

## **Data Reduction Guide – Understanding EPMA Analytical Results**

### **The goal of data reduction:**

The goal of data reduction is to transform raw, collected X-ray intensities into concentrations or mass-fraction *estimates* of analyzed elements, and to be able to assign some statistical significance to the analytical results (e.g. precision, accuracy, standard deviation, detection limit, lower limit of determination). In this guide, we will be addressing only those data obtained via WDS (wavelength-dispersive spectroscopy) analysis, and not from EDS (energy-dispersive spectroscopy) analysis. This guide assumes that the reader has a basic understanding of how an electron microprobe functions, and of X-ray generation and behaviour as stipulated by the Bragg Law.

In order to understand what the data mean, we need to understand, a) the assumptions for analysis and the validity of said assumptions, and b) how X-rays are generated within the sample, i.e., what events occur within the target as high-energy electrons from the electron source interact with elements in the target.

### **General assumptions of EPMA (electron-probe microanalysis)**

These are some of the general assumptions that are made regarding electron-probe microanalysis, although this list is not exhaustive: a) One assumes that the specimen to be analyzed is homogeneous to a depth and width that exceeds the penetration depth and excitation volume of the electron beam. b) The intensities obtained from all elements in the sample will be compared to those obtained from primary standards. c) Both the samples and standards are smooth and flat, having no pore spaces, inclusions, voids or fractures; standards are assumed to be compositionally homogeneous and not zoned with respect to composition. d) Both the samples and standards are similar in composition and crystallographic structure, so as to minimize matrix effects. e) Both samples and standards are conductive at the surface and have equally thick carbon-coats, and experience no net static charging phenomena, i.e., both are adequately grounded, electrically. f) The incident electron beam strikes the target at an angle of 90°.

### **Validity of assumptions in EPMA**

We will briefly examine the validity of each of the assumptions listed above, to give the reader a glimpse of the problems encountered when trying to understand the reliability of the collected intensity data. A) Penetration depth – When analysing typical petrographic thin

sections of standard thickness (30  $\mu\text{m}$ ), one can find two or more distinct mineral species present through the thickness of the sample, thereby making it possible for the electron beam to excite X-rays in more than one phase, simultaneously. The intensity data from such a sample will produce a composite analysis, at best. This problem can be magnified in microprobes where only reflected light images are observed, and the user is forced to depend upon the SEM image only, which image reveals almost nothing about the superposition of vertically adjacent mineral species. This problem is not as common as one might initially think it is. With respect to thin-film samples (e.g., 100 to 300 nm thick), the shape of the excitation volume has to be altered by choosing the correct accelerating voltage and beam diameter to as to not have the beam penetrate the whole thickness of the film.

B) The use of standards – In the case of quantitative WDS analysis, the intensities of X-rays from the sample normally are compared to those from a primary standard, or to an internal control or secondary standard. Not always are primary standards available, although they are the preferred materials upon which to make a comparison of X-ray intensity data. Ideally, the concentrations of trace elements in a standard should be as close as possible to those of the unknown, but owing to the very nature of the analysis, we don't always have *a priori* knowledge about the concentrations of trace elements in our unknowns.

C) Sample and standard homogeneity – Although we try to use standards that are homogeneous with respect to composition and texture, sometimes there are internal flaws, such as cracks, that we can't always see, especially for natural materials that are somewhat brittle, such as garnet. We assume that the standards are homogeneous in composition on the micron scale, and this is generally true for major elements. Volatile species, such as F, Cl, OH and  $\text{H}_2\text{O}$  and trace elements are not always uniform in distribution over the range of several microns, even within standards. Natural samples are almost always physically imperfect, and have internal flaws such as pores and inclusions, both of which can lead to anomalous diffraction of X-rays and even composite analytical data. Some samples are also inhomogeneous on the micron scale with respect to composition; in those cases, it is inevitable that the data will reduce to a composite analysis.

D) Compositional and structural similarities – Ideally, in order to reduce or eliminate matrix effects in the data reduction process, it is best if the unknown sample and the applicable standards be as close as possible to one another in chemical composition, relative elemental ratios, elemental valence states and molecular structure. For most major elements, mineral standards generally work well if they are of the same structural class as the minerals being analyzed, as the bonding environments are similar for most of the elements involved. In some cases, we may not have a standard that is a structural analog of the types of unknowns we are analysing, and for those cases, matrix effects may have a strong effect on data reduction.

E) Electrical conductivity - Many types of samples are poor electrical conductors, and must, therefore, be coated with a conductive medium like carbon. Although a carbon-coated sample may be conductive overall, there is no guarantee that every grain is electrically grounded; there may still be localized charging that can lead to

reduced X-ray intensities from that grain, giving rise to low analytical totals. These deleterious effects are made worse by rough samples that aren't well-polished and clean before being coated, or samples that have lots of pores exposed at the surface. Differences in thicknesses of the carbon coat between the standard and unknown can lead to anomalous X-ray intensities and totals; this problem is enhanced when analysing minerals like monazite, where the unknown sample may have a double-thick carbon coat to minimize sample damage at high beam currents and high counting times. F) Angle of incidence of electron beam – In most cases, we assume (or hope) that the electron beam meets the sample at a 90° angle, such that the take-off angle (TOA) of the X-rays is about 40°, which TOA allows for optimal X-ray detection by the spectrometers. In order to have the electron beam meet the sample at a normal angle, the sample has to be both flat (well polished), of even thickness and sit with its surface parallel to the stage mount. To put the situation into perspective, features with a relief of more than about 2 microns above or below the focal plane of the electron beam will yield anomalously low X-ray intensities owing to the angular difference in the TOA resulting from said relief. The same net result occurs from samples that have surfaces not parallel to the focal plane of the electron beam.

### **Electron interaction and related events within the sample**

Electrons interact with the sample in both inelastic and elastic modes. Inelastic collisions between incoming electrons and the sample result in the slowing down of the incident electrons – this phenomenon is called the 'stopping power' of the sample (or more specifically the stopping power of whichever element the electrons collide with). The inner shell electrons of target atoms are directly ionized by incident electrons colliding inelastically with the target to produce characteristic X-rays, which X-rays we then measure – this is the source of the main observed X-ray intensities. Elastic collisions between source electrons and target atoms result in the deflection of electrons from their original trajectories (called the 'back-scatter factor'), thus diminishing the energy available for the generation of X-rays. Both the stopping power and back-scattering effects are proportional to the average atomic number (Z) of the elements in a standard or unknown sample. Inner electron shells of atoms indirectly ionized by characteristic and continuous X-rays and by fast, secondary electrons cause a secondary, fluorescent X-ray emission (commonly called 'fluorescent X-rays', denoted 'F'), and the intensities of these X-rays can add to those produced by primary, inelastic collisions. Primary, characteristic X-rays from the target can be absorbed on their way to the surface of the target, before they can reach the X-ray detector (called 'absorption effects', denoted 'A'); said absorption effects reduce the 'expected', measurable X-ray intensity for given elements.

These three factors, Z (atomic number factor), A (absorption factor) and F (the fluorescence factor) taken together, represent what are called the 'matrix effects' of a sample, and they do affect the measured X-ray intensities.

Any adequate procedure of data reduction must take into account, either implicitly or explicitly all of the above-listed phenomena in order to preserve any sense of statistical significance, i.e., the physical phenomena must be as well characterized as possible, and described algebraically or included as some kind of constant within a process-algorithm. In other words, the observed X-ray intensities have to be corrected for the above-described 'matrix effects', Z, A and F in order to yield data that make sense.

### **The ZAF Model for Correction of Matrix Effects Upon Measured X-ray Intensities**

The prime method used for data reduction by the JEOL 8200 EPMA to account for the effects described above is the *ZAF Model*. This model assumes an initially linear relation between generated X-ray intensity and concentration of a given element (Castaing's approximation, 1951). Any deviation from the initial, linear relation is 'corrected' by a series of multiplicative factors that account for the effects of atomic number (Z – stopping power, back-scattering factor and X-ray production power), absorption (A) and fluorescence (F), each of which are calculated. Mathematically, the 'net', matrix-effect-corrected X-ray intensities start with a comparison of raw, measured and background-corrected X-ray intensities for an element in an unknown sample,  $I_{unk}$ , against that of a standard,  $I_{std}$ , with a known concentration of a given element, 'X'. The ratio of these two intensities is called the 'K-ratio', and is commonly denoted as 'K'. The ratio of the concentration of element 'X' in the unknown relative to that in the standard is proportional to the K-ratio; by using simple algebra, we can solve for the concentration of element 'X' in the unknown sample, or so it would seem. Thus,

$$\frac{I_{unk}}{I_{std}} \propto \frac{[X]_{unk}}{[X]_{std}}$$

where the K-ratio is equal to  $\frac{I_{unk}}{I_{std}}$ . The key issue to understand here, is that the K-ratio is only *proportional* to the ratio of respective concentrations in the unknown and standard. In order to achieve 'equality' (more accurately described as 'best fit') between the 2 sides of the equation, we need to mathematically integrate the matrix effect corrections, by multiplying the K-ratio by the Z, A and F corrective factors, as follows:

$$\frac{I_{unk}}{I_{std}} \times ZAF \cong \frac{[X]_{unk}}{[X]_{std}}$$

In essence, the ZAF corrections are only calculated and iteratively determined quantities – they are iterated over several cycles until they reach self-consistency. The Z-correction accounts for the fact that heavier elements produce and absorb more X-rays than lighter elements. From a physical point of view, this makes sense, especially in the case of binary compounds such as PbS, where there is a large disparity in Z between the constituent atoms. Even though the stoichiometry of PbS is 1:1 atomically, the Pb will produce more X-rays than S will in a PbS crystal, simply because the Pb is much larger than and has a more complex electronic configuration than S. Therefore, the uncorrected K-ratio would tell us, erroneously, that there is more Pb than S, on an atomic basis, and that the K-ratio needs to be corrected.

