

Quantifying heterogeneity of small test portion masses of geological reference materials by PXRF: implications for uncertainty of reference values.

Peter D Rostron and Michael H Ramsey
University of Sussex, Falmer, Brighton, United Kingdom BN1 9QG

There is an increasing use of analytical macro-beam techniques (such as portable XRF, PXRF) for geochemical measurements, as a result of their convenience and relatively low cost per measurement. Reference materials (RMs) are essential for validation, and sometimes calibration, of beam measurements, just as they are for the traditional analytical techniques that use bulk powders. RMs are typically supplied with data sheets that tabulate uncertainties in the reference values by element, for which purpose they also specify a minimum recommended mass of material to be used in the chemical analysis. This minimum mass may not be achievable using analytical beam techniques. In this study, the mass of the test portion interrogated by a handheld PXRF within pellets made from three silicate RMs (SdAR L2, M2 and H1) was estimated using a theoretical approach. It was found to vary from 0.001 g to 0.3 g for an 8 mm beam size, and 0.0001 g to 0.045 g for a 3mm beam. These test portion masses are mainly well below the recommended minimum mass for these particular RMs (0.2g), but were found to increase as a function of atomic number (as might be expected). The uncertainties caused by heterogeneity (U_{HET}) in PXRF measurements of the three RMs were experimentally estimated using two different beam diameters for 18 elements. The elements showing the highest levels of heterogeneity ($U_{\text{HET}} > 5\%$) seem generally to be those usually associated with either an accessory mineral (e.g. Zr in zircon, As in pyrite) or low test portion mass (associated with low atomic number). When the beam size was changed from nominally 8 mm to 3 mm, the uncertainty caused by heterogeneity was seen to increase for most elements by an average ratio of 2.2. These values of U_{HET} were used to calculate revised uncertainties of the reference values that would be appropriate for measurements made using a PXRF with these beam sizes. The methods used here to estimate U_{HET} in PXRF measurements have a potential application to other analytical beam techniques.

Introduction

A range of portable ‘macro-beam’ analytical devices have been developed in the last ten years. They typically have beam size in the *cm* range, are capable of making quantitative field measurements, and include FTIR, VIS-NIR and PXRF (Chen et al 2015, Horta *et al.* 2015). One example of this trend towards acquiring rapid measurements in the field is an increasing use of portable X-ray fluorescence (PXRF) instruments (for brevity, the acronym PXRF has been used to represent “hand-held” XRF throughout the text). The advantages of minimal sample preparation, relatively low start-up costs, and the ability to acquire measurements on site has led to an increased use of PXRF instruments by environmental consultancies and other organisations with lower levels of analytical expertise than traditional geochemical laboratories. Technical advances in instrumentation now enable quantitative measurements by PXRF of many elements, including trace elements, particularly when these can be supported by measurements of certified reference materials (CRMs) and cross-validation with other measurement methods (Parsons *et al.* 2013).

Reference materials (RMs) are primarily used to estimate the bias of measurements made by an analytical method, frequently as part of the validation of that method and to establish traceability. This approach works well when the mass of the test portion used in the analytical method exceeds a minimum specified on the certificate of analysis or reference material information sheet. A problem can arise when the analytical method specifies or interrogates a test portion that is smaller than that minimum. Contemporary PXRF instruments use nominal X-ray beam diameters in the region of 10 mm or less. The mass of the test portion interrogated depends on the penetration depths of the primary and fluorescent X-rays, as well as the density and attenuation properties of the matrix. For elements with low X-ray penetration depths the mass of the test portion interrogated by PXRF may drop well below the mass that would be used in a laboratory-based analytical method.

The reference values of concentration quoted for all analytes in reference materials also have a quoted uncertainty value. When combined with the measurement uncertainty obtained during validation measurements this effectively expresses the limited range of measured values around the reference value that can arise without indicating significant bias in the analytical method. The value of that uncertainty is derived from statistical analysis of the many measurements made at the time of the characterisation of the RM (ISO Guide 35 2006). It includes contributions from sources such as the analytical variability within each contributing lab, bias between the different labs, bias between the different analytical methods used, and the heterogeneity of the RM, both within-bottle and between-bottle (Kane *et al.* 2003). Every candidate RM goes through a 'homogeneity test' to ensure that it is 'sufficiently' homogeneous within-bottle and that there is no significant extra inhomogeneity between different bottles of that RM. This latter sufficiency only applies when the test portion mass exceeds the minimum value stated on the RM data sheet. When an analytical method, such as PXRF, interrogates a test portion mass that is less than the specified minimum, then 'sufficient homogeneity' cannot be assumed.

There is a precedent for calculating and publishing enlarged uncertainty values for small test portion masses that occur due to reduced beam sizes (Jochum *et al.* 2011). That work lists tables of uncertainties on reference values estimated for very low test portion masses, e.g. 0.02 – 1.0 µg for glass certified reference materials (NIST SRM 610-617). These materials are used in microbeam-analytical methods for procedures such as validation, calibration and quality control, but had previously been described as heterogeneous for some elements (Eggins and Shelley 2002). For PXRF, the effective test portion mass measured in an individual measurement is not immediately obvious but may well be less than the minimum recommended on the RM data sheet. This would be expected to result in increased uncertainties in the reference concentration values, compared to those quoted for 'bulk' analytical techniques on the data sheet.

The principle objective of this current work is to investigate the magnitude of the contribution of heterogeneity to the uncertainty in elemental concentrations within reference materials, when small test portion masses are measured as part of a verification or calibration process. More specifically, this heterogeneity component of the uncertainty needs to be quantified for measurements made when a contemporary PXRF unit is used with reduced beam sizes. This issue is likely to become increasingly important as analytical

technology is further miniaturized, with possible further decreases in beam size, particularly when measurements are made on materials that are increasingly heterogeneous at smaller scales.

The SdAR series of blended RMs used to exemplify this approach has been prepared by the US Geological Survey and characterized by the International Association of Geoanalysts (IAG). They are designed to resemble soils or sediments with different levels of environmental contamination from mining operations. They range from low contamination levels (SdAR-L2, Webb *et al.* 2015a) to medium contamination levels (SdAR-M2, Webb *et al.* 2015b) to high contamination levels (SdAR-H1, Webb *et al.* 2014). They will subsequently be referred to as L2, M2, and H1. The data sheets for all three SdAR RMs recommend a minimum test portion mass of 0.2 g for analysis.

Analysts who use analytical methods such as XRF or PXRF might be measuring test portion masses of reference materials that are below the recommended minimum. In these cases approaches are needed that enable evaluation of the significance of analytical bias in measurements on these smaller test portions. For example, the quoted reference values and uncertainties (U_{RV}) of the SdAR materials used in this study were originally derived from the results of a proficiency test *GeoPT*¹. The design of that PT makes it impossible to quantify the within-lab repeatability, and hence separate some of the other key components of this uncertainty, such as the between-lab bias and between-method bias. There is a requirement therefore to devise a new method to estimate U_{RV} for small beam sizes (e.g. $U_{RV \text{ beam}} = U_{RV \text{ 8 mm}}$). One potential approach adds the uncertainty caused by the heterogeneity of the RM for small beam sizes ($U_{HET \text{ beam}}$) to the published values of U_{RV} . This approach assumes that there is no significant contribution from heterogeneity ($U_{HET \text{ bulk}}$) within the quoted value of U_{RV} for test portion masses greater than the quoted minimum.

Detailed Objectives

1. Estimate theoretically the mass of the test portion interrogated by PXRF, and comment on its relationships with the atomic number of the analyte, and with the uncertainty due to analyte heterogeneity.
2. Estimate experimentally the component of measurement uncertainty that is caused by within-bottle heterogeneity (U_{HET}) for a number of elements in three silicate RMs. Compare this uncertainty when two different PXRF primary beam diameters of 8 mm and 3 mm are used, and test whether the observed differences are statistically significant.
3. Estimate the uncertainty of the reference value of the RMs for each element when analyzed using these small test portion masses, when these masses may be below the specified minimum (in this case 0.2 g). Consider whether these values are applicable when these RMs are prepared in different ways, analyzed by different calibration methods or on different manufacturers' instruments.

Theoretical estimates of test portion mass

As previously stated, the minimum test portion mass recommended on the SdAR data sheets is 0.2 g. The nature of PXRF measurements means that in many cases the test portion analyzed may be much lower than this figure. It is possible to make an estimate of the

maximum (critical) depth within a notionally infinitely thick sample that will make a significant contribution to the measured signal. If the test portion is assumed to be a regular shape, the critical depth can be used to further estimate the mass of the test portion.

Estimation of the critical depth (x) has been described in detail (Potts, 1987).

$$x = 4.605 / \rho \cdot \mu_{net} \quad \text{Equation 1}$$

where ρ = density of the sample, and μ_{net} is the net bulk mass attenuation coefficient for the primary X-ray beam and the fluoresced X-rays, taking into account the angles of the X-ray tube and spectrometer.

