EXPLORE

NUMBER 167 JUNE 2015

Newsletter for the Association of Applied Geochemists

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots – Advantages and Disadvantages

Robert G. Garrett, *Emeritus Scientist, Natural Resources Canada, Geological Survey of Canada, 601 Booth Street, Ottawa, ON CANADA, K1A 0E8, Email: garrett@NRCan.gc.ca*

INTRODUCTION

The need, in truth necessity, to monitor the abundant analytical data that applied geochemistry projects generate cannot be over stressed. If all the interpretation, mapping and statistical analysis is to be credible it requires sound sampling design and reliable geochemical analyses. Without a solid foundation later work is potentially at risk.

The most frequently used form of analytical quality control (QC) is the Shewhart plot, where data are displayed against date or time of acquisition, or the order in which the determinations being monitored were received back from the laboratory. The latter is based on the belief that data will be reported in the order they were generated, something that can only be ascertained by discussion with the laboratory. These control charts are named after Walter A. Shewhart, a statistician who worked for Bell Labs in the 1920s. He introduced them to meet the needs of engineers involved in the monitoring the reliability of telephony transmission systems (http://en.wikipedia.org/wiki/Control_chart). In applied geochemistry, Shewhart charts can be employed for monitoring Control Reference Material (CRM) results for analytical drift or occasional excursions due to failure in the systemic analytical process, analogies being chronic and acute medical problems, respectively.

The use of routine duplicate determinations for QA/QC, and the associated estimation of analytical variability, was introduced by Youden (1951), a statistician employed by the National Institute of Standards and Technology. A form of Shewhart plot widely used by applied geochemists displays the absolute difference between duplicate determinations from a single laboratory against their order of accession, or their mean, to be viewed in the context of prior control criteria. In 1973 Thompson and Howarth (1973, 1978) introduced a variant, employing logarithmic scaling, that accommodated the large range, often several orders of magnitude, encountered in duplicate determinations from applied geochemistry projects. Youden's displays of determination-2 vs. determination-1, with equal x- and y-scaling, became known as Youden plots (http://www.medcalc.org/manual/youdenplot.php). In addition to estimating the precision of the analyses statistically from the duplicates, the plot permitted visual assessment of duplicate quality through the closeness of the plotted duplicates to the 1:1 line. When used for inter-laboratory comparisons the departures from the 1:1 relationship could be interpreted in terms of the different laboratories performance.

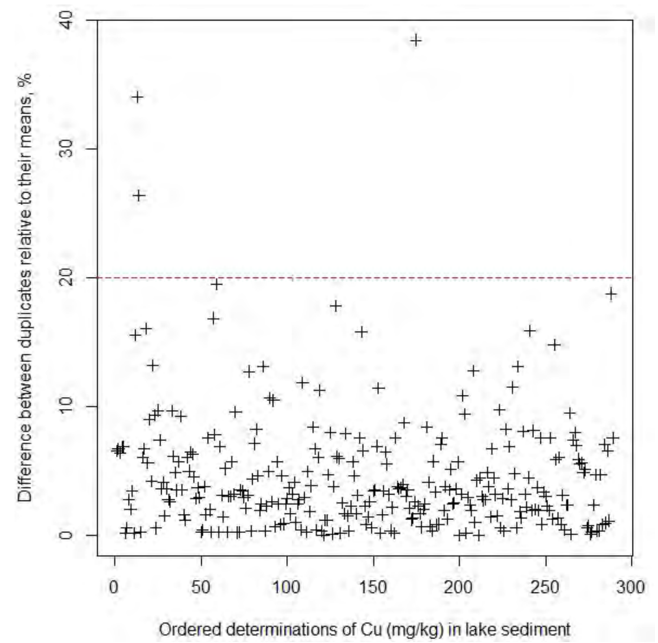

Figure 1. Shewhart plot for incoming laboratory Cu data in order of accession, the dashed line at 20% indicates the acceptance level, three duplicate pairs above the dashed line require investigation

EXAMPLES

Several examples of Shewhart and Youden plots, and a Thompson-Howarth plot, are displayed in this article using duplicate determination data (N = 289) from a 2014 re-analysis program (McCurdy et al., in press) of samples retrieved from the Geological Survey of Canada (GSC) sample archive for two National Geochemical Reconnaissance lake sediment surveys undertaken in Northern Saskatchewan in 1977 and 1993. The graphics presented were prepared using the Open Source R Project for Statistical Computing software (R Core Team 2015) and 'rgr', 'The GSC Applied Geochemistry EDA Package' (Garrett 2013, 2015), that sits on top of R and provides the functionality required for many QA/QC tasks.

continued on page 5

https://doi.org/10.70499/KPAZ9597

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 1

SHEWHART PLOTS

Figure 1 is a typical Shewhart plot as used for QA/QC in the GSC's regional geochemistry programs, where the absolute difference between duplicates is expressed as a percentage of their mean and plotted in the order of data acquisition. Data acceptability is assessed, as analytical batches are received back from the laboratory, against a criterion of difference between the duplicates relative to their means not exceeding some value, 20% in this instance.