Both the absorption correction factor, A, and the fluorescence correction factor, F, are dependent on Z, and are highly influenced by crystalline structure and orientation of the structure in the electron beam. The absorption correction, A, is required to account for absorption of primary, characteristic X-rays that we want to measure on their way out of a sample – without the absorption correction, the intensities of certain X-rays will be apparently reduced, giving us anomalously low concentration results. The fluorescence correction, F, accounts for the secondary fluorescence of target atoms by primary X-rays (i.e. X-rays produced from collisions of electrons from our source with target atoms) or high-energy, secondary electrons (electrons ionized from target atoms). The fluorescence effects cause our measured X-ray intensities to be higher than expected, and result in falsely high concentrations and totals. In most cases, absorption effects outweigh fluorescence effects, owing to the concept of mass-absorption coefficients. In most substances, the absorption correction commonly dominates the fluorescence correction, especially in the case of heavy trace elements such as REEs, whose mass absorption coefficients (MACs) tend to be very large compared with those of lighter elements that may make up the bulk of the matrix in which the REEs are present (e.g., La, Ce in apatites). MACs will be discussed in more detail, later in this guide.

*The iteratively corrective ZAF data reduction model is used to transform raw data into results that are as **accurate** and as **precise** as possible.* Much of the remainder of this guide will discuss accuracy and precision of results, and the factors that affect those two parameters, as well as the basic counting statistics that apply to quantitative analyses. Ultimately, primary standards must analyze within analytical error in order to gauge the accuracy and precision of the analytical method and suitability of the correction procedure. The question then becomes, “What is an acceptable analytical error?”. In order to answer this question, we need to recognize sources of error and find ways to minimize them, or at least account for them mathematically.

## Accuracy and Precision

Before we discuss statistical, analytical error, we will set out definitions for *accuracy* and *precision*. Accuracy is the absolute description of the ‘truth’ of the analysis – how close the reported analytical value of the concentration of a given element is to the ‘real’ concentration in a sample. In a target-shooting analogy, if you are aiming at the bull’s eye, your accuracy is measured by how close your shot comes to the bull’s eye – the closer your shot is to the bull’s eye, the better your accuracy is. In contrast, precision is defined as the relative description of how reproducible a result is, and is directly tied to counting statistics. In our target shooting analogy, precision can be defined as how tight the grouping of several shots is – the closer the shots are to one another, the higher is the precision. Ideally, we would like for our analytical data to be as accurate and as precise as possible, or as close to the ‘real’ value as possible, while simultaneously as reproducible as possible (i.e., with as little difference as possible between repeated trials on the same analytical point). Commonly, the precision will be expressed in terms of *standard deviation* when our sample size ( $n$ ) is 1, and as *standard variance* when the sample size is greater than 1 (and most commonly where  $n \geq 3$ ). These terms will be defined algebraically later in this document. Before we discuss the statistics in detail, it is best to have at least a qualitative understanding of all the common factors that effect the intensities of our measured X-rays. It is of vital importance to keep the idea of the intensities of the X-rays in mind, and the phenomena that affect said intensities, also, rather than getting lost in the algebraic manipulations. Once we have established an understanding of the factors that affect the X-ray intensities, we can assign a prioritized, statistical magnitude or importance of said factors to their effects on X-ray intensities.

### Factors to consider in achieving good accuracy and precision

Achieving the best possible accuracy and precision doesn’t happen by accident, rather it happens by understanding physical phenomena, modelling said phenomena mathematically and designing the experimental analysis in an intelligent way. The following list describes many of the primary factors that need to be considered when optimizing accuracy and precision – when these factors are not well-controlled, errors in counting X-rays may result (***N.B.*** *Much of the following information has been compiled from various sources that are referenced at the end of this document, and the remaining sections are based on public domain PowerPoint presentations constructed by and rightfully attributed to faculty members at the University of Wisconsin, Madison, Dept. of Earth Sciences*) as follows:

- *Standards* – The choice of standards is critical in X-ray analysis. Ideally, a standard is of the same structural and compositional type as the unknown, is compositionally

homogeneous and is well-characterized so as to minimize the magnitude of matrix effects. One of the problems with natural standards is that they are not always compositionally homogeneous, especially with respect to volatile species, such as F, Cl, OH and H<sub>2</sub>O. In some cases, such as with apatite, the crystallographic direction along which the electron beam ionizes the target atoms influences the resultant, net X-ray intensities. If the direction along which standard data are collected differs from that along which an unknown is analysed, an inaccurate concentration may result.

- *Standard and sample physical condition* – The physical conditions of standards and samples influence the homogeneity of X-ray production, particularly where the quality of the surface polish (or roughness) is concerned. Excessive roughness or undue surface relief will cause anomalous scattering of X-rays, potentially reducing the accuracy of the analysis. Also, the setting of the sample in the sample-holder may lead to the tilting of the sample – any tilt will cause the sample to point preferentially towards some WDS spectrometers whilst pointing away from others, thus altering the measured X-ray intensities. This type of error may appear to be systematic, as elements are assigned to be measured in one spectrometer only, under most common analytical conditions (ACO). The thickness of the carbon-coat is commonly uniform across small surfaces and is likely to be uniform across a given sample, but it may not be the same on a standard compared to an unknown sample. A difference in coating thickness between the standard and unknown may result in differences of absorption of emergent X-rays from the respective surfaces, thus altering the K-ratio and the accuracy of the analyses. Such differences in thickness accentuate differential X-ray absorption for lighter elements and the problem is further compounded in cases where low accelerating voltages are being used. Also, some samples are more sensitive to beam damage than others, and this can also result in anomalous absorption or emission of X-rays.
- *Instrumental stability* – The stability of a microprobe generally refers to the beam stability, and more specifically the constancy of the beam current and positional stability (i.e., the beam hits the same relative spot with respect to the stored X-,Y- and Z-coordinates of a sample). The beam stability is affected by stability of the power source, humidity effects on electrical components and age of the filament. The shape of the filament and its electron-emission profile change as the filament ages, both of which factors affect the shape and relative location of the excitation volume in the target, which in turn affects X-ray production. Another aspect of instrument stability is the spectrometer reproducibility, which can have a huge effect on both the accuracy and precision of the measurements taken. The thermal stability of the environment in the laboratory is crucial, as temperature fluctuations in the room will cause metallic parts such as spectrometer drive belts to expand and contract over several hours, thus causing the diffractors and detectors to be off-peak when measuring peak X-ray

intensities and background intensities (i.e., peak drift). Such instabilities will have detrimental effects on the accuracy and precision of measured data. Even something as simple as turning off the lights in the lab overnight can induce enough of a temperature change to affect X-ray peak positions and intensities.

- *Beam effects on sample* – In general, the electron beam does more to the sample than just generating X-rays. The beam can heat the sample locally and drive off volatile species. The beam can also change the oxidation state of sensitive elements, as well as causing light alkali elements such as Na to migrate away from the centre of the beam, both of which result in reducing the X-ray intensities generated by the sample over the time the intensities are being collected. Whereas the beam is made from electrons, certain chemical moieties can capture some of these electrons and spontaneously become reduced with respect to oxidation state. One example of this process is the reduction of 2 adjacent  $\text{CO}_3^{2-}$  anions in calcite into  $2 \text{CO}_2(\text{g}) + 1 \text{O}_2(\text{g})$  molecules, resulting in the physical damaging of the calcite structure.
- *Spectral line- and background-overlaps*- Specific X-ray spectral lines are unique in wavelength for every element, e.g., all first-order  $\text{K}_\alpha$  X-ray-wavelengths are unique for each element in the periodic table. However, that is not to say that all X-rays produced by all elements in a sample are unique and don't overlap to some degree. Elements in our sample that are irradiated by a sufficiently energetic electron beam will produce all possible X-rays for said elements that have absorption-edge energies below the accelerating voltage being used. Some of these X-rays, especially higher-order ones, will almost always overlap the wavelengths of X-rays that we are trying to measure. We refer to these interfering X-rays as 'peak overlaps' or 'spectral interferences', and they inevitably produce false counts for the X-ray lines we are measuring. For example, in a stainless steel sample that has both Ti and V, basic quantum X-ray theory tells us that the Ti  $\text{K}_\beta$  X-ray overlaps the V  $\text{K}_\alpha$  X-ray to a significant degree. Even though we aren't actively measuring the Ti  $\text{K}_\beta$  X-ray as part of our analysis, because it overlaps the V  $\text{K}_\alpha$  X-ray position in our spectrometer, it will be measured as V  $\text{K}_\alpha$  X-ray intensity if any Ti is present and even if V is absent from part of the steel alloy, resulting in the false reporting of V in the analysis. Commonly, 'unwanted' X-rays of both low and high order, usually minor peaks, will overlap the positions at which we are measuring background counts for the X-ray of interest that we are currently measuring. These are called 'background overlaps', and they can result in sloping backgrounds by making one side of the background 'curve' higher than the other. A sloping background commonly results in the reduction of measured counts for the X-ray we are trying to quantify, which then affects the accuracy of our data.
- *Counting statistics and analytical strategy* – Given that the 'resolution' or 'sensitivity' of an analysis is proportional to  $1/\sqrt{n}$ , where  $n$  = the number of samples or counting time,



we have to choose adequate counting times, especially for the backgrounds, to achieve the desired level statistical significance. Commonly, the statistical significance is quoted as the number of standard deviations ( $\sigma$ ) above the background that an X-ray intensity is; normally, in order for the X-ray intensity to be called 'real' (or statistically significant), it must be above  $3\sigma$  above the background noise. So, we must choose our counting times carefully, in order that we may have statistical confidence in the collected data. Such decisions become more important when we analyse trace elements that are present in quantities below 1 wt %, where the level of precision can approach that of the accuracy of the analysis, and therefore analyses of trace elements can require counting times an order of magnitude greater than those used for major elements.

