PREPRINT INSTRUMENT SOCIETY of AMERICA 19th Annual ISA Conference and Exhibit • October 12-15, 1964, New York
Preprint Number 12.3-2-64

UNCERTAINTIES ASSOCIATED WITH PROVING RING CALIBRATION

by Thomas E. Hockersmith
Mechanical Engineer
Harry H. Ku
Mathematical Statistician
National Bureau of Standards
Washington, D. C.

ABSTRACT

A method of error analysis is presented using data obtained from dead-weight calibration of various capacity proving rings. A breakdown of the errors into components by statistical methods and their combination into a final uncertainty statement is discussed in detail. Graphical representations are used in several places to help in the exposition.

Extension of the analysis and method of handling calibration data for multiple proving ring setups is discussed in an effort to show that the same general method of analysis should be adequate.

INTRODUCTION

A proving ring is a compact and dependable force measurement device developed at the National Bureau of Standards by H. L. Whittemore and S. N. Petrenko for the original purpose of calibrating testing machines. A typical proving ring is shown in Figure 1. It consists basically of the following components; an elastic steel ring with diametrically opposed integral loading bosses, a vibrating reed, and a micrometer dial and screw assembly. The reed and micrometer screw assembly are mounted along the diameter concentric with the bosses. When a load is applied to the ring a deflection is measured by turning the micrometer screw until positive contact is made with the vibrating reed. This deflection value is read in terms of the arbitrary scale inscribed in the face of the micrometer dial. For details on the design, use, and calibration of proving rings, see Circular of the National Bureau of Standards C 454 [1]*.

In recent years a significant increase in the use of the proving ring as a secondary transfer standard, in the field of force measurement, has precipitated the need for information dealing with accuracy of the calibration process. The purpose of this paper is to discuss methods of extracting such information from the calibration data and to present the results in a useable form.

UNCERTAINTY

Calibration may be generally thought of as the process of comparing an unknown with a standard and determining the value of the unknown from the accepted value of the standard. The accuracy of the reported values are usually given in terms of bounds to inaccuracy, or limits of uncertainty.

In any calibration process there are three possibilities available in dealing with the uncertainties. These are:

1. Report only the values obtained and make no statement about their uncertainty.
2. Make some statement of the uncertainties affecting the calibration process based on personal judgement and general experience.
3. Through the use of error analysis form an objective estimate of the uncertainties affecting the reported values.

The uncertainty of a measurement process may be characterized by giving (1) the imprecision, and (2) limits to the overall systematic error. Imprecision means the degree of mutual disagreement, characteristic of independent measurements of a single quantity, yielded by repeated applications of the process under specified conditions. The accepted unit for the imprecision of a calibration process is the standard deviation, σ, which provides a measure of how close a particular calibration result in hand is likely to agree with the results that might have been (or might be) obtained by the same calibration process in this (or other) instance(s). The larger the value of σ, the more imprecise the method of measurement, and the greater the disagreement to be anticipated between strictly comparable calibrations.

The systematic error of a calibration process refers to the more or less consistent deviations of the values observed, from the standard, or from the value intended to be measured. If the direction and the magnitude of systematic error were known with sufficient accuracy, a correction could be applied to render the reported values free from bias. Usually only limits of systematic error can be given, e.g., resulting from the uncertainty in the deter-

*The numbers in brackets refer to similarly-numbered references at the end of this paper.

mination of the mass of weights in a dead-weight load calibrating machine. Limits of systematic error are generally based on knowledge and experience with similiar measurements, information available from special studies, and judgement. In calibration the sources of systematic errors are usually studied carefully, and their effect on the final results minimized or eliminated if possible.

The total uncertainty of a calibration process places limits on its probable inaccuracy. It includes both the imprecision and the systematic error. Accuracy requires precision but precision does not necessarily imply accuracy. For example, a calibration process may be highly precise and yet when applied to a standard yield values consistently greater, or consistently less, than the accepted value of the standard.

The present method of reporting proving ring calibration employed by the NBS does not give explicitly a single expression of the overall uncertainty involved, but instead, gives estimates of the imprecision and systematic error from which the total uncertainty can be derived. This practice is in keeping with the recommendations on "Expression on the Uncertainties of Final Results" in Chapter 23 of NBS Handbook 91 [2]. The estimate of imprecision of the calibration process is given by the standard errors of the tabulated load values, which measure the combined performance of the calibration process and the particular ring. Bounds for the systematic errors are given in percent error of applied load for both dead-weight loads, and loads measured by means of a multiple ring setup.

