

# Homogeneity Testing at the Micrometer Scale

Dennis Harries

Department of Analytical Mineralogy, Institute of Geosciences, Friedrich Schiller University, Jena, Germany

dennis.harries@uni-jena.de

**Abstract:** The testing of homogeneity and the quantification of chemical heterogeneity in reference materials for microanalysis is an important step in obtaining reliable standards and correct analytical results. This article describes a statistical method for evaluating homogeneity at the micrometer scale using electron probe x-ray analysis. Operation of the method is demonstrated by analysis of a natural Ca phosphate mineral standard. The results and discussion highlight several important factors and concepts, in particular the importance of a sufficiently high number of spot analyses, the role of an uncertainty budget, and the critical level at which heterogeneity can be detected.

## Introduction

Homogeneous reference materials (standards) are a key to accurate and reproducible elemental microanalysis by energy- and wavelength-dispersive x-ray spectrometry in the scanning electron microscope (SEM) and the electron probe microanalyzer (EPMA). Standards are needed for quantification (converting x-ray counts to mass fractions) and for quality control of analysis results (testing on known compositions). Because the spatial resolution of the electron beam allows analysis of volumes down to a few  $\mu\text{m}^3$ , any reference material with a known composition used as a microanalysis standard ideally must have the same composition regardless of where on the standard the measurements are made: it must be homogeneous.

In principle, perfect homogeneity in materials does not exist. Modern aberration-corrected scanning transmission electron microscopes equipped with energy-dispersive x-ray detectors and electron energy loss spectrometers are capable of detecting elemental differences at atomic levels. For the more common use of x-ray microanalysis of bulk specimens, the scale and level of heterogeneity that can be accepted for a given purpose must be defined. Tests for homogeneity are then employed to answer the question of whether a material is “fit for the purpose.” This question can be answered positively when significant heterogeneity cannot be detected with the measurement method under its chosen conditions of operation.

While standards are often pure elements, homogeneous multi-element standards are relatively easy to produce if stoichiometric compounds with a limited number of components are used (for example, pure oxides). However, as soon as components substitute for each other, as in glasses or solid solutions, chemical heterogeneity becomes a concern. This is the case for natural minerals that may have formed under changing physicochemical conditions, but it also concerns synthetic materials where homogeneous doping or alloying may be difficult. For these reasons, testing for homogeneity is an important step in qualifying a material as an analysis standard.

Homogeneity testing of a potential standard requires the investigator to discriminate between the elemental variations

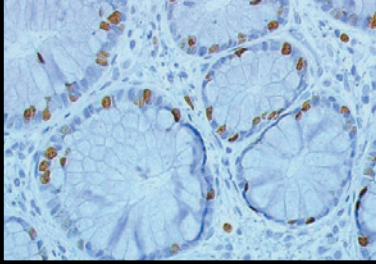
due to true compositional differences in the sample and those variations that are related to the instrument and the measurement process itself. The latter variations are the “precision” of the analysis, which represents the statistically random scatter in the results. Whether or not heterogeneity among different locations on the sample can be detected depends on the magnitude of the chemical variations relative to this omnipresent scatter; hence, the instrumental precision has to be known.

The easiest approach to determining the instrumental precision is to repeatedly analyze the same spot on the sample. Then the comparison of different locations on the sample by “analysis of variance” (ANOVA) allows the statistical separation of instrumental scatter from compositional variations with a given degree of confidence [1, 2]. However, in practical microanalysis this approach is often problematic because repeated analyses of a single spot may result in progressive degradation (or even loss) of the analyzed volume. In electron probe microanalysis this degradation may result from the build-up of surface contamination, induced diffusion, or structural damage and decomposition. In consequence, the apparent instrumental variances from repeated measurements on single spots are often not suitable for a sound statistical analysis. This was recognized early in the history of electron probe microanalysis, and counting statistics traditionally have been used to obtain the instrumental precision [3]. The method outlined here follows this approach: The instrumental precision of a single measurement is the standard deviation solely derived from measured x-ray counts; the variance of the measurement is the standard deviation squared. Calculations are usually more conveniently done with variances, while the standard deviations are usually stated because they preserve the units of the measured quantity. Both are expressions of the uncertainty of the measurement.

Beginning with [3], geological and technological reference materials have been characterized by an index derived as the ratio of the observed standard deviation to the standard deviation (instrumental precision) predicted by counting statistics [4, 5, 6]. However, the interpretation of this “sigma ratio,” or “homogeneity index,” has been based on subjective experience: usually materials with a homogeneity index of  $H < 3$  have been accepted as suitable standards. However, this is statistically not sound, and a revised interpretation of the homogeneity index has been given only recently [7]. Based on this new evaluation, this article describes an improved statistical method for determining the degree of homogeneity in a bulk specimen at the micrometer level of spatial resolution, given the instrumental precision of the analysis based on counting statistics.

