

Two Macros for Producing Graphs to Assess Agreement Between Two Variables

David W. Meek

USDA-ARS-MWA National Soil Tilth Laboratory
Ames, IA 50011

ABSTRACT

This paper is intended for those familiar with PROC GPLOT from SAS/GRAPH^{®1} and PROC UNIVARIATE and PROC CORR from base SAS[®]. Via macro procedures, two graphs are developed to aid in the assessment of agreement between two variables. The first is a variant on a mean-difference plot, the second is a variant on a bivariate plot. A published data set will be used to generate an example set of graphs.

INTRODUCTION

In many fields of science and engineering a researcher often has to assess agreement between two variables of interest. This assessment may have different names in different disciplines. In some cases, the analysis may also be called *concordance*, *equivalence*, or *reproducibility*. Typical examples are: (1) Do two different instruments give the same measure over a specified range of values? (2) Are measurements of an environmental variable monitored at two different locations the same throughout the year? (3) Are the chemical concentrations of a series of split samples run at independent laboratories in agreement? (4) Are model predictions consistent with independent observations? In the first example, the researcher may want to use a cheaper, smaller, or more available instrument or independently test selected readings. In the second example, the similarity of ambient or controlled factors or conditions may be of interest. The third example is a common quality control/quality assurance practice. In the fourth example, model assessment is a fundamental practice throughout science and engineering. There are many different terms used for the assessment of a model. Two common ones are *model performance* and *model validation*.

Literature in numerous scientific publications contains a multitude of ad-hoc and formal analyses as well as a variety of graphics. One of the most widely used procedures is an assessment of bias which may also be called the *mean difference* (often denoted *d*) or the *mean bias error* (often denoted *mbe*). With a formal assessment of a specific data set, Bland and Altman (1986) provide a very thorough example. The graphic in this assessment centers on what Cleveland (1993) calls a *mean-difference* plot or *m-d plot* for brevity. This graph is a scattergram of the paired observation difference of the two variables on the vertical axis and the mean of the each pair of observations for the two variables on the horizontal axis. The plot shows the scatter of the difference with respect to that of the mean. In addition it may reveal an undesirable relationship between the difference and the size of the measure. Based on the central limit theorem, the difference variable can often be considered reasonably normally distributed about the mbe. Although this next assumption should be evaluated, the difference variable is considered to have independent and identically distributed observations (formally the i.i.d. criterion). In some data sets the difference series may be spatially or temporally autocorrelated. Inferential statistics, based on the normal and i.i.d. assumptions, are included as horizontal reference lines in Bland and Altman's m-d plots. Let *n* be the number of observed pairs, *mbe* be the mean difference, δ_{mbe} be the standard deviation, and *s.e.* the standard error of the mbe. Specifically included are the mbe with its 95% confidence limits ($mbe \pm T_{n-1, 0.975} s.e.$) and the $mbe \pm 1.96\delta_{mbe}$. Ideally, the first set of limits should include zero if there is no serious bias. The actual acceptable significance level for the bias depends on the phenomenon as well as the goals of the analysis. In general, this analysis is like the diagnostics used on regression residuals (Ch. 3 in Montgomery and Peck, 1982). The second set of limits helps to reveal the number and size of large differences. Again, phenomenon, use, and subject specific needs dictate the acceptability of large differences.

Another widely used tool is a graphic that Cleveland (1993) calls a *bivariate plot* or *b-v plot* for brevity. The graph is a simple scatter plot of the paired observations, ideally on axes of identical length and scale (aspect ratio of 1) with 45° reference line included. The latter may be called many different names like the *1-to-1 line* or the *line of equivalence*. Lin (1989) developed an insightful and very widely used test statistic to assess the model, $y=1x$, called the concordance correlation coefficient, a.k.a., r_c or CCC. The r_c statistic is an adjusted version of the well-known Pearson product-moment correlation coefficient, r , which can be formally evaluated. Let the *y* variable have a mean, μ_y , and standard deviation, δ_y , and the *x* variable have a mean, μ_x , and standard deviation, δ_x . Then $r_c = rC_b$ where $C_b = [(v+1/v+u^2)/2]^{-1}$. Here $v = \delta_x/\delta_y$ and is called a *scale shift* while $u = (\mu_x - \mu_y)/(\delta_x\delta_y)^{1/2}$ and is called a *location shift relative to scale*. Notice $\mu_x - \mu_y$ is another way to denote the mbe. Some statisticians like Atkinson and Nevill (1997) and Bland and Altman (1986) argue against using this sort of plot and associated analysis. Lin and Chinchilli (1997) prepared a thorough and insightful reply to their concerns. Both correlation coefficients alone and t-test procedures alone have well-known and well-studied cases that reveal each test alone can be misleading or inadequate (Lin, 1989). The r_c statistic is an omnibus procedure