For a formal report the percent relative differences can be plotted against their means. As the range of the 289 duplicate means, representing the range of Cu levels encountered in the surveys, is in excess of two orders of magnitude it is appropriate to use logarithmic scaling for the x-axis, see Figure 2. A variation on this display is to plot the Relative Standard Deviations (RSDs) of the duplicate pairs rather than their absolute differences on the y-axis. For the case of a duplicate its RSD is the square root of the square of the difference divided by two, divided by the duplicate pair mean, i.e. $\sqrt{[(x1 - x2)2/2]} / (x1 + x2)/2$, expressed as a percentage. This is essentially a linear transformation of the difference (x1 - x2) and explains the similarity of the point distributions in Figures 2 and 3. All that has changed is the y-axis scaling, and in the case of these Cu data Figure 3 offers no benefit over Figure 2. Where plotting the RSDs can be of benefit is when presenting data to illustrate the problems in acquiring precise analyses as the detection limit of an analytical procedure is approached. Figure 4 presents the data for the Cd duplicates from the same re-analysis program.

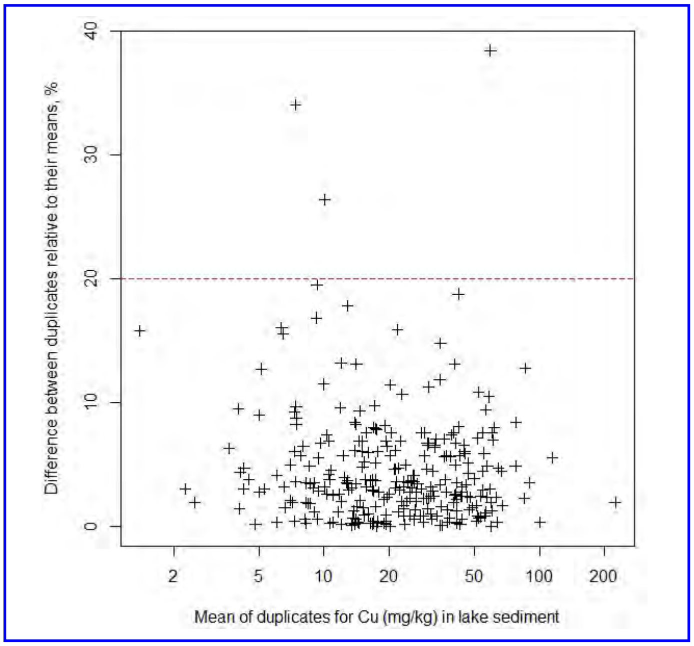

Figure 2. Shewhart plot where the percent relative differences are plotted vs. duplicate pair means for Cu, again with a 20% acceptance level indicated by the dashed line

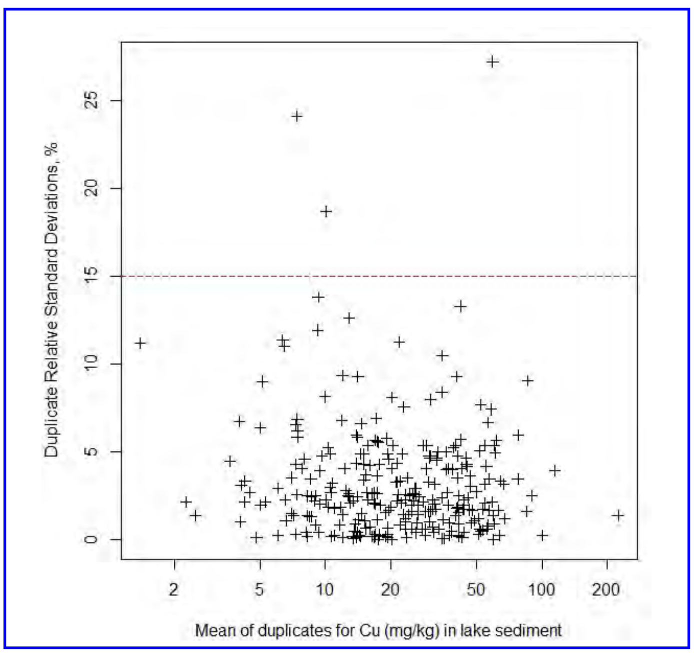

Figure 3. Relative Standard Deviations of individual duplicate pairs vs. duplicate pair means for Cu. An acceptable RSD of 15% is approximately equal to a 20% relative difference acceptance