The basis on which **Equation 1** is derived is the penetration distance from the sample surface which will contribute 99% of the measured signal, if the sample were assumed to be infinite in extent and assuming homogeneity of analytes and attenuation properties.

The assumption made in these calculations is that the primary X-ray source can be represented by a single energy level (22.16 KeV) corresponding to the $K\alpha$ emission line from the silver anode used in this PXRF. For the purposes of these approximations incident tube and detector take-off angles have been assumed to be 45°. These may differ between different PXRF models, but are often considered proprietary information.

The volume of the test-portion will be a complex shape. The approach used has been to assume that the test portion volume will be larger than a cone (with height equal to the critical depth, and diameter equal to the nominal diameter of the primary radiation beam), but smaller than a cylinder with the same dimensions. Estimates of test-portion volume were made for each element in each RM by averaging the volumes of these two solid shapes. These were converted to test-portion masses using the measured densities of the pellets (or powder).

Estimates of test portion mass were found to vary from 0.001 g (Al) to 0.32 g (Ba) for the 8 mm beam, and from 0.0001 g (Al) to 0.045 g (Ba) for the 3mm beam. Individual estimates, which take into account the mineral composition of individual RMs, were found to vary by 1 - 17% between different RMs. Average estimates of test portion masses across the three SdAR RMs for each element are shown graphically in **Figure 1**. It is clear that most elements are likely to have much lower test portion masses than the minimum recommendation of 0.2g. Only the element Ba shows an estimated mass equal or greater to 0.2 g when the 8 mm primary beam diameter is used. The test portion masses for all elements are well below this figure when the 3 mm beam is used. The estimated test portion mass is clearly generally a function of the atomic number of the element (**Figure 1**), as would be expected theoretically, with the exception of Pb (not shown in **Figure 1**). This is because the identification and quantification of Pb by this PXRF uses an $L\beta$ line whereas all the other elements shown are identified by $K\alpha$ lines.

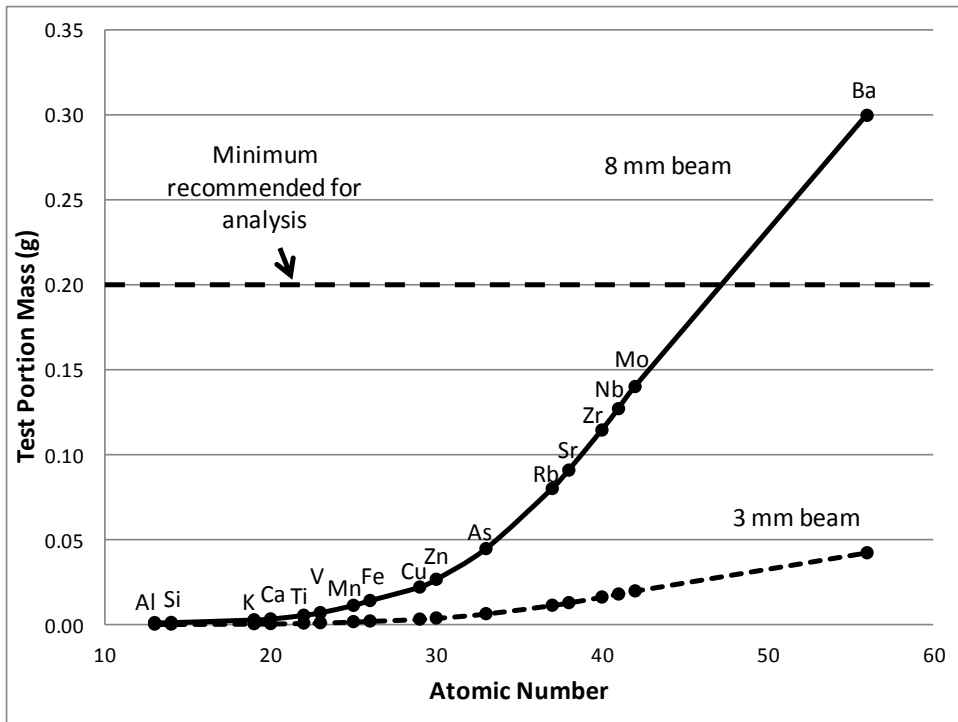


Figure 1. Estimates of test portion mass by atomic number (excluding Pb), averaged for the three SdAR RMs.

Experimental method

Six pellets of 13 mm diameter and 7 mm thickness were made from each of the 3 SdAR RMs, by mixing approximately 2.2 g of each RM with a few drops of 2% PVA solution and pressing to a pressure of 8 tonnes. For comparison, 3 powder samples of each RM were also prepared by placing approximately 12 g of material into a 25 mm x 25 mm closed-cell sample cup, using a sample-retaining film made of 4 μm gauge polypropylene.

Measurements were made using a Portable X-Ray Fluorescence spectrometer, model Niton XL3t Ultra PXRF. Measurements were carried out with the equipment set in Niton's "Mining Mode". This mode includes 4 filter settings and allows both 8 mm and 3 mm nominal beam sizes to be used. The XL3t uses a silver X-ray tube which can operate at up to 50 keV, allowing discrimination of all of the subsequently reported elements from K_{α} lines in the spectra, with the single exception of Pb which was identified using L_{β} lines (Caria and Hurley, 2016). In order to produce comparable limits of detection, measurement times for the 3 mm (216 s per filter) beam were increased by a factor of 3.6 above those for the 8 mm beam (60 s per filter). The factor 3.6 was designed to equal out the signal for the 8 and 3mm beam diameters over a wide range of elements (Caria and Hurley, 2016).

Prior to beginning the 8 mm beam and 3 mm beam measurement sets, the manufacturer's fundamental parameter (FP) calibration was adjusted using two measurements made on each side of a single pellet (or in two different positions on a powder sample) for each of the three SdAR materials. These measurements were plotted against the quoted reference or information values, in order to derive a manual slope adjustment factor that could then be applied to the calibration for each element in the PXRF software.

Each of the 6 pellets, for each SdAR material, was designated a code A-F. Both sides of each pellet were measured in the sequence L2A-M2A-H1A...L2F-M2F-H1F. This structured sequence design has some disadvantages compared with a quasi-random design, but has the overall advantage that it minimizes the effects of any instrumental drift within the replicated pellets and between the three different SdAR types, and is also far less likely to result in a recording error. The analyte concentrations were measured on each side of the pellet, and each of these measurements was immediately duplicated without moving the pellet, resulting in a total set of 72 measurements per beam size. An additional pellet was made from each of three National Institute of Standards and Technology (NIST) reference materials (coded SRM 2709, 2710 & 2711). These CRMs were measured with a single measurement on each side, and used for the estimation of bias against the certified values of concentration, and also for monitoring instrumental drift.

In the case of the powder samples, the 25 mm diameter sample cups were rotated by 90 degrees between each duplicated measurement. Measurements were made towards the outer edge of the sample cup to avoid overlap, also resulting in a total set of 72 measurements for each beam diameter.

The components of variance caused by heterogeneity within the pellets/powders, and analytical repeatability between duplicated analyses, were separated by analysis of variance (ANOVA). This was achieved using the computer program RANOVA2 (AMC 2014). In its simplest mode this gives ANOVA results for a 2 x 2 balanced design comprising 2 samples and 2 analyses as described in the Eurachem Guide (Ramsey and Ellison 2007). The following variances are calculated: a) between-target; b) between duplicated samples; c) between duplicated analyses. In these experiments just two components of variance are of interest, that due to heterogeneity within the pellets, and that due to analytical repeatability between duplicated analyses. In order to adapt the balanced design to this purpose, each measurement value was replicated, corresponding to the two analyses in the balanced design. The actual duplicated analyses then corresponded to the two samples in this design. It was then possible to use RANOVA2 with a total of 12 'sampling targets' (6 pellets on 2 sides, or 3 powder cups in 4 positions) in order to perform the ANOVA on each element in each SdAR type.

Results

Analytical repeatability

Analytical repeatability (U_{ANA}) is the random component of uncertainty caused by short-term variations in instrument response between measurements of the same quantity. It needs to be quantified and separated from the total variability to enable the estimation of uncertainty caused by heterogeneity. It was estimated by duplicating each measurement without moving the pellet between duplicate measurements. The standard deviation component in the analytical repeatability can be estimated using ANOVA. U_{ANA} %, with a coverage factor of 2, is then calculated:

$$U_{ANA} \% = \frac{200 \times \text{Standard Deviation of Analytical component from ANOVA}}{\text{Mean of measurements for the element/RM}} \quad \text{Equation 2}$$

Expressing uncertainty as a percentage of the mean measurements allows comparisons to be made between the uncertainties on different elements, and against external criteria. Analytical uncertainties for each element/RM combination are shown in **Table 1**.