- *Matrix correction factors* – Although we have discussed briefly the concept of matrix corrections factors, i.e., ZAF factors, we have to mention the concept of mass absorption coefficients, MACs for short. MACs basically represent the rate at which given elements can absorb radiation, and are typically expressed in units of  $\text{cm}^2/\text{g}$ . MACs vary with the average density of the target substance and with the wavelength ( $\lambda$ ) of the incident radiation. In general terms, for a given series of X-ray emission lines below an certain absorption edge energy (which varies with increasing atomic number), MACs increase with increasing  $\lambda$ . Basically, the total secondary absorption of a given matrix is given by the equation  $\mu_{\text{matrix}} = \sum_i (\mu_i w_i)$ , where  $\mu_i$  and  $w_i$  are the individual MACs and weight fractions, respectively, for each element  $i$  that make up the matrix. The overall absorption,  $\mu_{\text{matrix}}$  has to be multiplied by the average density,  $\rho$ , of the matrix in order to account for the absorptive behaviour of the sample volume on X-ray intensities. Given that the MACs for light elements are generally much smaller than those for heavy elements, in samples where there exists gross disparity in atomic numbers of constituent atoms, the effect of the lighter atoms will have little influence on the overall X-ray absorptive characteristics of the sample unless they are present in large concentrations. In the case where trace elements are of relatively low  $Z$  and low concentration, such as with fluorine in certain silicates, the ZAF correction process may less-than-accurately estimate the correction factors required to achieve overall accuracy in the analysis.
- *Sample size relative to excitation volume* – In most cases where the sample thickness exceeds 30 microns, and the sample isn't made up of mostly very light elements packed in loose structures, the sample thickness will exceed the depth of penetration of the electron beam, as well as the depth of the excitation volume produced by the incident electron beam at relatively high accelerating voltages (e.g., 15 kV). In the case of thin films or thin mineral 'slices' that may be only 100 – 200 nm thick, a relatively high accelerating voltage will cause the electron beam to penetrate the sample thickness entirely, causing excitation of whatever material is below the intended target. At best,

this phenomenon will result in a composite analysis of the target and the substrate underneath the target; at worst, the beam will destroy the target area. Various workers have modelled the X-ray generation characteristics of substances by altering the thickness of successive layers of a target substance laid down on a substrate of known composition that is thicker than any penetration depth associated with the maximum accelerating voltage of their electron probe. In doing so, they have measured what is called the 'mass-depth profile' of given elements, or the amounts of X-rays generated at specific analytical conditions with increasing depth – such methods empirically account for absorption and fluorescence phenomena. This general type of correction process is called the 'phi-rho-zeta' ( $\phi(\rho z)$ ) correction model – the ZAF data reduction model is a  $\phi(\rho z)$ -based model, with specific values being assigned to the 3 correction factors (not to be confused with the very specific  $\phi(\rho z)$ -correction model detailed in scientific literature, also commonly referred to as 'PROZA').

### **Categorization of the types of errors affecting accuracy and precision**

Now that we have examined many of the factors that affect accuracy and precision, we will categorize the types of errors that result from said factors, as well as mentioning other, practical implications. In general, there are several sub-types of errors, and sometimes, the individual factors we have already discussed can belong to more than one category. The subtypes of errors stem from the 3 main sources of error, namely instrumental issues, sample- and preparation-related sources, and user error. The subtypes and examples of errors are discussed, as follows:

**Random Errors** – Random errors are not mathematically or statistically predictable, nor are they generally able to be corrected. They are like lightning strikes, in that their timing, intensity and location aren't able to be well predicted. Among random errors in microanalysis are such factors as:

- *X-ray generation* – There is no guarantee that a given electron-atom collision is going to result in the emission of a characteristic X-ray wavelength. Only electrons on the exact ideal trajectory that retain enough kinetic energy are able to penetrate into the inner electron-shell of a target atom and ionize it such that a characteristic X-ray is produced. Incident electrons that strike a target atom at random, glancing angles will produce non-characteristic, 'continuum' X-rays almost up to the energy level of the incident beam (i.e., accelerating voltage) with a Gaussian distribution. Other phenomena occur during random electron collisions, such as the generation of cathodoluminescent photons, phonons, electron capture, generation of Auger X-rays, charging and other vibrational modes.

- *Electronic instrumental instability* – We have discussed this issue previously, but will mention a few relevant points. X-ray intensity peaks are actually 3-dimensional, and we sample a 2-dimensional slice of those 3-D objects/peaks. The orientation of the slice we sample is governed in part by the orientation of the diffractor crystals and detectors in our spectrometers, which are in turn partly controlled by the alignment of the ‘baseplate’ in the spectrometer; in most JEOL microprobes, the baseplate is a large, brass plate to which the moving parts of the spectrometer are attached. If the alignment of the baseplate changes between the time that the standards were last calibrated and the current samples are analysed, the 2-D X-ray ‘slice’ we sample may change shape and orientation, resulting in erroneous X-ray intensities being collected, and both the accuracy and precision will be detrimentally affected. The baseplate may change alignment in response to the spectrometer being bumped, for example, resulting in the generation of random errors.
- *User error* – This is one of the most subjective errors to judge, as they vary from individual to individual. For example, an inexperienced user may select ‘poor’ points, e.g., pitted points, poorly polished points, chemically altered points, grains that are too small, points that are too close to a grain margin. The most common user error is setting the optical focus of the sample such that the TOA is optimized for the Bragg diffraction condition to be satisfied for the detection of in-phase X-rays. Unfortunately, errors in focussing the optical microscope may be a result of user needing glasses, or because the user is overtired or even ill. The human factor is the most difficult one to rectify, but can’t be ignored, despite assertions from a prideful user or analyst.
- *Surface roughness* – We have discussed this issue previously, but it bears repeating here, in that surface roughness on the micron-scale can affect the ability to focus the optical microscope well, and may therefore introduce a random error in the intensity data.
- *Two or more phases within the excitation volume* – Given that the secondary, back-scatter and topographic electron images generated by the probe only see physical phenomena at the surface of a sample, and are unable to reveal sub-surface features such as inclusions, shallow grain-boundaries or internal fractures and flaws, it is not uncommon that a given excitation volume may contain more than one compositional entity, resulting in a composite analysis. Also, internal flaws such as fractures, dislocations, coalesced voids and convoluted grain boundaries will cause anomalous deflections of incident electrons and anomalous diffraction of X-rays, thus resulting in erroneous intensity data.
- *Secondary fluorescence* – Secondary X-ray fluorescence caused by primary X-rays, including self-absorption-and-re-emission is generally a randomly occurring phenomenon, and is not readily corrected. One example of false secondary

fluorescence is that caused by radioactive elements in the sample – radiation from U, for example, may be energetic enough to induce secondary fluorescence in surrounding, lighter elements.

- *Counting errors* – Counting errors can arise owing to random, electronic phenomena in the signal-processing hardware, and even within the detector itself, for example, stray cosmic rays can be detected by the detector and reported as analytical data. This may sound trivial, but can be a common effect when one is operating a microprobe near the earth's magnetic pole. I have seen this effect, personally, when I worked at the University of Manitoba, and the lab was hit by a burst of cosmic rays while the *aurora borealis* were in play.

**Systematic Errors** – Systematic error is a series of errors in accuracy that are consistent in a certain direction, or over time. In general, systematic errors cause a bias in measurements that result in an 'offset' of measured data points from the 'true' value of the quantity being measured. Systematic errors are commonly harder to detect than random errors, and fall under several sub-categories, as follows:

*Instrumental errors:*

- *Thermal instability of instrument* – Linear variation in temperatures in the laboratory can have systematic effects on the d-spacing of the diffractor crystals in the WDS spectrometers, potentially affecting the accuracy of the analysis. Thermal changes can also affect the gas pressure of the Ar-CH<sub>4</sub> gas flowing through the flow-proportional counters, thereby changing the efficiency of the detector, and changing the accuracy and precision of the data. Temperature variations can also cause significant drift in Z-dimensional control of the stage, which in turn can lead to the optical defocussing of the sample and subsequent decreasing of the measured X-ray intensities.
- *Less than ideal matrix correction* – If an inappropriate standard has been selected for a given element, the resultant matrix correction may not be optimal, leading to an erroneous calculation of concentration from a measured X-ray intensity.
- *Charging of sample or standard* – If the carbon-coat is insufficient to permit proper grounding of the sample, the measured X-ray intensities will be systematically reduced for all elements being analyzed, and will, therefore, affect the accuracy of the analyses.
- *Beam damage* – As discussed above, the electron beam can damage samples, depending on the composition of the samples, for example, the reduction of carbonates, and alkali migration in Na- and K-rich samples. In such cases, reduction of the X-ray signal with increasing time results in lower average intensities and calculated concentrations. Such phenomena can be corrected using time-dependent-intensity

(TDI) corrections, whereby theoretical 'temporal titration' curves are calculated to extrapolate back to what an X-ray intensity would have been at measurement-time = 0.