¹The mention of a trade name is for informational purposes only and is not an endorsement by the USDA-ARS.

that is in general more comprehensive than either of the specified single tests. The careful analyst, however, will always consider the underlying assumptions, study design, inference space, data range and representativeness, distribution, measurement error, and so on. In another work, Lin et al. (2002) expound upon the concordance correlation concept and offer publically available SAS macro code on the WEB (<http://www.uic.edu/~hedayat/>). Whether or not formal assessment of the relationship in a b-v plot is carried out, many analysts recommend always looking at both of the plots in all cases. This paper, therefore, shows the graphs produced by two SAS macros, one that generates an m-d plot and one that generates a b-v plot.

THE DATA SET

The PERF data (peak expiratory flow rate in liters per min, l/min), published in the first table of Bland and Altman (1986), are used in the examples presented in this paper. The set has $n = 17$. The SAS variable named Meter_1 is the first column for the Wright peak flow meter while Meter_2 is the first column for the Mini-Wright peak flow meter. For the formal analyses associated with the m-d plot, the Meter_2 - Meter_1 difference has $mbe = 2.1 \pm 9.4$ l/min and 95% confidence limits of -17.8 and 22.0. The lower data limit, $mbe - 1.96\delta$, is -73.9. The upper data limit, $mbe + 1.96\delta$, is 78.2. For the concordance correlation analysis associated with the b-v plot, $r = 0.9433 \pm 0.0287$ ($p < 0.0001$), $C_b = 0.9994$, $r_c = 0.9428$ with $u = 0.0185$ and $v = 0.9725$. In their analysis of these data, Bland and Altman believe the variability of the difference is too large to accept agreement for medical purposes.

THE GRAPHS

An award winning graph produced from SAS macros that partially combine features from the m-d and b-v graphs on a single plot is readily available (Berg, 1992). Figure 1 shows a Berg plot for the chosen example data set. Notice it has box-plots for each of the two-variables and for the difference of the two. The graph, as is, may suit the needs of many but lacks a box-plot for the mean of the two variables and the inferential statistics on the difference. Moreover, while a box-plot can reveal symmetry or asymmetry in a variable's distribution, it does not reveal possible multi-modality that a histogram or stem-leaf plot offers. Assessing the modality is a crucial part of the analysis.