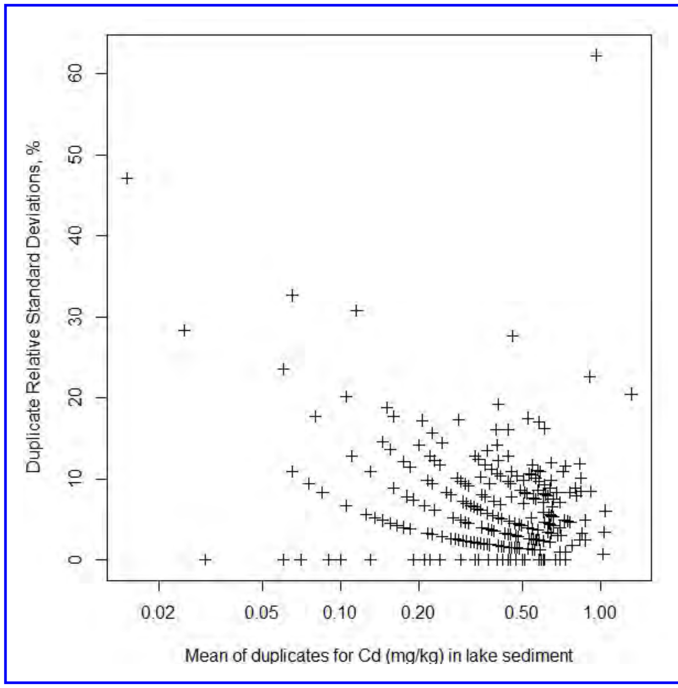

Figure 4. Plot of duplicate RSDs versus duplicate means for Cd demonstrating the effects of quantization and proximity to the lower detection limit of the analytical procedure

The eye-catching features are the trends of decreasing RSDs with increasing means, demonstrating the improvements in precision as levels increase above the method detection limit. These trends occur as discrete 'curves' due to quantization in the analytical data reporting. The detection limit for Cd is 0.02 mg/kg, with the one value below the limit, a non-detect, having been set to 0.01 to permit processing; between 0.02 and 0.09 there are only 9 possible reporting values, thus the number of possible differences is constrained. If the data were reported to an additional decimal place the discrete curves would not be apparent, and the 'curves' would merge into a 'band'. The line of points at zero RSD indicate instances where the duplicate

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 5

determinations were identical, and none of these were due to artifacts from setting non-detects to an arbitrary value of half the detection limit. The impact of proximity to the lower detection limit can be best illustrated by presenting the data on the x-axis as the ratio of the duplicate mean to the lower detection limit, see Figure 5. The single point with a ratio of less than one is the duplicate pair where one measurement was a non-detect. What is apparent from Figure 5 is that the combined effects of quantization in laboratory reporting and proximity to the lower detection limit for Cd are not lost until the duplicate means exceed the lower detection limit of 0.02 mg/kg by a factor of 30, i.e. reach 0.6 mg/kg. The distribution of the points in Figures 4 and 5 is identical and the extra insight is gained by choice of the axis scaling which is explicitly related to the detection limit.

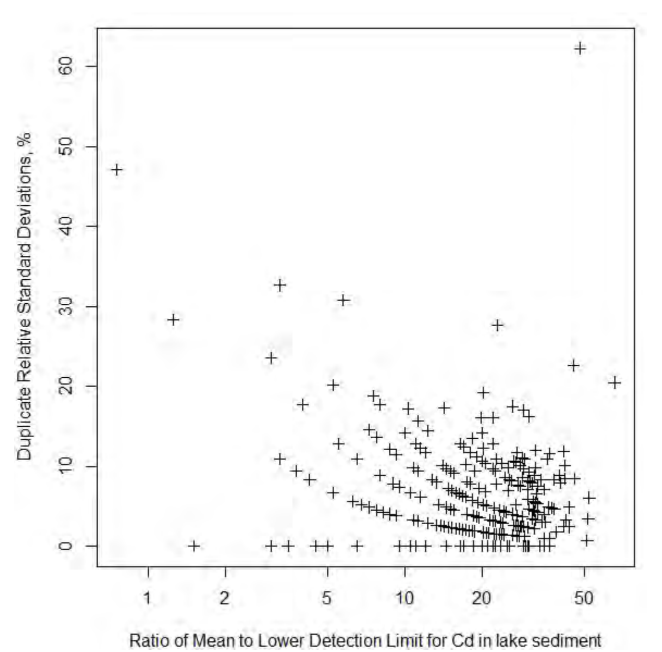

Figure 5. The Cd data of Figure 4, but with the duplicate means expressed as a multiplier of the lower detection limit

THE THOMPSON-HOWARTH PLOT

Figure 6 is the Thompson-Howarth plot for the Cu duplicate determinations displayed in Figures 1 to 3. A target precision for the analyses, expressed as a RSD of 10% was selected. The assumption that the differences are normally distributed permits the addition of the 95 percentile limit for that RSD to the plot and counting the number of duplicates that fall beyond the limit (Fletcher 1981, Garrett & Grunsky 2002, Garrett 2013). There are only two extreme points, and the probability that the data met the 10% RSD criterion exceeds the 0.9999 level. The difference between the 15% RSD acceptance level in Figure 3 and the acceptance that the RSD is better than 10% by the Thompson-Howarth procedure (Fig. 6) is that the Thompson-Howarth procedure is a population estimate, while in Figure 3 the RSDs apply to individual duplicates in the population.