DEAD-WEIGHT CALIBRATION

A dead-weight calibration of a proving ring consists of ten nearly equally spaced loads applied in either the 10,100-lb or the 111,000-lb capacity testing machines presently in use at the National Bureau of Standards.

Three runs of ten loads are taken on each ring to make up a calibration. Before and after each load reading a no-load reading is taken and recorded. The average of the two no-load readings is subtracted from the load reading to yield a deflection value of the ring under that load. This yields a total of thirty deflection values, three values for each load point from ten-percent of capacity to capacity load. These thirty deflections are punched on computer data cards with their corresponding load values and are fed into an electronic digital computer. A second degree equation of the form

$$D = a + bL + c(L)^2$$

is fitted to the averages of the three deflection values for each load, where

D = average deflection value
L = load in pounds

and a, b, c, are coefficients. The computer program performs the task of statistically analyzing the data, fitting the data by the method of least squares, and printing out a load versus deflection table as well as the various statistical quantities included in the report. The thirty deflection values obtained during the calibration of a 100,000-lb capacity proving ring are given in Table 1. A sample of the load versus deflection table printed out as a result of the computer fit of these data is found in Table 2.

The selection of a second degree equation in terms of load was decided upon as a result of preliminary investigation, both theoretical and experimental, to determine the proper degree of the calibration curve to represent the characteristics of the proving ring as evidenced by the raw data. At the same time it was necessary to keep in mind the many problems associated with applying an error analysis to such data.

Figure 2 shows the three deflection values at each of the ten load points for proving ring A, with most of the linear trend removed from the deflection values. The smooth curve represents the plot of the computed deflection values derived from the second degree fit with the same linear trend removed. This figure shows how well the second degree curve fits the observed deflections.

Several interesting and useful comparisons resulting from the error analysis and fitting techniques employed are as follows. From the dispersion of the three deflection values at each load point about their average, the standard deviation of a deflection value can be computed with two degrees of freedom. Since these standard deviations computed over the range of loads are comparable in magnitude, the ten values may be pooled together. This pooled value of the standard deviation, denoted as s_w, can be compared to its long run average value over many previous calibrations to determine if the calibration process is under control, i.e., stable with respect to precision.

A standard deviation s associated with the calibration of this particular ring can be computed from the residuals of the ten average deflection values about the second degree curve. This value of the standard deviation, s, can be compared with the pooled standard deviation of an average deflection value, $s_w/\sqrt{3}$, obtained from the ten sets of triplicate deflection values (i.e., the pooled estimate of the standard deviation of an individual deflection divided by $\sqrt{3}$). If the two standard deviations s and $s_w/\sqrt{3}$ are of nearly the same magnitude then the ring is in good condition and the scatter of the points is due mainly to the inability of the calibration process as a whole to repeat. Conversely, if the standard deviation s computed from the deviations of the ten average deflection values from the curve is considerably larger than $s_w/\sqrt{3}$, the estimated standard deviation of an average deflection value, then the condition of the ring is not good and reconditioning by the manufacturer is indicated. An example of this can be seen in figure 2. This ring is apparently not in good condition since the broken curve connecting the averages of the three deflection values at each load point does not follow the fitted curve closely. For rings in good condition, the two curves are practically indistinguishable on a graph to this scale. In the future this type of reasoning may be used as a basis for acceptance or rejection of a particular device.

Previously a calibration graph was included with the calibration certificate as shown in figure 3. This graph was a plot of the calibration factor for the ring in pounds per division versus the deflection in divisions. The straight line through the points was drawn for "best fit". Because the calibration factors were computed by dividing each deflection into its corresponding load, the points of the plot near the lower end of the load range, of the device, show considerably more dispersion than the points near the upper end. Therefore the upper points were considered to be better indicators for the drawing of the "best fit" line through the plotted points. In the case of the second degree fit of deflection versus load by the method of least squares, the individual points are treated with equal weight, a more accurate fit of the calibration data is obtained, and no possibility of personal bias is introduced.

The above can be illustrated as the by-products of a simple test designed and suggested by W. J. Youden of NBS. This test consisted of several operators taking readings with a proving ring under various known dead-weight loads. These loads were then computed as if they were unknown using first the table of load values from the second degree fit and second the load values derived from the "best fit" curve. Comparison showed that over the range of the ring, the load values computed from the second degree fit were closer to the actual known loads applied to the ring. Therefore, if the ring is to be used over its entire calibrated range the second degree fit gives more accurate load values. The same data were also used to check the computed limits of uncertainty for the particular ring and in no case did the difference between the actual and computed load value exceed these limits. A sample of the values determined during this experiment can be found in Table 3.