## Materials and Methods

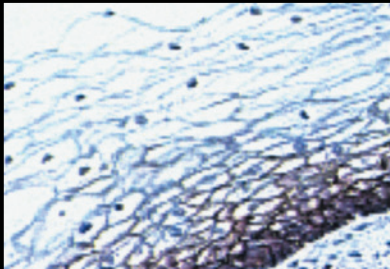
As a candidate reference material for use as a microprobe standard, natural crystals of the mineral fluorapatite from



Ki-67 (antibody MIB1)

### Now you can:

- Run antigen unmasking in 6 various buffers at once.
- Perform gentle antigen retrieval that does not damage the tissue morphology.
- Get identical results every time.
- Compare series of slides treated in independent sessions.



E-cadherin (Membranous, extracellular domain)

## ...EMS has it!

### CONTACT US FOR MORE INFORMATION...

## Electron Microscopy Sciences

P.O. Box 550 • 1560 Industry Rd.  
Hatfield, Pa 19440  
Tel: (215) 412-8400  
Fax: (215) 412-8450  
email: [sgkcck@aol.com](mailto:sgkcck@aol.com)  
or [stacie@ems-secure.com](mailto:stacie@ems-secure.com)

OUR MAIN INTERACTIVE WEBSITE:  
[www.emsdiasum.com](http://www.emsdiasum.com)

 TO REQUEST A COPY  
OF OUR CATALOG:  
[www.emsdiasum.com/  
requests/catalog](http://www.emsdiasum.com/requests/catalog)

 TO VIEW OUR  
DIGITAL CATALOG:  
[catalog.emsdiasum.com](http://catalog.emsdiasum.com)

# the solution for antigen unmasking

## The Retriever

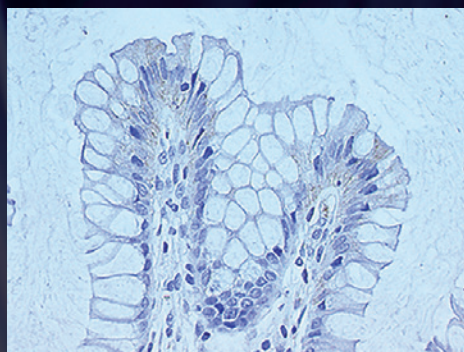
Our Unique Retriever solves the problem of staining formalin fixed tissues. This is an affordable solution to all known major problems with immunohistochemistry on paraffin sections. Ease of use combined with high reproducibility of the results will give you the best quality immunostaining.

### The Retriever is ideal for:

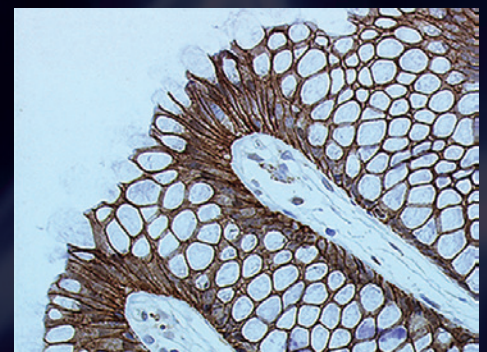
- Investigative Pathology, where the high quality of staining (a picture may be published!) is required.
- Any lab that is short on technical personnel: any student or post-doc can process slides for the staining without using much time
- Small routine pathology labs, where a limited number of slides should be processed daily
- Anyone who uses highly valuable samples, such as tissue arrays or unique tissue samples: Simplicity and reliability of the unit ensures the safety of your sample, and a high quality antigen unmasking