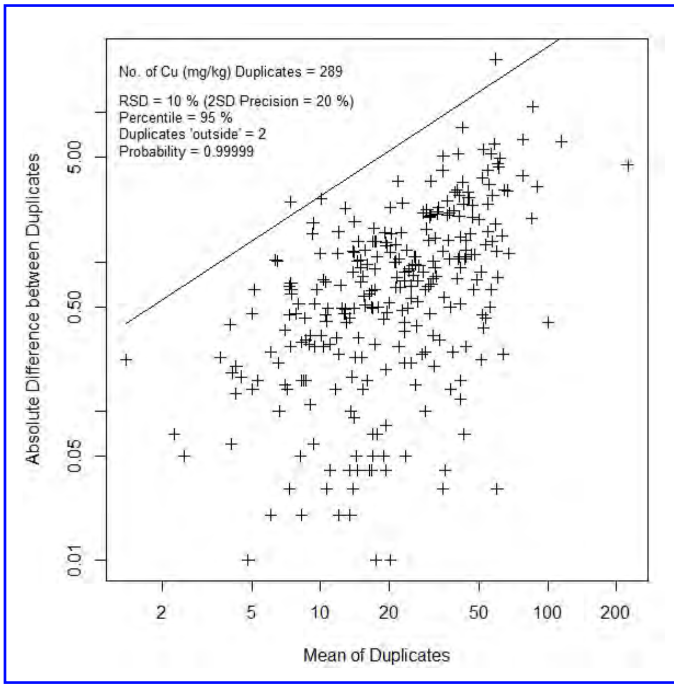

Figure 6. Thompson-Howarth plot of the Cu data with a target RSD of 10%. The diagonal line indicates the expected 95th percentile of the duplicates achieving a 10% RSD

From the point of view of making statements concerning the population statistics the Thompson-Howarth procedure is superior as a QC procedure. But for illustrating duplicate analysis behavior in reports plotting the individual RSDs in various ways has demonstrable advantages, as shown above.

YOUDEN PLOTS

Frequently geochemical reports present duplicate data as Youden plots, where the values for the duplicate determinations are plotted against the values for the sample from which the duplicate was split during sample preparation. In the instance of field duplicates, the analyses would be for two samples closely co-located on the scale of the survey. Figure 7 is the Youden plot for the Cu duplicate determinations in Figures 1 to 3, with the two axes equally logarithmically scaled, as the data span in excess of two orders of magnitude, and the 1:1 line plotted. Adequacy of the data is estimated by proximity to the 1:1 line. In some reports the correlation of the duplicates across the range of the data is presented. However, one could still have a good correlation between the duplicates and a systematic drift of greater divergence between the duplicates with increasing or decreasing levels. The correlation coefficient is not an informative statistic in the QA/QC context. Additionally, a regression line is sometimes estimated, and the gradient tested to determine if it is statistically equal to one, and the intercept is statistically equal to zero. As the determinations plotted against each other are independent of each other ordinary least squares regression is inappropriate, and orthogonal regression, a.k.a. the reduced major axis, should be used to quantify the interrelationship between the two sets of determinations (Garrett, 2013). In fact, the correct

continued on page 7

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... *continued from page 6*

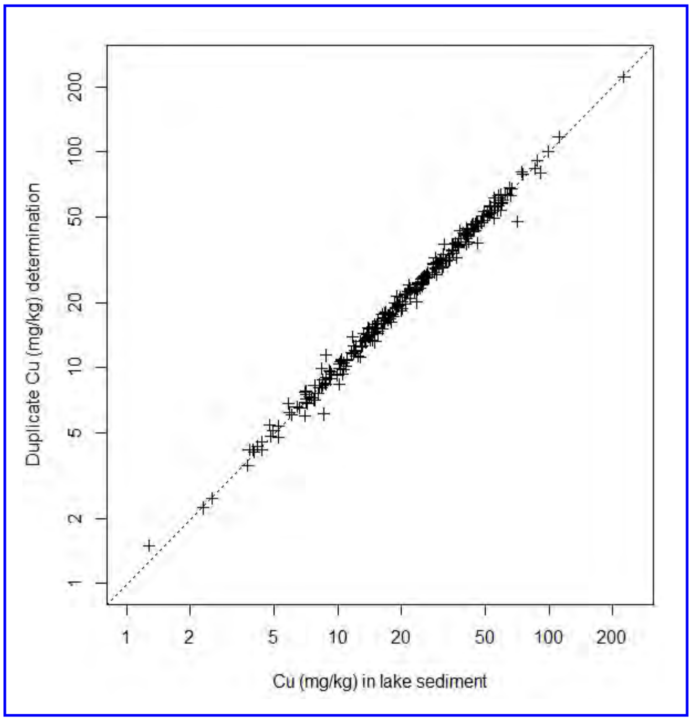

Figure 7. Traditional Youden plot for the duplicate Cu data. The diagonal line indicates the 1:1 relationship along which all data would ideally plot

statistical test does not involve regression modeling and is for the average difference between all the pairs of determinations being zero, the paired t-test (Reimann et al. 2008).

When the difference between two determinations is zero it follows that the ratio of their values will be one. This suggests an alternative plot, where the ratios between the determinations are plotted against the means of the duplicates on the y- and x-axes, respectively, see Figure 8.

Logarithmic scaling is used on both axes, firstly, for the x-axis as the data span more than two orders of magnitude, and secondly, and more importantly, for the y-axis so that similar-fold differences from unity plot the same linear distance above and below unity. Thus a value of 2 will plot the same distance above the unity dashed line as a value of 0.5 will plot below.