Table 1. Expanded relative analytical uncertainties (2 * relative standard deviation) for 8 mm and 3 mm beam diameters. Uncertainties shown in bold are below 5%.

Element- SdAR ID	$U_{ANA}\%$ 8mm	$U_{ANA}\%$ 3mm	Element- SdAR ID	$U_{ANA}\%$ 8mm	$U_{ANA}\%$ 3mm
AL2O3-L2	2.1	2.7	Cu-L2	14.4	11.8
AL2O3-M2	2.3	2.4	Cu-M2	3.4	5.3
AL2O3-H1	3.0	2.8	Cu-H1	1.8	2.0
CaO-L2	2.4	1.1	Mo-M2	6.1	12.4
CaO-M2	2.5	1.7	Mo-H1	3.2	2.6
CaO-H1	1.5	0.5	Nb-L2	2.2	2.2
Fe2O3-L2	0.6	0.8	Nb-M2	5.0	3.6
Fe2O3-M2	0.8	0.8	Nb-H1	5.3	5.0
Fe2O3-H1	0.8	0.4	Pb-L2	3.2	3.3
K2O-L2	1.0	0.8	Pb-M2	1.4	1.2
K2O-M2	0.8	0.7	Pb-H1	1.1	0.7
K2O-H1	1.0	0.5	Rb-L2	1.5	2.1
MnO-L2	6.5	5.6	Rb-M2	1.7	2.1
MnO-M2	6.5	3.1	Rb-H1	2.7	1.9
MnO-H1	2.5	2.1	Sr-L2	1.4	1.3
SiO2-L2	0.5	1.1	Sr-M2	1.5	2.0
SiO2-M2	0.5	0.8	Sr-H1	1.3	1.4
SiO2-H1	0.6	1.2	V-L2	17.7	9.9
TiO2-L2	2.0	1.2	V-H1	15.1	7.2
TiO2-M2	2.9	1.8	Zn-L2	4.7	3.2
TiO2-H1	2.6	0.9	Zn-M2	1.9	1.9
As-M2	23.0	18.0	Zn-H1	1.1	0.8
As-H1	11.4	8.6	Zr-L2	0.6	0.9
Ba-L2	3.3	5.6	Zr-M2	1.6	1.3
Ba-M2	3.3	5.1	Zr-H1	1.1	1.2
Ba-H1	3.8	5.0			

The $U_{ANA}\%$ value is low (i.e. < 5%) for the majority of elements (**Table 1**), but for some element/RM combinations (e.g. both SdARs for As) it can be much higher than 5%. In these few cases the estimates of elemental concentrations may be less reliable than when a low analytical uncertainty is achieved. The main reason for high values of $U_{ANA}\%$ appears to be the proximity of the measured value to the limit of detection (LOD) of the instrument. For 12 out of the 21 element/RM combinations with $U_{ANA}\% > 5\%$, the reference value is below 10 times the LOD quoted by the instrument manufacturer. Of the remaining 9, most have $U_{ANA}\%$ values that only just exceed 5%, e.g. Ba values of $U_{ANA}\%$ for the 3 mm beam size. The notable exception is As-M2 and As-H1 where the analytical uncertainty significantly exceeds 5%, despite the average measured values being well above LOD.

There are also some significant differences between $U_{ANA}\%$ (8 mm) and $U_{ANA}\%$ (3 mm), for example Mo-M2 increases from 6.1% (8 mm) to 12.4% (3 mm). Overall, there is no general tendency for $U_{ANA}\%$ to increase or decrease with beam diameter. However the elements

measured by the nominal light (no-filter) setting on the Niton PXRF (Al, Si) and the high filter setting (Ba) do show a moderate general increase for the 3 mm beam size. This may suggest that the increase in counting time for the 3 mm measurements by a factor of 3.6 (refer to **Methods** section) may not have been quite sufficient for these particular elements.

Analytical bias

The bias is the systematic component of uncertainty and was estimated by comparing measured values against the certified values of the NIST CRMs. This allowed an independent and traceable estimate of the bias, because the NIST CRMs had not been used in the calibration adjustments. While it is important to estimate systematic bias in normal practice, low bias values would be expected to have minimal impact on the estimates of the random component of uncertainty caused by heterogeneity. Bias tables have been included as they are part of analytical quality control and support the traceability of the measurements.

Bias and Relative Bias have been calculated as follows:

Bias = Average SdAR measured value – Reference value

$$\text{Relative bias (Bias\%)} = \frac{100 * \text{Bias}}{\text{Reference value}} \quad \text{Equation 3}$$

Estimates of analytical bias are summarized in **Table 2**. The majority (72%) of the relative bias values are below 10%. In many cases where the relative bias exceeds this figure, e.g. Pb, As, Mo in 2709, V in all NISTS, the certified values are less than 10 * LOD and the numerical bias values in mg kg⁻¹ are relatively low. In others, e.g. Ca, Mn in NIST 2709, the certified values are outside the ranges of the SdAR reference values that were used in the initial calibration adjustment.

Table 2. Relative biases (expressed as percentages) made using NIST CRMs, with certified concentration values (in mg kg⁻¹) for comparison. Values of bias in bold are lower than 10%. Note certified values marked with a “*” are non-certified or information values.

	Certified value mg kg ⁻¹			8mm diameter beam			3mm diameter beam		
	2709	2711	2710	2709	2711	2710	2709	2711	2710
Al	75000	65300	64400	-2.0	1.8	0.9	-0.7	6.8	7.0
Ca	18900	28800	12500	-9.4	-10.1	-15.2	-11.0	-9.5	-15.0
Fe	35000	28900	33800	1.2	-0.9	4.4	2.2	1.6	5.2
K	20300	24500	21100	-5.4	-2.5	-0.7	-5.2	-1.6	-0.5
Mn	538	638	10100	-12.6	-4.7	1.4	-13.9	-1.5	7.4
Ti	3420	3060	2830	11.3	8.3	11.2	9.9	8.3	8.3
As		105	626		-8.6	25.6		-17.0	18.8
Ba	968	726	707	-7.3	0.7	9.6	-7.7	4.5	8.6
Cu	34.6	114	2950	-8.4	3.7	-1.0	-67.8	-11.1	1.7
Mo	2	1.6	19		2.8	1.1		-8.0	5.3
Nb		No CRM			No CRM			No CRM	
Pb	18.9	1162	5532	-29.9	-1.3	-4.8	-32.9	0.7	-3.8
Rb	91.3*	110*	120*	-1.1	2.9	10.0	0.1	3.0	9.4
Si	296600	304400	289700	-5.0	-1.2	1.2	-5.3	-0.2	3.0
Sr	231	245.3*	330	-6.1	-4.1	-5.2	-5.0	-3.4	-5.1
V	112		76.6	-30.5		-21.2	-27.4		-13.5
Zn	106	350.4	6952	-4.5	2.4	-3.8	-5.4	1.3	-3.3
Zr	160	230	No CRM	-10.0	19.0	No CRM	-17.1	11.4	No CRM

Uncertainties caused by heterogeneity of analytes in the RM pellets (U_{HET})

The component of variance caused by heterogeneity for each element/RM combination was calculated using classical ANOVA. This was then converted with a coverage factor of 2 to an expanded (2s) relative uncertainty $U_{HET}\%$:

$$U_{HET}\% = \frac{200 \times \text{Standard Deviation of Heterogeneity component from ANOVA}}{\text{Mean of all measurements for the element/RM}} \quad \text{Equation 4}$$

This component of uncertainty is shown for each of 51 element/RM combinations in **Table 3**. In a number of cases $U_{HET}\%$ has been calculated as zero. This could be the result of either an undetectable level of heterogeneity, or of a relatively high level of analytical uncertainty, which it has not been possible for the ANOVA to separate from the variation in heterogeneity, and so sets to zero. It is possible that the experimental sensitivity may not have been sufficient to adequately quantify the heterogeneity for these particular element/RM combinations. The evident variability in the values of $U_{HET}\%$ between different elements, and the particularly high values for CaO (12-25%) are discussed later.

Table 3. Expanded relative uncertainties (i.e. 2s%) due to heterogeneity in pellet of the SdAR materials for the 8 mm beam diameter.

