Quantifying Heterogeneity with Microbeam Analysis

Michael Shaffer, University of Oregon

Electron probe microanalysis (EPMA) is not only an excellent technique for measuring the chemical composition of a wide variety of materials, but also provides an effective means to quantitatively evaluate the chemical homogeneity of the material. In this contribution[1] I present an easy approach for determining the chemical homogeneity of specimens in a statistically meaningful way.

The degree of chemical heterogeneity may be determined by a variety of methods varying both in ease and rigor. One might apply a simple test for the sake of classifying the material as *homogeneous* or *heterogeneous*. One very useful method has long been to simply relate the observed standard deviation, s_n, for many microanalyses, n, to the statistical counting error, *i.e.*, The square root of the average counts, $(N_{ave})^{1/2}$. This simple test is what is usually termed the *sigma ratio*:

$$\frac{s_n}{\sqrt{N_{ave}}} \qquad \text{eq. 1}$$

A ratio of approximate unity would imply homogeneity, however, this approach does not really tell us anything about heterogeneity, or of the specific relationship of a given sigma ratio to a particular amount of heterogeneity (or degree of homogeneity). For example, what does a sigma ratio of 2 imply for heterogeneity?

Another "yes or no" type of evaluation may be made by comparing all measurements to the range defined by a 3 sigma counting error:

$$(N_{ave} - 3\sqrt{N_{ave}}) \leq N \leq (N_{ave} + 3\sqrt{N_{ave}}) \qquad \text{eq. 2}$$

Using this approach, if all measurements fall within the defined range, the material could be considered homogeneous. As analysts we often require a technique to quantitatively determine homogeneity so that meaningful comparisons can be made. Goldstein[2] makes reference to a *t-distribution* technique which would apparently calculate something more meaningful. That is, it would yield a number which would allow the analyst to say with confidence that his sample was *at least* some weight percent heterogeneous. Such a number would be valuable and should be considered a measurement which could be shared amongst other similar measurements. However, what if the analyst wanted to describe his sample as *no more than*, as well as *at least* some weight percent heterogeneous. The analyst would also require technique to be based on statistical principles, to take into account the number of samples, n, the precision associated with N, and to be associated with a desired level of *confidence*. We might also ask the technique to separate the compositional variance from other variables, such as counting statistics.

Principle #1: Individual variances can contribute to the observed.

An observed variance can be considered the result of several specific variances summed in quadrature. Thus:

$$\sigma^2_{observed} = \sigma^2_a + \sigma^2_b + \ldots \qquad \text{eq. 3}$$

The above equation relies on the universal acceptance of an *observed variance* being the sum of individual absolute variances ... that is, if not interdependent, variances can be individualized. Assuming we have justification for ignoring variances directly attributable to the instrument and/or analyst, we already have a quantitative handle on heterogeneity if we assign one variance to fundamental counting statistics and another to heterogeneity,.

Principle #2: The chi-square distribution

While it is convenient to discuss the ideal case of *a restricted number of "perfectly" executed replicate counting experiments*, in truth we never really know whether or not a set of analyses truly meets this ideal expectation. In practice, all we can do is compare the frequency distribution of our sample analyses to the frequency distribution of its assumed parent population, and then try to decide whether the ideal assumption is indeed justified. This is what we do intuitively when we use the sigma ratio and state ... "when it is approximately equal to unity, our sample (characterized by N_{ave} and s_n) comes from a parent population characterized by σ^2 and σ^4.

The *chi-square* function, $\chi^2 = s_n^2 / \sigma^2$, is a statistical parameter which compares the observed variance of a sample to the variance of its parent population. Furthermore, there is a probability that can be calculated for drawing a sample with variance s^2 from a population with variance σ^2, *i.e.*, a probability of achieving a certain value, χ^2. The probability of drawing a sample with s_n^2 variance from a population of variance σ^2 [denoted by $P(\chi^2)$] depends on the value of χ^2 and on the degrees of freedom remaining after s_n^2 has been calculated from the sample. For a case of a sample of n replicate counts, the degrees of freedom remaining is $f = (n - 1)$.

The usefulness of the *chi-square* distribution is its probability function. The analyst can ask this function to yield two numbers both associated with a desired confidence level. For the purpose of determining homogeneity, we can, at a given confidence level, evaluate a range defined by lower and upper limits. Functionally, these limits allow the analyst to characterize the specimen's heterogeneity with the terms at least or no more than, or both.