### Excellent epitope recovery. Always.

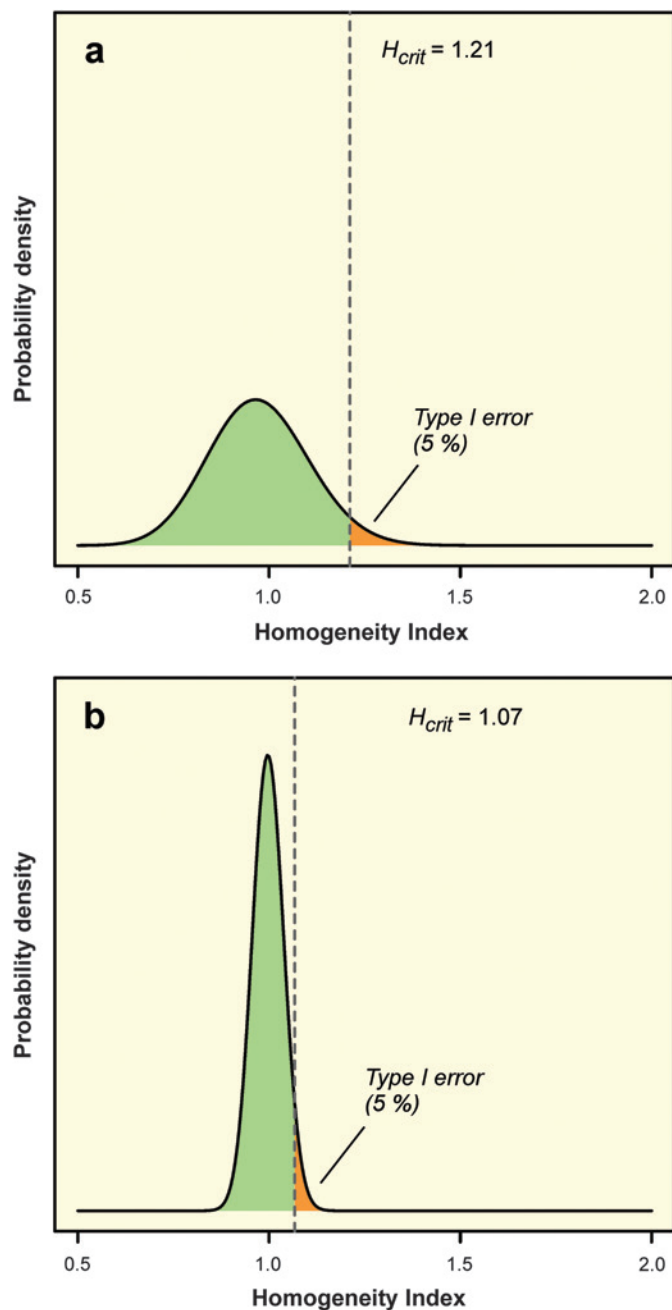


without *the Retriever*



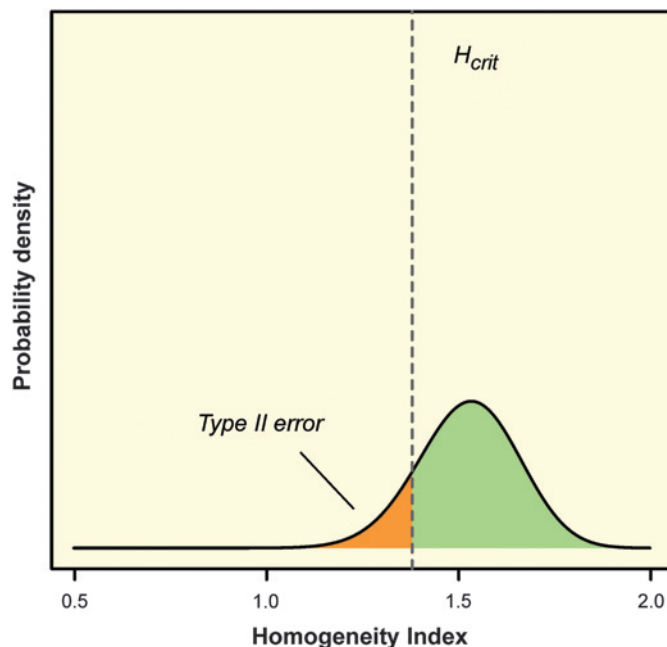
with *the Retriever*

A proper epitope recovery may be critical for your conclusions on protein expression. The pictures above show staining for E-cadherin on human small intestine tissue without processing in Retriever (above left) and after the processing in Retriever (above right). One pathology lab from Belgium reported that they had up to 30% of false-negative staining for oestrogen receptor with microwave epitope recovery, and less than 2% after they switched to Retriever.



**Figure 1:** Schematic probability density functions of the homogeneity index for a perfectly homogeneous material (maximum probability density at  $H = 1$ ). A significance criterion (critical homogeneity index  $H_{crit}$ ) is chosen by accepting that a certain number (usually 5%) of all tests fail despite the null hypothesis is true (type I error). (a) A curve assuming  $N = 30$  measurements. The width of the curve indicates the inherent uncertainty of the homogeneity index. (b) A curve assuming  $N = 300$  measurements. The uncertainty of the homogeneity index is strongly reduced and  $H_{crit}$  is reduced as well.

Imilchil in Morocco were investigated. Fluorapatite is a halogen-bearing calcium phosphate, ideally  $\text{Ca}_5(\text{PO}_4)_3\text{F}$ , and can substitute a wide range of elements into its structure. It is also one of the more beam-sensitive minerals and known for loss of fluorine during electron beam analysis [8]. It was analyzed using wavelength-dispersive electron probe x-ray spectrometry on a JEOL JXA-8900 (15 kV, 30 nA beam current, 14  $\mu\text{m}$  spot diameter). A worked



**Figure 2:** Schematic probability density functions of the homogeneity index for a heterogeneous material (maximum probability density at  $H > 1$ ). Most of the tests fail the homogeneity criterion, but some tests are below the critical homogeneity index  $H_{crit}$ . This erroneous non-detection of heterogeneity is a type II error. Increasing the number of measurements  $N$  decreases the width of the curve (its center stays fixed) and decreases  $H_{crit}$ . This decreases the probability of a type II error.

example of a different material is found in the electronic annex of [7].

**Instrumental precision.** In the case that only one measurement per analysis spot is practically feasible in order to avoid specimen degradation, significance criteria for detectable heterogeneity can be derived in a statistically sound fashion if the counting statistics of x-ray photons is the only significant contribution to the instrumental precision. Hence, the instrumental precision then would be the standard deviation derived from the number of counting events for a particular element x-ray peak. This limitation of instrumental precision to counting statistics is only valid if other sources of instrumental random noise are insignificant. Such a source may be, for example, the alignment of the analyzer crystal in a mechanical wavelength-dispersive spectrometer or variations in the beam current measurement—basically any parameter that is set and reset before and after the acquisition of each data point. Because technological development has achieved highly reproducible mechanics and electronics, these sources of scatter in the results are usually negligible.