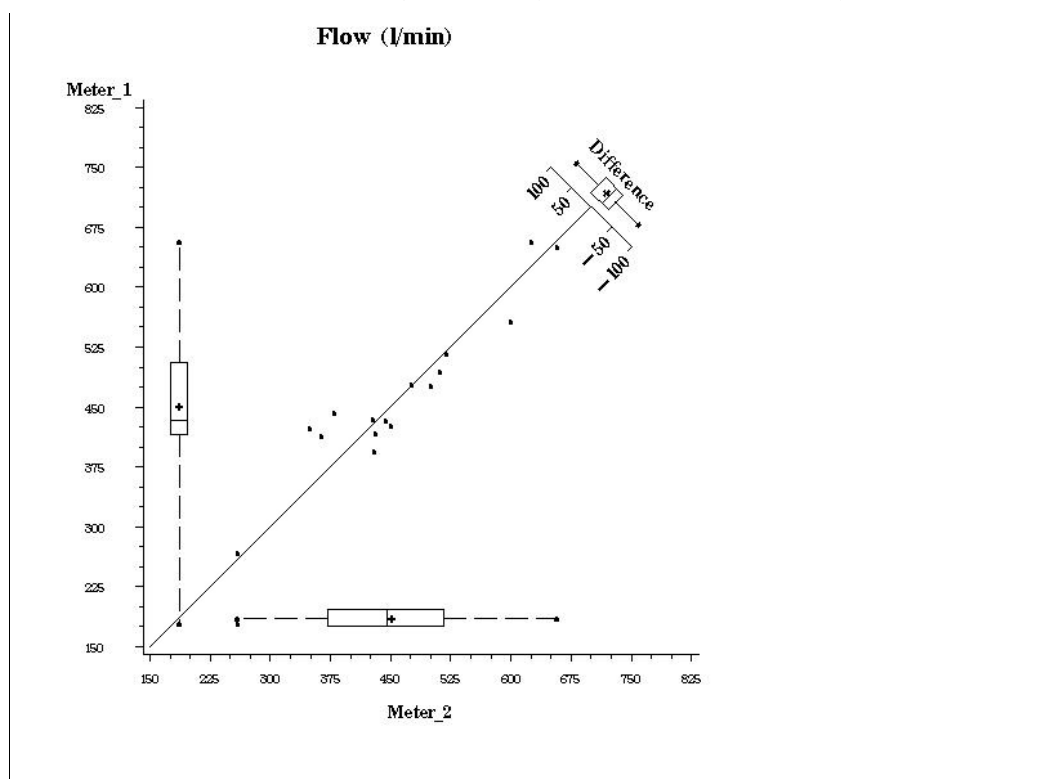


Figure 1. A Berg plot for the Bland-Altman (1986) data set featuring box-whisker plots for each variable and for the difference of the two.

A careful assessment of agreement would include examining descriptive statistics for the two variables as well as their paired mean and difference along with the distributional behavior of each of these variables. The descriptive statistics and distributional properties of all these variables can and should be formally assessed with options available in PROC UNIVARIATE. For insight and ease, trying to reveal as much as possible in the graphs is a worthy goal. To this end, principles and ideas from the statistical graphics literature are employed to reduce clutter and only include essential features (Cleveland, 1993 and Tufte, 1983). For simplicity and to assist the color blind, the graphs are plotted entirely in gray-scales. Macros to plot each graph separately are developed for several reasons. Including all the desired

features in a combined plot would make the graph much busier and the features for the difference assessment could be highly compressed and consequently difficult to plot clearly. The confidence band on the difference as well as a possible relationship between the mean and difference are easier to see on an m-d plot. In addition, as previously stated, some researchers only consider the m-d plot and related analysis while others like to consider the two plots sequentially. The graphs are principally meant for visual analysis so output is to the CRT in a Windows OS environment.

The M-D PLOT

This graph uses PROC Gplot and the %annomac routines available in SAS/GRAPH®. Figure 2 shows the m-d plot for the Bland and Altman (1986) data. Following the idea from Tufté (1983), each axis is replaced with a box-plot that serves both as an axis and data distribution description. The whiskers in this box-plot extend over the entire data range. The "+" sign is the arithmetic mean. The thin black lines are the whiskers. The thick gray lines with the gap are middle 50% of the data from the first to third quartile with the gap representing the median value. In addition a crude histogram of the distribution appears over the top of the horizontal axis and to the right of the vertical axis. The idea was inspired by Hintze and Nelson's violin plot (1998). Each was constructed by dividing the respective range into 16 equal segments and plotting the relative scaled number of observations within each segment at the midpoint of the segment. The data are represented by black dots. A black dashed line is the zero difference reference line; the black "0" on the right hand side is the label. A solid gray band with a white line gap in it represents the 95% confidence interval for the mbe with the white gap line representing the mbe. The value of the mbe is labeled in gray on the right hand side of the gap line. The gray dotted lines labeled "U" and "L" near the top and bottom of the plot area are the $mbe \pm 1.96\sigma_{mbe}$ data limits. Graphically it is easy to see that the distributions of the data, especially that for the difference are reasonably unimodal, there seems to be no relationship with size, the mbe is not different from zero, but the range of the difference is fairly large with respect to that of the mean, being about 0.4. Recall Bland and Altman found this level of variability unacceptable for their purposes.