Element-SdAR ID	$U_{HET}\%$	Element-SdAR ID	$U_{HET}\%$	Element-SdAR ID	$U_{HET}\%$
AL2O3-L2	4.75	TiO2-L2	4.50	Pb-L2	2.23
AL2O3-M2	5.10	TiO2-M2	5.90	Pb-M2	0.77
AL2O3-H1	5.63	TiO2-H1	3.76	Pb-H1	0.41
CaO-L2	15.98	As-M2	10.75	Rb-L2	0.70
CaO-M2	25.34	As-H1	0.00	Rb-M2	1.95
CaO-H1	12.09	Ba-L2	3.12	Rb-H1	0.00
Fe2O3-L2	3.07	Ba-M2	2.10	Sr-L2	0.65
Fe2O3-M2	1.86	Ba-H1	0.00	Sr-M2	0.55
Fe2O3-H1	0.37	Cu-L2	0.00	Sr-H1	1.45
K2O-L2	1.02	Cu-M2	4.80	V-L2	0.00
K2O-M2	1.62	Cu-H1	2.93	V-H1	9.39
K2O-H1	1.30	Mo-M2	9.85	Zn-L2	0.00
MnO-L2	3.20	Mo-H1	2.60	Zn-M2	2.50
MnO-M2	1.64	Nb-L2	1.27	Zn-H1	2.34
MnO-H1	2.02	Nb-M2	5.89	Zr-L2	8.84
SiO2-L2	2.09	Nb-H1	8.05	Zr-M2	6.60
SiO2-M2	2.06			Zr-H1	8.18
SiO2-H1	3.50				

Differences in uncertainties caused by heterogeneity (U_{HET}) for different beam sizes

In order to compare U_{HET} between different beam sizes, the $U_{HET}\%$ ratio is defined as:

$$U_{HET}\% \text{ Ratio} = \frac{U_{HET}\% \text{ 3mm}}{U_{HET}\% \text{ 8mm}} \quad \text{Equation 5}$$

This ratio can be used as a quantification of the change in $U_{HET}\%$ as the beam size of the PXRF is reduced from 8 mm to 3 mm. It is considered to be a significant increase if this ratio exceeds a critical ratio of 1.7. The derivation of the critical ratio is based on the square root of the critical F-ratio for 12 replicates for both beam sizes, and therefore 11 degrees of freedom. For the single sided hypothesis that the variance of the 3 mm measurement is *greater* than that for the 8 mm, the critical value of $F_{0.05(1),11,11}$ is 2.818.

We have

$$U_{HET}\% = \frac{200 \times \text{Standard Deviation of } \mathbf{Heterogeneity} \text{ component from ANOVA}}{\text{Mean of all measurements for the element/SdAR}}$$

Substituting and simplifying Equation 5 by cancelling out mean values top and bottom:

$$U_{HET} \text{ Critical Ratio} = \frac{s_{3mm}}{s_{8mm}} = \sqrt{F_{critical}} = \sqrt{2.818} = 1.7$$

Out of a total of 51 element/RM combinations, 41 (80%) show a nominal increase in the $U_{\text{HET}}\%$ ratio, while 21 (41%) of those with non-zero $U_{\text{HET}}\%$ ratios for both beam sizes show an increase that exceeds the critical ratio. This strongly suggests that heterogeneity of several elements results in a larger component of uncertainty when smaller beam sizes are used. This effect may not be quantifiable for all elements, or for all concentrations of the heterogeneous elements, due to dominance of the analytical uncertainty $U_{\text{ANA}}\%$ in some cases. An overview of the changes in $U_{\text{HET}}\%$ is graphically presented in **Figure 2**, after removal of element/RM combinations with zero estimates of $U_{\text{HET}}\%$ for one or both of the beam sizes. Only five out of the 41 element/RM combinations show a zero or negative change in $U_{\text{HET}}\%$, only one of which is statistically significant. This also suggests that there is typically an increase in the $U_{\text{HET}}\%$ Ratio when the PXRF beam size is reduced from 8 mm to 3 mm.

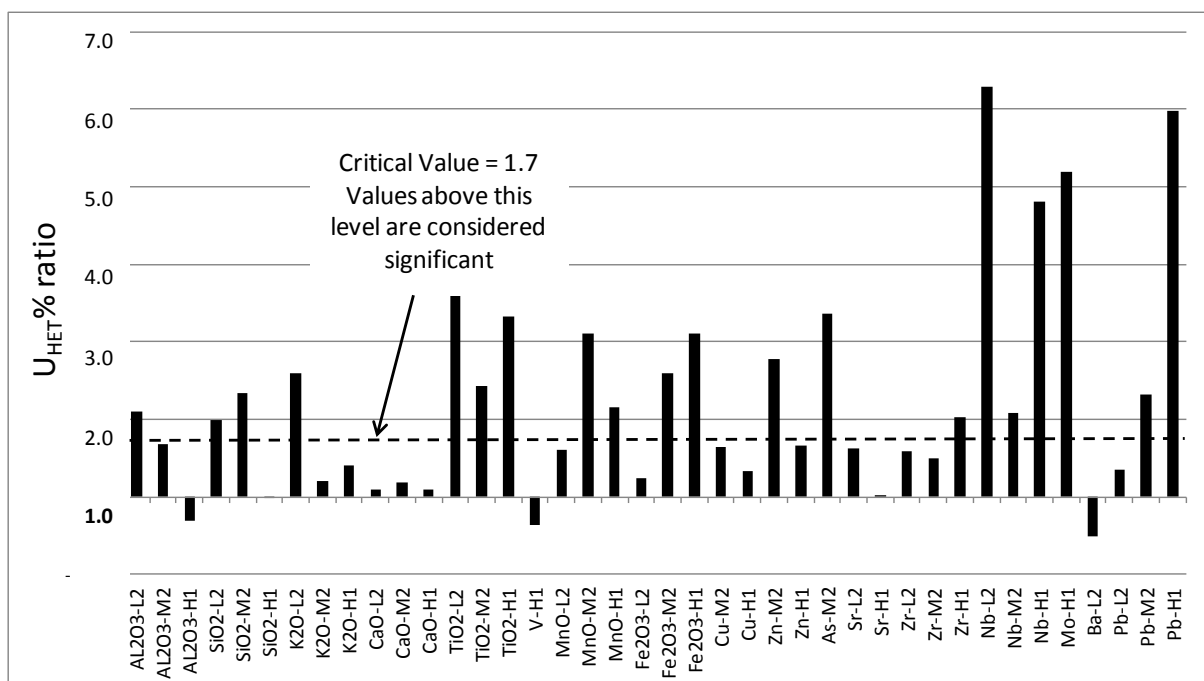


Figure 2. The change in heterogeneity when the beam size reduces from 3mm to 8mm, for elements in order of increasing atomic number. This $U_{\text{HET}}\% \text{ 3 mm} / U_{\text{HET}}\% \text{ 8 mm}$ ratio changes for 40 element/RM combinations, where U_{HET} was estimated at > 0 for both beam sizes.

It is possible to make an approximate prediction of the change in the random component of uncertainty caused by heterogeneity based on simple geometry, and sampling theory (Gy, 1979). An approximation of the relationship between uncertainty U , standard deviation of the sampling error s , and mass m of the test portion is given by:

$$U \propto s \propto 1/\sqrt{m} \quad \text{Equation 6}$$

Making an approximate assumption that the volume of the test portion is directly proportional to the radius r of the PXRF spot, we can say:

$$m \propto r^2$$

And

$$\frac{U_{3mm}}{U_{8mm}} = \frac{\sqrt{m_{8mm}}}{\sqrt{m_{3mm}}} = \frac{r_{8mm}}{r_{3mm}} \quad \text{Equation 7}$$

Making the further assumption that the beam sizes are exact, and with no edge effects, then the ratio $U_{\text{HET}}\% \text{ 3 mm} / U_{\text{HET}}\% \text{ 8 mm}$ could be predicted to be $4 \text{ mm}/1.5 \text{ mm} = \underline{2.7}$.

A histogram of the U_{HET} ratios with this predicted value shown is given in **Figure 3**. The majority of the $U_{\text{HET}}\%$ ratios are below the predicted ratio of 2.7 (Median = 2, Average = 2.2). This discrepancy may partly result from limitations in the assumption that the test portion volume is proportional to the radius of the primary X-ray beam. It will also be determined by other factors such as the angles of both tube and spectrometer, and the differential attenuation of primary and secondary X-rays on different path-lengths.

Note that although for the purposes of comparisons between heterogeneity for different elements the $U_{\text{HET}}\%$ ratio has been used (i.e. uncertainties relative to the mean measurement values) similar conclusions are arrived at if the ratio of U_{HET} values is used.

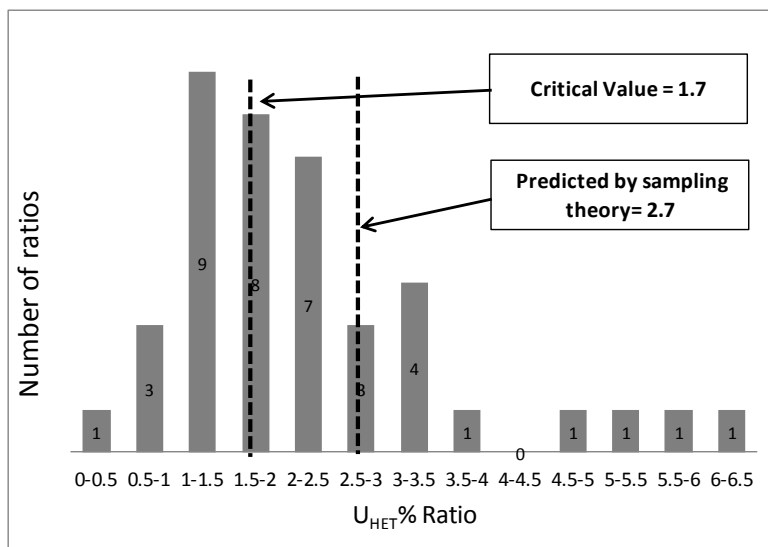


Figure 3. Histogram of $U_{\text{HET}}\%$ ratio for 40 element/RM combinations with non-zero values of $U_{\text{HET}}\%$. The predicted value of 2.7 has been estimated by a simplified geometry and an approximated relationship between uncertainty and the mass of the test portion modelled in sampling theory.

