Supplementary Data 2
R Script for Automatic Error Calculation of the Endogenous Concentration Correction

The functional version of the “Error_Endogenous_Correction.R” script can be found here: https://www.dropbox.com/sh/8agalof79qfwe81/AADYzwHQqD4z9Ix8qSsVYd1Sa?dl=0. Users can watch a video tutorial detailing its use in R Studio here: https://youtu.be/_gpwQSKDWeC.

To summarize, the user should copy to the clipboard a table containing the type of sample (“Cal” or “QC”), the spiked concentration and measurement (area, area ratio or response ratio). Columns can be in any order, but the exact headers “Type”, “Spiked_Conc” and “Measure” should be present. For example, in Excel, the zone in grey below should be copied (Ctrl+C):

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td><strong>Sequence A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td><strong>Name</strong></td>
<td><strong>Type</strong></td>
<td><strong>Spiked_Conc</strong></td>
</tr>
<tr>
<td>3</td>
<td>QC1</td>
<td>QC</td>
<td>60</td>
<td>0.3222</td>
</tr>
<tr>
<td>4</td>
<td>QC1</td>
<td>QC</td>
<td>60</td>
<td>0.5295</td>
</tr>
<tr>
<td>5</td>
<td>QC2</td>
<td>QC</td>
<td>150</td>
<td>0.7687</td>
</tr>
<tr>
<td>6</td>
<td>QC2</td>
<td>QC</td>
<td>150</td>
<td>0.7585</td>
</tr>
<tr>
<td>7</td>
<td>QC3</td>
<td>QC</td>
<td>375</td>
<td>1.7962</td>
</tr>
<tr>
<td>8</td>
<td>QC3</td>
<td>QC</td>
<td>375</td>
<td>1.7336</td>
</tr>
<tr>
<td>9</td>
<td>STD0</td>
<td>Cal</td>
<td>0</td>
<td>0.051</td>
</tr>
<tr>
<td>10</td>
<td>STD1</td>
<td>Cal</td>
<td>10</td>
<td>0.0919</td>
</tr>
<tr>
<td>11</td>
<td>STD2</td>
<td>Cal</td>
<td>20</td>
<td>0.1353</td>
</tr>
<tr>
<td>12</td>
<td>STD3</td>
<td>Cal</td>
<td>50</td>
<td>0.2791</td>
</tr>
<tr>
<td>13</td>
<td>STD4</td>
<td>Cal</td>
<td>100</td>
<td>0.5001</td>
</tr>
<tr>
<td>14</td>
<td>STD5</td>
<td>Cal</td>
<td>200</td>
<td>0.9991</td>
</tr>
<tr>
<td>15</td>
<td>STD6</td>
<td>Cal</td>
<td>425</td>
<td>1.9131</td>
</tr>
<tr>
<td>16</td>
<td>STD7</td>
<td>Cal</td>
<td>500</td>
<td>2.0555</td>
</tr>
</tbody>
</table>

The script is then sourced in R Studio. At the end of the compilation procedure, the calculated parameters will be displayed in the console as a tibble:
With $EM$ being the error on the uncorrected estimated concentration (in concentration units), $Bias$ being the percent bias on each QC level if the endogenous concentration was not corrected, $Corrected\_Conc$ being the corrected concentration, i.e. $x_c = x_s + x_e$ and $EC$ being the error on the corrected estimated concentration (in concentration units).
# Load the function OptParam
# (Taken from CodeV5.R of the calibration project, written by FCL)
OptParam <- function (X, Y, poids, indice)
{
  # Finds the calibration parameters for
  # Weight (poids) 1 (=0) or 1/x (=1) or 1/(x^2) (=2).
  # Model order (indice) linear (1) or quadratic (2).
  W <- diag (1/abs (X)^poids)
  xf <- matrix (rep(X, each = indice +1), ncol =indice +1, byrow=T)
  ex <- matrix (rep(0:(indice), length(X)), ncol =indice +1, byrow=T)
  Xp = xf^ex
  matInv <- solve (t(Xp)%*%W%*%Xp)
  param.optimaux <- rev (matInv%*%t(Xp)%*%W%*%Y)
  return (param.optimaux)
}

# Calculate the error generated by the endogenous concentration correction according to the
# weighting scheme selected.
# Weighting: none (A)
if (WS == "A")
{
  # Perform the linear regression and isolate the regression coefficients.
  Reg_Params <- OptParam(Data_Cal$Spiked_Conc, Data_Cal$Measure, 0, 1)
  B0 <- Reg_Params[2]
  B1 <- Reg_Params[1]

  # Calculate the individual (xi-xbar)^2 values for each level.
  Data_Cal$X_Gap_Sq <- (Data_Cal$Spiked_Conc - mean(Data_Cal$Spiked_Conc))^2

