RELIABILITY OF CALIBRATIONS AND COMPARISONS OF REFERENCE STANDARDS WITH AN APPLICATION TO A GROUP OF LINE STANDARDS

By E. G. Thwaite*

[*Manuscript received April* 26, 1965]

Summary

The problem discussed is that of combining interrelated data which refer to the calibration and intercomparison of a group of standards, to give best estimates for the values of the standards together with their reliabilities. A résumé is given of least squares in matrix notation, which is particularly suited to computations using automatic computers, and this technique is applied to results obtained over a period of 15 years at the Bureau International des Poids et Mesures and the National Standards Laboratory, in the calibration and intercomparison of five 1 m reference line standards of length. One of these scales was a platinum-iridium prototype (No. 20), while the remainder were in nickel or 58% nickel-steel. The analysis yields time-dependent values, where appropriate, for the lengths of the scales together with their variances, and an estimate of the observational variance. The estimate of the observational variance, $(0\cdot077\ \mu\mathrm{m})^2$, is based in part on "external consistency", as the data used in the analysis were obtained in more than one laboratory.

I. Introduction

Where a number of reference standards have been in use for some time, and have been individually calibrated on a number of occasions and compared with each other from time to time, the questions naturally arise of the best values to be associated with the standards, and the reliabilities of these values. The answers to these questions are of particular interest when the observations come from more than one laboratory, since the results can then lead to reliable estimates of consistency between laboratories and to more realistic estimates of accuracy.

The interrelationships between the various observations can be very complex, particularly in the case where the standards are unstable. To arrive at the "best" values for the standards at any time, it is necessary to combine all the information available, while making allowance for particular circumstances such as instability, unequal reliability, or correlations between the observations.

The appropriate way of arriving at the "best" estimates in the situation described above would seem to be by a least-squares treatment. Further, and this is perhaps more relevant here, the sizes of the residuals from any least-squares solution, or statistics calculated from them, are an excellent guide to the reliability of the measurements.

The technique of least squares has been used for decades in adjusting and combining observations. However, solutions involving even small numbers of

* Division of Applied Physics, National Standards Laboratory, CSIRO, University Grounds, Chippendale, N.S.W.

unknowns require formidable computations and, while for small numbers it is reasonable to undertake solutions using a desk calculator, the use of an automatic computer considerably lightens the task and makes it possible readily to undertake solutions involving very large numbers of unknowns.

The size of problem that can be handled on a digital computer is limited by the memory capacity of the machine and by the accuracy required in the solution, together with the amount of machine time available. There are several ways in which limitations of memory capacity and accuracy can be overcome, and so the effective limitation is that of machine time.

For the purpose of computation on a digital computer, the formulation of the least-squares solutions in matrix notation is ideal. Not only does the matrix approach lead to a highly systematic and general set of relationships for the design and for the solutions, but also it gives the variances and covariances of the improved estimates with very little additional calculation once the solutions have been obtained.

In the following sections, a brief résumé of the matrix formulation of least squares is given, and an analysis is made of a comprehensive set of data relating to observations made over a period of 15 years, at the Bureau International des Poids et Mesures and the National Standards Laboratory, on a group of 1 m reference line standards.

The approach adopted here is a general one and suitable for application to any set of data that shows the type of dependencies already discussed. The data could refer to any one of a number of different types of standards, such as resistance, capacitance, mass, etc.

II. Least Squares in Matrix Notation

Let the results $y_1, y_2, \ldots, y_m$ of m observations be related to the quantities $x_1, x_2, \ldots, x_n$ by the expectations

$$\epsilon(y_i) = \sum_{j=1}^{n} a_{ij} x_j \qquad (i = 1, 2, \ldots, m),$$

where the a_{ij} are constants and form the design matrix $[a_{ij}]$, $\mathbf{A}$ say, of order $m \times n$.