Effects of Instrumental drift

Instrumental drift was generally very small and had an insignificant effect on the estimates of heterogeneity. However, some drift was apparent for elements Al and Si, which were analyzed with the 'no filter' PXRF setting, and were subject to higher analytical uncertainty. For 37 out of the 51 element/RM combinations, drift was calculated to have a potential effect of less than 10% on the estimates of $U_{\text{HET}}\%$. Of all elements only Al, Si and K show

significant drift using a paired sample t-test (df=2, p < 0.05, **Table 4**). Drift adjustments have not been applied to the reported data, so drift must be considered an additional potential source of uncertainty in these cases.

Table 4. Element/RM combinations where drift was found to be significant for the 8 mm beam diameter.

Element	SdAR	U _{HET} %	
		Reported	Drift Adjusted
Al2O3	L2	4.7	4.2
Al2O3	M2	5.1	4.2
Al2O3	H1	5.6	3.0
K2O	H1	1.3	0.7
SiO2	M2	2.1	1.5
SiO2	H1	3.5	2.2

Powder samples

Similar findings were made for samples prepared as powders, as for pellets. These measurements again show strong evidence of a general increase in U_{HET}% as the beam size is changed from 8 mm to 3 mm. A total of 46 (85%) of the element/RM combinations show an increase in the U_{HET}% ratio, while 31 (57%) exceed the critical ratio. Again, similar conclusions are arrived at if the ratios of U_{HET} values are used instead of U_{HET}%. There are some significant differences between the U_{HET}% values for pellets and powders, in particular CaO was found to have much lower values of U_{HET}% in powders compared with pellets for all three SdAR RMs. The reason for these differences is not clear

Uncertainty on the reference values

There is currently no established procedure for estimating the uncertainty of a reference or certified value for masses of test portion below the specified minimum. One approach taken by Jochum *et al.* (2011) had U_{RV} of glass RMs unchanged for masses down to 0.02 µg for elements classified as 'homogeneous' (RSD_{inhom} <1%, i.e. U_{HET} <2%). For elements that showed moderate or gross inhomogeneity (RSD_{inhom} >1%, i.e. U_{HET} >2%), variance from heterogeneity (VAR_{inhom}) was added using part of the equation from Kane *et al.* (2003):

$$u_{RV}^2 = \frac{VAR(Y_{mean})}{N} + VAR_{inhom} + VAR_{bias} \quad \text{Equation 8}$$

As already discussed, the uncertainty of a reference or certified value (u_{RV}) is usually stated for a minimum test portion mass, e.g. ≥ 0.2 g, of the reference material. This uncertainty value includes an explicit variance caused by any in-homogeneity of the analyte being measured at this mass (VAR_{inhom}). Here Y_{mean} is the mean of the concentration values measured by N laboratories, and VAR_{bias} is the variance caused by the bias between the different analytical methods used. VAR_{inhom} has been expressed below as the expanded measurement uncertainty arising from analyte heterogeneity (U_{HET}²).

A similar approach is taken in this study. It is assumed initially that the minimum recommended test portion mass of 0.2 g is not affected by heterogeneity of analytes, i.e. the reference materials are “sufficiently homogeneous” at this mass. It is then possible to estimate the standard uncertainty on reference values for a reduced mass $U_{RV\ beam}$, by adding the experimentally determined $U_{HET\ beam}$ to the quoted standard uncertainty (U_{RV}), by sum squares, using:

$$U_{RV\ beam} = \sqrt{(U_{RV}^2 + U_{HET\ beam}^2)} \quad \text{Equation 9}$$

For this purpose it is necessary to use absolute U_{HET} values, not the previously discussed relative $U_{HET}\%$ values.

In order to convert the standard uncertainty into the expanded uncertainty on the reference value ($U_{RV\ beam}$) required for the user on the information sheet, it is necessary to multiply by the coverage factor (k). For this particular purpose, where the number of measurements used to determine an uncertainty value is low (e.g. $n < 30$) it is necessary to use a value from the student's t -distribution for k , rather than the value of 2 for 95% confidence that is used more generally (e.g. Equations 2 & 4). To extract this t -value (from tables or a computer program) the number of degrees of freedom (ν) needs to be calculated. Where the uncertainty arises from the sum of other uncertainties, as for Equation 9, then the effective degrees of freedom ($\nu_{RV\ beam}$) can be calculated using the Welch-Satterthwaite formula (JCGM, 2008). Applied to Equation 9, this gives

$$\nu_{RV\ beam} = \frac{u_{RV\ beam}^4}{\frac{u_{RV\ bulk}^4}{\nu_{RV\ bulk}} + \frac{u_{HET}^4}{\nu_{HET}}}$$

The degrees of freedom $\nu_{RV\ bulk}$ are calculated as $n_L - 1$, where n_L is the total number of lab-measurements used to derive RV, listed for each element on the relevant RM information sheet. The value of ν_{HET} , is 11, which is $n_s - 1$, where n_s is the number of pellet-sides used in this experiment for each RM (i.e. 12). The coverage factor k for each element/RM is the t -value for the stated probability (p) and degrees of freedom (ν) using the Excel function $TINV(p, \nu)$. This k value is rounded to two figures (i.e. 2.0, 2.1 or 2.2) to reflect the precision in its estimation. The value of k remains as 2.0 for element where the uncertainty budget is dominated by the value of $U_{RV\ bulk}$ (and $\nu_{RV\ beam} \approx n_L - 1$), but increases up to 2.1 or 2.2, where it is dominated by U_{HET} (and $\nu_{RV\ beam} \approx n_s - 1$).

These new estimates of expanded uncertainty on the reference values for reduced beam sizes are given in **Table 5a-Table5c**. The values of $U_{RV\ beam}$ ($= k * U_{RV\ beam}$) are generally rounded to one digit (and two digits between ‘10’ and ‘25’) to reflect both the precision of the estimate, and also the rounding scheme used in the published values of U_{RV} .

Table 5a. $U_{RV\ beam}$ values calculated for 2 beam sizes. SdAR-L2 pellets.

	RV from data sheet	U_{RV} from data sheet	Recalculated U_{RV} 8 mm beam	Recalculated U_{RV} 3 mm beam
$g\ 100g^{-1}$				
Al ₂ O ₃	11.58	0.05	0.6 ^b	1.4 ^b
CaO	1.06	0.01	0.17 ^b	0.19 ^b
Fe ₂ O ₃	3.63	0.02	0.12 ^b	0.15 ^b
K ₂ O	4.1	0.02	0.05 ^a	0.12 ^b
MnO	0.099	0.001	0.003 ^a	0.005 ^b
SiO ₂	74.48	0.11	1.8 ^b	4 ^b
TiO ₂	0.62	0.003	0.03 ^b	0.12 ^b
$mg\ kg^{-1}$				
Ba	809	10	30 ^a	16
Cu	50.8	1.1	1.1	9 ^b
Nb	63	1.5	1.7	6 ^b
Pb	183	3	5 ^a	7 ^a
Rb	120	1	1.3	1
Sr	150	2	2.2	3
V	35	0.8	0.8	4 ^b
Zn	201	3	3	9 ^a
Zr	618	10	60 ^b	90 ^b

Coverage factor k = 2 unless superscripted; ^aCoverage factor k = 2.1; ^bCoverage factor k = 2.2

Table 5b. $U_{RV\ beam}$ values calculated for 2 beam sizes. SdAR-M2 pellets.

	RV from data sheet	U_{RV} from data sheet	Recalculated U_{RV} 8 mm beam	Recalculated U_{RV} 3 mm beam
$g\ 100g^{-1}$				
Al ₂ O ₃	12.47	0.06	0.7 ^b	1.3 ^b
CaO	0.84	0.01	0.25 ^b	0.30 ^b
Fe ₂ O ₃	2.63	0.02	0.05 ^a	0.14 ^b
K ₂ O	5	0.03	0.09 ^a	0.11 ^b
MnO	0.134	0.002	0.003	0.007 ^b
SiO ₂	73.45	0.17	1.7 ^b	4 ^b
TiO ₂	0.3	0.003	0.02 ^b	0.05 ^b
$mg\ kg^{-1}$				
As	76	5	10 ^a	30 ^b
Ba	990	12	24 ^a	12
Mo	13.3	0.4	1.5 ^b	0.4
Cu	236	4	13 ^b	18
Nb	26.2	0.4	1.8 ^a	4 ^b
Pb	808	14	15	21
Rb	149	2	4 ^a	2
Sr	144	3	3	3
Zn	750	13	24 ^a	60 ^b
Zr	259	7	20 ^a	30 ^b

Coverage factor k = 2 unless superscripted; ^aCoverage factor k = 2.1; ^bCoverage factor k = 2.2

Table 5c. $U_{RV\ beam}$ values calculated for 2 beam sizes. SdAR-H1 pellets.