**Drift.** A much bigger problem is instrumental drift. In this case instrumental variations occur not randomly but in a time-constrained manner, for example in relation to changing temperature. Instrumental drift must be avoided or properly corrected, and the time series of analyses acquired must be screened meticulously for the presence of variations that correlate with time. Statistical methods for assessing this have been suggested [9]. The technique presented here assumes that drift is absent and any instrumental contribution

**Table 1:** Results of the EPMA homogeneity studies of 36 fluorapatite crystals ( $N = 30$  for each crystal).

Crystal ID	Mean Ce concentration (wt%)	$s_c$ (wt%)	$s_{Pois}$ (wt%)	$s_h$ (wt%)***	$s_{h,rel}$ (%)***	Homogeneity index $H$	Heterogeneity detected?*
s1.1	0.347	0.0478	0.0203	0.0433	90.5	2.36	yes
s1.2	0.336	0.0320	0.0201	0.0249	77.9	1.59	yes
s1.3	0.216	0.0208	0.0190	(0.0083)	(39.9)	1.09	no
s1.4	0.242	0.0172	0.0193	---*	---*	0.89	no
s1.5	0.320	0.0485	0.0199	0.0442	91.1	2.43	yes
s1.6	0.216	0.0177	0.0191	---*	---*	0.93	no
s2.1	0.244	0.0185	0.0194	---*	---*	0.96	no
s2.2	0.410	0.0308	0.0207	0.0228	74.1	1.49	yes
s2.3	0.210	0.0233	0.0191	0.0134	57.5	1.22	yes
s2.4	0.211	0.0305	0.0190	0.0239	78.3	1.61	yes
s2.5	0.214	0.0271	0.0191	0.0193	71.1	1.42	yes
s2.6	0.205	0.0236	0.0190	0.0139	59.1	1.24	yes
s2.7	0.207	0.0217	0.0190	(0.0105)	(48.4)	1.14	no
s2.8	0.205	0.0228	0.0191	(0.0125)	(54.6)	1.19	no
s2.9	0.220	0.0234	0.0192	0.0133	56.8	1.22	yes
s3.1	0.218	0.0247	0.0192	0.0156	63.0	1.29	yes
s3.10	0.218	0.0228	0.0192	(0.0124)	(54.3)	1.19	no
s3.2	0.210	0.0204	0.0192	(0.0070)	(34.3)	1.06	no
s3.3	0.215	0.0258	0.0192	0.0173	66.9	1.35	yes
s3.4	0.204	0.0171	0.0190	---*	---*	0.90	no
s3.5	0.423	0.0512	0.0210	0.0467	91.2	2.44	yes
s3.6	0.210	0.0208	0.0193	(0.0076)	(36.7)	1.08	no
s3.7	0.208	0.0227	0.0192	(0.0123)	(53.9)	1.19	no
s3.8	0.221	0.0221	0.0192	(0.0109)	(49.3)	1.15	no
s3.9	0.217	0.0244	0.0193	0.0149	61.3	1.27	yes
s6.1	0.217	0.0195	0.0191	(0.0039)	(20.1)	1.02	no
s6.2	0.210	0.0211	0.0191	(0.0089)	(42.4)	1.10	no
s6.3	0.215	0.0195	0.0190	(0.0043)	(22.2)	1.03	no
s6.4	0.203	0.0229	0.0189	(0.0129)	(56.2)	1.21	no
s6.5	0.208	0.0195	0.0190	(0.0041)	(20.8)	1.02	no
s6.6	0.440	0.0539	0.0209	0.0497	92.2	2.58	yes
s7.1	0.323	0.0611	0.0200	0.0577	94.5	3.06	yes
s7.2	0.239	0.0212	0.0195	(0.0084)	(39.4)	1.09	no
s7.3	0.219	0.0224	0.0193	(0.0115)	(51.2)	1.16	no
s7.4	0.363	0.0207	0.0203	(0.0037)	(17.7)	1.02	no
s7.5	0.378	0.0380	0.0206	0.0319	84.0	1.84	yes

\*In these cases  $s_{Pois}^2$  is larger than  $s_c^2$  due to statistical scatter, and the resulting  $s_h^2$  is negative (and the homogeneity index smaller than unity). \*\*At  $N = 30$  the critical homogeneity index is 1.21; a measured homogeneity index larger than this indicates detected heterogeneity. \*\*\*Values in parentheses are below the detection limit of heterogeneity of  $s_{h,rel} = 56.4\%$  (calculated from the critical homogeneity index at  $N = 30$ ) and should not be stated because they are statistically not significant. Instead the upper limit of heterogeneity (i.e. "<56.4%") and the corresponding value of  $s_h$  (i.e. "<56.4% of  $s_c$ ") should be stated. They are shown here to illustrate the principle.

to measurement uncertainty, other than the statistical noise of x-ray counting, is insignificant.