  # Calculate the squared residual (yi-yhat)^2 for each level.
  Data_Cal$RES <- (Data_Cal$Measure - ((B1*Data_Cal$Spiked_Conc) + B0))^2

  # Calculate the squared standard error (se^2) of the regression.
  SE2 <- sum(Data_Cal$RES)/(n-2)

  # Calculate the average response (ybar).
  YMoy <- mean(Data_Cal$Measure)

  # Calculate the n (calibration levels) value.
  n <- as.numeric(length(Data_Cal$Spiked_Conc))

  # Calculate XE, the estimated endogenous concentration.
  XE <- abs(-B0/B1)

  # Calculate EE, the error on the estimated endogenous analyte concentration.
  EE <- (sqrt(SE2)/B1)*sqrt(((1/n)+((YMoy^2)/((B1^2)*((sum(Data_Cal$X_Gap_Sq)))))))

  # Identify the different QC levels that need to be analyzed.
  QC_LVL <- as.character(unique(Data_QC$Name))

  # Create a DF to harbour the results.
Results <- tbl_df(QC_LVL)
Results$Spiked_Conc <- as.numeric(unique(Data_QC$Spiked_Conc))

for (i in 1:length(QC_LVL)){
  # Segregate the data for the QC level analyzed.
  Data_QC_Temp <- filter(Data_QC, Name == QC_LVL[i])

  # Calculate the number of measurement replicates for this level.
  m <- length(Data_QC_Temp)

  # Calculate EM, the error on the estimated concentration in the sample (uncorrected).
  EM <- (sqrt(SE2)/B1) * sqrt((1/m) + (1/n) + (((mean(Data_QC_Temp$Measure) - Y_Moy)^2)/((B1^2)* (sum(Data_Cal$X_Gap_Sq)))))

  # Store this value in the Results DF.
  Results$EM[i] <- EM

  # Calculate the bias incurred if no correction is made.
  Results$Bias[i] <- (XE/Results$Spiked_Conc[i]) * 100

  # Calculate the corrected expected concentration.
  Results$Corrected_Conc[i] <- Results$Spiked_Conc[i] + XE

  # Calculate EC, the error on the corrected concentration.
  Results$EC[i] <- Results$EM + EE
}

# Weighting: 1/x (B)
if (WS == "B"){
  # Generate the wt vector to perform linear regression.
  Data_Cal$WF <- 1/Data_Cal$Spiked_Conc

  # Perform the linear regression and isolate the regression coefficients.
  Reg_Params <- OptParam(Data_Cal$Spiked_Conc, Data_Cal$Measure, 1, 1)
  B0 <- Reg_Params[2]
  B1 <- Reg_Params[1]

  # Calculate the concentrations weighted mean (XW).
  XW <- sum(Data_Cal$WF*Data_Cal$Spiked_Conc)/sum(Data_Cal$WF)

  # Calculate the measurements weighted mean (YW).
  YW <- sum(Data_Cal$WF*Data_Cal$Measure)/sum(Data_Cal$WF)

  # Calculate the residuals for each calibration level.
  Data_Cal$RES <- (Data_Cal$Measure - ((B1*Data_Cal$Spiked_Conc) + B0))^2

  # Calculate the n (calibration levels) value.
  n <- as.numeric(length(Data_Cal$Spiked_Conc))

  # Calculate the squared standard error (SE2) of the regression.
SE2 <- (sum(Data_Cal$WF*Data_Cal$RES))/(n-2)

# Calculate EE, the error on the estimated endogenous analyte concentration.
EE <- (sqrt(SE2)/B1)*sqrt(((1/sum(Data_Cal$WF))+(YW^2)/((B1^2)*(sum(Data_Cal$WF*Data_Cal$Spiked_Conc^2))-sum(Data_Cal$WF*XW^2))))

# Calculate XE, the estimated endogenous concentration.
XE <- abs(-B0/B1)

# Identify the different QC levels that need to be analyzed.
QC_LVL <- as.character(unique(Data_QC$Name))

# Create a DF to harbour the results.
Results <- tbl_df(QC_LVL)
Results$Spiked_Conc <- as.numeric(unique(Data_QC$Spiked_Conc))

for (i in 1:length(QC_LVL)) {
  # Segregate the data for the QC level analyzed.
  Data_QC_Temp <- filter(Data_QC, Name == QC_LVL[i])

