyses where the true value fell outside the three standard deviation confidence interval. The next step was to determine if these results were related to a matrix effect on a class (volatile, non-purgeable, or semivolatile compounds), or a matrix effect on subset of a class, concentration, or sample size.

Outliers that were identified were not uniquely linked to a quality control measurement. Typically the outlier condition could only be eliminated by requiring a stringent QC range that also disqualified the majority of analyses regardless of sample size or level of spike. It became evident that the mixing of soil and spikes, equilibration times, and analyte degradation were the major factors causing the observed outliers and these effects went undetected by surrogate or internal standard behaviour.

Ensuring internal standards, surrogates, or matrix spikes are in equilibrium with a soil matrix is not prescribed in Methods 8260C or 8261A. In this study however, attempts were made to add internal standards, surrogates and analytes to soil prior to adding water. Initially it was thought that the spiking of soil (before adding water) allowed sorption by the matrix to be complete and the vigorous boiling that takes place during vacuum distillation would distribute the spikes throughout the sample. However, distributing the spikes throughout the sample and uptake of the analytes by the soil both proved problematic. It was evident that there needed to be assurance internal standards, surrogates, and analytes, had been distributed throughout the soil. When spikes were not distributed throughout the soil, results were inconsistent and analytes and their labelled analogues appeared to behave differently.

One way to detect poorly distributed spikes was through a comparison of recoveries of labelled internal standards and their natural analogues for each analysis. It was expected that if the recoveries of an analyte and its analogue were grossly different it would be due to insufficient mixing of the soil after spiking. The relative responses (sample response per mass unit/standard response per mass unit) would be identical if they were well distributed throughout the sample. The relative response of an unlabelled compound would be allowed to vary from the relative response of its labelled analog by a percentage (maximum variation limit). The relative responses of vinyl chloride (representing gases), bromobenzene (VOAs), 1,2-dichlorobenzene and 1,2,3-trichlorobenzene (semi-VOAs) were compared to the relative responses of their deuterium analogs with the maximum variation limits of 75, 50, 25 and 25%, respectively. When a maximum variation limit was exceeded for an analysis, those results for analytes in the respective class of compounds were not used in assessing accuracy.

The procedure for distributing the spikes throughout a sample became more rigorous with each matrix examined. The frequency of exceeding maximum variation limits decreased as mixing (after spiking soil before adding water) went from none (sand), manual shaking (NV), to mechanical mixing (GA and OR). There is not a QC measure in either Method 8260C or 8261A that indicates the adding and mixing of internal standards is complete. One way to assess mixing would be to add a compound such as 1,3,5-trichlorobenzene as a surrogate and its labelled analogue as an additional internal standard. Monitoring the recovery of both to ensure equivalence would identify when mixing was inadequate as long as the mixes were being added to samples separately and at different locations. However, in following studies, the use of a mechanical stirrer to mix the samples after being attached to the vacuum distiller demonstrated the maximum variation limits used in this study were easily met without losing the gas compounds.

The relative response of internal standards were monitored (the response of an internal standard in a sample divided by its response in the day's continuing calibration standard). A subset of Method 8261 internal standards was monitored for consistency as performed in Superfund's Contract Laboratory Program (CLP) methods [15]. The CLP limits on acceptable variation in the internal standards (as %) are presented in Table 1. If an internal standard relative response (high or low) corresponds with the occurrences of outlier analytes, it would be assumed to affect its class of compounds. The only relative responses of internal standards that were found to be related to outlier data was when a sample vessel seal leaked during a distillation impacting the non-purgeable class and gases.