**Homogeneity index.** A traditional approach to test for homogeneity is the recently refined homogeneity index ( $H$ ) [3, 7]:

$$H^2 \equiv \frac{s_c^2}{s_{Pois}^2} \quad (\text{Equation 1})$$

The homogeneity index compares the variance expected from Poisson counting statistics ( $s_{Pois}^2$ ) to the combined variance actually observed ( $s_c^2$ ). The combined variance  $s_c^2$  is simply the variance (standard deviation squared) of the mass fraction values obtained from a number of analysis points (designated  $N$ ) across the sample and includes the variance due to Poisson noise and any compositional variance. For each of the  $N$  analysis points an individual variance is calculated based on counting statistics and these are then combined into  $s_{Pois}^2$  as the average of the  $N$  individual values. Equations for this are given in [7].

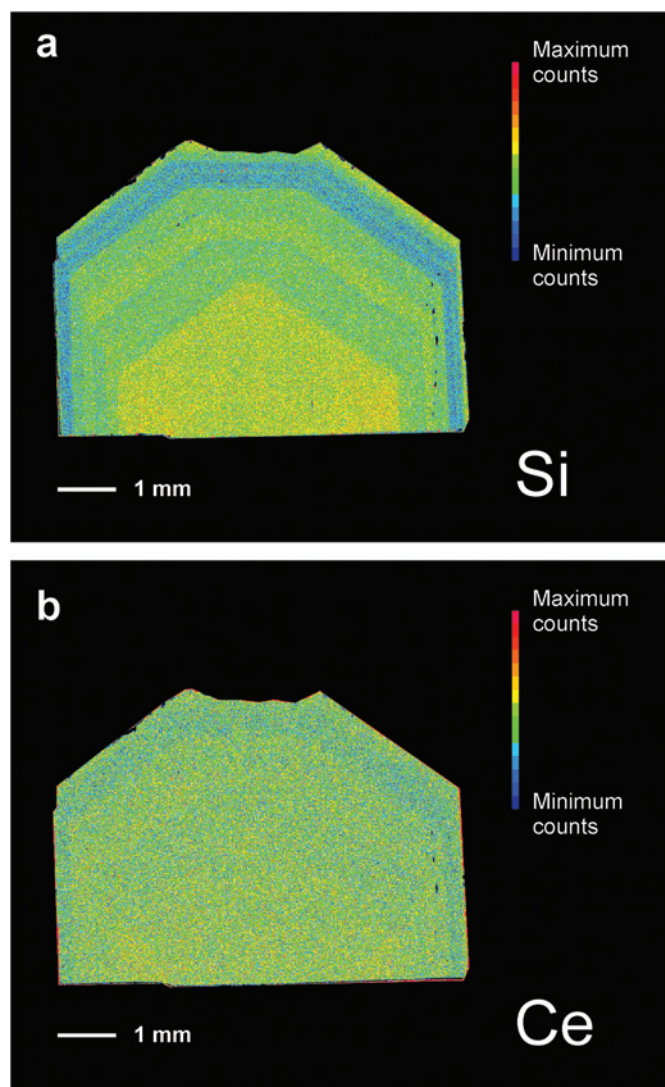
If the observed variance is larger than the expected variance from counting statistics ( $H > 1$ ), compositional heterogeneity may be present. This significance of this heterogeneity can be tested based on F or chi-squared statistics because  $H^2$  is a ratio of variances. The answer is strictly valid only for the specific set of measurement conditions that were used for the data acquisition. For the null hypothesis that no heterogeneity can be detected,  $H = 1$ , the combined variance observed is equal to the variance due to counting statistical noise.

The homogeneity index on its own is not suitable for deciding whether heterogeneity is detected or not, because  $H$  itself has an inherent uncertainty. This is because the variances used to calculate  $H$  are statistical estimates and will scatter around true but unknown values. This scatter will decrease by increasing  $N$ . Hence,  $H > 1$  may indicate significant heterogeneity, or it may just be larger than unity because of statistical scatter, even when there is no heterogeneity present. In order to decide which case applies, a critical homogeneity index ( $H_{crit}$ ) can be calculated based on the chi-squared distribution [7, 9].

$$H_{crit}^2 = \chi_{(\alpha, N-1)}^2 / (N - 1) \quad (\text{Equation 2})$$

If  $H > H_{crit}$ , the null hypothesis has to be rejected and significant heterogeneity was detected. However, because statistical testing cannot provide a definitive answer, a level of significance  $\alpha$  has to be chosen before  $H_{crit}$  is computed; this level is usually 0.05 (5%). At this level, a false rejection of the null hypothesis may occur in 5% of all tests when there is actually no detectable heterogeneity present, a type I error (false detection of heterogeneity) shown in Figure 1. Note that  $\chi_{(\alpha, N-1)}^2$  is available in Excel as function CHISQ.INV where the significance level is stated as  $1 - \alpha$ , that is, 0.95 for a 5% level.