Three counting experiments

It is instructive at this point to introduce three actual examples of replicate counting experiments, and introduce some terms:

(1) Replicate counts for 100 different EPMA spot analyses on a optically homogeneous, synthetic glass, for which the counting period was relatively short and only ~900 counts were obtained:

$$N_{ave} = 901, \text{ and } s_n^2 = 993 \text{ and } s_n = 31.5$$

(2) Replicate counts for 100 different EPMA spot analyses on a optically homogeneous synthetic, glass, for which the counting period was long enough to obtain ~9000 counts:

$$N_{ave} = 9005, \text{ and } s_n^2 = 10{,}609 \text{ and } s_n = 103$$

(3) Replicate counts for 100 different EPMA spot analyses on an unknown glass specimen, for which the counting period was long enough to obtain ~9000 counts:

$$N_{ave} = 8974, \text{ and } s_n^2 = 247{,}596 \text{ and } s_n = 497$$

A calculation of the sigma ratio for each of the 3 cases yields 1.05, 1.08 and 5.25, respectively. Based on the sigma ratio criterion, the first two glasses appear to be "homogeneous".

Putting quantitative limits on sample heterogeneity

The third counting experiment, however, suggests heterogeneity. There is practically no chance that our set of observations could have sampled a parent population with $\sigma^2_{observed} \sim N_{ave}$. The assumption that it did would run counter to overwhelming odds and we must abandon it. Accordingly, we must describe the sample as being in all probability heterogeneous. We can be more quantitative, however, by using Principle #1 to describe the sample as having a parent population characterized by:

$$\sigma^2_{observed} = \sigma^2_{heterogeneity} + \sigma^2_{counting} \qquad \text{eq. 4}$$

where all units are counts for a given element (the final product being a conversion of *absolute* variances to *relative*, and applying the relative error to absolute weight percent values).

The first step in what we believe to be the correct approach to this problem is to exploit the relationship between the observed population and the parent as predicted by the *chi-square* distribution:

$$\chi^2 = \frac{s_n}{\sigma^2} = \left(\frac{s_n^2}{\sigma^2_{counts} + \sigma_{heterogeneity}} \right) = \left(\frac{s_n^2}{N_{ave} + \sigma^2_{heterogeneity}} \right) \qquad \text{eq. 5}$$

The next step is to decide at what probability (confidence) level we wish to specify the upper and lower limits of heterogeneity, thereby allowing us to state with a defined degree of confidence that "the heterogeneity exceeds *p* weight percent, but it does not exceed *q* weight percent". Common confidences are 95% assured or absolutely assured at 99%.

Next, we utilize the integral x^2 probability distribution as tabulated in most references on statistics (*e.g.*, the table titled *percentage points, chi-square over degrees of freedom distribution* in the CRC *Handbook of Tables for Probability and Statistics*)[2]. For this example we choose the absolute confidence interval 99%, *i.e.*, the tabulated values for $P(x^2)$ for $(1 - P) = 1.0$ and $P = 99.0$, and (100 - 1) as the degrees of freedom. Table 1 lists these as well as other comparative values.

N	f	χ^2 @ $(1 - P) = 1\%$	χ^2 @ $P = 99\%$	χ^2 @ $(1 - P) = 5\%$	χ^2 @ $P = 95\%$
5	4	.0742	3.3192	.1778	2.3720
10	9	.2320	2.4073	.3694	1.8799
30	29	.4916	1.7099	.6106	1.4675
50	49	.5906	1.5290	.6924	1.3539
100	99	.6993	1.3600	.7782	1.2447

Table 1. Values of χ^2 for useful examples of P and n replicate samplings

The table yields $x^2_1 = 0.6993$ (This value sets the lower limit of X^2 and therefore the upper limit of $\sigma^2_{heterogeneity}$), and $x^2_{99} = 1.360$ (which sets the lower limit of $\sigma^2_{heterogeneity}$).