	RV from data sheet	U_{RV} from data sheet	Recalculated U_{RV} 8 mm beam	Recalculated U_{RV} 3 mm beam
$g\ 100g^{-1}$				
Al ₂ O ₃	11.83	0.07	0.9 ^b	0.70 ^b
CaO	1.46	0.01	0.20 ^b	0.22 ^b
Fe ₂ O ₃	6.45	0.04	0.05	0.09 ^a
K ₂ O	4.17	0.03	0.07 ^a	0.09 ^a
MnO	0.515	0.005	0.01 ^a	0.03 ^b
SiO ₂	65.45	0.18	3 ^b	3 ^b
TiO ₂	0.56	0.004	0.02 ^b	0.08 ^b
$mg\ kg^{-1}$				
As	396	24	24	60 ^a
Ba	866	12	12	12
Cu	1160	20	40 ^a	50 ^a
Mo	64	3	3	10 ^b
Nb	21.9	0.9	2.2 ^a	11 ^b
Pb	3890	80	80	130
Rb	152	3	3	3
Sr	182	3	4	4
V	73.2	1.7	7 ^b	5 ^a
Zn	3680	60	109 ^a	170 ^a
Zr	258	60	60	70

Coverage factor $k = 2$ unless superscripted; ^aCoverage factor $k = 2.1$; ^bCoverage factor $k = 2.2$

Discussion

Estimates of As concentrations derived through its K_{α} line are likely to be subject to interference by the proximity of the Pb L_{α} line in the spectra, although good discrimination using this model of PXRF has been observed where these two elements exist at low concentrations (Caria and Hurley, 2016). Although there is a possibility of positive bias to As measurements from interference by high Pb concentrations, it is the random error component that has been used to calculate U_{HET} and therefore $U_{RV\ beam}$ in **Tables 5a-5c**.

The levels of heterogeneity in the three SdAR RMs may be tabulated using three heterogeneity categories, similar to those previously used for silicate glass RMs (Jochum *et al.* 2011). The ranges of $U_{HET}\%$ selected for each category in this study have been chosen to reflect the results of these experiments (**Table 6**).

Table 6. Broad categorisation of element heterogeneity levels as $U_{HET}\%$ in SdARs using 8 mm and 3 mm beam sizes.

	8mm beam									3mm beam									
	Effectively Homogeneous ($U_{HET} < 1\%$)			Moderately heterogeneous ($U_{HET} > 1\%, < 5\%$)			Grossly heterogeneous ($U_{HET} > 5\%$)			Effectively Homogeneous ($U_{HET} < 1\%$)			Moderately heterogeneous ($U_{HET} > 1\%, < 5\%$)			Grossly heterogeneous ($U_{HET} > 5\%$)			
	L2	M2	H1	L2	M2	H1	L2	M2	H1	L2	M2	H1	L2	M2	H1	L2	M2	H1	
Al				Al				Al	Al	Al	Al					Al	Al	Al	Al
Ca							Ca	Ca	Ca	Ca	Ca					Ca	Ca	Ca	Ca
Fe			Fe	Fe	Fe						Fe	Fe	Fe						
K				K	K	K					K	K	K						
Mn				Mn	Mn	Mn					Mn		Mn	Mn	Mn				
Si				Si	Si	Si					Si	Si	Si						
Ti				Ti		Ti		Ti	Ti	Ti					Ti	Ti	Ti	Ti	
As			As					As	As	As					As	As	As	As	
Ba			Ba	Ba	Ba						Ba	Ba	Ba						
Cu	Cu				Cu	Cu							Cu	Cu	Cu				
Mo						Mo		Mo	Mo	Mo	Mo							Mo	
Ni			Ni															Ni	
Nb				Nb				Nb	Nb	Nb					Nb	Nb	Nb	Nb	
Pb		Pb	Pb	Pb								Pb	Pb	Pb					
Rb	Rb		Rb		Rb					Rb	Rb	Rb							
Sr	Sr	Sr				Sr				Sr	Sr		Sr	Sr					
V	V								V	V							V	V	
Zn	Zn				Zn	Zn							Zn	Zn			Zn	Zn	
Zr							Zr	Zr	Zr	Zr						Zr	Zr	Zr	

The elements classified as grossly heterogeneous in some RMs ($U_{HET} \% > 5\%$) include those that might be expected to only occur in ‘accessory’ minerals with low abundance, such as zircon for Zr, rutile for Ti (and sometimes Nb), and pyrite for As and perhaps calcite for some of the Ca. The minerals quantified by X-ray Diffraction and reported on the reference material data sheets include pyrite and calcite, but not zircon and rutile (IAG, 2016). This explanation does not apply to all of these elements, such as Al, which would be found in many of the main minerals present in the RMs, and listed on their data sheets. The cause of the apparent high heterogeneity for Al may be related to the very low mass of the test portion (<0.01 g, **Figure 1**). The reason for the high heterogeneity of Ca in pellets (12-25%) is unknown, but may include both its low test portion mass, and its presence in accessory calcite.

In order to show how these factors affect the measurements directly, samples of raw data for 3 elements have been selected, from the categories Effectively Homogeneous ($U_{HET} < 1\%$) and Grossly Heterogeneous ($U_{HET} > 5\%$). These are shown in **Table 7**. The variability between the different ‘samples’ (A-H, both sides 1 & 2), relative to their mean value, are greater for the grossly heterogeneous analytes. However, this effect is seriously obscured by the analytical variability (e.g. 2 replicates of A1), which is removed by the ANOVA.

Table 7. Sample raw data for 3 typical elements on two sides (1 & 2) of 6 pellets (A-F) at the 8 mm beam size, which are either 'Effectively Homogeneous ($U_{\text{HET}}\% < 1\%$) or 'Grossly Heterogeneous ($U_{\text{HET}}\% > 5\%$).

Pellet code	$U_{\text{HET}} < 1\%$	$U_{\text{HET}} > 5\%$	$U_{\text{HET}} < 1\%$	$U_{\text{HET}} > 5\%$	$U_{\text{HET}} < 1\%$	$U_{\text{HET}} > 5\%$
side1/2	Cu-L2	Zr-L2	Pb-M2	Zr-M2	Pb-H1	Zr-H1
A1	63.5	600	813	266	3819	271
A1	57.8	602	823	267	3867	272
A2	64.4	617	823	256	3894	259
A2	57.8	616	829	258	3860	259
B1	55.9	656	824	251	3850	264
B1	58.5	654	834	251	3844	265
B2	64.9	598	830	270	3856	262
B2	61.6	605	825	268	3860	258
C1	61.2	649	832	245	3869	266
C1	55.6	649	833	246	3885	269
C2	59.9	635	834	254	3900	256
C2	57.1	639	823	254	3853	255
D1	59.3	591	827	253	3881	259
D1	62.8	591	819	251	3863	258
D2	59.5	650	822	258	3851	289
D2	56.1	650	810	264	3898	293
E1	63.7	600	827	270	3837	261
E1	55.7	599	821	266	3849	261
E2	68.0	642	830	263	3910	273
E2	68.2	643	831	268	3890	273
F1	67.7	678	822	271	3879	256
F1	52.8	679	827	271	3853	256
F2	60.1	610	812	251	3850	249
F2	56.4	614	825	248	3881	251

A low test portion mass might be expected to be a contributing factor to increased heterogeneity, from the proportionality in sampling theory $s \propto 1 / \sqrt{m}$. It therefore might be expected that the typical mass of test portion would be lower for the highly heterogeneous elements than it is for the moderately heterogeneous or the effectively homogeneous elements, although this was not clear from these findings, probably as a result of additional factors such as mineralogy and abundance levels. The slope of a regression of $U_{\text{HET}}\%$ on atomic number was found to be not significant, although it does appear that higher values of $U_{\text{HET}}\%$ are associated with elements with lower atomic numbers (and therefore low test portion mass, see **Figure 1**), and also occurrence in accessory minerals (**Figure 4**). Similar conclusions apply when $U_{\text{HET}}\%$ is modelled against the key-line X-ray energies, i.e. higher values of $U_{\text{HET}}\%$ were found for elements with lower X-ray energies. This is probably at least partially because low X-ray energy lines have lower penetration depth and the test portion mass is therefore smaller for these elements. However, the higher $U_{\text{HET}}\%$ ratios, which represent the difference in $U_{\text{HET}}\%$ between the 8 mm beam and the 3 mm beam diameters ($U_{\text{HET}}\% > 4.0$), only occur for the elements with atomic numbers higher than Zr (See **Figure 2**).

