Concentration Dependent Confirmation Model of Quartz X-Ray Diffraction Analysis

Ilya B. Reznik and Daniel N. Johansen

August 31, 2015

Abstract

A study to develop a quantitative model of variation between analytical peaks in quartz standards on Rigaku Dmax 2500 XRD instrument at Salt Lake Technical Center was undertaken using close to three years of calibration data. The analysis was carried out separately on two goniometers (right and left). The equations $range = (0.11 \pm 0.01) + (3.3 \pm 0.8) \times x^{-0.9 \pm 0.1}$ and $range = (0.10 \pm 0.02) + (5 \pm 3) \times x^{-1.1 \pm 0.3}$, where x is the amount of quartz in μg, were found to fit the historical data and provide a possible model to determine quantitative confirmation limit between peaks of unknown quartz samples on the right and left goniometers respectively.

Introduction

The Occupational Safety and Health Administration's Salt Lake Technical Center (SLTC) uses four X-ray diffraction angles for respirable quartz analysis, which are treated as independent measurements for a given sample. This paper will refer to integrals of X-ray intensity signals as peaks.

All four peaks when multiplied by their respective calibration coefficients should produce similar result for the mass of quartz. Various effects can account for the variability in these results. The focus of this study is to determine how far those peaks can vary between the four measurements using calibration standards. This range will then be used as a threshold to determine if an unknown sample has a less than or equal variance.

Analysis

Data Acquisition

SLTC uses Rigaku Dmax 2500 X-ray Diffractometer (XRD) with a dual goniometer for quartz analysis using ID 142 version 2. Each goniometer needs to be calibrated about two times per year. In order to do so, a set of standards is prepared with a known amount of quartz using National Institute of Standards and Technology (NIST) standard reference material 1878a. Nearly three years of calibration data were acquired for statistical significance.

Because the standards get similar treatment in the laboratory as other samples, but contain a known amount of quartz, the calibration data represent peak variability when controlled for particle size distribution and the absence of interferences. For the purposes of this study it is assumed that the calibration samples represent all other sources of error.

Methodology

The data from ten calibrations between February 2012 and July 2014 were analyzed using the R language and environment for statistical computing.

```
R.version.string
```

```
## [1] "R version 3.1.2 (2014-10-31)"
```

1. Calibration data were compiled and then read into R with the following commands.

```
## Import needed library
library(stringr)
## Get the names of the files
file.list <- list.files(path = "./Data/")
## Figure out the full list of dates
dates.of.analysis <- str_extract(file.list, "[0-9]{8}")
## Figure out which side of the instrument the callibration was analyzed on
sides.of.instrument <- str_extract(file.list, "L|R")
```

2. The peaks have individual sensitivities, thus an intrinsic variability. In analysis this effect is taken into account with a constant derived during the calibration process. In this study, means were calculated for each theoretical amount of quartz per peak in each calibration. Then the peaks were normalized by dividing each peak by the mean of all peaks at the same angle range for the same theoretical mass (or level) of quartz in the same calibration. The result is a unit-less normalized peak.

3. The range of normalize peaks was then determined by subtracting the smallest normalize peak from the largest and dividing by the smallest within each individual standard.

```
all.observations <- data.frame() ## Empty container
for(i in 1:length(file.list)){ ## Go through each file
  df <- read.csv(file = paste("./Data/",file.list[i],sep="/"))
  ## Calculate averages
  ms <- aggregate(x=df[c(2:6)], by = list(df[[2]]), FUN = "mean", na.rm =TRUE)
  df$Range <- 0
  ## Calculate ranges
  for(j in 1:nrow(df)){
    df[j,c(3:6)] <- df[j,c(3:6)]/ms[df[j,2]==ms[[2]],-c(1,2)]
    df$Range[j] <- (max(df[j,c(3:6)])-min(df[j,c(3:6)]))/min(df[j,c(3:6)])
  }
  ## Keep the side and date information
  df$Date <- dates.of.analysis[i]
  df$Side <- sides.of.instrument[i]
  all.observations <- as.data.frame(rbind(all.observations,df))
}
```

4. Ranges from the right and left goniometers were compared using a Welch t-test. Since several concentrations have significant differences at 90% confidence interval, the two sides will be treated as different data sets.

5. The data were split into two sets.

```
left <- all.observations[which(all.observations$Side == "L"),]
right <- all.observations[which(all.observations$Side == "R"),]
```

	P-values
10 ug	0.27
20 ug	0.01
50 ug	0.21
100 ug	0.05
250 ug	0.39
500 ug	0.04
1000 ug	0.99
2000 ug	0.06

6. Because range measures maximum-minimum over minimum and the study aims to look at maximum confirming range, a one-sided 95% confidence interval of ranges ($\bar{x} + 1.645 \times s$) at each combined level was chosen as the dependent variable for regression with the mass as a predictor.

```
## Find means of each level of quartz on both sides
ave.left <- aggregate(Range~Theoretical.ug.,data=left,FUN=mean)
ave.right<- aggregate(Range~Theoretical.ug.,data=right,FUN=mean)

## Find standard deviation times 1.645
ave.left$ci <- aggregate(Range~Theoretical.ug.,data=left,FUN=sd)$Range*1.645
ave.right$ci <- aggregate(Range~Theoretical.ug.,data=right,FUN=sd)$Range*1.645

## Find 95% one-sided confidence interval
ave.left$y <- ave.left$Range+ave.left$ci
ave.right$y <- ave.right$Range+ave.right$ci
```

7. A power model was fitted to both sides to the following equation:

$$y = a + b \times x^{-c}$$

where y is the maximum range between peaks that would be said to confirm, x is the theoretical amount of quartz in the sample, and **a**,**b**,and **c** are regression parameters. One parameter, **b**, is found not significant on the left at 95% confidence level. Theoretically it can be removed, but given the level of significance of the result on the right and a desire to keep the equations in the same form, it will be retained.

```
model.right <- nls(y ~ a + b * Theoretical.ug. ^ -c,  data=ave.right,
                   start = list(a=0.1,b=2,c=0.8))
summary(model.right)
```

```
##
## Formula: y ~ a + b * Theoretical.ug.^-c
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## a   0.1094     0.0116   9.434 0.000226 ***
## b   3.3109     0.8407   3.938 0.010978 *
## c   0.9031     0.1098   8.228 0.000432 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01774 on 5 degrees of freedom
```

3

```
##
## Number of iterations to convergence: 5
## Achieved convergence tolerance: 7.574e-06
```

```
model.left <- nls(y ~ a + b * Theoretical.ug. ^ -c,  data=ave.left,
                  start = list(a=0.1,b=2,c=0.8))
summary(model.left)
```

```
##
## Formula: y ~ a + b * Theoretical.ug.^-c
##
## Parameters:
##    Estimate Std. Error t value Pr(>|t|)
## a  0.09592    0.01771   5.417  0.00290 **
## b  4.52379    2.87823   1.572  0.17682
## c  1.10659    0.27295   4.054  0.00979 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03154 on 5 degrees of freedom
##
## Number of iterations to convergence: 11
## Achieved convergence tolerance: 4.826e-06
```

8. The results were plotted.

Concentration–Depended Confirming Ranges
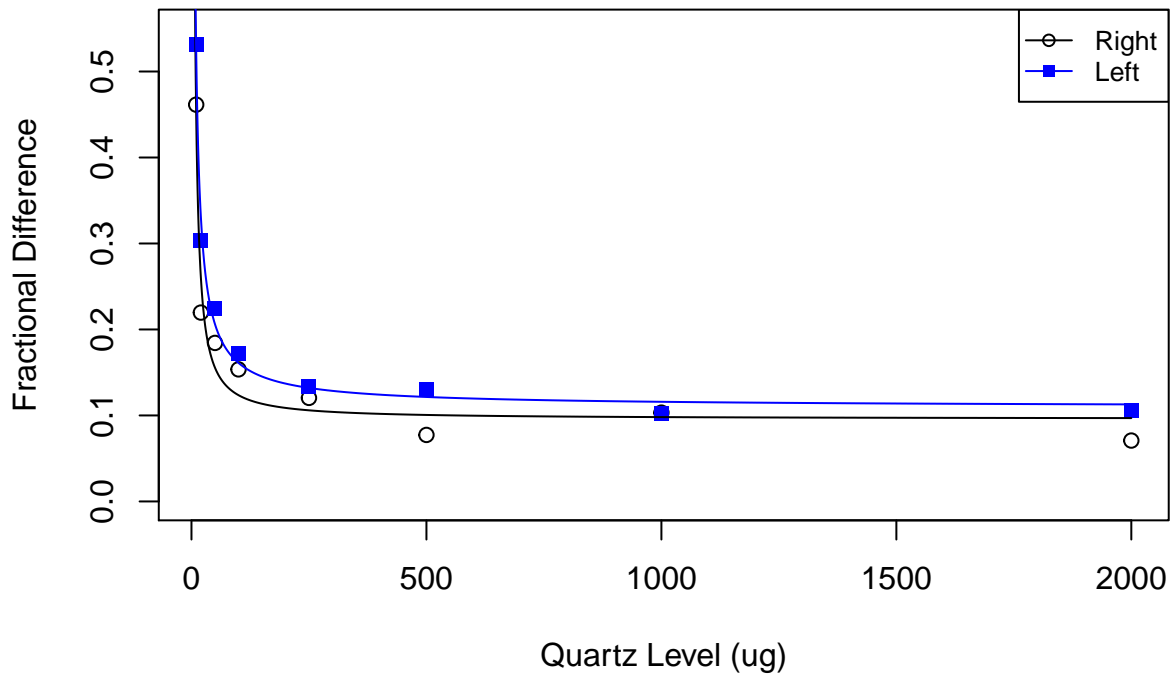
Validation

Validating this model will require a data set that contains known concentrations of quartz with and without introduced known interferants. Such a dataset is not readily available at each level, however it is available at 40μg.