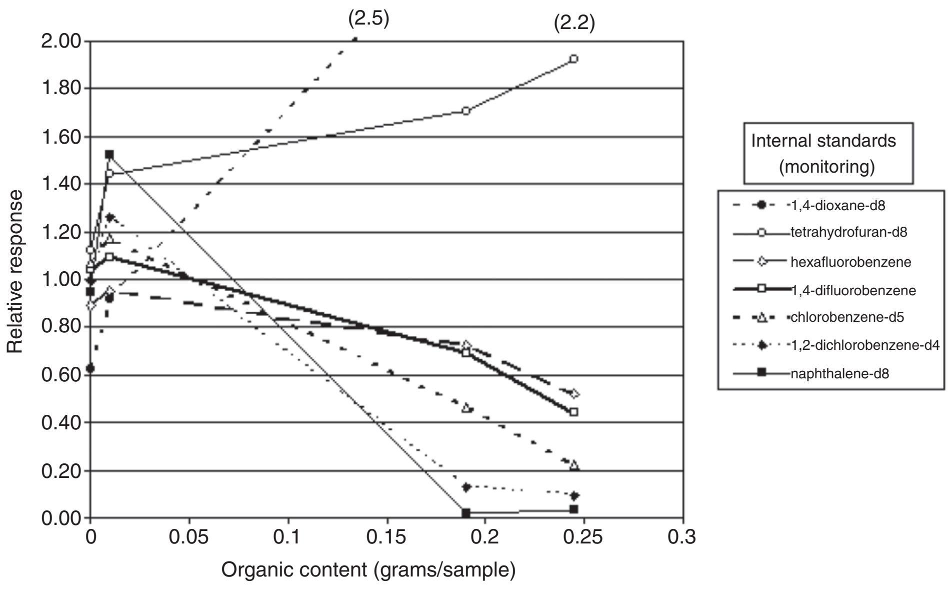

Figure 1. Relative response of internal standards vs. organic content.

In general, the relative response of the internal standards closely related to the organic content in the sample (Figure 1). For low-organic content soil samples, relative responses are greater than those for the day's continuing calibration standard. This is likely due to the vigorous boiling of water mixed with soil during vacuum distillation. As organic content increases, the internal standards responses decrease with the more lipophilic compounds decreasing the most.

The surrogate compounds, their recommended recovery ranges from Method 8261A and the range of experimental recoveries are listed in Table 2. The only surrogate that was linked to an outlier analyte was vinyl chloride-d_3. The surrogate ethyl acetate-^{13}C was observed to degrade quickly in some soils and this tendency was previously reported [11]. The sample that was not properly sealed for analysis (NV soil) had several surrogate recoveries outside their normal range (Table 2).

Three new compounds were evaluated for use as semivolatile surrogates. Recovery of azulene was nil for most analyses other than sand when applied directly to dry soil. The recovery of the other compounds (3,5-di-tert-butyltoluene and *a,a′*-dichloro-*o*-xylene) were at times much greater than 100% yet their extreme recoveries could not be linked to an outlier.

Two analytes were found to be outliers in one of the matrices but were not outliers when samples were analysed a day after spiking. Hexachlorobutadiene was bound tightly when spiked directly on sand with a little over 50% recovered if the analysis was within 3 hr of spiking. With an overnight equilibration time (or spiked after water added to sand) the compound recovery was $102 \pm 8\%$ (Table S4, Supplementary Material). Dibromomethane behaved similarly with the OR soil matrix (Table S7, Supplementary Material).

There was difficulty mixing the GA samples with the initial practice of mixing the samples external from the vacuum distiller resulted in losing the most volatile compounds. This loss was remedied with the improved mixing procedure (mechanical stirring of sample while container attached to vacuum distiller) and the early results for the gases were not included in accuracy determinations (Table S6, Supplementary Material). With the exception of mentioned compounds, the results were near experimental error expectations when spiking the soils before adding water (Table 3) suggesting the reporting error for most analytes is random. Therefore the confidence intervals reported with Method 8261A results are a good measure of the analytical error for spiked soils and sediments even when there are efforts to equilibrate spikes with the sample.

The same conditions were applied to the analyses of the LOQ-spiked samples. An additional criterion was that a result was not used when the analyte was also found to be in blanks at >½ the concentration of the LOQ spikes. The frequency these confidence intervals included the true values did not match theoretical prediction as closely as the mid-concentration spikes. This is likely to be due less accurate integrations or background contributions for measuring responses at the lowest calibration point. It is interesting to note that by raising the minimum one standard deviation confidence intervals to 15% as was done in the previous study [12], the confidence intervals included the true values near empirical rule frequencies (Table 3).