The mean ratio and its standard deviation may be estimated and the standard error of the mean computed. If the difference between the mean ratio and unity is within the 95% two-sided confidence interval on the mean there is no statistically significant difference between the duplicate determinations at the 95% confidence level. The 'rgr' function that prepares Figure 8 displays the result of this test, the equivalent of a paired t-test. The 95% confidence interval on the ratios is estimated and plotted, and optionally the information placed as a text block on the plot at a position of the user's choice (see Fig.8). Both classical and robust estimates of the 95% confidence interval on the range of the ratios are made, the latter using the median and Median Absolute Deviation (MAD) rather than the mean and standard deviation. For Cu in Figure 8, there are 13 duplicates that fall outside the classical 95% limits, as

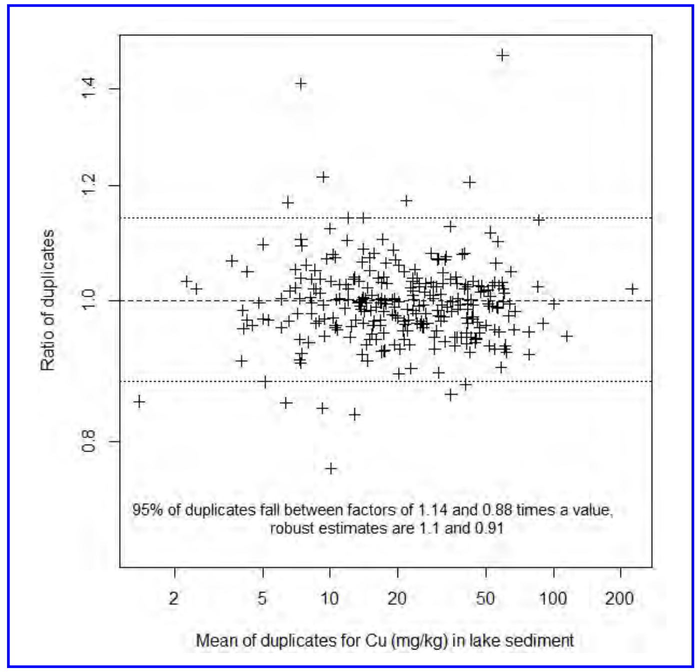

Figure 8. Plot of duplicate pairs as ratios vs. the duplicate mean. The dotted horizontal lines indicate the upper and lower 95% confidence bounds for the duplicates

would be expected for a data set of size 289 with normally distributed data.

Figures 9 to 11 display the Youden and ratios plots for the Mn data. The data span in excess of three orders of magnitude, making plotting with log-log scaling essential; four duplicates require follow-up, with the duplicates reporting uncharacteristically low levels relative to the routine determinations (Fig. 9).

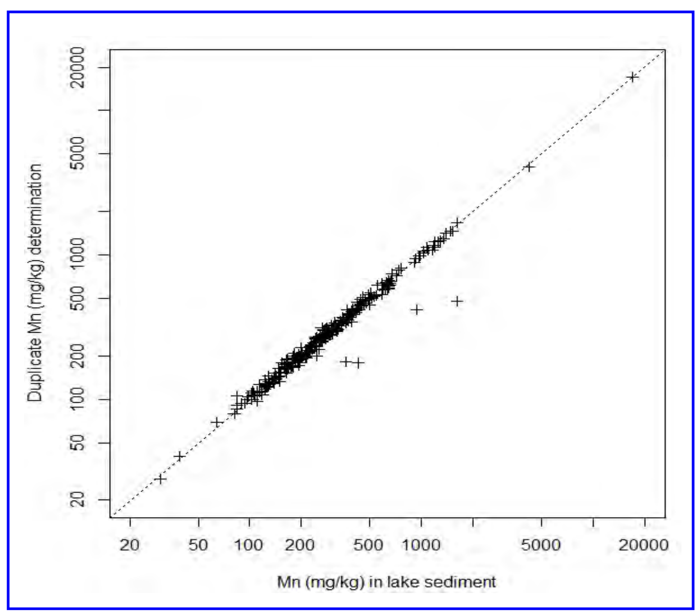

Figure 9. Traditional Youden plot for the duplicate Mn data, with the 1:1 line indicated

continued on page 8

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 7

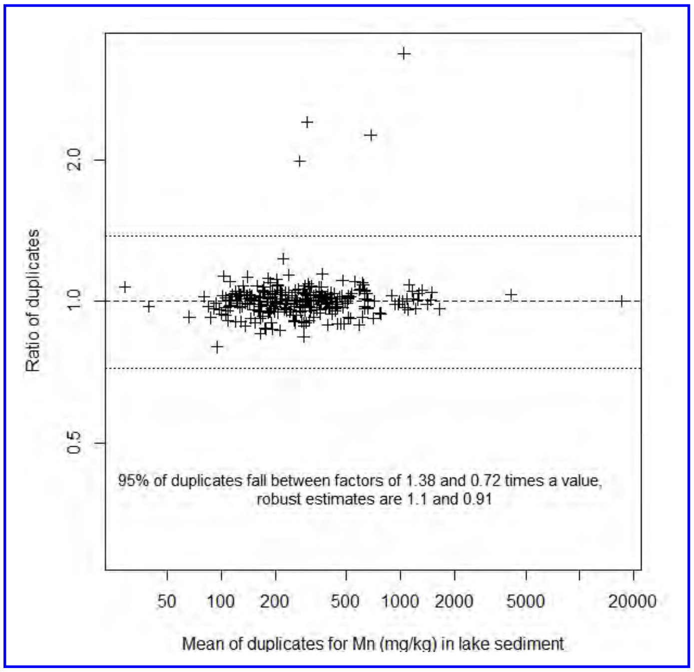

Figure 10. Ratio plot for the Mn duplicates. The dotted horizontal lines indicate the upper and lower 95% confidence bounds for the duplicates