- *Peak shift* – Over long analysis times, the peak positions of X-rays can drift or shift slightly; in order to systematize the errors induced by such drift, users collect X-ray data by analysing standards as though they are unknowns. If the concentrations yielded for an element(s) in the standard vary significantly over time, one may deduce that the peak-position(s) has drifted and can correct for the loss in signal intensities.
- *Beam current stability* – The stability of the beam current can change over time for a variety of reasons, some of which are out of the user's control, for example, the changing of a heating/cooling cycle by a building's climate-control system. Commonly, this problem is fixed (or at least tracked) by the software used to control the microprobe, as the X-ray intensities are expressed in units of counts/second/unit area (beam spot-size)/nano-amp, i.e., the counts are normalized based on a live measurement of the beam current.
- *Spectrometer reproducibility* – As the moving parts of a spectrometer age, they will experience mechanical damage owing to the repetition of duty-cycles. As wear-and-tear takes its toll, the goniometers may drive to slightly different positions from the 'ideal' positions, potentially resulting in loss of signal in the long-term, thereby affecting both the accuracy and precision unfavourably.
- *Pulse height adjustment* – Periodically, the high-voltage and gains of the detectors must be adjusted owing to the aging of electronic components – this results in a change of measured intensities as signals are electronically amplified and 'reshaped', hopefully for the better, but this process still can introduce errors from time to time.
- *Dead time* – "Dead time is the time interval after a photon enters the detector during which the system cannot respond to another pulse...In order to obtain accurate values of the peak intensities, corrections must be applied for the dead time associated with the measurement of an X-ray...In the WDS spectrometer, the dead-time correction is applied after the intensity is measured. For the proportional counter used in WDS, [the following equation is used to correct for dead time, and this correction must be used]...before that intensity is used to form a K-ratio:  $N = N' / (1 - \tau N')$ , where  $N'$  is the measured count rate,  $N$  is the true count rate which we wish to calculate, and  $\tau$  is the dead time in seconds...." (Goldstein et al. 1992).
- *Spectrometer focus* - We have already discussed this subject above with respect to baseplate alignment, as the baseplate alignment does affect the focus of the spectrometer itself. If one of the steel belts inside the spectrometer stretches too much, or the tensioner wire that holds the weight of the detector stretches, the X-rays from the diffractor crystal will not focus properly through the slit in front of the detector, thus affecting both accuracy and precision of the intensity data.

- *Diffraction crystal deterioration* – Commonly, the diffractor crystals become scratched, cracked, contaminated and suffer beam damage over the long-term. These types of issues reduce the crystal's ability to diffract incident X-rays properly, which can lead to reduced intensities, and reproducibility of measurements (usually associated with a few elements, systematically, that have very similar Bragg angles/L-values).
- *Filament change* - Given that every gun-filament is different in physical shape, responds differently to being heated by the electricity applied to it, ages differently and has slightly different emission characteristics and therefore need to be individually aligned, errors in both accuracy and precision can occur as the filament ages. If a filament is changed in the middle of an analytical run, it is wise to recalibrate the primary standards, as intensity counts commonly improve with a new, well-aligned filament.
- *Window/crystal/detector/column contamination* – As the microprobe is used and samples are changed regularly, there are opportunities for contamination to enter the main chamber, and to contaminate Be-windows, diffractor crystals, detector windows, and the column itself. One of the most common contaminants is finger oils from users who handle their samples or sample-holders without wearing rubber gloves. These oils evaporate in the vacuum of the microprobe, and re-condense on the inside of the probe parts. To minimize these effects, the column needs to be cleaned about twice a year. Contamination within the column can affect the aim of the beam and focussing qualities of the beam, leading to errors in both accuracy and precision with time.

*Physical Standard and sample errors:*

- *Surface irregularities* – We have discussed this already, in a section, above.
- *Etched samples* – Samples that have been chemically etched or etched by a plasma beam/cleaner are commonly rougher than unetched samples, which leads to anomalous diffraction. This type of roughness is induced and controllable to a certain extent, which is why etching qualifies as a systematic error.
- *Polish* – In order to minimize surface effects, samples should be polished with a diamond, alumina or silica abrasive with a grain-size of  $< 1 \mu\text{m}$ .
- *Specimen homogeneity* – We have already discussed the effects of having more than one phase present in the excitation volume, above.
- *Non-orthogonal geometry* – As sometimes happens with grain-mounts that are made into an epoxy 'puck' format, and especially for those that are automatically polished, the surface of the mounts tends to slope in one direction, i.e, the surface is not perpendicular to the central axis of the puck. When these pucks are placed in ring holders, they almost always tilt as the screw is tightened, and the surface of the sample ends up inclined at an angle to the horizontal plane. X-rays emerging from the tilted sample won't proceed towards the diffractor crystal properly, as the Bragg condition is

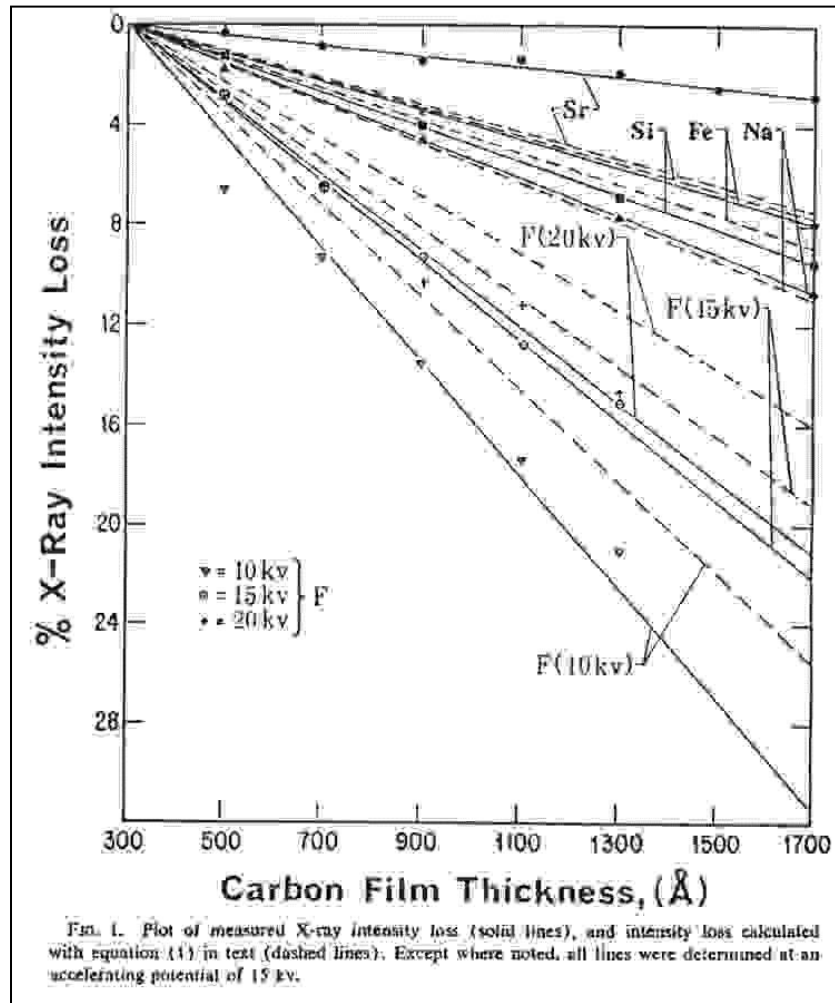
violated by the angle of tilting. Therefore, X-ray intensities will be reduced relative to a non-inclined, ideal geometry, giving rise to errors in the accuracy of the analysis.

- *Edge effects* – Materials mounted in epoxy and then polished with loose abrasives commonly develop channels or grooves in the epoxy at the sample-epoxy interface, which leads to rounding of the sample edge. The rounding of the sample results in a reduced absorption path-length of the beam from ideal/assumed, resulting in an erroneous analysis.
- *Oxide/oxidation coating* – For samples and standards that readily oxidize during storage (before analysis), such as Ti, Al, Mg, Mn, an oxide coating can reach a thickness of fractions of a micron, and result in a reduced intensity being collected for a standard – this results in an over-estimate of the amount of the appropriate element in the unknown, as the denominator of the K-ratio will be reduced below what it would be if the standard was not oxidized.
- *Polishing artifacts* – Polishing media can become trapped in pores in the sample surface, or embedded in samples that are considerably softer than the polishing media. I have seen this with 0.25-micron diamond particles from a polish becoming stuck in the surface of Sn-metal. Also, if Pb polishing laps are used to polish the sample, Pb metal from the lap can become trapped in pores or along grain boundaries – this leads to a nightmare if one is analyzing zircons or monazites, because the Pb from the lap may smear across the surface of a sample, giving rise to ‘false’ X-ray intensities for Pb, and potential reduced counts for lighter elements whose counts get absorbed by the Pb.
- *Smear coating* – Smear coating refers to the smearing of softer minerals across the surfaces of adjacent ones during the polishing process, causing contamination. This can happen with Sn and Pb in certain samples. This effect is more relevant for trace-element analysis than for major elements.
- *Carbon-coat thickness* – Ideally, to remove yet another variable/source of error from our analyses, it is best if the thickness of the carbon-coat is the same for our standards as it is for our unknowns. Kerrick *et al.* (1973) discovered that this effect of carbon-coating thickness was worse for light elements, and that in the case of fluorine, a difference in coating thickness of about 200 Å between the sample and standard resulted in a 4 % difference in X-ray intensity for the F K<sub>α</sub> X-ray (see next page for illustration from Kerrick’s 1973 paper).