Writing

$$\mathbf{x} = [x_1, x_2, \ldots, x_n]^T,$$

$$\mathbf{y} = [y_1, y_2, \ldots, y_m]^T,$$

both column matrices, we have in matrix notation

$$\epsilon(\mathbf{y}) = \mathbf{Ax}.$$

Making the restriction that the observed values y are independent and have equal variances σ^2, we may write the dispersion matrix in the form

$$\mathbf{D} = \sigma^2 \mathbf{I};$$

the least-squares solutions are then

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y},$$

with the variance–covariance matrix given by

$$\mathbf{D}_x = \sigma^2(\mathbf{A}^T\mathbf{A})^{-1}.$$

In general σ^2 is unknown, and its unbiased estimate s^2 is given by

$$s^2 = \frac{\mathbf{y}^T\mathbf{y}-\hat{\mathbf{x}}^T\mathbf{A}^T\mathbf{y}}{m-n}.$$

Rigorous treatments using matrix notation of least squares in the linear case will be found in the works of Plackett (1960) and Guest (1961), and these include the important case of dependent observations.

III. Least-Squares Analysis for a Group of Reference Line Standards

Since 1948 the National Standards Laboratory (N.S.L.) has maintained a number of 1 m line standards as reference standards of length. Until the introduction of the wavelength standard these line standards embodied the primary standard of length for the Commonwealth of Australia.

The information relating to the lengths of these scales is contained in a complex set of relationships between the lengths of the scales, which are the result of direct comparisons made from time to time at N.S.L., and also in their relationships with the prototype metre, determined through calibrations at the Bureau International des Poids et Mesures (B.I.P.M.) and on one occasion by the National Physical Laboratory (N.P.L.). The complexity of the information lies in the fact that the scales are of unequal age, and the periods they have been under surveillance vary. Further, the majority of the scales are known to be unstable in length, and this adds greatly to the difficulty of determining the appropriate value for their length at any time.

At the beginning of the period, the set consisted of a pair of 58% nickel-steel scales, 4028 and 4756(I), both calibrated by B.I.P.M. The original set was augmented, firstly by a nickel scale 79, then by another 58% nickel-steel scale 172/62, and finally by the reruled platinum-iridium prototype No. 20. The nickel scale and, of course, the prototype, were initially calibrated by B.I.P.M. Owing to corrosion damage, one of the original pair of nickel-steel scales required to be reruled, and this resulted in an unfortunate discontinuity in the knowledge of its behaviour. The reruled scale is designated 4756(II).

Details of the calibration and intercomparison of the set of scales are given in Table 1, together with additional information which will be discussed later.

We now proceed to obtain the solutions for the data given in Table 1.

The error in the length of a scale is assumed to be given by an equation of the form

$$L_i(d) = a_i+b_i(d-1948),$$

where L_i is the error at 20°C in the length of the ith scale at date d (expressed in years and decimals of a year), and the a_i and b_i are constants of the expression that are to be estimated by the least-squares solution. The lengths of all the scales are assumed to be adequately represented by this form of expression, with the exception of prototype No. 20 which is assumed to be stable and so lacks the time-dependent term. It is known that the nickel-steel scales are unstable, and it would be unwise to assume stability for the nickel scale. In the absence of information other than

TABLE 1

OBSERVED AND CALCULATED VALUES* (ELEMENTS OF **y**)

Observational standard deviation 0·077 μm

Scale	Authority	Date	Observed Value (μm)	Calculated Value (μm)	Residual (μm)
4028	B.I.P.M.	1948·96	−2·19	−2·253	−0·063
4028	B.I.P.M.	1957·04	−3·02	−2·988	0·032
4756(I)	B.I.P.M.	1952·71	−0·67	−0·635	0·035
4756(II)	B.I.P.M.	1961·96	−0·13	−0·130	0·000
79	B.I.P.M.	1957·25	2·11	2·115	0·005
79	N.P.L.	1957·54	2·06	2·108	0·048
20	B.I.P.M.	—†	1·31	1·253	−0·057
4028–4756(I)	N.S.L.	1957·21	1·94	1·956	0·016
4028–4756(I)	N.S.L.	1955·87	2·05	1·957	−0·093
4028–4756(I)	N.S.L.	1957·96	1·93	1·956	0·026
4028–79	N.S.L.	1957·96	−5·18	−5·170	0·010
79–4756(I)	N.S.L.	1957·96	3·23	3·214	−0·016
4028–4756(II)	N.S.L.	1962·37	−3·31	−3·261	0·049
4028–172/62	N.S.L.	1962·37	−5·62	−5·566	0·054
4028–79	N.S.L.	1962·37	−5·55	−5·468	0·082
4756(II)–172/62	N.S.L.	1962·37	−2·26	−2·305	−0·045
4756(II)–79	N.S.L.	1962·37	−2·30	−2·206	0·094
172/62–79	N.S.L.	1962·37	0·09	0·098	0·008
172/62–79	N.S.L.	1961·62	0·24	0·206	−0·034
172/62–4028	N.S.L.	1961·62	5·59	5·624	0·034
20–79	N.S.L.	1963·79	−0·64	−0·708	−0·068
20–4028	N.S.L.	1963·79	4·73	4·855	0·125
79–4028	N.S.L.	1963·79	5·51	5·564	0·054