Figure 10 is the equivalent ratios plot, where the ratios are computed as determination-1 divided by determination-2. The routine determination precedes the duplicate in the data file, and as a result the four errant duplicates plot above the unity line. It is not known whether the duplicate determinations were relatively low, or the routine determinations relatively high, or even some combination of these two situations. It is worth noting the effect of the four 'outliers' in both Figures 10 and 11. In Figure 10 they exert a major effect on the estimated 95% limits for the range of ratios, for a value of 100 mg/kg there is a 95% chance the duplicates were in the range of 138 to 72 mg/kg. The robust estimates of the expected range, which have 'ignored' the 'outliers' by using the median and MAD, are 110 and 91 mg/kg, indicating the benefits of following-up on the four duplicate determinations and rectifying whatever problems may have occurred in the batches of analyses where they occurred. Finally, Figure 11 presents a probability plot of the ratios in addition to the ratio plot. This confirms that, as one would hope, that the ratios approximate a normal distribution and the four errant duplicates are true 'outliers' requiring attention.

As noted above, the optimal parametric statistical test for the equivalence of duplicate determinations is the paired t-test, which compares the mean difference for the duplicates to zero in the context of the spread of the duplicate differences about their mean. If a confidence band, at some stated level, around the mean difference includes zero one accepts the hypothesis that the mean difference is zero and that the duplicates are 'identical' in the context of the study. If zero falls outside the confidence bounds, there are serious problems that need to be understood before proceeding further with data interpretation.

For three reasons the paired t-test for duplicate geo-

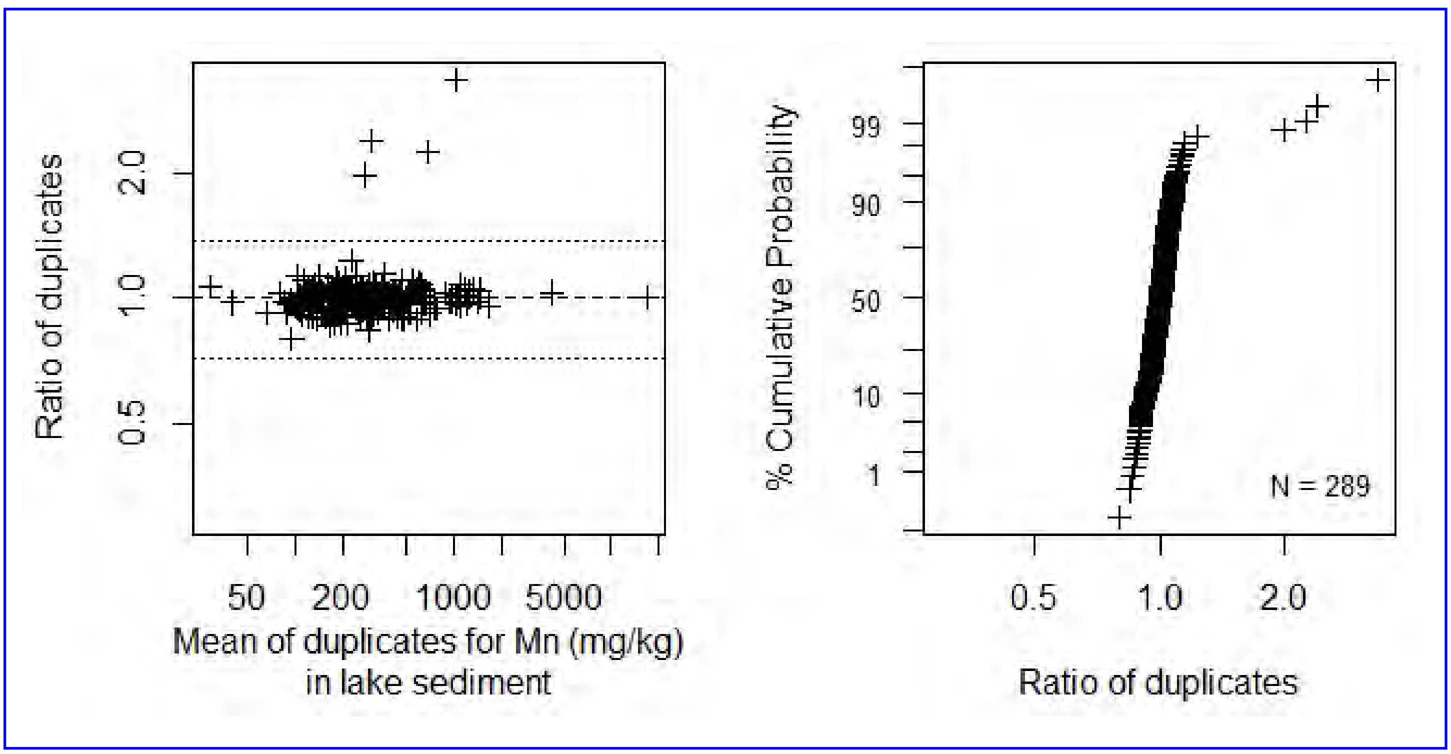

Figure 11. Combined ratio plot and cumulative normal probability plot of the ratios for Mn. The dotted horizontal lines indicate the upper and lower 95% confidence bounds for the duplicates