#### *Procedural errors:*

- *Peak interferences* - We have already discussed the effects of interfering peaks (see “Factors to consider...”, above) and discovered that peak interferences result in falsely high X-ray intensities. The ‘false intensity’ caused by the overlapping X-ray peak must be removed by using an interference standard that has the interfering element in it,

without the interfered element, and an overlap coefficient may be determined by collecting X-ray signals from the interference standard at the peak position for the interfered element. The intensity of the interfering element is then subtracted from the measured intensity at the interfered peak position.



Kerrick *et al.* (1973) *American Mineralogist* 58:920-925

- *Background interferences* – As previously stated, background overlaps give rise to inflated background intensity counts, and given that background counts are subtracted from peak intensity counts to yield a net intensity count for a given element, an inflated background will lead to an erroneously reduced net intensity.
- *Peak shifts/shape differences* – Although we have previously discussed peak shifts/drifts over time within an instrument, there is a separate issue of peak shifts caused by the bonding environment of the element in question. It is well documented that the peak positions for Si and Al, for example, will be different between the pure metal, oxide and alumino-silicate phases. These shifts may have significant effects on both accuracy and



precision, especially when using standards that are dis-similar in bonding environment compared to unknowns.

- *PHA settings* – We previously discussed the readjustment of peak-height analyzer parameters owing to variation with time/aging of the electronics, but there is also a potential for systematic errors to creep in when using structurally or chemically disparate standards and unknowns. “Bias, gain, and baselines should be checked. Gross errors in them could produce significant errors in the analytical results. Pulse height depression occurs mainly where there is a large discrepancy in count rate between standard and unknown.” ([www.geology.wisc.edu/~johnf/g777/ppt/70\\_Errors.ppt](http://www.geology.wisc.edu/~johnf/g777/ppt/70_Errors.ppt))
- *Resetting goniometer* – Occasionally, the goniometer may jam as a result of trying to drive the spectrometer to an unreachable position, in which case the position of the goniometer may have to be re-initialized by doing a peak search on a given standard. This can introduce a difference in peak positions if the re-initialization procedure is done incorrectly, or if the error is not cleared properly and the goniometer is at one physical address/angle while the software value is different from the physical address/angle. This can result in a serious reduction of the accuracy of the analysis.
- *Resetting operational conditions* – On occasion, there may be a need to reset or reboot the microprobe hardware, firmware or software systems to clear an error, or because a catastrophic operational error has occurred, such as a power failure of the system. When the hardware reboots, it is possible that residual charges within the electrical system cause certain mechanical components (e.g., crystal-flipper or stage motors) to ‘twitch’ slightly and introduce positional errors in the spectrometers or stage positions, for example. Such small twitches are enough to introduce errors in X-ray intensities.

### **Counting statistics:**

Now that we have examined the causes of statistical errors both in accuracy and precision in some detail, we will now discuss the counting statistics themselves, and how they are derived from our intensity data to allow us to attach significance to our analyses – in essence, we are going to look at how we use statistics to tell us how good or bad our data really are. The following sections assume that the reader already has a basic understanding of statistics, and will not spend time discussing derivations of basic terms.

For those users who have data processed using the JEOL software, the following section talks about the algebraic and statistical treatment of intensity data to yield standard deviations and detection limits.

- Algebraic definition of ‘Standard Deviation’ in JEOL software (S.D. =  $\sigma$ ) expressed as %

$$S. D. (\%) = \frac{100}{I_{net}} \sqrt{\frac{I_{peak}}{t_{peak}} + \left(\frac{\overline{L_{PBH}}}{L}\right)^2 \frac{I_{PBL}}{t_{PBL}} + \left(\frac{\overline{L_{PBL}}}{L}\right)^2 \frac{I_{PBH}}{t_{PBH}}}$$

$$I_{net} = I_{peak} - \frac{I_{PBH}\overline{L_{PBL}} + I_{PBL}\overline{L_{PBH}}}{\overline{L_{PBL}} + \overline{L_{PBH}}}$$

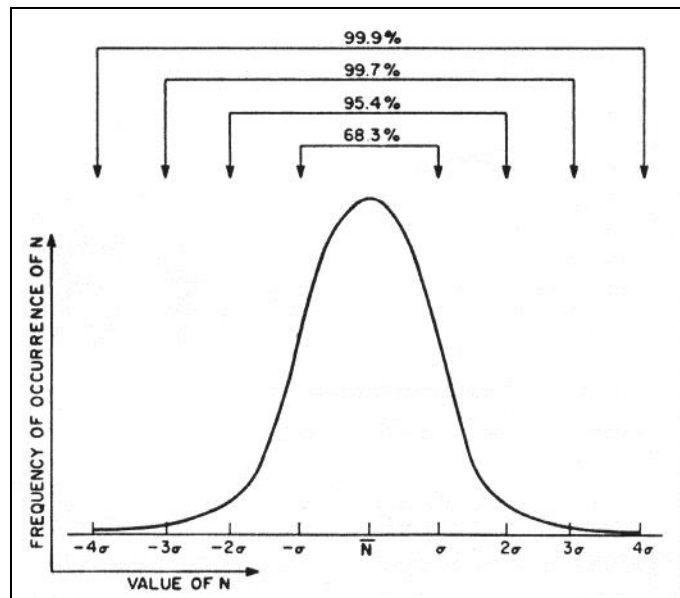
- where  $I_{peak}$  is the intensity of the X-rays measured at the peak position in question expressed in counts/second;  $I_{PBL}$  and  $I_{PBH}$  are the intensities of the background at both the lower and higher background positions, respectively, expressed in counts/second;  $\overline{L_{PBH}}$  and  $\overline{L_{PBL}}$  are the background offset positions (i.e., how far the backgrounds were measured away from the peak position) expressed in mm;  $I_{net}$  is the background-corrected intensity of the measured X-ray peak;  $t_{peak}$  is the counting time for the X-ray peak;  $t_{PBL}$  and  $t_{PBH}$  are the counting times for the lower and higher backgrounds, respectively;  $L$  is the total distance between the lower and higher background positions
- Detection Limit – The detection limit is defined as the smallest signal that is distinguishable from the background noise; in general, the signal must be  $3\sigma$  times greater than the background noise. The JEOL software calculates the Detection Limit (D.L.) as follows (quoted at  $1\sigma$  by default):

$$D. L. = \frac{1}{\frac{I_{NetSTD}}{mass(\%)_{STD}}} \sqrt{\frac{2 \times I_{back}}{t_{back}}}$$

- where  $I_{NetSTD}$  is the net intensity of a given X-ray in the standard;  $mass(\%)_{STD}$  is the mass % of the appropriate element in the standard;  $I_{back}$  is the average intensity of the lower and higher backgrounds;  $t_{back}$  is the total counting time of the lower and higher backgrounds.
- Lower Limit of Determination (LLD) – The LLD is defined as the lowest amount of analyte in a sample that can be quantitatively determined with a stated, acceptable precision and accuracy under stated experimental conditions. Generally, the LLD is quoted at  $6\sigma$  for most analytical methods, although the JEOL software does not utilize this parameter, rather it calculates a D.L. at a software-default of  $1\sigma$ , which can be misleading in that a user may assign undue significance to a concentration of an element that has been assigned an overly optimistic D.L. based on single-digit, Net Intensity counts that may not be above  $3\sigma$  of the background noise. The LLD is mentioned here, mostly because most users don't take this quantity into account, and take the DL as being the be-all-and-end-all, without assigning a statistical importance to the DL.

For those users who have used or will use the newly-implemented Probe for EPMA software package (PfEPMA/PfE), the following information is part of the public domain and presented by an expert on the use of PfEPMA software, and can be found on the web at [www.geology.wisc.edu/~johnf/g777/ppt/70\\_Errors.ppt](http://www.geology.wisc.edu/~johnf/g777/ppt/70_Errors.ppt), in its entirety with respect to the “Counting Statistics” section in this document.

We desire to count X-ray intensities of peak and backgrounds, for both standards and unknowns, with high precision and accuracy. X-ray production is a random process (Poisson statistics), where each repeated measurement represents a sample of the same specimen volume. The expected distribution can be described by Poisson statistics, which for large number of counts is closely approximated by the ‘normal’ (Gaussian) distribution. For Poisson distributions, 1 sigma = square root of the counts, and 68.3% of the sampled counts should fall within ±1 sigma, 95.4% within ±2 sigma, and 99.7 within ±3 sigma.