Rearranging eq.5 implies:

$$\sigma_{[heterogeneity(P)]} = \sqrt{\frac{s_n^2}{x(P)} - N_{ave}} \text{ and } \sigma_{[heterogeneity(1-P)]} = \sqrt{\frac{s_n^2}{x(P-1)} - N_{ave}} \qquad \text{eq. 6}$$

Using the data collected, our obviously heterogeneous specimen and eq. 6, we solve for $\sigma_{heterogeneity}$ where $x^2 = 0.6993$ and 1.360, which yields the absolute counting values of 587 and 416, respectively. If we desire relative values we normalize by the average counts, N_{ave}, which yields with 99% confidence our sample's heterogeneity. This varies at least 4.6%, but no more than 6.5%. These values can be converted into absolute weight percents by applying them to the concentration levels as measured with the microprobe.

Summarizing with comparisons

As previously described (*cf.* eq. 4), the absolute variance for counting x-rays is simply the counts obtained over time, N, or the average counts, N_{ave}, obtained for many points, n. $\sigma^2_{observed}$ can be replaced with the square of the observed standard deviation from averaging, s_n, thus yielding an estimate for the contribution heterogeneity makes to the observed error:

$$\sigma_{heterogeneity} = \sqrt{s^2_n - N_{ave}} \qquad \text{eq. 7}$$

For our heterogeneous specimen, a quick calculation yields 488 or 5.44%. Considering its simplicity, this is a reasonable value. It should be noted, however, that this calculation corresponds to confidence level of 68%. If the same sigma is calculated for a confidence level comparable to our *chi-square* evaluation, a value near 15% would result. This discrepancy is related to the fact that this equation does not take into account the large number of replicate samples.

Continued on following page

Quantifying Heterogeneity with Microbeam Analysis

Continued from preceding page

The number of replicate samples (n) has a large effect on the *chi-square* calculation. For example if $n = 5$ ($f = 4$), the values of *chi-square* for the same confidence level are 0.0742 and 3.3192, and the upper and lower values for heterogeneity are 20.3% and 2.9%. This heterogeneity range is clearly of little use compared to the more limited range obtained from a larger number of replicate analyses, and because N and s remain similar in spite of n being large or small. Recognizing that 100 random spot analyses on a single specimen may be considered excessive and not an efficient use of instrument time, consider $n = 30$; the heterogeneity range for our sample (3) based on this number of replicate analyses is 7.8% and 4.1%. Compared with 6.5% and 4.6% (for $n = 100$), and for less than a third of the instrument/operator time, 30 replicate samplings seems like a good compromise.

The counting experiments presented above for optically homogeneous glasses (1 and 2) appear similar, in that application of the simplistic *sigma ratio* implies both materials to be "homogeneous". More rigorous analysis using the technique described above shows that the statistical precision is different and analysis (1) either represents a case for a minor element not being analyzed, or a case where the analyst chose too short a counting time. Using case (1) and a 99% confidence level, this *chi square* technique suggests the range of heterogeneity to be between 2.5% and "zero", whereas for case (2) the calculated range is 0.87% and "zero". That is, for both cases homogeneity is a possibility, but the statistical range is significantly larger for sample (1) than for sample (2). Even though the sigma ratio is better for sample (1), statistically it is less homogeneous than sample (2), as a result of the poorer precision due to smaller N. The bottom line is that application of the *chi square* technique, and other robust statistical techniques will be limited, or misleading for both small n and small N. For studies where quantitative information on the homogeneity of samples is required, the analyst must take the time to collect sufficient data for a reasonable result.

Such evaluation of homogeneity usually suffices for major elemental constituents. However, for minor and trace constituents, I need mention the more general case of eq. 4 would include the variance associated with the:

$$\sigma^2_{counting} = \sigma^2_{peak} + \sigma^2_{background} \qquad \text{eq. 8}$$

That is, we had been able to ignore the contribution to the counting error as long as $N_p >> N_b$, but this is no longer true for minor elements. This contribution would need be considered in the equations which follow eq. 4, but it may take one of several forms because the analyst might use one of several methods of measuring the background more accurately and more appropriately than for every N, thus minimizing its variance and optimizing the contribution to the accuracy of N_{ave}. ■

1. Rice, J., Shaffer, M., Weill, D., 1998 Course Notes for Electron Beam Analysis in Mineralogy and Petrology, Department of Geological Sciences, University of Oregon.

2 Goldstein, J., Newbury, D., Echlin, P., Joy, D., Fiori, C., Lifshin, E., 1984 *Scanning Electron Microscopy and X-ray Microanalysis.*, Plenum Press, (p.432)

3. See also Bevington, P.R., 1969. *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill