Low-Mg Carbonate XRD data analysis with Xd,Xq; 0-5% calcite

John M. Jaeger

July 10, 2017

## Code Description

This code is designed to import and analyze X-ray diffraction data collected at the University of Florida on a Rigaku Ultima IV diffractometer. The data are from carbonate mineral standards created to calibrate the diffractometer to provide quantitative measures of low- and high-Mg calcite.

The general calibration approach follows: Sepulcre, S., Durand, N., and Bard, E., 2009, Global and Planetary Change: Global And Planetary Change, v. 66, p. 1–9, doi: 10.1016/j.gloplacha.2008.07.008.

Establishing detection and quantificiation limits follows: Evard, H., Kruve, A., and Leito, I., 2016, Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review: Analytica Chimica Acta, v. 942, p. 23–39, doi: 10.1016/j.aca.2016.08.043.

Code generated by: John M. Jaeger Associate Professor 241 Williamson Hall P.O. Box 112120 Dept. of Geological Sciences University of Florida Gainesville FL 32611-2120, USA (352) 846-1381 ORCID ID# orcid.org/0000-0003-0248-489X <http://people.clas.ufl.edu/jmjaeger/>

## Load packages

# if necessary uncomment and install packages.  
# install.packages("dplyr")  
# install.packages("knitr")  
library(dplyr)  
require(knitr)  
library(ezknitr)  
library(chemCal)  
library(reshape2)  
library(RColorBrewer)

## Install functions

rnorm2 <- function(n,mean,sd) { mean+sd\*scale(rnorm(n)) }

## Import the data

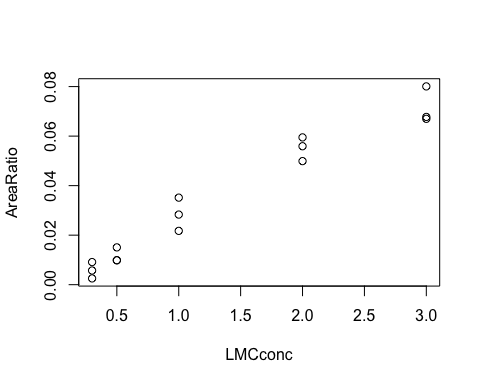
### Load peak area values for calibration standards

Low-Mg calcite peak at 29.4° 2-theta; Aragonite peak at 26.2° 2-theta (copper anode)

# load the End-member Dataset and give the dataset a shorter name  
  
XRD.data <- read.table("../raw\_data/Sepulcure\_2009\_LMC.csv",header=T,sep=",")   
AreaRatio <- XRD.data$AreaRatio  
LMCconc <- XRD.data$LMCConc  
AR.sig <- XRD.data$X1sig  
n <- 3 #number of replicates of each calibration sample  
calib.levels <- as.integer(nrow(XRD.data))  
test.data <- matrix(nrow = n, ncol = calib.levels)  
for(u in 1:calib.levels){  
 test.data[,u] <- rnorm2(n,AreaRatio[u],AR.sig[u])  
}  
  
test.data <- t(test.data)  
Sep.data <- as.data.frame(cbind(LMCconc,test.data))  
Sep.data.melt <- melt(Sep.data, id=2:4)  
 ff <- melt(Sep.data[,c("V2","V3","V4")])

## No id variables; using all as measure variables

C <- rep(LMCconc, n)  
 Sep.data.melt <- as.data.frame(cbind(C,ff[,2]))  
colnames(Sep.data.melt) <- c("LMCconc","AreaRatio")  
plot(Sep.data.melt)



## Create ordinary least squares (OLS) regression model

We first develop an OLS linear calibration model for low-Mg calcite. We create a residual plot to examine for homoscedasticity and whether a linear model is an appropraite fit (see Edavrd et al., 2016 for rationale of these inital tests).

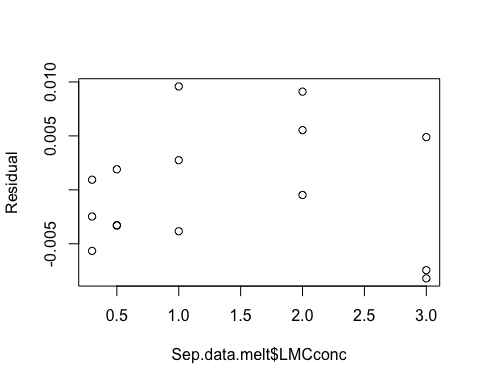
Sep.lm= lm(AreaRatio~ LMCconc, data=Sep.data.melt)  
vcov(Sep.lm)

## (Intercept) LMCconc  
## (Intercept) 6.340049e-06 -3.006439e-06  
## LMCconc -3.006439e-06 2.210617e-06

summary(Sep.lm)

##   
## Call:  
## lm(formula = AreaRatio ~ LMCconc, data = Sep.data.melt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0082099 -0.0035770 -0.0004748 0.0038163 0.0095869   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0007564 0.0025179 0.30 0.769   
## LMCconc 0.0248115 0.0014868 16.69 3.68e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.005811 on 13 degrees of freedom  
## Multiple R-squared: 0.9554, Adjusted R-squared: 0.952   
## F-statistic: 278.5 on 1 and 13 DF, p-value: 3.68e-10

Bcalc = Sep.lm$coefficient[[2]]  
Acalc = Sep.lm$coefficient[[1]]  
Residual = Sep.data.melt$AreaRatio - Bcalc \* Sep.data.melt$LMCconc - Acalc  
# Data is plotted  
plot(Sep.data.melt$LMCconc, Residual)

 Residuals show no trend and cluster around zero, so linear model is appropriate. Variance in residuals does not have distinctive "funnel" shape, but variance is not constant with concentration, so data are considered heteroscedastic. OLS model is not approprate, so WLS is preferred. Results using an OLS approach are preseted at end for comparision with WLS results.

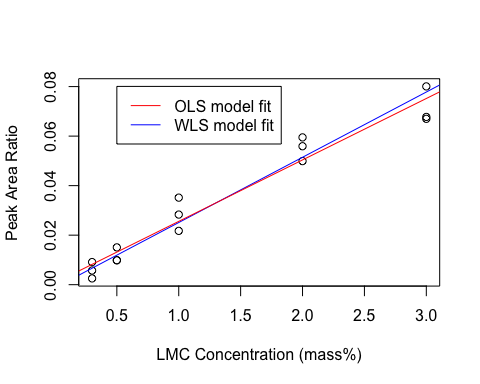
## Create weighted least squares (WLS) regression model

We determine WLS calibration regression according to methods and R code of: Evard, H., Kruve, A., and Leito, I., 2016, Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review: Analytica Chimica Acta, v. 942, p. 23–39, doi: 10.1016/j.aca.2016.08.043.

Y <- Sep.data.melt[,2]  
C <- Sep.data.melt[,1]  
  
# Calculating weights for calibration samples  
 SdRow <- NULL  
 for(k in 1:nrow(test.data)){  
 SdRow[k] <- sd(test.data[k, ])^-2  
 }  
 Nsamp <- nrow(test.data) \* ncol(test.data)  
 sumCal <- sum(SdRow)  
 sumvar <- (sumCal) / (Nsamp)   
 w <- NULL  
 for(v in 1:nrow(test.data)){  
 w[v] <- sd(test.data[v, ])^-2 / sumvar  
 }  
 W <- rep(w, n)  
   
  
 WLS <- lm(Y ~ C, weights = W)  
 bw\_calc = lm(Y ~ C, weights = W)$coeff[2]  
 aw\_calc = lm(Y ~ C, weights = W)$coeff[1]  
 summary(WLS)

##   
## Call:  
## lm(formula = Y ~ C, weights = W)  
##   
## Weighted Residuals:  
## Min 1Q Median 3Q Max   
## -0.010644 -0.004963 -0.002038 0.006062 0.012019   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.001248 0.001689 -0.739 0.473   
## C 0.026346 0.001437 18.338 1.13e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.007723 on 13 degrees of freedom  
## Multiple R-squared: 0.9628, Adjusted R-squared: 0.9599   
## F-statistic: 336.3 on 1 and 13 DF, p-value: 1.132e-10

sigmodel <- bw\_calc\*C+aw\_calc  
 plot(C,Y, ann=FALSE)  
 abline(aw\_calc, bw\_calc,col = "blue")# WLS fit  
 abline(Acalc,Bcalc, col="red")#OLS fit  
 title(xlab="LMC Concentration (mass%)")  
 title(ylab="Peak Area Ratio")  
 legend (0.5, max(Y), c("OLS model fit","WLS model fit"),col=c("red", "blue"),lty=c(1,1))

 The slope and intercept of the two regression lines are slightly different, which will affect the resulting limit calculations that depend on these values. See text for further discussion.

## Establish limits of detection and quantification

We determine the limits of decision, detection, and quantification (Yc, Xd, Xq, respectively) from calibration regression according to: Burdge, J.R., MacTaggart, D.L., and Farwell, S.O., 1999, Realistic detection limits from confidence bands: J Chem Educ., v.76(3), p. 434-439  
and Zorn, M.E., Gibbons, R.D., and Sonzogni, W.C., 1997, Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration: Analytical chemistry, v. 69, p. 3069–3075.

The decision limit (Lc or Yc as used here) is the peak area ratio above which the signal is deemed at a 95% probabilty to rise above noise. The intersection of this y-value with the regression line is the Xc, or critical level in the concentration domain; however, it is suggested by Edvard et al. (2016) that the crtiical level be expressed only in the signal domain. The limit of detection (Xd here but often refered to as LoD) is the detection limit in concentration units (Yd here in signal units, sometimes denoted as Ld).

# Equations for Yc and Xd used here are from Burdge, J.R., MacTaggart, D.L., and Farwell, S.O., 1999, Realistic detection limits from confidence bands: J Chem Educ., v.76(3), p. 434-439   
  
#If no blank replicates are measured, it is necessary to create a model of signal st deviation as a function of concentration, which is then used to model weights for intercept and unknown values  
#stval.lm <- lm(sdvalm~poly(LMCconc,2,raw=TRUE)) #use if necessary to model St dev in calibration samples to assign weights; See Zorn et al. (1997)  
p <- 0 #parameters in model of st dev versus concentration, which is zero for this because we use measured values applied to ranges closest in % LMC. Use 2 for linear model or 3 for 2nd order polynomial; See Zorn et al. (1997)  
t.quantiles <- qt(c(.05, .95), Nsamp-p-2)# degrees of freedom in regression model  
tval <- t.quantiles[2]# degrees of freedom in regression model  
m <- 1# number of replicate measurements of unknown, usually 1 for establishing limits  
#See Burge et al. (1999) for explantion of variables below:  
xbarw <- sum(C\*W)/sum(W)  
Sxxw <- sum(W\*C^2)-((sum(C\*W))^2/sum(W))  
Syyw <- sum(W\*Y^2)-((sum(Y\*W))^2/sum(W))  
Sxyw <- sum(W\*Y\*C)-(sum(W\*C)\*sum(W\*Y))/sum(W)  
Swr <- sqrt((Syyw-bw\_calc\*Sxyw)/(Nsamp-2))  
SwrN <- Swr/sqrt(sum(W)/Nsamp)  
gw <- ((tval\*Swr)^2)/(bw\_calc^2\*Sxxw)  
  
Q <- sqrt((1/(m\*W[1]))+(1/sum(W))+xbarw^2/Sxxw)  
A <- ((tval\*Swr\*Q)/bw\_calc)-(xbarw\*gw)  
B <- xbarw\*gw\*(xbarw\*gw-(2\*tval\*Swr\*Q/bw\_calc))  
Cv <- 1/sum(W)+xbarw^2/Sxxw+gw/(m\*W[1])  
Yc.wls <- aw\_calc+tval\*Swr\*(sqrt(1/(m\*W[1])+(1/sum(W))+(xbarw^2/Sxxw)))# decision limit in signal units  
Xd.wls <- (A+(sqrt(B+(gw\*Sxxw\*(((1-gw)/(m\*W[2]))+Cv)))))/(1-gw) # limit of detection in concentration units. We do not iterate to find precise weight at Xd as recommended by Burdge et al. and simply assume the weight at the calibration point closest to estimate of Xd (i.e., W[2] here)  
  
# We can also esitmate limits of quantification  
Yq.wls <- 10\*(sqrt(1/SdRow[2]))+aw\_calc #limit of quantification in signal units; From Zorn et al. (1997), eqn. 23  
Xq.wls <- (Yq.wls-aw\_calc)/bw\_calc #limit of quantification in concentration units; From Zorn et al. (1997), eqn. 24  
  
  
cat("WLS limit of detection (%LMC): ", round(Xd.wls,digits = 1)," ")

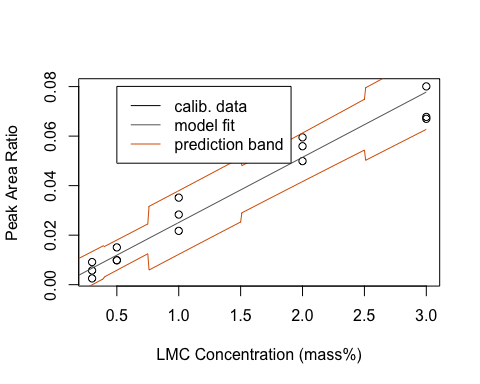
## WLS limit of detection (%LMC): 0.5

cat("WLS limit of quantification (%LMC): ", round(Xq.wls,digits = 1)," ")

## WLS limit of quantification (%LMC): 1.1

## Establish prediction bands for WLS calibration regression

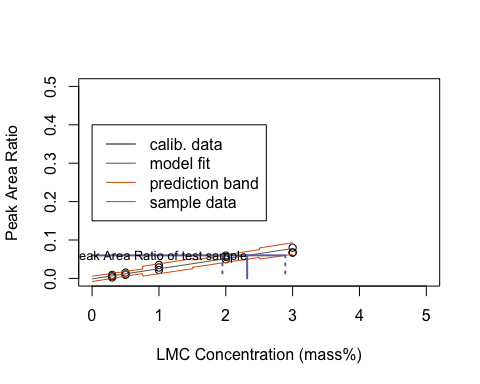
wls.predval <- matrix(nrow = calib.levels, ncol = m)  
  
LMC.model <- seq(0, 3, by=0.01)# set up a range of LMC values spanning limits at 0.01% LMC intervals  
PAR.model <- bw\_calc\*LMC.model+aw\_calc #model peak area ratio (PAR) as a function of LMC.model  
weight.modin <- cbind(LMCconc,w) # link calibration concentration measurement weights to concentrations  
  
#instead of modeling weights as a polynomial function of %LMC, we choose to set weights based on measured values. Weights for LMC.model values are assigned based on nearest measured concentration replicates.  
j <- length(PAR.model)  
wls.predvaltest <- matrix(nrow = j, ncol = 1)  
for(jv in 1:j){  
 lookvalue <- LMC.model[jv]  
 weightx <- which(abs(LMCconc-lookvalue)==min(abs(LMCconc-lookvalue)))  
 wls.predvaltest[jv,1] <- weight.modin[weightx[1],2]  
}  
Inverse.LMCdata <- cbind(LMC.model,PAR.model,wls.predvaltest)  
wls.predval2 <- matrix(nrow = j, ncol = m)  
  
#We develop 95% prediction band equation from Zorn et al., equation 13  
for(pv in 1:j){  
 wls.predval2[pv,1] <-tval\*Swr\*(sqrt(1/(m\*wls.predvaltest[pv,1])+(1/sum(W))+((Inverse.LMCdata[pv,1]-xbarw)^2/Sxxw)))  
}  
predband.uM <- PAR.model+wls.predval2  
predband.lM <- PAR.model-wls.predval2  
  
Inverse.LMCdata <- cbind(LMC.model,PAR.model,wls.predvaltest,wls.predval2,predband.uM,predband.lM)  
colnames(Inverse.LMCdata) <- c("LMC.model", "PAR.model","weight.model","uncert.model","upper.uncert","lower.uncert")  
cols<-brewer.pal(n=8,name="Dark2")  
  
plot(C,Y, ann=FALSE)  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,2], col = cols[8])  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,5], col = cols[2], lty=1)  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,6], col = cols[2], lty=1)  
#segments(0, Yc.wls, LMC.xupperC, unknown.PAR,col =cols[3],lty=1,lwd=2)  
title(xlab="LMC Concentration (mass%)")  
title(ylab="Peak Area Ratio")  
legend(0.5, max(Y), c("calib. data","model fit","prediction band"),col=c("black", cols[8],cols[2]), lty=c(1,1,1))



## Create an inverse model for establishing %LMC from a peak area ratio value

We create a model of fit and uncertainity to perform inverse modeling. We use the intercept of the P.A.R. with the regression line to determine the most likely %LMC value. We use the intercept of the P.A.R. with the prediction bands from the calibration curve to establish the 95% predition limits on the % LMC.

unknown.PAR <- 0.06 # value chosen to compare with model fits of Sepulcure et al. (2009) and Smodej et al. (20XX)  
  
if(unknown.PAR > max(PAR.model)) {  
 stop("Your unknown sample value is outside calibration range")  
 }  
LMC.x <- which(abs(PAR.model-unknown.PAR)==min(abs(PAR.model-unknown.PAR)))  
LMC.inverse <- LMC.model[LMC.x ]# inverse fit of PAR that provides best fit to %LMC  
  
unk.PARline <- rep(unknown.PAR,j)  
  
#First, we find intersection of PAR value with upper prediction band  
x1=unk.PARline  
x2=Inverse.LMCdata[,5]  
# Find points where x1 is above x2.  
above<-x1>x2  
# Points always intersect when above=TRUE, then FALSE or reverse  
intersect.points<-which(diff(above)!=0)  
# Find the slopes for each line segment.  
x1.slopes<-x1[intersect.points+1]-x1[intersect.points]  
x2.slopes<-x2[intersect.points+1]-x2[intersect.points]  
# Find the intersection for each segment.  
x.points<-intersect.points + ((x2[intersect.points] - x1[intersect.points]) / (x1.slopes-x2.slopes))  
LMC.xlower <- LMC.model[x.points]  
LMC.xlowerC <- LMC.xlower[which(abs(LMC.xlower-LMC.inverse)==min(abs(LMC.xlower-LMC.inverse)))]  
  
#Next, we find intersection of PAR value with lower prediction band  
x1=unk.PARline  
x2=Inverse.LMCdata[,6]  
# Find points where x1 is below x2.  
below<-x1<x2  
# Points always intersect when below=TRUE, then FALSE or reverse  
intersect.points<-which(diff(below)!=0)  
# Find the slopes for each line segment.  
x1.slopes<-x1[intersect.points+1]-x1[intersect.points]  
x2.slopes<-x2[intersect.points+1]-x2[intersect.points]  
# Find the intersection for each segment.  
x.points<-intersect.points + ((x2[intersect.points] - x1[intersect.points]) / (x1.slopes-x2.slopes))  
LMC.xupper <- LMC.model[x.points]  
LMC.xupperC <- LMC.xupper[which(abs(LMC.xupper-LMC.inverse)==min(abs(LMC.xupper-LMC.inverse)))]  
  
if (length(LMC.xlowerC) == 0){  
 stop("Your unknown sample value is outside prediction band")  
 }   
if (length(LMC.xupperC) == 0){  
 stop("Your unknown sample value is outside prediction band")  
 }   
  
  
  
#display.brewer.pal(8,"Dark2")  
#pdf("unknownLMCsample\_UF.pdf", encoding = "MacRoman", width=7, height=5, useDingbats=FALSE)  
plot(C,Y, ann=FALSE,ylim=c(0,0.5),xlim=c(0,5))  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,2], col = cols[8])  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,5], col = cols[2], lty=1)  
lines(Inverse.LMCdata[,1],Inverse.LMCdata[,6], col = cols[2], lty=1)  
segments(0, unknown.PAR, LMC.xupperC, unknown.PAR,col =cols[3],lty=1,lwd=2)  
text(1, unknown.PAR-unknown.PAR/20,"Peak Area Ratio of test sample",cex=0.8)  
segments(LMC.inverse, unknown.PAR, LMC.inverse, 0,col =cols[3],lty=1,lwd=2)  
segments(LMC.xlowerC, unknown.PAR,LMC.xlowerC, 0,col =cols[3],lty=3,lwd=2)  
segments(LMC.xupperC, unknown.PAR,LMC.xupperC, 0,col =cols[3],lty=3,lwd=2)  
title(xlab="LMC Concentration (mass%)")  
title(ylab="Peak Area Ratio")  
legend(0,0.4, c("calib. data","model fit","prediction band","sample data"),col=c("black", cols[8],cols[2],cols[3]), lty=c(1,1,1,1))



#dev.off()

Provide the user with best fit and uncertainty for unknown sample from WLS regression model

cat("WLS best fit, mass %LMC: ", round(LMC.inverse,digits = 1), " ")

## WLS best fit, mass %LMC: 2.3

cat("WLS lower %LMC (95% CI): ", round(LMC.xlowerC,digits = 1)," ")

## WLS lower %LMC (95% CI): 2

cat("WLS upper %LMC (95% CI): ", round(LMC.xupperC,digits = 1)," ")

## WLS upper %LMC (95% CI): 2.9

## Alternative ordinary least squares calibration approach

It is possible to determine Yc,Xd,and Xq from OLS calibration regression if data are considered homoscedatic according to: Evard, H., Kruve, A., and Leito, I., 2016, Tutorial on estimating the limit of detection using LC-MS analysis, part I: Theoretical review: Analytica Chimica Acta, v. 942, p. 23–39, doi: 10.1016/j.aca.2016.08.043.

Nsamp <- length(LMCconc)  
  
#using method of Lavagnini & Magno (2006)  
newdata = data.frame(LMCconc=0)  
Yc.all <- predict(Sep.lm, newdata, interval="predict", level = 0.95, df=Nsamp-n)  
Yc <- Yc.all[3]  
Ld\_data <- data.frame(AreaRatio=Yc)  
ivtest=inverse.predict(Sep.lm,Yc,alpha=0.05)  
Xc <- ivtest$Prediction  
inverse.Sep.lm<- lm(LMCconc~AreaRatio, data=Sep.data.melt)  
Xd.all <- predict(inverse.Sep.lm, Ld\_data, interval="predict", level = 0.95, df=Nsamp-n)  
Xd <- Xd.all[3]  
  
