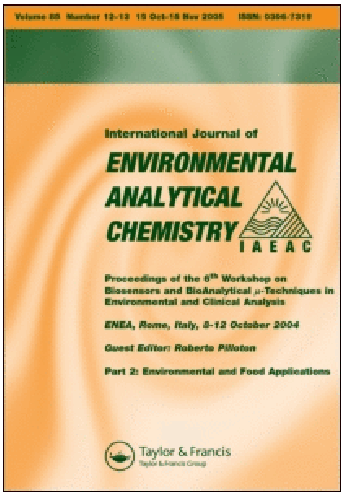

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Semi-Empirical Equations for Interlaboratory Studies in Water Chemistry

J. E. Gaskin[a]
[a] Integrated Monitoring Branch, Atmospheric Environment Service, Ottawa, Ontario, Canada

To cite this Article Gaskin, J. E.(1995) 'Semi-Empirical Equations for Interlaboratory Studies in Water Chemistry', International Journal of Environmental Analytical Chemistry, 60: 2, 281 — 287
To link to this Article: DOI: 10.1080/03067319508042883
URL: http://dx.doi.org/10.1080/03067319508042883

Intern. J. Environ. Anal. Chem., Vol. 60, pp. 281–287
Reprints available directly from the publisher
Photocopying permitted by license only

SEMI-EMPIRICAL EQUATIONS FOR INTERLABORATORY STUDIES IN WATER CHEMISTRY

J. E. GASKIN

Integrated Monitoring Branch, Atmospheric Environment Service, Environment Canada, 373 Sussex Drive, Ottawa, Ontario, Canada, K1A 0H3

(Received, 5 July 1994; in final form, 8 September 1994)

Recently, a number of semi-empirical equations have been used effectively to assist in the design and the analysis of interlaboratory studies for water quality parameters. In general, these equations are simple and straightforward in their applications to environmental data, covering a wide variety of chemical and physical water quality variables.

The equations permit: (a) the calculation of the minimum number of laboratories that should participate in an interlaboratory quality control study; (b) the estimation of the between-laboratory variability in interlaboratory QC studies; and, (c) the determination of the bivariate relationship between paired data points in a Youden plot.

KEY WORDS: Semi-empirical equations; interlaboratory quality control; between-laboratory variability; within-laboratory variability; Youden plot; contour ellipse.

INTRODUCTION

In designing an 'Interlaboratory Quality Control Study', emphasis is placed on a number of factors inorder to ensure an effective program. Particular consideration is given to:

(a) the number of laboratories that will participate in the study;
(b) the number of samples that will be sent to each laboratory;
(c) the specific water quality variables that must be tested for;
(d) the number of repeat measurements that are required for each water quality variable in each sample;
(e) the evaluation of the analytical data; and,
(d) the overall cost of the study in both time and resources.

In the interlaboratory QC study, it is always desirable to obtain the maximum amount of information from the anlytical results. The greater the number of participating laboratories, the greater the chance of achieving the goal of maximum information. Thus it becomes imperative that a minimum limit be placed on the number of participants that will make the study useful and evaluative.

To evaluate interlaboratory performance, Youden[1] suggested that two similar samples (x and y) be sent to each participating laboratory, and the results of the measurements for one sample be plotted with respect to those obtained for the other. This approach would allow an evaluation of both systematic and random errors. On the basis of random error alone, the paired data values would be expected to be distributed in a circular pattern around the expected result[2]. The dimensions of the circle would depend on the probability level chosen. When bias is the predominant source of error, the data will be distributed along the 45° line and enclosed in an ellipse, the major axis of which is related to bias (systematic error) while the minor axis is related to precision of measurement (random error).

Estimating the number of laboratories participating in an interlaboratory QC study

There are no set rules regarding the optimum number of laboratories that is necessary for any given interlaboratory QC study. However, it is recommended[3] that the number of participating laboratories should never be less than 8, and if there is only one sample, the number of laboratories should preferably be higher, say 15 or more. The general recommended level of analytical replication per sample is 2.

While some coordinators of interlaboratory quality assurance programs prefer to run studies that involve a limited number of participating laboratories, other coordinators strive for the maximum participation by well-qualified laboratories. In some cases, 100 or more laboratories are asked to participate in order to generate the maximum level of comparable data.

In keeping with the views of Youden[4] and Parkany[3] concerning the level of replication (r) and the minimum number of laboratories (L) that should participate in interlaboratory QC studies, we have developed a number of equations that serve to give rough estimates of L. The structures and uses of the equations (1.1 to 1.4) are explained in the following comments:

$$L_1 = (r + v)^3/v; \qquad r = 1,2;\ v \leq 5 \tag{1.1}$$

where L_1 is the estimated number of participating laboratories required for the interlaboratory study, r is the level of replication and v is the number of water quality variables under investigation.

$$\begin{gathered} L_2 = (r_1 + v_1)^3/v_1 + (r_2 + v_2)^3/v_2 \\ r_1 = 1;\ r_2 = 2;\ v_1 + v_2 \leq 5 \end{gathered} \tag{1.2}$$

v_1 = # of variables requiring no replication (i.e. requiring just a single test); v_2 = # of variables requiring duplicate tests.