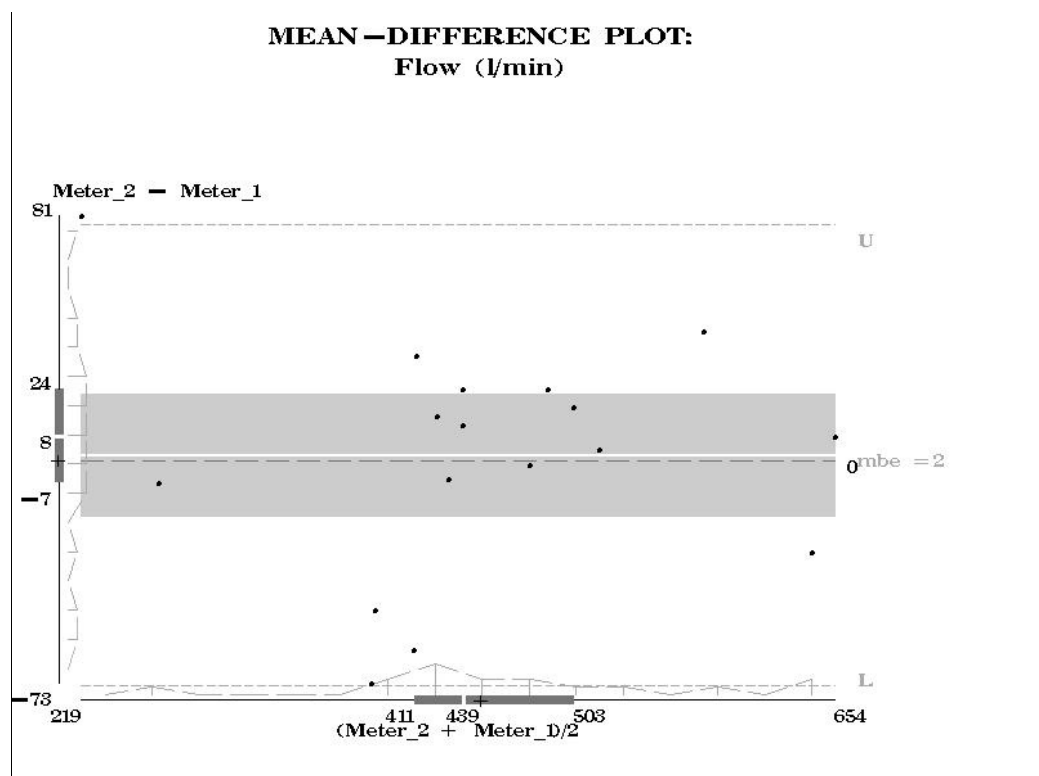


Figure 2. The m-d plot for the Bland-Altman (1986) data with box-plot axes, axial histograms, and gray-scale reference features.

To use the m-d plot macro, assume the SAS data set is named *a0*. Define the mean and difference in the data step. In this example the difference variable is named *dif* for Meter_2 - Meter_1. The mean variable is named *mu* for (Meter_2+Meter_1)/2. After both the macro code is compiled and data step is run, the routine is invoked as follows:

```
%mmdplot(a0,mu,1,dif,1,xtitle="(Meter_2 + Meter_1)/2",ytitle="Meter_2 - Meter_1",gtitle="Flow (l/min)")
```

The names of the variables, axes titles, and graph title are evident. The 1 after each variable name is to tell the macro the rounding unit for a SAS round function; if one decimal place is desired then a 0.1 would replace the 1.