#using method of Miller & Miller(1988) and Edvard et al. (2016)   
# Output parameters:  
Bcalc <- Sep.lm$coefficient[[2]] # estimated slope  
Syx <- ( sum(resid(Sep.lm)^2 ) / (Nsamp - 2) )^0.5 # Standard deviation of residuals  
Acalc <- Sep.lm$coefficient[[1]] # estimated intercept  
Aerror <- summary(Sep.lm)$coef[1,2] # estimated standard deviation of intercept  
CCaC <- (1.645 \* Syx)/Bcalc #also known as Yc  
CCaS <- (1.645 \* Syx)#also known as Lc or Xc  
Xd.ols <- 3.3\*Syx/Bcalc #CCb, also limit of determination in concentration, also known as LoD   
Xq.ols <- 10\*Syx/Bcalc #Limit of quantification in concentration   
Xd.int <- 3.3\* Aerror/Bcalc #limit of determination based only on varince at intercept  
cat("OLS Decision value (peak area ratio): ", round(CCaS,digits = 3), " ")

## OLS Decision value (peak area ratio): 0.02

cat("OLS Limit of detection (%LMC): ", round(Xd.ols,digits = 1)," ")

## OLS Limit of detection (%LMC): 1.6

cat("OLS Limit of quantification (%LMC): ", round(Xq.ols,digits = 1)," ")

## OLS Limit of quantification (%LMC): 4.9

The ordinary least squares decision limit (threshold) (CCa) based on peak area ratio is 0.024. The ordinary least squares limit of determination in concentration units is 1.1% LMC The ordinary least squares limit of quantification in concentration units is 3.3% LMC

We establish peak area ratio values from calibration associated with crtical 3% by mass low-Mg calcite cutoff, which is within of limit of quantification:

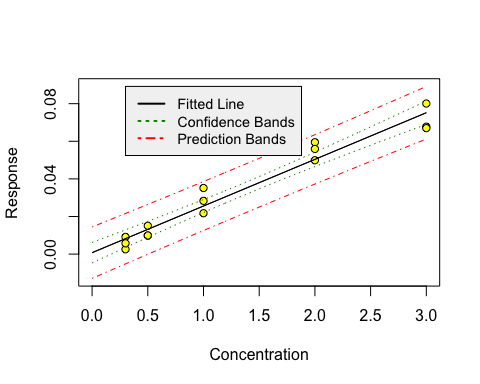
newdata = data.frame(LMCconc=3)  
predict(Sep.lm, newdata, interval="predict")

## fit lwr upr  
## 1 0.07519081 0.06119563 0.08918599

predict(Sep.lm, newdata, interval="confidence")

## fit lwr upr  
## 1 0.07519081 0.0690056 0.08137602

calplot(Sep.lm)



Calculate accuracy and precision for reverse model (95% CI) using a peak area ratio of 0.133 (predicited from linear model fit for a mass% of 3 for low-Mg calcite; see output above)

ivtest=inverse.predict(Sep.lm,0.133)  
cat("Inverse fit (%LMC): ", round(ivtest$Prediction, digits = 1), " ")

## Inverse fit (%LMC): 5.3

cat("Inverse confidence limits (%LMC): ", round(ivtest$`Confidence Limits`, digits = 1), " ")

## Inverse confidence limits (%LMC): 4.2 6.4

#ivtest$Prediction  
#ivtest$`Confidence Limits`

Accuracy and precision for linear calibration model at crtical cutoff is 3.0 +/- 0.7% based on calibration data of April 2017.

Use linear model to predict low-Mg calcite concentration from peak area ratio of unknown sample (e.g., 0.100):

ivtest=inverse.predict(Sep.lm,0.100)  
cat("Inverse fit (%LMC): ", round(ivtest$Prediction, digits = 1), " ")

## Inverse fit (%LMC): 4

cat("Inverse confidence limits (%LMC): ", round(ivtest$`Confidence Limits`, digits = 1), " ")

## Inverse confidence limits (%LMC): 3.1 4.9

Sample concentration is above limit of detection. Caution should be used aoplying absolute value because it is below limit of quantification.