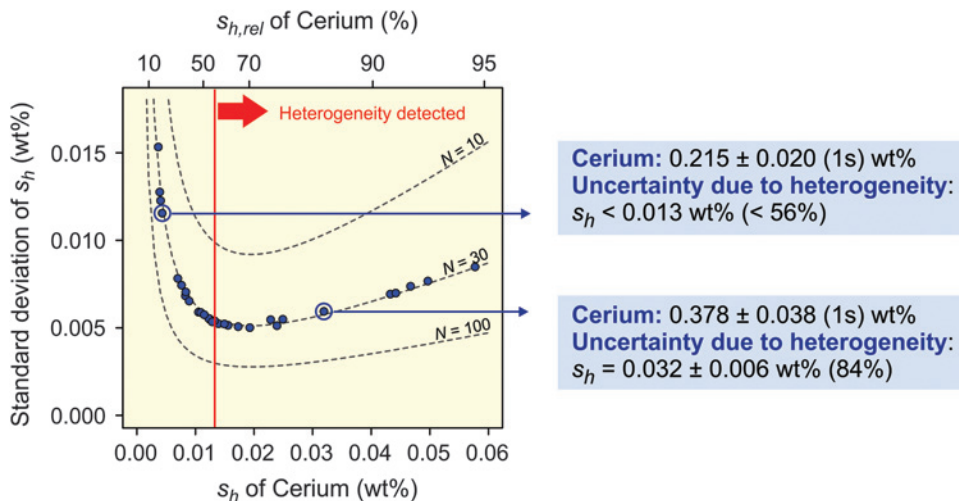
Because the uncertainty of the homogeneity index depends on the number of measurements  $N$ , the critical homogeneity index  $H_{crit}$  depends on  $N$ . At  $N = 30$  measurements and  $\alpha = 0.05$ , the critical homogeneity index is 1.21. At  $N = 300$ , it is reduced to 1.07 (Figure 1). To be considered homogeneous, the determined



**Figure 3:** Electron microprobe x-ray distribution maps by wavelength-dispersive spectrometry obtained on a polished apatite crystal investigated for its suitability as reference material. (a) Distribution of silicon (Si K $\alpha$ ) shows chemical zoning within the crystal. The material is obviously heterogeneous and, in its present form, not suitable as a reference material. (b) Distribution of cerium (Ce L $\alpha$ ) shows barely visible heterogeneity. In this case, statistical testing based on quantitative analyses has to be conducted.

homogeneity index must be below this critical value. Therefore, increasing the number of measurements increases the power to detect heterogeneity. Increasing  $N$  also decreases the probability of erroneously accepting the null hypothesis, the type II error (false non-detection of heterogeneity) illustrated in Figure 2. However, this probability cannot be quantified because the true heterogeneity present in the sample is not known—this is what the test is trying to infer statistically. Moreover, increasing  $N$  has practical limitations because it requires time and increases the risk that instrumental drift may become a significant contribution to the variations observed. Thus, choosing a reasonable  $N$  is an important task.

**Uncertainty budget.** A useful concept in homogeneity testing and helpful in choosing a reasonable  $N$  is the “uncertainty budget.” In its simplest form this relates the combined variance ( $s_c^2$ )



**Figure 4:** Results of the homogeneity studies on 36 apatite crystals. The composition of each crystal (shown as dots) was measured 30 times along a diagonal profile across the crystal. Shown is the contribution of heterogeneity to the total combined uncertainty of the average Ce mass fraction, either as absolute value ( $s_h$ ) or relative value ( $s_{h,rel}$ ). In 16 crystals significant heterogeneity was detected; they plot to the right of the critical values of  $s_h$  and  $s_{h,rel}$  (red solid vertical line), which can be computed from the critical homogeneity index  $H_{crit}$ . The ordinate shows the uncertainty of the uncertainty due to heterogeneity. It becomes very large below the critical values, which serve as a detection limit for heterogeneity. Increasing the number of measurements  $N$  decreases the expected uncertainty of  $s_h$  (dashed curves).

to its components comprised of the variance due to heterogeneity of the sample ( $s_h^2$ ) and the variance due to the measurement process, which, in this case, is reduced to the Poisson variance:

$$s_c^2 = s_h^2 + s_{Pois}^2 \quad (\text{Equation 3})$$

Based on this simple relationship, it is possible to define the relative contribution of heterogeneity to the total uncertainty of a reference value ( $s_{h,rel}$ ), a parameter that can be related to the homogeneity index:

$$s_{h,rel} = \frac{s_h}{s_c} = \sqrt{1 - \frac{1}{H^2}} \quad (\text{Equation 4})$$

The parameter  $s_h$  provides information on the extent of the compositional variations of the sample. The interval  $\pm 2s_h$  around the measured mass fraction of an element covers about 95% (2 sigma) of the sample's compositional variations. Because  $s_h$  and  $s_{Pois}$  cannot be separated with absolute certainty,  $s_h$  contains some uncertainty that depends on the number of measurements  $N$  (this is shown in the Results section). This approach is limited by the non-linear relationship between  $H$  and  $s_{h,rel}$  resulting in a bias that yields apparently large contributions of compositional heterogeneity even if materials are almost perfectly homogeneous [7]. In such a case only an upper limit of possibly present heterogeneity can be stated: a detection limit of heterogeneity. At  $N=10$  a homogeneous sample that passes the homogeneity test ( $H < H_{crit}$ ) may show an apparent contribution of heterogeneity to the uncertainty budget of up to 68%. Similar to stating the detection limit for a non-detected element, this  $s_{h,rel}$  indicates the upper limit of relative heterogeneity that may