The B-V PLOT

This graph again uses PROC Gplot and the %annomac routines available in SAS/GRAPH® along with the same type of axes used in the m-d plot. Figure 3 shows this b-v plot for the Bland and Altman (1986) data. The data in the scattergram are shown as black dots. The only unique reference line in this kind of plot is the gray 45° line. For the convenience of the user who wants them, the r_c and its component u and v values are printed in black on the right hand side of the graph. Differences in the empirical distribution for each variable are revealed. While the range of Meter_1 to that of Meter_2 is about 1.2, the interquartile range of Meter_1 to that of Meter_2 is about 0.6. The difference of the medians is 11 l/min. The lowest value on the vertical axis appears to be a severe outlier. At a glance, neither empirical distribution appears to be unimodal. More importantly, the distributions do not appear to be the same. The dissimilarities should encourage the careful researcher to consider more extensive and detailed univariate analysis, available in PROC UNIVARIATE output; doing so confirms most of the visual judgements. Consideration of the b-v plot and concordance analysis could be reasonable for this data set if the variability of the difference were acceptable because the mbe was not found to be significant in the m-d analysis. If Meter_1 or mbe variability were smaller then u would be closer to 0 and v closer to 1. Acceptable standards for variability of the difference in an m-d analysis or r_c , u , and v in a concordance analysis are not statistically set but based on considerations such as knowledge of the experimental design, familiarity with the procedures or instrumentation, the reproducibility of the data, and most importantly, the application needs. Note that Lin and Chinchilli (1997) offer guidelines for planning and using the concordance analysis.

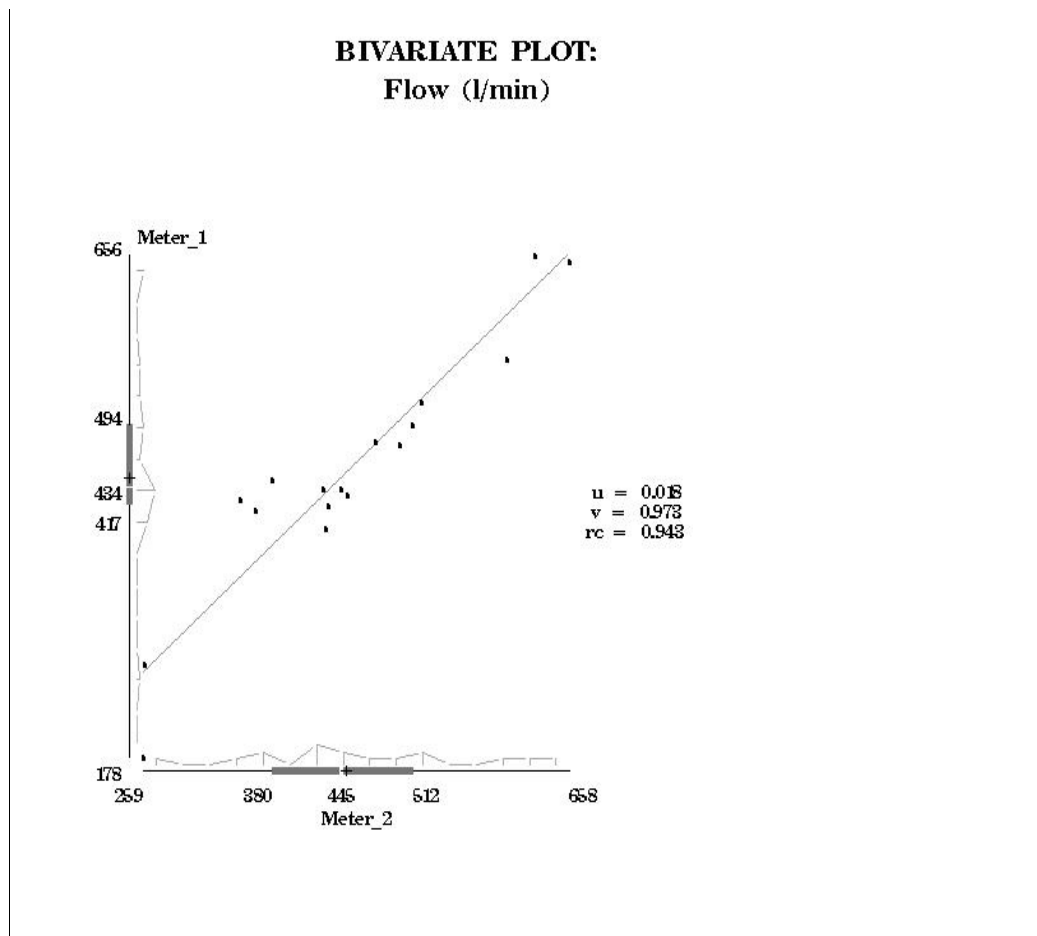


Figure 3. The b-v plot for the Bland-Altman (1986) data with box-plot axes, axial histograms, gray 45° reference line, and correspondence correlation coefficient statistics.