* Referred to 20°C.

† See text.

the data alone about the degree of the curve to be fitted, the justification of a model is its success in fitting the data points, the model of least degree that adequately fits the data being the one chosen. While in the present case the data are hardly adequate to fit anything but a linear model, as will be seen later, the fit is good. Under these conditions extrapolation will be as good as is possible.

Prototype No. 20 is of the same melt as the original prototypes, and it can be reasonably assumed that, in the very short period between the B.I.P.M. calibration (June–September 1960 and January 1963) and its inclusion in the N.S.L. series (1963·79), any drift will be negligible.

From the data of Table 1, the design matrix is the 23 row by 11 column matrix

$$\mathbf{A} = \begin{bmatrix}
1 & 0\cdot 96 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 9\cdot 04 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 4\cdot 71 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 13\cdot 96 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 9\cdot 25 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 9\cdot 54 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -9\cdot 21 & 1 & 9\cdot 21 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -7\cdot 87 & 1 & 7\cdot 87 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -9\cdot 96 & 1 & 9\cdot 96 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 9\cdot 96 & 0 & 0 & 0 & 0 & -1 & -9\cdot 96 & 0 & 0 & 0 \\
0 & 0 & -1 & -9\cdot 96 & 0 & 0 & 1 & 9\cdot 96 & 0 & 0 & 0 \\
1 & 14\cdot 37 & 0 & 0 & -1 & -14\cdot 37 & 0 & 0 & 0 & 0 & 0 \\
1 & 14\cdot 37 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -14\cdot 37 & 0 \\
1 & 14\cdot 37 & 0 & 0 & 0 & 0 & -1 & -14\cdot 37 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 14\cdot 37 & 0 & 0 & -1 & -14\cdot 37 & 0 \\
0 & 0 & 0 & 0 & 1 & 14\cdot 37 & -1 & -14\cdot 37 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & -14\cdot 37 & 1 & 14\cdot 37 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & -13\cdot 62 & 1 & 13\cdot 62 & 0 \\
-1 & -13\cdot 62 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 13\cdot 62 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & -15\cdot 79 & 0 & 0 & 1 \\
-1 & -15\cdot 79 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -15\cdot 79 & 0 & 0 & 0 & 0 & 1 & 15\cdot 79 & 0 & 0 & 0
\end{bmatrix}$$

The column vector $\mathbf{x}$ is given by

$$\mathbf{x} = [a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_5, b_5, a_6]^T,$$

where the correspondence between the scales and the constants is as given in Table 2.

TABLE 2

ESTIMATES OF COEFFICIENTS AT 20°C

Scale	i	a_i (μm)	b_i (μm m^{-1} yr^{-1})
4028	1	$-2\cdot 166$	$-0\cdot 0910$
4756(I)	2	$-0\cdot 204$	$-0\cdot 0915$
4756(II)	3	$2\cdot 649$	$-0\cdot 1991$
79	4	$2\cdot 333$	$-0\cdot 0235$
172/62	5	$4\cdot 497$	$-0\cdot 1672$
20	6	$1\cdot 253$	—

The observational vector is

$$\mathbf{y} = [-2\cdot 19, -3\cdot 02, -0\cdot 67, \ldots, 5\cdot 51]^T.$$

It is assumed that the elements of y have equal weights, that is, that they are independent and have equal variances. While the data have been obtained over a considerable length of time and at different laboratories, techniques applied in the comparison of line standards differ little, and should have remained sufficiently stable for the assumption of equal variances to be reasonably made. The results of the solution can be used to test this assumption, but unfortunately in this case the test is necessarily inexact. The solutions are then given by the last three equations of Section II.