3.2 *Additional tests to clarify observations*

Verification of degradation, evaluation of spiking soils after addition of water and evaluation of impacts for extending equilibration times were performed at the high spike level (1/2 upper point of calibration range Table 3) with 5 g sample amounts. Two to six replicates were analysed and calibration was performed at the same amount as the sample spiking. Using the high spike and a smaller sample size minimised impact from background.

Degradation of analytes was determined by analysing samples spiked with analytes, surrogates, and internal standards and held at room temperature for extended times. Many analytes were found to degrade in a matrix over a 120-hour period and these analytes were not used in the evaluation for a given matrix (Table S3, Supplementary Material).

There was also an assessment of how standard practice of adding spikes with or after adding water to soil matrix impacted results. These samples were analysed the same day. The recovery of the non-polar internal standards was generally greater than when the internal standards were added to the soils before adding water (Table 1). There was not a lot of difference in the range of surrogate recoveries for when adding compounds to dry soil and adding them after the water was added (Table 2). The analyte results by matrix (Tables S4–8, Supplementary Material) did not indicate that the 'wet spike' analyses were less accurate. Of course, accurate analyte results may only occur when analytes are present in a soil sample through wet spiking, and not already present in the soil. Understating of analyte concentrations should be lessened if the internal standards and surrogates were at equilibrium with the sample matrix.

Overnight equilibration (averaging near 20 hrs) of the soil/water/spike slurry yielded results similar to when spikes were added to soils before adding water, but some degradation of analytes became evident. Extending the equilibration to 120 hours demonstrated even more sorption of the internal standards indicating the equilibration of spikes and matrices could be a slow process for the higher boiling compounds like 1,2-dichlorobenzene (Figure 2). In general, the sorption of volatile compounds by soil through a quick process and a slow process is similar to fast and slow sorption rates observed in the environment [4,16]. Relying on only a quick equilibration (e.g. analyses on

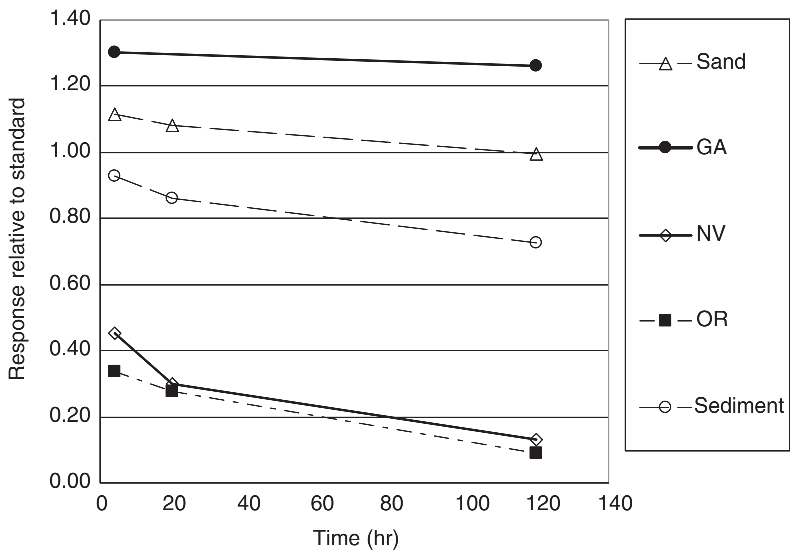

Figure 2. Relative response of 1,2-dichlorobenzene-d_4 vs. equilibration time.

day of spiking) for internal standards understates the concentrations of analytes that reach equilibrium through the slower sorption processes (and at equilibrium with the matrix).