To use the b-v plot macro, assume the SAS data set is named *a0*. Define the input variables in a data step. In this example the y variable is named *Meter_1* and the x variable is named *Meter_2*. After both the macro code is compiled and data step is run, the routine is invoked as follows:

```
%mbvplot(a0,Meter_2,1,Meter_1,1,xtitle="Meter_2",ytitle="Meter_1",gtitle="Flow (l/min)")
```

The names of the variables, axes titles, and graph title are again evident. Also, as before, the 1 after each variable name is to tell the macro the rounding unit for a SAS round function.

DISCUSSION AND CONCLUSIONS

These graphs can be exported to many different standard formats through the export wizard. The export may change the 1-to-1 aspect appearance of the b-v plot changing the slope from 45°. If exported, the import of the graph needs to be properly sized and scaled in the application program. In addition, the procedures have been tested and work well on variable scales for ranges from 0 to 10, 0 to 100, and 0 to 999. Most scientific and engineering applications can fit in these intervals with a suitable change of units. An attempt to graph the data in the 0 to 1 range was unacceptable because the axes and other features were too large, obscuring some of the data. The macros, however, worked well when the data were changed to percent. Alternatively, in some applications, a user can edit the macros.

These macros were developed to help researchers at the author's agency assess agreement between two variables for data from many different observational studies, analytical laboratory practices, and designed experiments. The proper use and assessment of agreement is the responsibility of the individual researcher. Moreover, the researcher should have a sound and more complete understanding of the data set and the research goals involved. In the given example with the Bland-Altman (1986) data, the practical needs of the medical clinician defined the limits of acceptable variability. Further information and the code for both of the macros are available from the author.

REFERENCES

- Atkinson, G. and Nevill, A. 1997. Comment on the Use of Concordance Correlation to Assess the Agreement Between Two Variables. *Biometrics* 53(2): 775-777.
- Berg, R.L. 1992. First Place Best Presentation of Data-Monochrome, P. 1521-1527. Sugi 17 - Proceedings of the Seventeenth Annual SAS® Users Group International Conference. Apr. 12-15, 1992. Honolulu, HI.
- Bland, J.M. and Altman, D.G. 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet* i: 307-310.
- Cleveland, W.S. 1993. *Visualizing Data*. Hobart Press, Summit, NJ.
- Hintze, J.L. and Nelson, R.D. 1998. Violin Plots: A Box Plot-Density Trace Synergism. *Am. Statis.*, May 1998 V. 52(2): 181-184.
- Lin, L.I. 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45: 255-268.
- Lin, L. and Chinchilli, V. 1997. Rejoinder to the Letter to the Editor from Atkinson and Nevill. *Biometrics* 53(2): 777-778.
- Lin, L.I., Hedayat, A.S., Sinha, B., Yang, M. 2002. Statistical Methods in Assessing Agreement: Models, Issues, and Tools. *J. Am. Stat. Assoc.* 97(457): 257-270.
- Montgomery, D. and Peck, E. 1982. *Introduction to Linear Regression Analysis*. J. Wiley and Sons., New York, NY.
- Tufte, E.R. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.

ACKNOWLEDGMENTS

This work was supported by the USDA-ARS National Soil Tilth Laboratory, Ames, IA, Dr. J.L. Hatfield, Director. Thanks to Elsevier, Health Sciences Rights Department, Philadelphia, PA 19103-2899, USA for permission to use the Bland and Altman (1986) data set. Thanks to Deb Palmquist, USDA-ARS MWA, Peoria, IL, Dr. Jeremy Singer, USDA-ARS MWA, Ames, IA, and Delayne Stokke, Wells Fargo Bank, Des Moines, IA for comments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Name: David Meek
Enterprise: USDA-ARS-MWA National Soil Tilth Laboratory
2150 Pammel Dr.
Ames, IA 50011
Work phone: (515) 294-2246
Fax: (515) 294-8125
Email: Dave.Meek@ars.usda.gov
Web: None

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.