The whole of the least-squares solution was carried out on a digital computer. The core of the solution is the inversion of the matrix $\mathbf{A}^T\mathbf{A}$, and this was undertaken by the "elimination" method (see, for example, Kunz 1957).

On post-multiplication by $\mathbf{A}^T\mathbf{y}$ the inverse yields the estimates $\hat{\mathbf{x}}$, the components of which are listed in Table 2.

The estimate of the observational variance s^2 was $0 \cdot 005917$ $\mu\mathrm{m}^2$, giving a standard deviation of $0 \cdot 077$ μm, the estimate being based on 12 degrees of freedom.

Multiplication of the inverse by s^2 then gives the dispersion matrix $\mathbf{D}_{\hat{x}}$, the diagonal elements corresponding to the variances and the off-diagonal elements to the covariances.

$$
\begin{array}{ccccccccccc}
\multicolumn{2}{c}{\overbrace{\qquad\qquad\qquad}^{4028}} & \multicolumn{2}{c}{\overbrace{\qquad\qquad\qquad}^{4756(\mathrm{I})}} & \multicolumn{2}{c}{\overbrace{\qquad\qquad\qquad}^{4756(\mathrm{II})}} & \multicolumn{2}{c}{\overbrace{\qquad\qquad\qquad}^{79}} & \multicolumn{2}{c}{\overbrace{\qquad\qquad\qquad}^{172/62}} & 20 \\
a_1 & b_1 & a_2 & b_2 & a_3 & b_3 & a_4 & b_4 & a_5 & b_5 & a_6
\end{array}
$$

$$
\mathbf{D}_{\hat{x}} = (10^{-7}\mu\mathrm{m}^2)
\begin{bmatrix}
56608 & & & & & & & & & & \\
-5123 & 642 & & & & & & & & & \\
1290 & -1184 & 233872 & & & & & & & & \\
638 & 208 & -27592 & 3551 & & & & & & & \\
548924 & -131862 & 524589 & -119890 & 14492452^{*} & & & & & & \\
-39321 & 9446 & -37578 & 8588 & -10232902 & 722681 & & & & & \\
40171 & -5676 & -6692 & 742 & 1639543 & -117446 & 223235 & & & & \\
-3856 & 649 & -585 & 186 & -210687 & 15092 & -19357 & 1828 & & & \\
48389 & -5400 & -2700 & 689 & 5667879 & -406008 & 131698 & -11606 & 18268155 & & \\
-4489 & 645 & -884 & 197 & -507075 & 36323 & -12516 & 1238 & -1301547 & 92983 & \\
-15000 & 3193 & -11110 & 2534 & -1073465 & 76896 & -43956 & 5298 & -29475 & 4245 & 44764
\end{bmatrix}
$$

* The accuracy of the solution is not sufficient to give the value of this last place.

As the matrix $\mathbf{D}_{\hat{x}}$ is symmetric about the diagonal, only the lower triangular matrix has been reproduced here.

Substitution of the estimates $\hat{\mathbf{x}}$ in the expressions $\mathbf{Ax}$ gives estimates of the observed values, and these, together with the residuals from the observed values, are given in Table 1.