  # Calculate the weighting factor of the QCs.
  Data_QC_Temp$WF <- 1/Data_QC_Temp$Spiked_Conc

  # Calculate the number of measurement replicates for this level.
  m <- length(Data_QC_Temp$Measure)

  # Calculate EM, the error on the estimated concentration in the sample (uncorrected).
  EM <- (sqrt(SE2)/B1)*sqrt(((1/sum(Data_QC_Temp$WF*[1]*m))+(1/sum(Data_Cal$WF))+
    (((mean(Data_QC_Temp$Measure)-YW)^2)*(sum(Data_Cal$WF))/((B1^2)*((sum(Data_Cal$WF)*(sum(Data_Cal$WF*Data_Cal$Spiked_Conc^2)))-
    (sum(Data_Cal$WF*Data_Cal$Spiked_Conc)^2))))-((sum(Data_Cal$WF*Data_Cal$Spiked_Conc)^2)))))

  #Store this value in the Results DF.
  Results$EM[i] <- EM

  # Calculate the bias incurred if no correction is made.
  Results$Bias[i] <- (XE/Results$Spiked_Conc[i])*100

  # Calculate the corrected expected concentration.
  Results$Corrected_Conc[i] <- Results$Spiked_Conc[i] + XE

  # Weighting: 1/x^2 (C)
  if (WS == "C") {
    
  }
# Perform the linear regression and isolate the regression coefficients.
Reg_Params <- OptParam(Data_Cal$Spiked_Conc, Data_Cal$Measure, 2, 1)
B0 <- Reg_Params[2]
B1 <- Reg_Params[1]

# Sterilize data set to remove "missing" standards.
#Data_Cal <- tbl_df(Data_Cal) %>% filter(Spiked_Conc > 1E-8)

# Generate the wt vector to perform linear regression.
Data_Cal$WF <- 1/(Data_Cal$Spiked_Concˆ2)

# Calculate the concentrations weighted mean (XW).
XW <- sum(Data_Cal$WF * Data_Cal$Spiked_Conc) / sum(Data_Cal$WF)

# Calculate the measurements weighted mean (YW).
YW <- sum(Data_Cal$WF * Data_Cal$Measure) / sum(Data_Cal$WF)

# Calculate the residuals for each calibration level.
Data_Cal$RES <- (Data_Cal$Measure - ((B1 * Data_Cal$Spiked_Conc) + B0))ˆ2

# Calculate the n (calibration levels) value.
n <- as.numeric(length(Data_Cal$Spiked_Conc))

# Calculate the squared standard error (SE2) of the regression.
SE2 <- (sum(Data_Cal$WF * Data_Cal$RES)) / (n − 2)

# Calculate EE, the error on the estimated endogenous analyte concentration.
EE <- (sqrt(SE2) / B1) * sqrt((1 / sum(Data_Cal$WF)) + ((YW^2) / ((B1^2) * (sum(Data_Cal$WF * (Data_Cal$Spiked_Conc^2)) - sum(Data_Cal$WFs*XW^2))))

# Calculate XE, the estimated endogenous concentration.
XE <- abs(−B0/B1)

# Identify the different QC levels that need to be analyzed.
QC_LVL <- as.character(unique(Data_QC$Name))

# Create a DF to harbour the results.
Results <- tbl_df(QC_LVL)
Results$Spiked_Conc <- as.numeric(unique(Data_QC$Spiked_Conc))

for (i in 1:length(QC_LVL)) {
# Segregate the data for the QC level analyzed.
Data_QC_Temp <- filter(Data_QC, Name == QC_LVL[i])

# Calculate the weighting factor of the QCs.
Data_QC_Temp$WF <- 1/(Data_QC_Temp$Spiked_Conc^2)

# Calculate the number of measurement replicates for this level.
m <- length(Data_QC_Temp$Measure)
# Calculate EM, the error on the estimated concentration in the sample (uncorrected).
EM <- (sqrt(SE2)/B1) * sqrt(((1/(Data_QC_Temp$WF[1]*m)) + (1/sum(Data_Cal$WF)) +
  (((mean(Data_QC_Temp$Measure)-YW)^2)*(sum(Data_Cal$WF)))/
  ((B1^2)*((sum(Data_Cal$WF)*(sum(Data_Cal$WF*(Data_Cal$Spiked_Conec^2))))-
  ((sum(Data_Cal$WF+Data_Cal$Spiked_Conec)^2)))
  )

# Store this value in the Results DF.
Results$EM[i] <- EM

# Calculate the bias incurred if no correction is made.
Results$Bias[i] <- (XE/Results$Spiked_Conec[i])*100

# Calculate the corrected expected concentration.
Results$Corrected_Conec[i] <- Results$Spiked_Conec[i] + XE

# Calculate EC, the error on the corrected concentration.
Results$EC <- Results$EM + EE

# Print results.
print(Results)

# Export final results table in the clipboard (e.g. to paste in Excel).
write.table(Results, 'clipboard', sep='\t')