In Equation 1.2, variables for which duplicate tests are requested, are grouped together, and those for which only one test is requested, are also grouped together. In Equation 1.3, the symbols have the same meaning as in Equation 1.1.

$$L_3 = (r + v)^4/v^2; \qquad r = 1,2;\ v > 5 \tag{1.3}$$

Another equation which can be used for two different levels of replication, is Equation 1.4, in which:

$$\begin{gathered} L_4 = (r_1 + v_1)^4/v^2_1 + (r_2 + v_2)^4/v^2_2 \\ r_1 = 1;\ r_2 = 2;\ v_1 + v_2 > 5 \end{gathered} \tag{1.4}$$

Table 1 contains data and information for 5 interlaboratory quality control studies over an 18 month period. The calculations that follow indicate the minimum number of laboratories that should participate in each of the five studies.

The use of Equation 1.1 gives the following respective values for studies 1 and 2:

$$\text{Study 1: } L_1 = (1 + 1)^3/1 = 8$$

$$\text{Study 2: } L_2 = (2 + 1)^3/2 = 27/2 = 13.5 \approx 14$$

For Study 3, Equation 1.2 gives:

$$L_3 = (1 + 2)^3/2 + (2 + 3)^3/3 = 3^3/2 + 5^3/3 = 55.2 \approx 55$$

The application of Equation 1.3, provides the following result for study 4:

$$L_4 = (1 + 7)^4/7^2 = 8^4/7^2 = 83.59 \approx 84$$

Finally, the use of Equation 1.4 for study 5 gives:

$$L_5 = (1 + 9)^4/9^2 + (1 + 3)^4/3^2 = 10^4/9^2 + 4^4/3^2 = 151.96 \approx 152$$

For replications greater than 2, an additional set of equations have been proposed:

$$L_A = (r + v)^3/rv; \quad r > 2; v \leq 5$$

$$L_B = (r_1 + v_1)^3/r_1v_1 + (r_2 + v_2)^3/r_2v_2;$$
$$r_1 > 2; r_2 > 2; v_1 + v_2 \leq 5$$

$$L_C = (r + v)^4/(rv)^2; \quad r > 2; v > 5$$

$$L_D = (r_1 + v_1)^4/(r_1v_1)^2 + (r_2 + v_2)^4/(r_2v_2)^2;$$
$$r_1 > 2; r_2 > 2; v_1 + v_2 > 5$$

Analysis of Equations 1.1, 1.2, 1.3, 1.4, L_A, L_B, L_C and L_D by differential calculus, shows that these equations satisfy the requirements for the number of participating laboratories (L) to be a minimum[5].

Table 1 Data for determining the number of laboratories that should participate in specific interlaboratory QC studies.

Study	*# of variables*	*# of tests for each variable*	*Total # of tests*
(1)	1	Single test.	1
(2)	1	Duplicate tests.	2
(3)	5	(a) 2 variables, each with a single test.	2
		(b) 3 variables, each with diuplicate tests.	6
(4)	7	Single test for each variable.	7
(5)	12	(a) Single test for 9 variables.	9
		(b) Duplicate tests for 3 variables.	6

Estimation of the between-laboratory variability

At present, there is no completely valid method available to directly calculate the interlaboratory precision (between- laboratory variability) from the intralaboratory precision (within-laboratory variability)or vice-versa. This is unfortunate because frequently only one type of precision estimate is available for a method.

A semi-empirical formula that is useful in relating the between-laboratory variability (S_B) with the within-laboratory variability (S_W) is represented by the following relationship:

$$S^2_B = S^2_{\bar{X}} - S^2_W/N \tag{2.1}$$

$S_{\bar{X}}$ denotes the overall standard deviation (or the 'standard error' of the mean) and N is the number of replicate measurements conducted on a sample.

The data from 3 laboratories for a water quality sample containing arsenic are recorded in Table 2. From this data, the relationship between S_W and S_B can be exemplified.