be present. To obtain this number,  $H_{crit}$  is calculated as a scaled quantile of the chi-squared distribution with a significance level  $\alpha = 0.05$  and  $N - 1$  degrees of freedom via Equation 2. If  $H_{crit}$  replaces  $H$  in Equation 4, then the upper limit of  $s_{h,rel}$  results, and the percentage can be calculated by multiplying it with 100%. Because the uncertainty of the homogeneity index is primarily determined by the number of measurements  $N$ , this limit can be reduced by increasing  $N$ . In order to state that the contribution of compositional heterogeneity to the total uncertainty budget is less than 30%, a homogeneity test with  $N = 577$  measurements has to be passed ( $H_{crit} = 1.048$ ). For 20% this number increases to more than 3,000 measurements.

The homogeneity index and the uncertainty budget not only allow

decisions of whether or not heterogeneity has been detected—in the sense of a detection limit for elemental variations instead of the mere presence of a chemical element—they allow the quantification of heterogeneity if it is detected. Because  $s_c$  and  $s_{Pois}$  are obtained from the measurements,  $s_h$  can be calculated easily from Equation 3.

## Results

Figure 3 shows silicon and cerium x-ray intensity maps of one crystal of the fluorapatite candidate reference material. In this example the distribution of Ce was studied, which replaces Ca in a coupled substitution with Si (replacing phosphorus) or Na (replacing calcium) for charge balance. While the variation in Si is clearly visible, the significance of heterogeneity in the content of Ce is difficult to judge by the eye.

Figure 4 displays the wavelength-dispersive electron probe results of the Ce content of the apatite crystals. Each location on the polished crystal section was analyzed only once. Out of 36 crystals, each measured at  $N = 30$  locations, 16 samples showed detectable heterogeneity of Ce under the measurement conditions used. The critical homogeneity index and the corresponding values of  $s_h$  and  $s_{h,rel}$  define the level at which heterogeneity becomes significant at the  $\alpha = 0.05$  level of confidence. Below  $H_{crit}$ , the uncertainty of the heterogeneity contribution  $s_h$  to the total combined uncertainty strongly increases. A minimum of uncertainty at any given  $N$  is obtained if  $s_h$  equals  $s_{Pois}$  and the overall uncertainty can be reduced by increasing  $N$ .

One apatite crystal marked in Figure 4 showed an average Ce content of 0.215 wt% with a standard deviation ( $s_c$ ) of 0.020 wt% (crystal s6.3 in Table 1). In this case, heterogeneity among the 30 measurement locations could not be detected: the data point is to the left of the critical homogeneity index (red vertical line), which was derived from Equation 2 ( $N = 30$  and  $\alpha = 0.05$ ) and converted to  $s_{h,rel}$  using Equation 4. Therefore, it is possible to state that the uncertainty due to heterogeneity is less than

**Table 2:** EPMA homogeneity study of a single fluorapatite crystal (crystal s7.5).

Point No. ( $i = 1 \dots N$ )	Ce concentration (wt%)	$s_{Pois,i}^*$ (wt%)	$s_{Pois,i}^2$
1	0.385	0.0204	0.000417
2	0.374	0.0205	0.000419
3	0.408	0.0208	0.000434
4	0.355	0.0207	0.000426
5	0.432	0.0207	0.000430
6	0.431	0.0207	0.000428
7	0.394	0.0207	0.000430
8	0.406	0.0206	0.000423
9	0.383	0.0206	0.000425
10	0.337	0.0204	0.000414
11	0.389	0.0210	0.000439
12	0.340	0.0205	0.000421
13	0.407	0.0208	0.000432
14	0.469	0.0212	0.000447
15	0.431	0.0209	0.000438
16	0.401	0.0205	0.000421
17	0.362	0.0208	0.000434
18	0.410	0.0207	0.000430
19	0.404	0.0207	0.000428
20	0.358	0.0208	0.000432
21	0.378	0.0207	0.000430
22	0.375	0.0203	0.000412
23	0.371	0.0207	0.000428
24	0.346	0.0206	0.000423
25	0.341	0.0204	0.000414
26	0.338	0.0207	0.000427
27	0.343	0.0208	0.000431
28	0.339	0.0203	0.000412
29	0.315	0.0202	0.000410
30	0.309	0.0198	0.000391
Mean concentration (wt%)	0.378	$s_{Pois}^{2**}$	0.000425
Std. deviation $s_c$ (wt%)	0.0380	$s_{Pois}$ (wt%)	0.0206
$H = s_c / s_{Pois} = 1.84$ $s_h = \sqrt{s_c^2 - s_{Pois}^2} = 0.032 \text{ wt\%}$ $s_{h,rel} = 100\% \times s_h / s_c = 84\%$			

\*Individual standard deviations calculated from counting statistics (see [7]). \*\*Mean of all individual  $s_{Pois,i}^2$ .