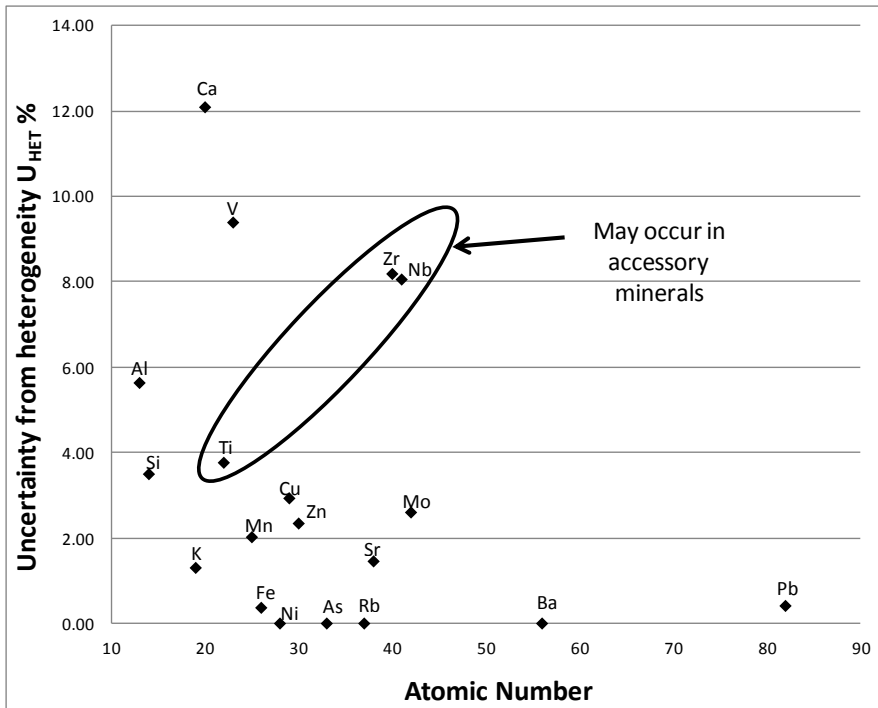


Figure 4. Plot of U_{HET} % against atomic number for SdAR-H1 measured with the 8 mm diameter X-ray beam. Suggests that higher heterogeneity may be generally associated with lower atomic number and/or occurrence of element in accessory minerals.

The experiments also found a general tendency for the uncertainty caused by heterogeneity (U_{HET} %) to increase with decreasing beam size (**Figure 2**). Estimates of the masses of the test portions are mostly less than the 0.2 g minimum recommendation (**Figure 1**). The actual test portion masses will depend on several factors including the characteristic wavelength used to identify each element in the PXRF spectra, and attenuation of the primary and fluorescent X-rays by the test material. It is unfortunately not possible to separate the uncertainty caused by heterogeneity from the uncertainty values (U_{RV}) that are quoted in the SdAR data sheets, because, as previously discussed, these include additional uncertainties such as between-lab and within-lab uncertainties. Where the test portion mass is below 0.2 g, it may be reasonable to suppose that the contribution to the total uncertainty caused by heterogeneity may often be increased. It would therefore be expected that the uncertainty in PXRF measurements is generally higher than would be achieved in laboratory measurements that use at least the recommended minimum test portion mass of 0.2 g.

The estimates of test portion mass make it theoretically possible to model changes in uncertainty with the mass of material that is analyzed in a single measurement. Examples of this approach show the heterogeneity declining as the mass increases, broadly as predicted by sampling theory (**Figure 5**). There are limitations to what can reasonably be achieved with estimates of uncertainty for only three different masses. A possible future study might measure uncertainty for a greater number of different test portion masses, either using laboratory methods or a larger beam XRF. Re-plotting the graphs in **Figure 5** for a greater number of test portion masses would also enable the relationship between uncertainty and mass to be investigated as the mass approaches 0.2 g, potentially providing greater

justification for the assumption made that heterogeneity is not a significant factor at the recommended minimum mass of 0.2 g.

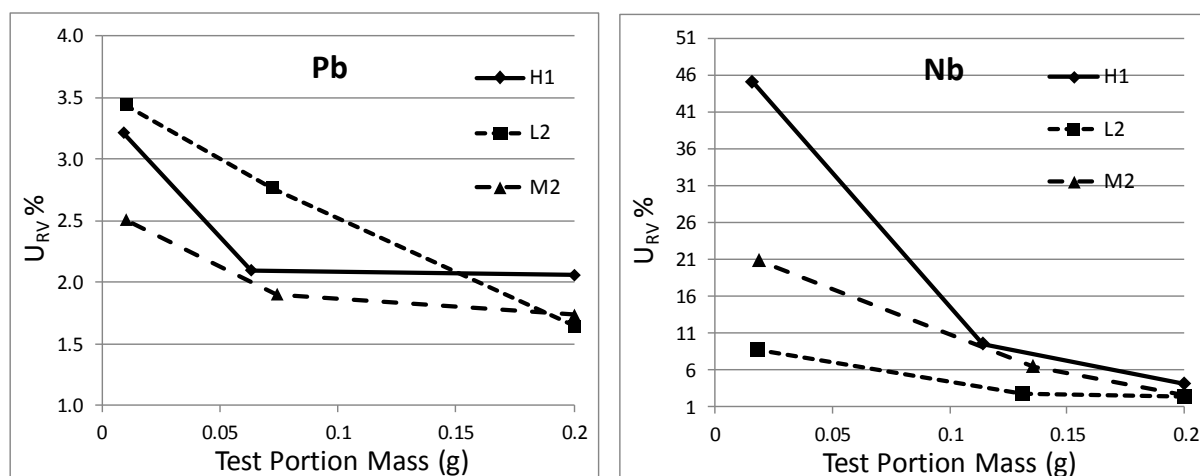


Figure 5. Examples of modelling uncertainty against estimated mass of the test portion. Uncertainties have been expressed as percentages of the quoted reference value, for comparability.

In some cases where values for U_{HET} were calculated as zero (**Table 3**) it could be speculated that these elements are sufficiently homogeneous, even at the smaller test portion sizes.

The estimates of $U_{\text{RV } 8 \text{ mm}}$ and $U_{\text{RV } 3 \text{ mm}}$ in **Tables 5a-5c** have been based on the assumption that heterogeneity does not contribute to the uncertainty when the minimum test portion mass of at least 0.2 g is used. This assumption may not be justified for some elements, although it is reasonable to suppose that uncertainty due to heterogeneity will be at a very much lower level when 0.2 g of material is analyzed than is the case for the very small test portions analyzed by PXRF in most cases. An alternative approach for making improved estimates of $U_{\text{RV beam}}$ could be to undertake fresh laboratory experiments to quantify the random component of uncertainty due to heterogeneity when the test portion mass is equal to 0.2 g. A similar experimental design to that used in the current study could be used for this quantification. It would then be possible to remove this 'bulk' heterogeneity component of uncertainty from U_{RV} , prior to calculating $U_{\text{RV } 8 \text{ mm}}$ and $U_{\text{RV } 3 \text{ mm}}$ using **Equation 9**.

Another potential approach allowing the use of unmodified U_{RV} values for the validation of beam techniques would be to make multiple measurements of each RM such that the total test portion mass analyzed equals the minimum mass specified on the RM data sheet. However, the approximated estimates of test portion mass made in this study suggest that the number of replicate measurements required for PXRF would be unrealistically large for some elements (e.g. 200 for Al, 37 for Ti) even when using the larger 8mm X-ray beam. If the practical maximum number of replicate measurements on each RM was considered to be ten, then this alternative approach could be realistically applied to 12 of the 18 elements for the 8 mm beam size, and 3-5 out of the 18 elements for the 3 mm beam size depending on SdAR type.

There are possible wider implications of these research findings. The methods used to estimate uncertainty and separate analytical repeatability (U_{ANA}) and the component of uncertainty due to heterogeneity (U_{HET}), also have a potential role in the estimation of uncertainty of the measurements obtained during routine sampling of any heterogeneous materials using PXRF instruments (i.e. not just RMs).

It is assumed that the reported values of U_{HET} and $U_{RV \text{ beam}}$ for pellets and powders are broadly applicable to (a) pellets and powders of these materials made by a range of different preparation procedures, and (b) measurements made on PXRF machines from other manufacturers. Although these assumptions could be tested experimentally, it might not always be possible to detect any significant differences between such estimates.