Assuming that the three standard deviations are estimates of one and the same population standard deviation, it is proper to 'pool' the variances, and take the square root of the pooled variance[6]. Using this procedure, the best estimate of the within-laboratory standard deviation (S_W) is obtained as:

$$S_W = [\{(1.01)^2 + (1.22)^2 + (0.51)^2\}/3]^{1/2} = 0.961$$

If the standard error of the mean $S_{\bar{X}}$ is found for the three averages 67.45, 66.75 and 65.95, we obtain:

$$S_{\bar{X}} = [\{(67.45 - 66.72)^2 + (66.75 - 66.72)^2 + (65.95 - 66.72)^2\}/2]^{1/2} = 0.751$$

If the laboratories displayed no systematic differences,the standard deviation that has been calculated from the averages (of 4 replicates each) should be equal to $S_W/\sqrt{4}$, and $S_W/\sqrt{4} = 0.961/\sqrt{4} = 0.486$

Table 2 Interlaboratory data for a water sample containing arsenic.

Laboratory #	*Reported concentrations (ng/L)*	*Average concentration (ng/L)*	*Standard deviation (ng/L)*
(1)	66.1 68.5 67.4 67.8	67.45	1.01
(2)	68.4 66.1 66.9 65.6	66.75	1.22
(3)	65.2 66.1 66.3 66.2	65.95	0.51

The fact that the calculated value of $S_{\bar{X}}$ is larger than $S_W/\sqrt{4}$, can be explained only through the presence of an additional component of variability. This component which is the between-laboratory variability (S_B) is calculated by subtracting the anticipated S^2_W from the observed variance $S^2_{\bar{X}}$, and taking the square root. That is,

$$S_B = [(0.751)^2 - (0.486)^2]^{1/2} = 0.573$$

In general, if $S_{\bar{X}}$ and S_W are known, then S_B can be easily calculated using the formula in Equation 2.1.

Determination of the bivariate relationship between paired data points in a youden plot

Invariably the mathematical or statistical analysis of paired data (x,y) from interlaboratory QC studies, leads to a distribution of points that closely conforms to an elliptical contour[7-11]

The surface of a contour ellipse for paired (x,y) data, has a 2-dimensional distribution of the form:

$$p(x,y) = [2\pi\sigma_x\sigma_y(1-r^2)]^{-1}.\exp-\{1/[2(1-r^2)][L^2-2rLM + M^2] \qquad (3.1)$$

where: $L = (x - \bar{x})/\sigma_x$ and $M = (y - \bar{y})/\sigma_y$.

The probability ellipse corresponding to this surface is given as:

$$(1-r^2)^{-1}.[(x - \bar{x})^2/\sigma^2_x - 2r(x - \bar{x})(y - \bar{y})/\sigma_x\sigma_y + (y - \bar{y})^2/\sigma^2_y] = \chi^2 \qquad (3.2)$$

where $\chi^2 = 5.99$ for two degrees of freedom.

The analysis of the water quality data for arsenic (As) in Table 3, gives the following statistical parameters:

$$\bar{x} = 7.10\ \mu g/L;\ \bar{y} = 7.30\ \mu g/L;\ \sigma_x = 6.46\ \mu g/L;\ \sigma_y = 6.43\ \mu g/L; \qquad r = 0.9963$$

Incorporation of the values for $\bar{x}$, $\bar{y}$, σ_x, σ_y and r in Equation 3.2, gives the following bivariate relationship for the elliptical contour:

$$10x^2 + 10y^2 + 4x - 4y - 20xy - 18 = 0 \qquad (3.3)$$

Table 3 Interlaboratory QC data for a sample containing arsenic*.

Lab#	*x (mg.L^{-1})*	*y (mg.L^{-1})*	*Lab#*	*x (mg.L^{-1})*	*y (mg.L^{-1})*
(1)	0.004	0.005	(8)	0.006	0.007
(2)	0.008	0.008	(9)	0.005	0.005
(3)	0.005	0.005	(10)	0.003	0.003
(4)	0.006	0.005	(11)	0.006	0.007
(5)	0.007	0.007	(12)	0.030	0.030
(6)	0.006	0.006	(13)	0.004	0.004
(7)	0.005	0.005	(14)	0.005	0.005

* Design Value = 0.006 mg.L^{-1}

Inspection of the resulting bivariate relationship such as Equation 3.3, provides a diagnostic tool for determining the shape of the contour which envelops the distribution of the interlaboratory QC data points. For example, if the coefficients of x^2 and y^2 are the same and those of x and y are also equivalent, then the contour will be an ellipse. Table 4 gives a summary of the various contours that are obtained from the relationship between the coefficients of x^2 and y^2, and between x and y in Equation 3.4.

$$ax^2 + by^2 + cx + dy + exy + k = 0 \tag{3.4}$$

A special and unique contour not described in Table 4, occurs when the data points merge into a single point. This situation would be the most desireable outcome for any interlaboratory QC study, in which the results from all participating laboratories could be coincident (i.e. the x and y values would be equal for each and all the data pairs).

For the occurrence of this unique situation, x = y; $\sigma_x = \sigma_y$; and the correlation coefficient r will be equal to unity.