A set of 3 samples was prepared with 2000μg of kaolinite and 40μg of quartz using NIST 1878a standard reference material. All samples were analyzed on the right goniometer. To see if the peaks are confirming, the concentration equal to the lowest seen peak, x_1, was assumed. Then the second lowest peak, x_2, was compared with x_1 as

$$r = (x_2 - x_1)/x_1$$

This r, depicted below in red, was larger than maximum confirming range predicted by the model.

The same procedure was employed with the same samples after acid-washing. After acid washing the peaks did confirm within the range predicted by the model, as shown in green.

Concentration–Depended Confirming Ranges
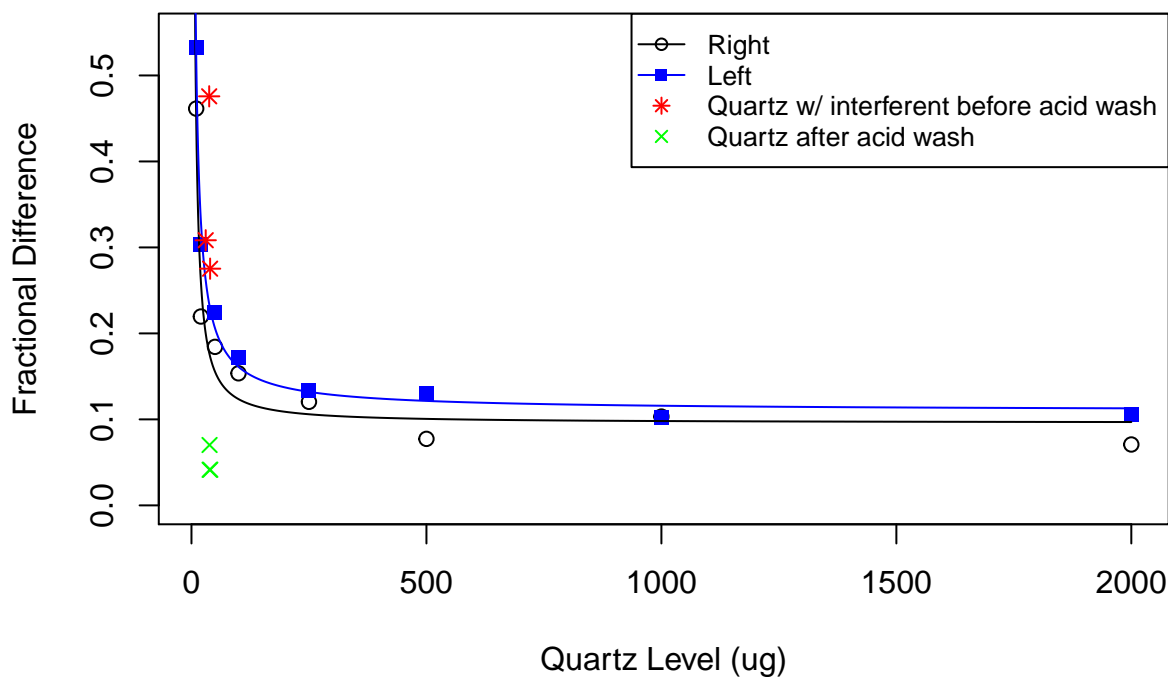
Note that two of the points are too close to resolve on this graph.

Discussion

The model shows a trend in the data. Peak agreement is a power function of the mass. Because different instruments have inherent differences, the right and the left goniometers of the XRD produce slightly different models as follows:

$$Right : range = (0.11 \pm 0.01) + (3.3 \pm 0.8) \times x^{-0.9 \pm 0.1}$$

$$Left : range = (0.10 \pm 0.02) + (5 \pm 3) \times x^{-1.1 \pm 0.3}$$

where x is the amount of quartz in μg.

All coefficients on the right and the left sides are within error of each other.

This model predicts that a sample after calibration has a confirming peak if there are two peaks x_1 and x_2 such that x_1 is the lowest predicted mass of quartz and $(x_2 - x_1)/x_1 \leq range$, where range is the result of applying the model above, using x_1 as μg.

This model should only be used in the range of quartz weights for which it was produced, 10 to 2000 μ g.