56% of the total uncertainty (corresponding to  $s_h < 0.013$  wt%). Table 2 shows the detailed analysis of another crystal (crystal s7.5 in Table 1) that showed an average Ce content of 0.378 wt% with a total uncertainty of 0.038 wt%. In this case heterogeneity was significant (in Figure 4 to the right of the critical value). The contribution of heterogeneity to the total uncertainty was determined to be 84% (corresponding to  $s_h = 0.032$  wt%). The dashed curves in Figure 4 display the standard deviation of  $s_h$ , which can be approximated as:

$$S(s_h) = \frac{s_c^2}{s_h} \frac{1}{\sqrt{2(N-1)}} \quad (\text{Equation 5})$$

These curves illustrate that below the critical value (left of the red line) the uncertainty strongly increases. Hence, the critical homogeneity index and the parameters derived from it, as explained above, serve as a detection limit for heterogeneity at a chosen level of significance given by  $\alpha$ .

### Discussion

The example given in Figure 4 shows that the quantification of heterogeneity needs to consider a detection limit. Obtaining the uncertainty due to heterogeneity  $s_h$  from the uncertainty budget offers a quantitative parameter, but the significance of  $s_h$  and its uncertainty strongly depend on the number of analysis spots and the analytical precision. For example, a short counting time leads to a poor precision and to a large contribution of counting statistics to the uncertainty budget. In this case the relative contribution of heterogeneity to the uncertainty budget is low, and heterogeneity may be below the detection limit discussed above. Because compositional heterogeneity is a fixed characteristic of the sample, an increase in counting time or beam currents will decrease the counting statistical uncertainty and increase the relative contribution of heterogeneity to a value above the detection limit. Hence, perception of homogeneity or heterogeneity depends on the analytical method of investigation and its conditions of operation. An EPMA routine tailored for a more precise determination of Ce in fluorapatite might detect heterogeneity even in those crystals that passed the homogeneity test in the example above. The statistical evaluation discussed is potentially applicable to a wide range of analytical techniques. A standard suitably homogeneous for electron probe x-ray analysis would be adequate for methods with poorer spatial resolution and similar precision but may require re-testing for use with a method with better spatial resolution (for example, thin specimen analysis in STEM) or better precision at low concentrations (for example, SIMS).

From a practical perspective the correct determination of  $s_{Pois}$  can be a challenge. In the case of electron probe microanalysis it is not sufficient to just consider the counts above the background because the counts of the background measurements also contribute to  $s_{Pois}$  [7]. At high precision, matrix effects may produce apparent heterogeneity of a homogeneously distributed element A if a heterogeneously distributed element B affects the measured x-ray intensity of A. Matrix correction procedures should account for this.

### Conclusion

In electron probe x-ray microanalysis, homogeneity testing using the homogeneity index based on counting statistics offers the advantage that only one analysis per sample spot is required. Methods requiring multiple measurements on the same spot are prone to errors from degradation of beam-sensitive samples such as many geological materials and glasses. The sound statistical evaluation shown here is applicable to all kinds of reference materials (alloys, glasses, natural and synthetic crystals, etc.) and may be applied to other microanalytical techniques in which photon or ion counting takes place. The level at which heterogeneity may be detected and quantified depends strongly on the number of measurements  $N$ .

### References

- [1] TPJ Linsinger et al., *Accredit Qual Assur* 6 (2001) 20–25.
- [2] RB Marinenko et al., *National Bureau of Standards Special Publication* 260–65 (1979).
- [3] FR Boyd et al., *Carnegie Institution of Washington Year Book* 67 (1967) 210–15.
- [4] E Jarosewich, JA Nelen, and JA Norberg, *Geostandard Newslett* 4 (1980) 43–47.
- [5] E Jarosewich and IG MacIntyre, *J Sediment Petrol* 53 (1983) 677–78.
- [6] P Carpenter et al., *J Res Natl Inst Stan* 107 (2002) 703–18.
- [7] D Harries, *Chemie der Erde* 74 (2014) 375–84.
- [8] JC Stormer et al., *Am Mineral* 78 (1993) 641–48.
- [9] V Liebich et al., *Fresen Z Anal Chem* 335 (1989) 945–53.

MT

## Do You Know Your SEM X-ray Analysis Accuracy?

- Most SEMs with X-ray analyzers in use today incorporate energy dispersive systems.
- Standardless analysis, the most often used method, produces results very quickly, but do you know how accurate it is (hint- this likely varies with different sample compositions)?
- One way to measure accuracy is to analyze calibration standards as unknown samples.

At Geller MicroAnalytical Laboratory, we offer some 295 calibration standards that include pure elements, compounds, minerals, alloys and NIST standards all properly prepared for immediate use. Please visit our web site for listings and prices then feel free to call us for advice.

We also manufacture and certify stage micrometers and magnification standards under ISO-17025 accreditation.

ISO-9001 certified for 16 years!



**GELLER  
MICROANALYTICAL  
LABORATORY, Inc.**

426e BOSTON ST., TOPSFIELD, MA 01983-1216  
TEL: 978 887-7000 FAX: 978-887-6671  
www.GellerMicro.com