Lifshin and Gauvin, 2001, Fig. 6, p 172

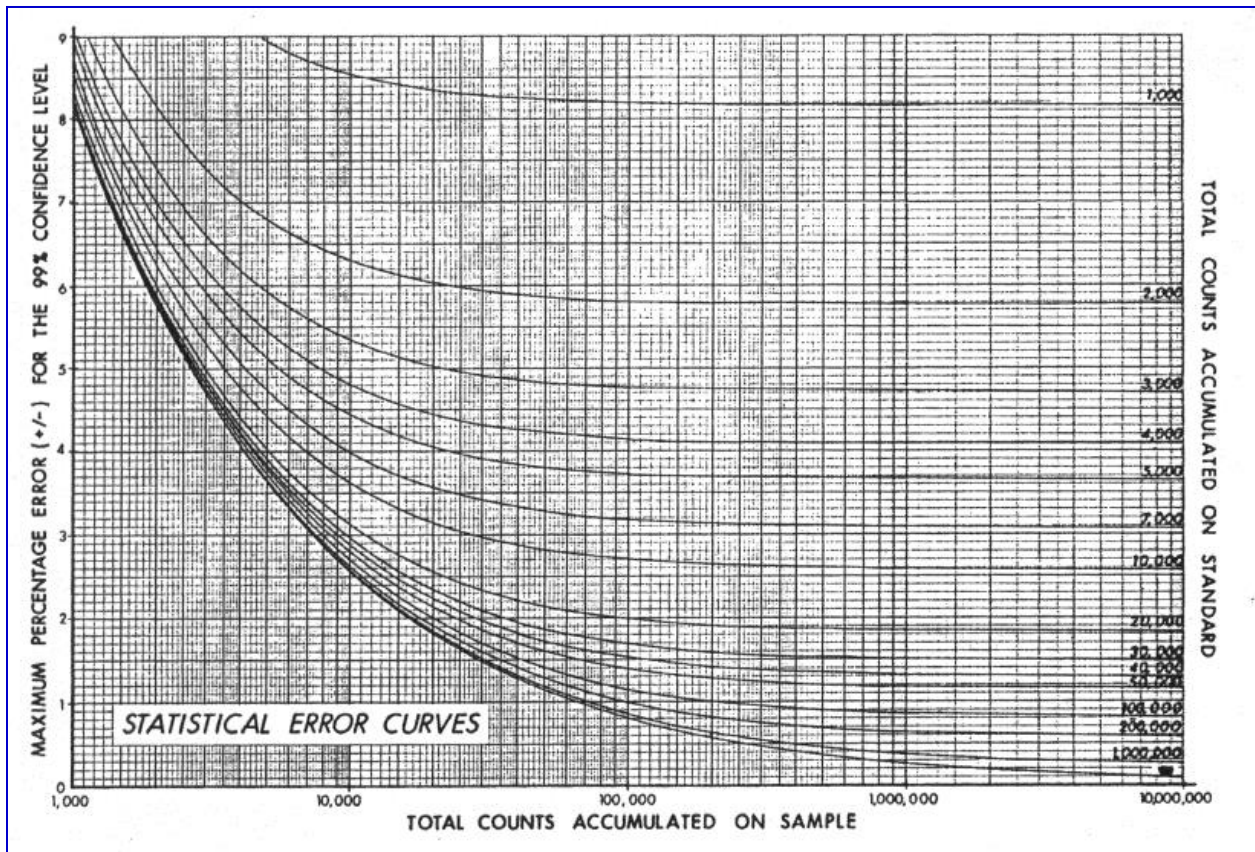
The precision of the composition ultimately is a combination of the counting statistics of both standard and unknown, and Ziebold (1967) developed an equation for it. Recall that the K-ratio is

$$K = \frac{I_P^{unk} - I_B^{unk}}{I_P^{std} - I_B^{std}}$$

where P and B refer to peak and background. The corresponding precision in the K ratio is given by

$$\sigma_K^2 = K^2 \left[ \frac{I_P^{unk} + I_B^{unk}}{n(I_P^{unk} - I_B^{unk})^2} + \frac{I_P^{std} + I_B^{std}}{n'(I_P^{std} - I_B^{std})^2} \right]$$

where n and n' are the number of repetitions of counts on the unknown and standard respectively. (The rearranged  $s_k/K$  -- with square roots taken -- term was sometimes referred to as the 'sigma upon K' value.)



Another format for considering cumulative precision of the unknown is the above graph. A maximum error at the 99% confidence interval can be calculated, based upon the total counts acquired upon both the standard and the unknown: e.g. to have 1% max counting error you must have at least ~120,000 counts on the unknown and on the standard; you could get 2% with ~30,000 counts on each.

#### Probe for EPMA Statistics

PfE provides several statistics in the normal default 'log window' printout for background subtracted peak counts: average, standard deviation, 1 sigma, std dev/1 sigma (SIGR), standard error, and relative standard deviation. For Si: the average is 4479 cps, and the average sample uncertainty (SDEV) for each of the 3 measurements is 15 cps. The counting error (1 sigma) is

somewhat larger (21 cps), and the ratio of standard deviation to sigma is <1, indicating good homogeneity in Si.

<b>ELEM:</b>	<b>Si ka</b>	<b>Fe ka</b>	<b>Mn ka</b>	<b>Mg ka</b>	<b>Ca ka</b>	<b>Ni ka</b>
<b>BGD:</b>	<b>OFF</b>	<b>OFF</b>	<b>OFF</b>	<b>OFF</b>	<b>OFF</b>	<b>OFF</b>
<b>MOTOR:</b>	<b>4</b>	<b>5</b>	<b>5</b>	<b>1</b>	<b>3</b>	<b>5</b>
<b>CRYST:</b>	<b>TAP</b>	<b>LIF</b>	<b>LIF</b>	<b>TAP</b>	<b>PET</b>	<b>LIF</b>
<b>ORDER:</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>3</b>
<b>132G</b>	<b>4494.6</b>	<b>328.1</b>	<b>7.7</b>	<b>7452.9</b>	<b>7.0</b>	<b>13.6</b>
<b>133G</b>	<b>4476.9</b>	<b>321.0</b>	<b>3.7</b>	<b>7368.5</b>	<b>6.5</b>	<b>12.8</b>
<b>134G</b>	<b>4465.5</b>	<b>316.1</b>	<b>4.3</b>	<b>7436.1</b>	<b>7.8</b>	<b>11.8</b>
<b>AVER:</b>	<b>4479.0</b>	<b>321.8</b>	<b>5.2</b>	<b>7419.2</b>	<b>7.1</b>	<b>12.8</b>
<b>SDEV:</b>	<b>14.6</b>	<b>6.0</b>	<b>2.1</b>	<b>44.7</b>	<b>.7</b>	<b>.9</b>
<b>1SIG:</b>	<b>21.2</b>	<b>5.7</b>	<b>.7</b>	<b>27.2</b>	<b>.8</b>	<b>1.1</b>
<b>SIGR:</b>	<b>.69</b>	<b>1.06</b>	<b>2.97</b>	<b>1.64</b>	<b>.79</b>	<b>.79</b>
<b>SERR:</b>	<b>8.4</b>	<b>3.5</b>	<b>1.2</b>	<b>25.8</b>	<b>.4</b>	<b>.5</b>
<b>%RSD:</b>	<b>.3</b>	<b>1.9</b>	<b>41.2</b>	<b>.6</b>	<b>9.4</b>	<b>7.0</b>

For homogeneous samples, we can define a standard error for the average: here, 8 cps. Finally, the printout shows the relative standard deviation as a percentage (0.3%, excellent).

## STANDARD DEVIATION

A calculation that essentially describes the error for a single point of a sample. The calculation is :

$$S.D. = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1}}$$

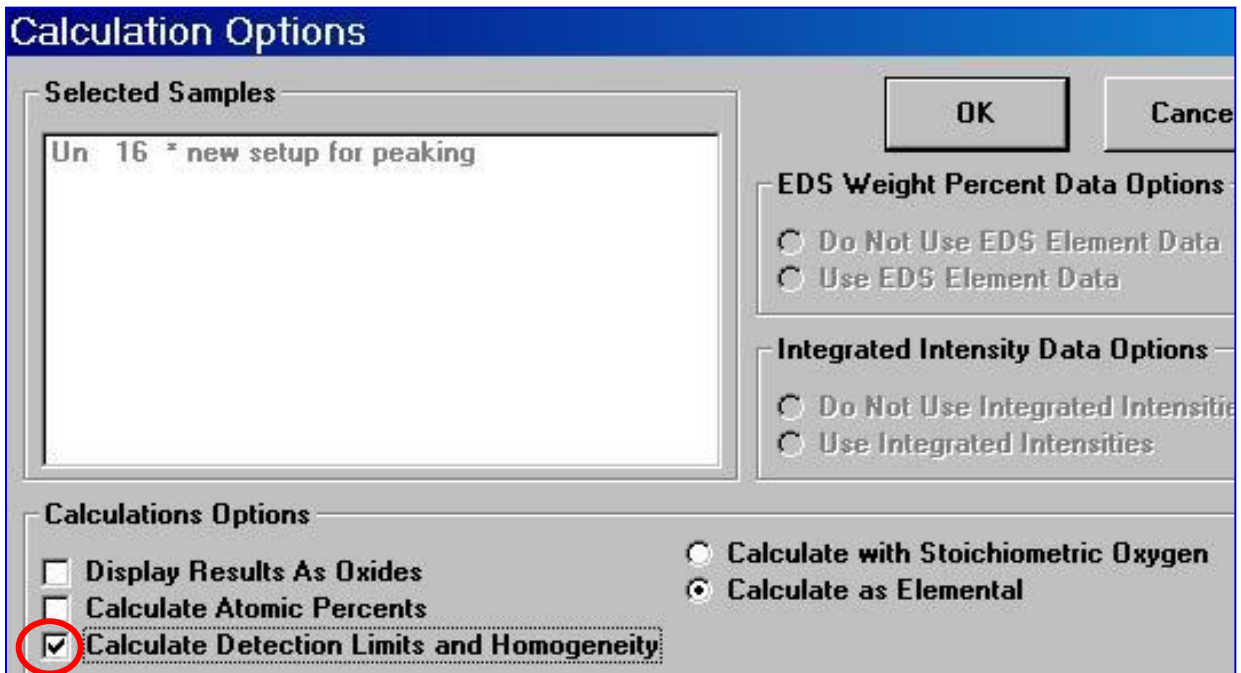
where :

- $X_i$  is the data observations
- $\mu$  is the mean of the data
- $n$  is the number of observations





After the raw counts, the elemental weight percents are printed, with some of the same statistics, followed by the specific standard (number) used (see the box above, "Results in Elemental Weight Percent"). Following that are the standard K-ratio, and standard peak (P-B) count rate. Below that are the unknown K-ratio, the unknown peak count rate, and the unknown background count. Below that are the ZAF ("ZCOR") for the element, the raw K-ratio of the unknown, the peak-background ratio of the unknown, and any interference correction applied ("INT%", as percentage of measured counts). PFE software provides for additional optional statistics. One set relates to detection limits, i.e. what is the lowest level you can be confident in reporting.