Figure 2 illustrates the sorption of the internal standard, 1,2-dichlorobenzene-d_4, in the different matrices from the wet spike with increasing equilibration times. For the OR matrix, the response of 1,2-dichlorobenzene-d_4 drops 66% over the first 3 hrs, drops another 6% overnight, and another 19% to the 120 hr endpoint where its response is 9%. If the 120 hr endpoint reflects how analytes would be taken up by soil through exposure at the site where the soil was collected, then there is a potential to dramatically overstate the recovery of analytes and therefore understate their true concentration in a soil. For instance, if 1,2-dichlorobenzene was present in the OR soil from a site exposure, same-day spiking with internal standards and analysis would only reflect 27% of the true concentration. Figure 2 shows only 9% of the compound would be released during analysis (at equilibrium) but the spiking of internal standards on the day of analysis indicate 34% of the compound was recovered (9/34 = 27%). Overnight equilibration improves the determination (33% of true concentration) but spiking soil before adding water was a better approach (47% of true concentration from Supplementary Material Table S9).

The range of relative responses (recovery) for internal standards found experimentally (Table 1) has application to other concentration methods that are less efficient [10]. That is, the more effective the interaction of internal standards with soils (with elevated organic content), the more likely an analysis will fail the lower limits of internal standard relative response. If Method 8260C with purge and trap concentration (Method 5035) were equivalent to vacuum distillation concentration, then adding internal standards to the NV soil prior to dilution with water would cause the analyses to consistently fail criteria for Superfund's Contract Laboratory Program [15]. However, by adding internal standards to soil after adding the water and then analysing quickly to minimise equilibration time an analysis is more likely to pass criteria. This produces a dilemma where enforcing a QC parameter encourages a practice of under reporting analyte concentrations. To address this issue, it is recommended that the time between adding internal standards to soil and analysis is reported along with the whether the internal standards were added to soil after or prior to adding water.

The sediment matrix was not included as a soil in summarised data. Even though surrogate recoveries for sediment samples consistently fell within the soil surrogate ranges recommended in Method 8261A for soils, there were still numerous outlier analytes (Table 3) as well as biased results (Table S3, Supplementary Material). Taking a smaller amount of sediment (1.0–1.7 g) for analyses improved results (Table 3) and surrogate recoveries met the surrogate recovery ranges found for water in a previous study [12]. Selecting a smaller sample size also generated less biased analytes (3 biased analytes) than for 5 g samples (12 bias analytes). The analyses of a smaller sample size yielded reporting errors that were much more consistent with random measurement errors than 5–15 g soil samples (Table 3). If taking a smaller sample size does not jeopardise meeting sensitivity requirements or does not result in a nonrepresentative sample, analysing a smaller sample size so that surrogates meet recovery limits found previously for water would be preferred.

As the level of organic content increases, so does the potential for understating results when internal standards and the sample are not equilibrated. Because both fast and slow sorption of internal standards appear to be closely related to organic content, the knowledge of organic content in soil can be useful in interpreting data (high organic content and high recovery of internal standards likely indicate spikes are not equilibrated with matrix and therefore results likely understated).

3.3 *Method 8260C compared to Method 8261A*

Method 8260C with a vacuum distillation concentration step (Method 5032) differs from Method 8261A (method specifically incorporates vacuum distillation as a concentration option) in the role of internal standards to quantify an analyte. The battery of internal standards used in Method 8261 is used to parse matrix effects into relationships relating to boiling point and relative volatility and the analyte is quantified as a function of its boiling point and relative volatility. Method 8260C uses one internal standard for quantitation of an analyte and assumes the analyte behaves the same as its internal standard. If an analyte recovery from a matrix does not behave as the internal standard, there can be a significant bias in determinations.

The same analyses (six samples of each study matrix) used to document optimised spiking and mixing conditions were used to compare the methods. The raw analyte responses of theses analyses (originally quantified using Method 8261A) were re-quantified using Method 8260C. The analytical conditions would not ensure equilibration of spikes with matrices but would be more representative of current analytical laboratory spiking practices. It would be expected that the more thorough the spike equilibration, the more Method 8261A results would be superior to those of Method 8260C.