The new uncertainty values ($U_{RV \text{ beam}}$) for smaller test portion masses have previously been tabulated by beam size (**Tables 5a-5c**). This is thought to be the most useful presentation of the measured data at this time, because a) it allows analysts to estimate the modified U_{RV} value when using a PXRF instrument without the need to know the test portion mass for the target analyte; b) it is independent of the assumptions made in the estimates of test portion mass. Listing the revised $U_{RV \text{ beam}}$ values by test portion mass would make these results more general because they would then be appropriate for analysts using different instrumentation, following an estimate of the test portion mass achieved by the particular method. A more accurate estimate of the test portion mass of a PXRF measurement for each element in a given medium could be achieved through the use of computer simulation.

Initial investigations have shown that the values of $U_{RV \text{ 8mm}}$ are not significantly different when a different calibration package (i.e. Niton 'Soils Mode') was used for this same experiment on this same PXRF instrument.

This study has shown that analyte heterogeneity potentially becomes an important factor when small beam sizes are used in PXRF measurements. It can reasonably be assumed that other analytical methods that depend on macro-beam methodologies will be similarly affected by analyte heterogeneity. For example FTIR (Chen *et al.* 2015) and VIS-NIR (Horta *et al.* 2015) have application as *in situ* methods that can be used for quantitative measurements in the field. At the micro scale, LA-ICPMS (Limbeck *et al.* 2015) can be considered as a laboratory-based method that uses beam technology, and where measurements may also be affected by heterogeneity of analytes. The experimental design described in this study would also be applicable for the estimation of uncertainties in these analytical methods, when suitably adjusted to the particular spatial scale of measurement.

Conclusions

The uncertainties caused by heterogeneity within pellets (U_{HET} and $U_{HET\%}$) were estimated for 18 elements within 3 silicate reference materials. These values ranged from 'effectively homogeneous' ($U_{HET\%} < 1$) for several major elements, to 'grossly heterogeneous' ($U_{HET\%} > 5$) for several other elements, particularly those in accessory minerals, or that had a very low estimated test portion mass. There was an overall tendency for $U_{HET\%}$ to increase when the PXRF beam diameter was changed from 8 mm to 3 mm, with 80% of the element/RM

combinations showing an increase. In some cases $U_{\text{HET}}\%$ was estimated at zero, which may indicate that some elements in some RMs are “sufficiently homogenous” for PXRF measurements. Of the non-zero values, 41% of the increases in $U_{\text{HET}}\%$ from 8 mm to 3 mm beam diameters were found to be statistically significant. It was generally found that those elements with higher test portion masses had lower values of $U_{\text{HET}}\%$. This is also a potential explanation of some cases where $U_{\text{HET}}\%$ was estimated to be zero (e.g. Ba in SdAR-H1).

Revised values for the uncertainties on the reference values for use of the RMs with PXRF ($U_{\text{RV beam}}$, for either pellets or powders) have been estimated. This was achieved by adding (by sum of squares) the estimated uncertainty due to heterogeneity (U_{HET}) to the quoted uncertainties (U_{RV}) from the SdAR data sheets. This approach has the limitation that it assumes the quoted uncertainties for a minimum test portion mass of 0.2 g do not include a significant contribution of uncertainty from within-bottle heterogeneity. This limitation could possibly be overcome in future by having an experimental estimate of U_{HET} at the 0.2g scale, which could then be subtracted before the calculation of $U_{\text{RV beam}}$.

Estimates of the masses of test portions measured by PXRF for an 8 mm PXRF beam size ranged between 0.001 g for Al and 0.3 g for Ba. Most elements were found to have a test portion mass below the recommended minimum of 0.2 g for the three SdAR reference materials investigated.

Analytical repeatability of PXRF measurements on pellets made from these RMs was estimated, primarily to enable the subsequent estimation of heterogeneity. For the major elements, the analytical repeatability was generally <3% (median = 2.3% for 8 mm, 2.1% for 3 mm), but somewhat higher for some of the trace elements.

Relative analytical bias was estimated as part of quality control, but did not affect the values of U_{HET} or $U_{\text{RV beam}}$. Bias values obtained against a single pellet made from each of three NIST certified reference materials were found to be less than 10% for 72% of the elements. However, some were higher than 10%, particularly when the reference values were closer to the quoted LOD of the instrument.

The use of the SdAR reference materials has demonstrated the effectiveness of this methodology for quantifying the effects of heterogeneity, it would be useful to extend this work to other reference materials, particularly those used for macro- and micro-beam techniques.

Acknowledgements

Thanks are given to the International Association of Geoanalysts (IAG) for funding; John Hurley, and other staff at Niton UK and Thermo-Fisher, for loan of, and advice on the use of PXRF; Prof Jeremy Field and Michael Henry of the University of Sussex providing, respectively, laboratory space and pellet preparation assistance for these experiments.

References

AMC (2014) RANOVA2 program (Robust Analysis Of Variance). <http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/RANOVA2.asp>

Caria, P. and Hurley, J. (2016) Thermo Fisher Scientific. **Personal Communication.**

Chen, Y., Zou, C., Mastalerz, M., Hu, S., Gasaway, C. and Tao, X. (2015) Applications of Micro-Fourier Transform Infrared Spectroscopy (FTIR) in the Geological Sciences—A Review. **International Journal of Molecular Sciences**, **16**, 30223–30250;

Eggins, S.M. and Shelley, J.M.G. (2002) Compositional Heterogeneity in NIST SRM 610-617 Glasses. **Geostandards Newsletter, The Journal of Geostandards and Geoanalysis**, **26**, **3**, 269-286.

Gy P.M. (1979) Sampling of Particulate Materials Theory and Practice. **Elsevier Scientific Publishing Company**, Amsterdam.

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T.F.A., McBratney, A.B., Pallasser, R. and Pozza, L. (2015) Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. **Geoderma** **241–242**, 180–209.

IAG (2016) Geochemical Reference Materials, Page on website of the International Association of Geoanalysts. <http://www.iageo.com/index.php/geochemical-reference-materials/14-products/65-sdar-reference-materials.html>

ISO Guide 35 (2006) Reference Materials – General and statistical principles for certification. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39269

JCGM 100 (2008) Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM 2008). Joint Committee for Guides in Metrology (Sèvres Cedex). <http://www.bipm.org/en/publications/guides/gum.html>

Jochum K.P., Weis U., Stoll B., Kuzmin D., Yang Q., Raczek I., Jacob D.E., Stracke A., Birbaum K., Frick D.A., Günther D. and Enzweiler J. (2011) Determination of Reference Values for NIST SRM 610–617 Glasses Following ISO Guidelines. **Geostandards and Geoanalytical Research**, **35**, **4**, 397–429.

Kane J.S., Potts P.J., Wiedenbeck M., Carignan J. and Wilson S. (2003) International Association of Geoanalysts' Protocol for the Certification of Geological and Environmental Reference Materials. **Geostandards Newsletter, The Journal of Geostandards and Geoanalysis**, **27**, **3**, 227–244.

Limbeck, A., Galler, P., Bonta, M., Bauer, G., Nischkauer, W. and Vanhaecke, F. (2015) Recent advances in quantitative LA-ICP-MS analysis: challenges and solutions in the life sciences and environmental chemistry. **Analytical and Bioanalytical Chemistry** **407**, 6593–6617.

Parsons, C., Grabulosa, E.M., Pili, E., Floor, G.H., Roman-Ross, G. and Charlet, L. (2013) Quantification of trace arsenic in soils by field-portable X-ray fluorescence spectrometry: Considerations for sample preparation and measurement conditions. **Journal of Hazardous Materials** **262**, 1213-1222.

Potts, P.J. (1987) Handbook of silicate rock analysis. **Blackie & Son Ltd**, Glasgow.

Ramsey M.H. and Ellison S.L.R. (eds.) (2007) Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches **Eurachem (2007)**.

https://eurachem.org/images/stories/Guides/pdf/UfS_2007.pdf

Webb, P. C., Thompson, M., Potts, P.J. and Wilson, S.A. (2015a) GeoPT37A — AN INTERNATIONAL PROFICIENCY TEST FOR ANALYTICAL GEOCHEMISTRY LABORATORIES — REPORT ON ROUND 37A (Blended sediment, SdAR-L2).

Webb, P. C., Thompson, M., Potts, P.J. and Wilson, S.A. (2015b) GeoPT36A — AN INTERNATIONAL PROFICIENCY TEST FOR ANALYTICAL GEOCHEMISTRY LABORATORIES — REPORT ON ROUND 36A (Metal-rich sediment, SdAR-M2).

<http://www.geoanalyst.org/geopt/GeoPT36AReport.pdf>

Webb, P. C., Thompson, M., Potts, P.J. and Wilson, S.A. (2014) GeoPT35A — AN INTERNATIONAL PROFICIENCY TEST FOR ANALYTICAL GEOCHEMISTRY LABORATORIES — REPORT ON SUPPLEMENTARY ROUND 35A (Metalliferous sediment, SdAR-H1).

<http://www.geoanalyst.org/geopt/GeoPT35AReport.pdf>