Equation 3.2 can be re-written as:

$$(x - \bar{x})^2/\sigma_x{}^2 - 2r(x - \bar{x})(y - \bar{y})/\sigma_x\sigma_y + (y - \bar{y})^2/\sigma_y{}^2 = (1 - r^2)\chi^2 \tag{3.5}$$

$$\text{or, } \Sigma[x - \bar{x})^2]\sigma_x{}^2 - \Sigma[2r(x - \bar{x})(y - \bar{y})]/\sigma_y{}^2 + \Sigma[(y - \bar{y})^2]/\sigma_y{}^2 = \Sigma(1 - r^2)\chi^2 \tag{3.6}$$

Dividing both sides of Equation 3.6 by n and taking into consideration that $\sigma_x = \sigma_y$, $\bar{x} = \bar{y}$, and x values are equivalent to y values, lead to:

$$\Sigma[x - \bar{x})^2]/n\sigma_x{}^2 - 2r\Sigma[(x - \bar{x})(y - \bar{y})]/n\sigma_x\sigma_y + \Sigma[y - \bar{y})^2]/n\sigma_y{}^2 = \Sigma(1 - r^2)\chi^2/n \tag{3.7}$$

$$\text{and, } \sigma_x{}^2/\sigma_x{}^2 - 2r\sigma_x{}^2/\sigma_x{}^2 + \sigma_x{}^2/\sigma_x{}^2 = n(1 - r^2)\chi^2/n \tag{3.8}$$

When $r = 1$, $1 - 2 + 1 = (0)\chi^2$; i.e. $0 = 0$

This is the condition for a single point to be the generated locus of a given set of points in 2-dimensional cartesian space[12].

Table 4 Geometric contours obtained from comparing the x^2, y^2, x and y coefficients in the bivariate relationship exemplified in equation 3.4.

Relationship between a and b in Equation 3.4	*Relationship between c and d in Equation 3.4*	*Status of the coefficients e and k*	*Resulting geometrical contour*
a = b	c = d or c ≠ d	e ≠ 0; k > 0	Elliptical contour
a ≠ b	c = d or c ≠ d	e ≠ 0; k > 0	Elliptical contour
a = b	c = d or c ≠ d	e = 0; k > 0; r (Correl. Coeff.) ≈ 0	Circular envelope
a = b	c = a; d = –a	e = –2a; k = 0	Straight line
a = b	c = d = a	e = 2a; k = 0	Straight line

Acknowledgements

This paper was presented at the 24th International Symposium on Environmental Analytical Chemistry, Carleton University, Ottawa, Ontario, Canada, May 16–19, 1994, and the author gratefully acknowledges the encouragement and support given to this effort by Mr. Ian McLaurin, Head, Surface Water Standards Section, Atmospheric Environment Service, Environment Canada.

References

1. W. J. Youden, *J. Qual. Technol.* **4**, 1–67 (1982).
2. J. K. Taylor, *Quality Assurance of Chemical Measurements* (Lewis Publishers, Inc., Chelsea, Michigan, U.S.A., 1987), 1st ed., 348pp.
3. M. Parkany in *Collaborative Studies in Chemical Analysis* (H. Egan and T. S. West, eds., Pergamon Press, New York, 1981), pp. 7–14.
4. W. J. Youden, *J. Assoc. Offic. Agire Chem.* **46**, 55–62 (1963).
5. J. E. Gaskin, *Principles and Guidelines for Interlaboratory Quality Control Studies,* (Environment Canada, Ottawa, Ontario, 1991).
6. J. Mandel and L. F. Nanni, in *Quality Assurance Practices for Health Laboratories* (S. L. Inhorn, ed., American Public Health Association, 1015 Eighteenth Street, Washington, D. C. 20036, 1978) pp. 229–231.
7. A. Hald, *Statistical Theory with Engineering Applications* (Wiley Publications in Statistics, New York, U.S.A., 1952), 1st ed., pp. 585–602.
8. W. J. Youden, *Statistical Techniques for Collaborative Tests* (A.O.A.C., Washington, D. C., 1973), 1st ed., 68pp.
9. J. Mandel and T. W. Lashof, *Quality Technology,* **6**, 22–36 (1974).
10. J. K. Taylor, *Quality Assurance for Chemical Measurements* (Center for Analytical Chemistry, National Bureau of Standards, Gaithersburg, MD 20899, U.S.A., 1984), 2nd ed., 80pp.
11. J. E. Gaskin, *Graphical Diagnosis of Interlaboratory Quality Control Data for Surface Water Samples'; Analyst,* **119**, pp. 1531–1535 (1994).
12. J. E. Draper and J. S. Klingman, *Mathematical Analysis–Business and Economic Applications* (Harper and Row, New York, N. Y. 1972), 2nd ed., pp. 72–92.