continued on page 9

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 8

chemical analyses should be carried out with a logarithmic transform. Firstly, the data often span more than an order of magnitude which induces a property known as heteroscedasticity, implying that variances of groups of data across the data range are not constant. Constant variance is a critical assumption for parametric statistical tests like the t-test, in fact it is arguably more important than the requirement for normality. The standard solution to this problem is a logarithmic transform, see Bartlett (1947) and many statistical texts since. Secondly, and equally important, geochemical data are compositions and are constrained in the values they can take, i.e. they cannot be less than zero, if such could ever be determined, or greater than 1 million mg/kg or 100% (Aitchison 1986). The situation with univariate data has been discussed by Filzmoser et al. (2009), and at trace element levels, and in fact up to 10%, a logarithmic transform will 'open' the data satisfactorily. Finally, geochemists tend to think in ratios, levels are 'twice average background', or 'three times threshold'. Working after a logarithmic transform simply treats the data the way geochemists think about them, as ratios.

With logarithms a subtraction is a division, thus taking the difference of the logarithms of the duplicates, or computing their ratios is equivalent, and a paired t-test on the 'logs' of the data is equivalent to a t-test on the ratios. Table 1 presents the screen output when ratio plots are generated (Figs. 8, 10 and 11). The paired t-test results are in the second 'paragraph' for the plot results, they include the absolute difference from unity, the standard error of the mean of the ratios, and the 95% confidence interval. The software tests to see whether unity falls within the confidence interval, and if it is the following line is displayed informing the user that there is no statistical difference from unity, and thus there is no systematic bias. Should the test fail an appropriate line is displayed.

DISCUSSION

Several variations have been presented for Shewhart plots of analytical duplicate data from a regional geochemical survey. Each display has its own particular advantage. For monitoring incoming laboratory data plotting in acquisition sequence assists in detecting if a laboratory procedure has gone out of control for a short period of time, or in a single or series of batches – the GSC procedure is to insert a duplicate in each batch of 20 samples for analysis. The Thompson-Howarth plot is also well suited to monitoring data as it is acquired and assessing its acceptability. For summarizing the results for a report it is more informative to plot the duplicate means on the x-axis. It is a matter of choice whether the y-axis is scaled for the absolute differences between the duplicates, or their Relative Standard Deviations – it is the mean of all variances behind the individual RSDs that estimates the overall precision for the duplicate data being presented. As shown for Cd (Fig. 5), plotting the duplicate means as their ratios to the procedure detection limit can assist to displaying the relationship of analytical precision to proximity to the detection limit. If for some element or compound this relationship is unacceptable the display helps justify the need for an alternative procedure. In the case of nugget effects due to a miss-match between particulate/mineral grain size and aliquot

continued on page 10

Table 1. Text display associated with the ratio plots for Cu and Mn

```
> ad.plot4(Cu, "Cu (mg/kg) in lake sediment", ifalt = T, if.order = F)

  Range of ratios of the 289 duplicates: 0.767 1.48
  Median ratio = 0.9961   MAD of ratios = 0.04919    Mean ratio = 0.9989    SD of ratios = 0.06956    95% CI for ratios = 0.1369

  Absolute ratio difference from 1 = 0.001091    SE of Mean of ratios = 0.00409     95% CI = 0.00805
  Mean ratio is not different from 1 at the 95% level, no bias

  95% of duplicates will fall between factors of 1.14 and 0.88 times a value
     Robust factor estimates based on the MAD are 1.1 and 0.91

> ad.plot4(Mn, "Mn (mg/kg) in lake sediment", ifalt = T, if.order = F)

  Range of ratios of the 289 duplicates: 0.8 3.37
  Median ratio = 1   MAD of ratios = 0.05031     Mean ratio = 1.019     SD of ratios = 0.195     95% CI for ratios = 0.3837

  Absolute ratio difference from 1 = 0.01917   SE of Mean of ratios = 0.0115          95% CI = 0.0226
  Mean ratio is not different from 1 at the 95% level, no bias

  95% of duplicates will fall between factors of 1.38 and 0.72 times a value
     Robust factor estimates based on the MAD are 1.1 and 0.91
```

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 9

weight the solution could be as simple as scaling-up the procedure to accommodate an increased aliquot weight.

The traditional Youden plot is well established as a report display for duplicate data. However, it is extremely wasteful of plot space, especially when the data range over several orders of magnitude (see Figs. 7 and 9). Furthermore, it encourages the use of the correlation coefficient, and even a regression model, implicitly accepting either the initial determination or duplicate as the dependent variable. As noted previously, the correlation coefficient is not an informative statistic in terms of assessing all the aspects of bias between the duplicate analyses. Furthermore, a simple linear regression is not the appropriate regression model, the question being, which is the independent variable and which the dependent? There is no answer. The appropriate regression model is one for orthogonal regression (the reduced major axis), which accepts both variables as independent. This model minimizes the sums of the error squares at right angles to the regression line, and not parallel to either the routine or duplicate determination axis. The optimal statistical test is the paired t-test that directly tests for the average difference between the duplicates being zero.