IV. Discussion and Conclusion

The variances to be associated with the estimates from the regression lines are, of course, time-dependent and are given by the propagation of variance formula

$$\mathrm{Var}(L_i) = \mathrm{Var}(a_i) + (d-1948)^2 \,.\, \mathrm{Var}(b_i) + 2(d-1948) \,.\, \mathrm{Cov}(a_i, b_i).$$

Substitution then leads to the following:

$$\begin{aligned}
\mathrm{Var}(4028) &= 0{\cdot}005661+0{\cdot}0000642(d-1948)^2-0{\cdot}0010246(d-1948),\\
\mathrm{Var}(4756\ \mathrm{I}) &= 0{\cdot}023387+0{\cdot}0003551(d-1948)^2-0{\cdot}0055184(d-1948),\\
\mathrm{Var}(4756\ \mathrm{II}) &= 14{\cdot}492452+0{\cdot}0722681(d-1948)^2-2{\cdot}0465804(d-1948),\\
\mathrm{Var}(79) &= 0{\cdot}022324+0{\cdot}0001828(d-1948)^2-0{\cdot}0038714(d-1948),\\
\mathrm{Var}(172/62) &= 1{\cdot}826816+0{\cdot}0092983(d-1948)^2-0{\cdot}2603094(d-1948),
\end{aligned}$$

where the unit is μm^2.

On first inspection of the time-dependent variance equations, the impression may be gained that the variances, in some cases, are very large indeed. Since a common time, the year 1948, has been chosen as the origin for all the regression equations, the constant terms in the right-hand side of the variance equations will be large in those cases where the first observations occur at times much later than 1948. In these cases, however, the covariance term has a substantial negative value during the observational period, and the variances over these periods are much smaller than the values at the origin.

The expressions for the time-dependences of the lengths have a varying degree of usefulness that depends primarily upon the length of time that the scale has been under observation. The values of the variances for 4756(II) and 172/62 show that these two scales have been under observation for much too short a period for any reliability to be placed upon the predictions from their regression lines.

The results of the adjustments do not answer unequivocably the question of the stability of the nickel scale 79. A test of the slope of this regression line against the hypothesis of zero slope showed it to be almost significant at the 10% level. The normal interpretation of this result would be that there is insufficient evidence to say that the scale is unstable; the effect on the model, however, is small.

The sizes of the individual residuals are all satisfactorily small. The estimate of the observational standard deviation, $0{\cdot}077\ \mu m$, is to be compared with the value $0{\cdot}1\ \mu m$ which is usually spoken of as the ultimate accuracy with which line standards can be compared using visual microscopes. This estimate of the observational standard deviation is of some importance as it is based, in the main, on data from two laboratories, that is, it is an estimate based in part on "external consistency" rather than purely "internal consistency". Birge (1957) and Youden (1962) have discussed at some length the importance of estimates based on external, rather than internal, consistency.

The form of the design matrix in this case does not lend itself to a rigorous test of the assumption of equal variance for the results of the different laboratories. However, estimation of the variances for B.I.P.M. and N.S.L. separately from the residuals, using degrees of freedom based on a partitioning of the total number of degrees of freedom in the ratio 7 : 16, leads to an F ratio of $2{\cdot}16$. To be significant at the 5% level this ratio is required to be of the order 6, while to be significant at the 1% level it would have to reach 15. It is certain then that, whatever the actual degrees of freedom for the two estimates, the F ratio would not reach a significant value, and it must be concluded that the results show no evidence to contradict the assumption of equal variance for the observed values.

The result of the present adjustment indicates that it would be practicable and indeed desirable to undertake an adjustment on a much wider basis. The inclusion of the prototype metres in a least-squares adjustment would appear to be valuable at present, when the relationship between lengths based on the prototypes and on the wavelength definition are under discussion.

V. Acknowledgments

The least-squares solution was carried out on the I.B.M. 7090 computer, at Weapons Research Establishment, Salisbury, S. Aust. The author wishes to thank Mr. M. J. Puttock and Dr. R. T. Leslie for their comments on the manuscript.

VI. References

Birge, R. T. (1957).—*Nuovo Cim.* (10) **6** (Suppt.): 39–67.

Guest, P. G. (1961).—"Numerical Methods of Curve Fitting." (Cambridge Univ. Press.)

Kunz, K. S. (1957).—"Numerical Analysis." pp. 234–5. (McGraw-Hill: New York.)

Plackett, R. L. (1960).—"Regression Analysis." Ch. III. (Oxford Univ. Press.)

Youden, W. J. (1962).—N.B.S. Miscellaneous Publ. 248. (Pap. 5.1, Proc. 1962 Stds Lab. Conf.) (U.S. Dept. of Commerce.)