The occurrence of analyte bias was the means to compare the methods. The power of using the battery of internal standards per analyte in Method 8261A over the single internal standard per analyte used by Method 8260C is demonstrated by the number of analytes that had average results that differed by more than 25% from true values (Table S3, Supplementary Material). Method 8261A had six analytes differing by more than 25% from true values (7% of analytes) while Method 8260C had 20 analytes (24% of analytes). The average recovery of all Method 8261A results was nearer ideal at $100 \pm 20\%$ compared to $112 \pm 37\%$ for Method 8260C. Also the range of recoveries was generally greater for the method 8260C surrogates compared to Method 8261A (Table 2). The Method 8261A internal standard corrections were superior to the Method 8260C internal standard use resulting in a 70% decrease of the occurrence of results differing more than 25% from true values.

More specific information is available as a supplement to this article. Contained in the supplement are each analyte's results by matrix, average recovery and frequency that each analyte's confidence interval included the true value, and identification of those analytes impacted by continuing calibration limits and minimum confidence intervals. Please see the Supplementary Material available online.

Acknowledgements

A note of appreciation to John Zimmerman and Brian Schumacher of NERL-Las Vegas/ORD/EPA for providing the performance soils used in the study and to William Nelson of NERL-Narragansett/ORD/EPA for providing the Greenwich Bay sediment.

The US Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), funded and performed the analytical research described. This manuscript has been subjected to the EPA's review and has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

[1] W.S. Ingersoll, Environmental Analytical Measurement Uncertainty Estimation-Nested Hierarchical Approach (Defense Technical Information Center, #ADA396946, 2001).
[2] Environmental Laboratory Sector, Quality Systems: General Requirements, Vol. 1, Module 2 (NELAC, Weatherford, TX, 2007).
[3] S.M. Steinberg, J.S. Schmeltzer, and D.K. Kreamer, Chemosphere **33**, 961 (1996).
[4] J. Pignatello and J. Xing, Environ. Sci. Technol. **30**, 1 (1996).
[5] A. Hewitt, Am. Environ. Lab. **6**, 6 (1994).
[6] M.J. Minnich, J. Zimmerman, and B. Schumacher, J. Environ. Qual. **26**, 108 (1997).
[7] K.M. Maney, C.M. Davidson, and D. Littlejohn, Analyst **123**, 195 (1998).
[8] Environmental Protection Agency, Office of Solid Waste SW-846 Methods Manual (New Test Methods) Method 8260C, August 2006: http://www.epa.gov/epawaste/hazard/testmethods/sw846/new_meth.htm#8260C.
[9] Environmental Protection Agency, Office of Solid Waste SW-846 Methods Manual (New Test Methods) Method 8261A, October 2006: http://www.epa.gov/epawaste/hazard/testmethods/sw846/new_meth.htm#8261A.
[10] J.W. Hines and E.D. Pellizzari, Research Triangle Institute, 'Evaluation of Analytical Test Procedures for the Measurement of Organic Pollutants in Sediment', 1983. Unpublished report.
[11] M. Hiatt, Anal. Chem. **67**, 4044 (1995).
[12] M. Hiatt, Accred. Qual. Assur. Accreditation **13**, 247 (2008).
[13] F. Mosteller, R. Rourke, and G. Thomas, Probability with Statistical Applications (Addison-Wesley Publishing, Reading MA, 1961).
[14] EPA, Office of Research and Development, National Exposure Research Laboratory, SMCReporter 4.2 December 2008: http://www.epa.gov/nerlesd1/chemistry/vacuum/methods/software-pc.htm.
[15] EPA, Office of Solid Waste and Emergency Response, Multi-Media, Multi-Concentration Organics Analysis, SOM01.1/SOM01.2, June 2007: http://www.epa.gov/superfund/programs/clp/som1.htm.
[16] Th.E.M. Ten Hulscher and G. Cornelissen, Chemosphere **32**, 609 (1996).