The ratio plot, a Youden variant, is a far more efficient use of plot space. Only one axis (x) spans the whole range of the duplicate data and the other (y) focuses of the differences from the 1:1 line always plotted on the traditional Youden plot, and the real subject of interest. To give equal graphical 'weight' to ratios above and below unity it is necessary to employ logarithmic scaling. Due to the nature of the ratio it implicitly leads to a statistical test equivalent to a paired t-test on the logarithms of the duplicates – an appropriate test for acceptability of the duplicates, assuming they satisfy any initial prior quality control criteria. Furthermore, this approach leads to the estimation of 95% limits on the expected range of the duplicates across the range of the data, a useful way of presenting the analytical precision.

The importance of the graphical inspection of QA/QC data cannot be overstressed. In the examples provided any outliers can be readily recognized (Figs. 1-3, 6, 9-11). Simply calculating precisions from duplicate data can easily hide unpleasant truths. If outliers exist they need to be explained, they could be due to procedural errors in the analytical process, in which case they need to be understood and appropriate action undertaken. Alternately they could be due to sample inhomogeneity arising from the particulate nature of the mineral(s) containing the element being monitored. If the latter is the case it has to be accepted as a reflection of the geochemical/mineralogical reality and precision can only be rectified by adopting a different analytical procedure. If sample heterogeneity has to be accepted, robust estimators may be used to calculate precision which down-weight or remove the influence of outliers. In effect, a precision is estimated that applies to the main mass of the data, ignoring the outliers, see for example Table 1 and Figures 8 and 10. This may be comforting, but can also be misleading as to the realities of the analyses.

CONCLUSION

It is proposed that ratio plots are a useful addition to the displays available to the applied geochemist for displaying analytical duplicate data, noting that it could also be used for field sampling duplicates. Ratio plots are graphically efficient and provide for a rapid appraisal of duplicate quality across the data's range. While software is available in 'rgr' (Garrett 2013, 2015) to prepare all the plots and carry out the statistical computations presented here, the procedures can be implemented in other software packages.

ACKNOWLEDGEMENTS

The author thanks Martin McCurdy for drawing the author's attention to the issues around the graphical presentation of duplicate data and appropriate statistical expressions of analytical precision. He also provided the data set used here from one of his current projects. The author also thanks Eric Grunsky and Martin McCurdy, both of the Geological Survey of Canada, Ottawa, for their comments on a draft of this article for **EXPLORE**, and to the Editor of **EXPLORE** and Lynda Bloom for their constructive reviews. ESS Contribution No.: 20140496

REFERENCES

Aitchison, J. 1986. The statistical analysis of compositional data. Chapman and Hall, London, 416 p.

Bartlett, M.S. 1947. The use of transformations. *Biometrics,* **3**(1), 38-51.

Filzmoser, P., Hron, K. & Reimann, C. 2009. Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of the Total Environment,* **407**(2009), 6100-6108.

Fletcher, W.K. 1981. Analytical methods in geochemical prospecting. In: G.J.S. Govett (Ed.), *Handbook of Exploration Geochemistry, Vol. 1,* Elsevier, Amsterdam, 255 p.

Garrett, R.G. 2013. The 'rgr' package for the R Open Source statistical computing and graphics environment - a tool to support geochemical data interpretation. *Geochemistry: Exploration, Environment, Analysis,* **13**(4), 355-378. doi 10.1144/geochem2011-106.

Garrett, R.G. 2015. rgr: The GSC Applied Geochemistry EDA Package. http://cran.r-project.org/web/packages/rgr/index.html

Garrett, R.G. & Grunsky, E.C. 2002. S and R functions for the display of Thompson-Howarth plots. *Computers & Geosciences,* **29**(2), 239-242.

McCurdy, M.W., Pehrsson, S.J. & Campbell, J.E. in press. Regional lake sediment and water geochemical data, North-

continued on page 12

A Comparison of Shewhart, Thompson and Howarth, and Youden Plots... continued from page 10

ern Saskatchewan (NTS 64L, 64P, 74N, 74O and 74P). *Geological Survey of Canada,* Open File 7746.

R Core Team, 2015. The R Project for Statistical Computing. http://www.r-project.org/

Reimann, C., Filzmoser, P., Garrett, R. & Dutter, R. 2008. Statistical Data Analysis Explained: Applied Environmental Statistics with R. John Wiley & Sons Ltd., Chichester, 343 p.

Thompson, M. & Howarth, R.J. 1973. The rapid estimation and control of precision by duplicate determinations. *Analyst,* **98**(1164), 153-160.

Thompson, M. & Howarth, R.J. 1978. The rapid estimation and control of precision by duplicate determinations. *Journal of Geochemical Exploration,* **9**(1), 23-30.

Youden, W.J., 1951. Statistical methods for chemists. John Wiley & Sons Ltd., New York, 126 p.