The other set of statistics relates to the homogeneity of the unknowns as well as calculation of analytical error. We will now discuss these statistics.

This calculation is for analytical sensitivity of each line (= one measurement), considering both peak and background count rates (Love and Scott, 1974). It is a similar type of statistic as the 1 sigma counting precision figure, **but it is presented as a percentage.**

Percent Analytical Error (Single Line):						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni
132	.5	1.8	18.2	.4	17.3	12.7
133	.5	1.8	34.5	.4	19.0	13.3
134	.5	1.8	32.8	.4	15.9	14.2
<b>AVER:</b>	.5	1.8	28.5	.4	17.4	13.4
<b>SDEV:</b>	.0	.0	8.9	.0	1.5	.8
<b>SERR:</b>	.0	.0	5.2	.0	.9	.4

$$\epsilon_{P-B} = \frac{\sqrt{\frac{N_P}{t_P^2} + \frac{N_B}{t_B^2}}}{\left(\frac{N_P}{t_P} - \frac{N_B}{t_B}\right)}$$

Where :

- $N_P$  is the total peak counts
- $N_B$  is the total background counts
- $t_P$  is the peak count time
- $t_B$  is the background count time

Probe for EPMA provides a more advanced set of calculations for analytical statistics. The calculations are based on the number of data points acquired in the sample and the measured standard deviation for each element. This is important because although x-ray counts theoretically have a standard deviation equal to square root of the mean, the actual standard deviation is usually larger due to variability of instrument drift, x-ray focusing errors, and x-ray production.

A common question is whether a phase being analyzed by EPMA is homogeneous, or is the same or distinct from another separate sample. A simple calculation is to look at the average composition and see if all analyses are within some range of sigmas (2 for 95%, 3 for 99% normal probability).

A more exacting criterion is calculating a precise range (in wt%) and level (in %) of homogeneity. These calculations utilize the standard deviation of measured values and the degree of statistical confidence in the determination of the average.

The degree of confidence means that we wish to avoid a risk a of rejecting a good result a large per cent of the time (95 or 99%) of the time. "Student's t distribution" gives various confidence levels for evaluation of data, i.e. whether a particular value could be said to be within the expected range of a population -- or more likely, whether two compositions could be confidently said to be the same. The degree of confidence is given as 1-  $\alpha$ , usually .95 or .99. This means we can define a range of homogeneity, in wt%, where on the average only 5% or 1% of repeated random points would be outside this range.

The general problem, where the sample size is small and the population variance is unknown, was first treated in 1905 by W.S. Gossett, who published his analysis under the pseudonym "Student". His employer, the Guinness Breweries of Ireland, had a policy of keeping all their research as proprietary secrets. The importance of his work argued for its being published, but it was felt that anonymity would protect the company. (S.L. Meyer, Data Analysis for Scientists and Engineers, 1975, p. 274.)



**Table 9.16.** Values of Student *t* Distribution for 95% and 99% Degrees of Confidence<sup>a</sup>

<i>n</i>	<i>n</i> - 1	<i>t</i> <sub><i>n</i>-1</sub> <sup>95</sup>	<i>t</i> <sub><i>n</i>-1</sub> <sup>99</sup>
3	2	4.304	9.92
4	3	3.182	5.841
8	7	2.365	3.499
12	11	2.201	3.106
16	15	2.131	2.947
30	29	2.042	2.750
∞	∞	1.960	2.576

<sup>a</sup> Bauer (1971).

### Test for Homogeneity

1. The range of homogeneity in plus or minus weight percent.

$$W_{1-\alpha} = \pm C \frac{t_{n-1}^{1-\alpha}}{n^{1/2}} \frac{S_C}{\bar{N}}$$

2. The level of homogeneity in plus or minus percent of the concentration.

$$\pm \frac{W_{1-\alpha}}{C} = \pm \frac{(t_{n-1}^{1-\alpha}) S_C (100)}{n^{1/2} \bar{N}}$$

### Olivine Analysis Example of Homogeneity Test:

The following data-set of olivine analyses will be used to show how the test for homogeneity works – these are the original analytical data:

ELEM:	Si	Fe	Mn	Mg	Ca	Ni	SUM
132	18.956	6.267	.159	30.953	.094	.301	100.665
133	18.870	6.134	.077	30.568	.087	.284	99.532
134	18.834	6.042	.089	30.810	.105	.262	99.751
AVER:	18.886	6.148	.108	30.777	.095	.282	99.983

Range of Homogeneity in +/- Elemental Weight Percent (Average of Sample):						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni
60ci	.034	.071	.013	.115	.002	.007
80ci	.060	.126	.023	.204	.004	.013
90ci	.093	.195	.036	.316	.006	.020
95ci	.137	.287	.053	.465	.009	.030
99ci	.317	.662	.122	1.074	.021	.069

What this means: for Si, at highest level (95%), we can say that there is chance that only 5% of a number of random points will be 0.14 wt% greater or lesser than 18.89 wt% (or as a percent, 0.7%).

Level of Homogeneity in +/- Percent (Average of Sample):						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni
60ci	.2	1.2	12.0	.4	2.4	2.6
80ci	.3	2.0	21.4	.7	4.3	4.6
90ci	.5	3.2	33.2	1.0	6.6	7.2
95ci	.7	4.7	48.9	1.5	9.8	10.6
99ci	1.7	10.8	112.7	3.5	22.5	24.4

PfE also provides a handy table to show if the sample is homogeneous at the 1% precision level, and if so, at what confidence level.

Test of Homogeneity at 1.0 % Precision (Average of Sample):						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni
60ci	yes	no	no	yes	no	no
80ci	yes	no	no	yes	no	no
90ci	yes	no	no	no	no	no
95ci	yes	no	no	no	no	no
99ci	no	no	no	no	no	no

### Counting Statistics

Analytical sensitivity is the ability to distinguish, for an element, between two measurements that are nearly equal.

4. The analytical sensitivity in weight percent.

$$\Delta C = C - C' \geq \frac{2^{1/2} C (t_{n-1}^{1-\alpha}) S_C}{n^{1/2} (\bar{N} - \bar{N}_B)}$$

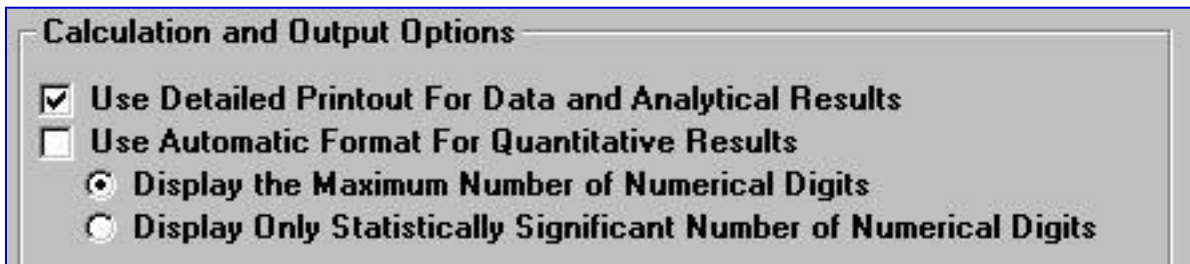
Analytical Sensitivity in Elemental Weight Percent (Average of Sample):						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni
60ci	.048	.100	.018	.162	.003	.010
80ci	.085	.178	.033	.288	.006	.019
90ci	.132	.275	.051	.447	.009	.029
95ci	.194	.406	.075	.658	.013	.042
99ci	.448	.936	.173	1.518	.030	.097

So here, at the 95% confidence level, two samples would have to have a difference in Si of > .20 wt% to be considered reliably different in Si.

There have been cases where people have taken reported compositions (i.e. wt % elements or oxides) from probe printouts and then faithfully reproduced them exactly as they got them. Once someone took figures that were reported to 3 decimal points and argued that a difference in the 3rd decimal place had some geochemical significance.

The number of significant figures reported in a printout is a “mere” programming format issue, and has nothing to do with scientific precision! (However, a feature of PfE is an option to output only the actual significant number of digits. This is not normally enabled.)

Having said that, it is “tradition” to report to 2 decimal places. However, that should not be taken to represent precision, without a statistical test, such as given before.



In the example of the olivine analysis above, where Si was printed out as 18.886 wt%, it would be reported as 18.89 -- but looking at the limited number of analyses and the homogeneity tests, I would feel uncomfortable telling someone that another analysis somewhere between 18.6 and 19.2 were not the same material. Nor would I be uncomfortable with someone reporting the Si as 18.9 wt% (though I stick to tradition.)

Considering silicate mineral or glass compositions, Si is traditionally reported with 4 significant figures. If we were to be rigorous regarding significant figures, we would follow the rule that we would be bound by the least number of figures in a calculation where we multiply our measurement (K-ratio, which will have thousands of counts divided by thousands of counts) by the ZAF. As you can appreciate there are many calculations that comprise each part of the ZAF, and it would be stretching it to argue that the ZAF itself can have more than 3 significant figures. Ergo, we should not strictly report Si with more than 3 significant figures.

When we enable the PfE Analytical Option “Display only statistically significant number of numerical digits” for the olivine analysis, here is the result:

<b>ELEM:</b>	<b>Si</b>	<b>Fe</b>	<b>Mn</b>	<b>Mg</b>	<b>Ca</b>	<b>Ni</b>	<b>SUM</b>
132	19.0	6.3	.2	31.0	.09	.3	100.665
133	18.8	6.1	.08	30.5	.09	.3	99.532
134	18.7	6.0	.09	30.7	.1	.3	99.751

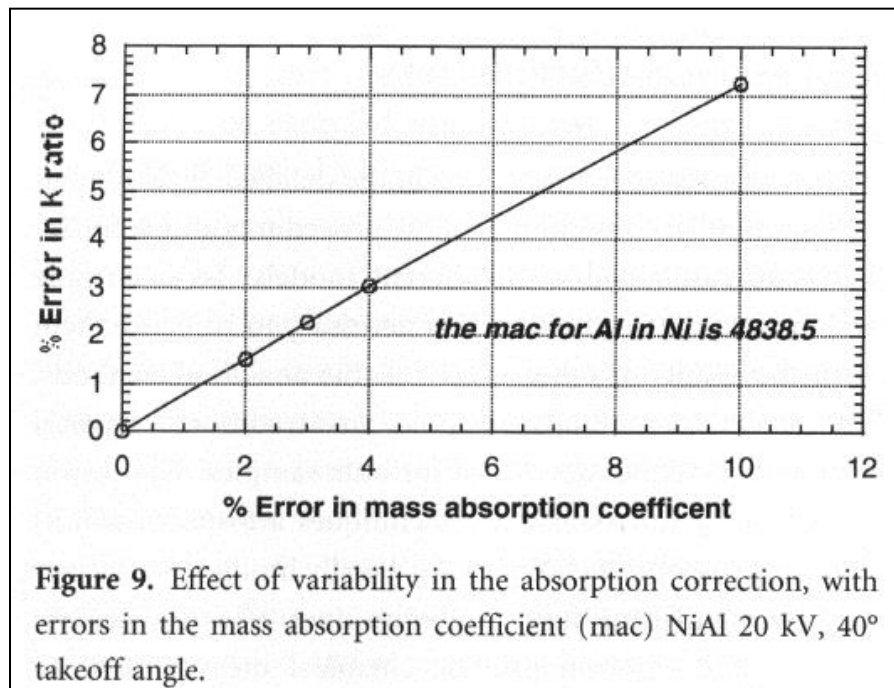
For comparison, here's the original printout:

ELEM:	Si	Fe	Mn	Mg	Ca	Ni	SUM
132	18.956	6.267	.159	30.953	.094	.301	100.665
133	18.870	6.134	.077	30.568	.087	.284	99.532
134	18.834	6.042	.089	30.810	.105	.262	99.751
AVER:	18.886	6.148	.108	30.777	.095	.282	99.983

### Errors in Matrix Correction

The K-ratio is multiplied by a matrix correction factor. There are various models – alpha, ZAF, f(rz) – and versions. Assuming that you are using the appropriate correction type, there may be some issues regarding specific parameters, e.g. mass absorption coefficients, or the f(rz) profile.

There is a possibility of error for certain situations, particularly for “light elements” as well as compounds that have drastically different Z elements where pure element standards are used. The figure above shows that a small (2%) error in the mass absorption coefficient for Al in NiAl will yield an error of 1.5% in the matrix correction.



This is a strong incentive to either use standards similar to the unknown, and/or use secondary standards to verify the correctness of the EPMA analysis.

## Summary: How to know if the EPMA results are “good”?

There are only 2 tests to prove your results are “good” – actually, it is more correct to say that if your results can pass the test(s), then you know they are not necessarily bad analyses:

**100 wt% totals** (NOT 100 atomic % totals). The fact that the total is near 100 wt%. Typically, a range from 98.5 - 100.5 wt% for silicates, glasses and other compounds is considered “good”. It extends on the low side a little to accommodate a small amount of trace elements that are realistically present in most natural (earth) materials. These analyses typically “do oxygen by stoichiometry” which can introduce some undercounting where the Fe:O ratio has been set to a default of 1:1, and some the iron is ferric (Fe:O 2:3). So for spinels (e.g.  $\text{Fe}_3\text{O}_4$ ), a perfectly good total could be 93 wt%.

**Stoichiometry**, if such a test is valid (e.g. the material is a line compound, or a mineral of a set stoichiometry).

### Checking the olivine analysis

ELEM:	Si	Fe	Mn	Mg	Ca	Ni	SUM
132	18.956	6.267	.159	30.953	.094	.301	100.665
133	18.870	6.134	.077	30.568	.087	.284	99.532
134	18.834	6.042	.089	30.810	.105	.262	99.751
AVER:	18.886	6.148	.108	30.777	.095	.282	99.983

SPEC:	0						
TYPE:	CALC						
AVER:	4.00000						
SDEV:	.00000						
ELEM:	Si	Fe	Mn	Mg	Ca	Ni	SUM
132	.98315	.16346	.00422	1.85513	.00341	.00748	7.01685
133	.98816	.16154	.00205	1.84978	.00319	.00711	7.01184
134	.98409	.15877	.00238	1.86029	.00383	.00654	7.01591
AVER:	.98513	.16126	.00288	1.85507	.00348	.00704	7.01487
SDEV:	.00267	.00236	.00117	.00525	.00033	.00047	
SERR:	.00154	.00136	.00068	.00303	.00019	.00027	
%RSD:	.3	1.5	40.6	.3	9.4	6.7	

The total is excellent, 99.98 wt%. The stoichiometry is pretty good (not excellent): on the 4 oxygens, there should be 1.00 Si atoms and we have .985. The total cations Mg+Fe+Ca+Ni should be 2.00, and we have 2.03. The analysis is OK and could be published. If this were seen at the time of analysis, it might be useful to recheck the Si and Mg peak positions, and reacquire standard counts for Si and Mg. If this were only seen after the fact, you could re-examine the standard counts and see if there are any obvious outliers that were included and could be legitimately discarded.

## References :

Castaing, R. 1951 PhD Thesis, University of Paris

Chatterjee, N. (unknown year) Electron Microprobe Analysis, Course 12.141 Notes, MIT  
[https://ocw.mit.edu/courses/earth-atmospheric-and-planetary-sciences/12-141-electron-microprobe-analysis-january-iap-2012/lecture-notes/MIT12\\_141IAP12\\_coursenotes.pdf](https://ocw.mit.edu/courses/earth-atmospheric-and-planetary-sciences/12-141-electron-microprobe-analysis-january-iap-2012/lecture-notes/MIT12_141IAP12_coursenotes.pdf)

Goldstein, J.I, *et al.* 1992 Scanning Electron Microscopy, A textbook for Biologists, Material Scientiststs, and Geologists. Plenum Press.

Gullickson, E.M. X-Ray Data Booklet Section 1.6 MASS ABSORPTION COEFFICIENTS  
[http://xdb.lbl.gov/Section1/Sec\\_1-6.pdf](http://xdb.lbl.gov/Section1/Sec_1-6.pdf)

Heinrich, K.F.J. and Newbury, D.E. eds. 1991. Electron Probe Quantitation. Plenum Press, N.Y., USA

Jenkins, R *et al.* 1995 Quantitative X-Ray Spectrometry, Second Edition. CRC Press, N.Y., USA

Jenkins, R *et al.* 1970 Worked Examples in X-Ray Analysis. MacMillan Press Ltd., UK

JEOL Manual XM-27330 WD/ED Basic Software, JEOL Ltd, Tokyo, Japan, 2001

Kerrick, D.M *et al.* 1973 The Role of Carbon Film Thickness in Electron Microprobe Analysis  
American Mineralogist 58:920-925

Lyman, C. *et al.* 2012 Scanning Electron Microscopy, X-ray Microanalysis, and Analytical Electron Microscopy, A Laboratory Workbook. Springer Verlag

Newbury, D.E. (unknown year) Quantitative Electron Probe Microanalysis (EPMA): Doing It Right and Doing It Wrong! (PPT file  
<http://www.ctcms.nist.gov/~cecamp/spring04/EPMA%20Right&Wrong.pdf> )

Potts, P.J. *et al.* eds. 1995 Microprobe Techniques in the Earth Sciences, Mineralogical Society Series #6, Chapman and Hall, UK

Reed, S.J.B, 1975. Electron Microprobe Analysis. Cambridge University Press, Cambridge.

Reed, S.J.B., 1990. Microbeam Analysis San Francisco Press, San Francisco, USA

University of Wisconsin website [www.geology.wisc.edu/~johnf/g777/ppt/70\\_Errors.ppt](http://www.geology.wisc.edu/~johnf/g777/ppt/70_Errors.ppt)

Victoreen, J.A. 1949 The Calculation of X-Ray Mass Absorption Coefficients. J. Appl. Phys 20: 1141- 1147 <http://scitation.aip.org/content/aip/journal/jap/20/12/10.1063/1.1698286>