In order to arrive at some measure of dependability of the values given in the load table the corresponding confidence interval is needed. To determine such an interval, the standard error of a deflection value for a given load is computed and some multiple of this value is used as limits of uncertainty on the imprecision.

To predict a deflection value D_i for a particular given load L_i, the deflection value can be expressed as $D_i = a + bL_i + cL_i^2$. Thus D_i is a linear combination of the coefficients estimated, and its standard error s_i can be expressed in terms of the standard deviation s (estimated from the residuals of the fit, with seven degrees of freedom), and the load L_i, and the variance-covariance matrix $[C_{ij}]$ of the estimated coefficients a, b, and c, as follows:

$$s_i^2 = \underline{L}'[C_{ij}]\,\underline{L}\,s^2$$

where the vector $\underline{L}' = (1, L_i, L_i^2)$.
(For details of the method of polynomial fitting used and the calculation of the standard error s_i the reader is referred to sections 6-3 and 6-5 of Chapter 6 of NBS Handbook 91 "Experimental Statistics" [2].)

The dependence of s_i on the value of L_i indicates that D_i values corresponding to L_i values at the two ends of the range of L have larger prediction errors than do D_i values corresponding to L_i in the center portion. For convenience, the largest value of the standard error s_i computed from the above expression is used for all values of L_i in a proving ring report, and for ten equal increments of equally spaced loads L_i, the value of the largest standard error, s_i, is approximately equal to 0.79s. This is converted into load in pounds by multiplying 0.79s by the maximum calibration factor, in pounds per division, for the particular ring.

Using the t statistic and the computed standard error a confidence interval for the deflection value on the curve for a single given load can be calculated. In calibration work, however, we require not merely the calculation of a confidence interval for the deflection value corresponding to a single load, but the calculation of a confidence band for the whole calibration curve. Therefore, a wider interval will be required for the same level of confidence. The confidence band for a line as a whole is discussed on pages 5-15 to 5-17 of reference [2] and for entire curves, in Chapter 28 of [3], where it is shown that in the general case, the half width at $L = L_i$ of the confidence band for the curve as a whole is:

$$\sqrt{k\,F_{.05}(k,\nu)}\ \times\ s_i\ ,$$

where k is the number of coefficients estimated, ν is the number of degrees of freedom in estimating s_i, and F is an appropriate upper percentage point of the distribution of the F statistic (as an illustration we are using the upper 5%

point). Thus for $k = 3$, $\nu = 7$, and ten equal increments of loads the half width of the 95% confidence interval is $\sqrt{3 \times 4.35} \times 0.79s = 2.86s$. (Since the largest value of the computed standard error is used, the confidence level is at least 95%.) Therefore the over-all limits of uncertainty for the calibration by this procedure could be expressed as 2.86 x s, (s is the standard deviation given in the report), plus the systematic error.

It may appear that the above procedure for determining the limits of uncertainty in the calibration of a proving ring by basing it on the prediction of a deflection value for a given load is a reverse procedure. However, for the method of calibration described this seemingly reverse procedure is the proper one. Figure 4 is a schematic diagram of the deflection - load curve obtained from a calibration with the corresponding confidence band sketched about it. For any given load, the true deflection value is expected to be situated within the band. Conversely, if a deflection value d is given, a horizontal line parallel to the load axis will intercept the curve at the corresponding load value $\hat{L}$; in addition this line will also intercept the band at two points L_l and L_u which give the corresponding lower and upper confidence limits for the load. This is true provided that the deflection value is known without error. If the uncertainty of the deflection value can be represented by D_l and D_u, then the corresponding confidence interval for the load will be wider, as given by L_l' and L_u'. In other words, the accuracy with which the deflection readings are obtained in using the ring must be taken into account by the user of the ring.

Each load value given in the table of load versus deflection is therefore the predicted value of the load, given a deflection value, and is expected to be within the uncertainty limits given for the calibration.

CALIBRATION OF RINGS USING MULTIPLE RING SETUPS

The present practice for calibrating proving rings with nominal capacities in excess of 110,000 lb is as follows:

1. to divide the nominal capacity into ten approximately equal increments,
2. to calibrate the ring by dead weights for the increments of load less than 110,000 lb, and
3. to calibrate the ring by either a 3, 4, or 5 proving ring setup for the remaining increments of load.

For a calibration using this procedure there are a number of problems relating to the analysis of data and interpretation of results. Some of these problems cannot be solved without considerable changes in the procedure of calibration. Since such changes are impractical, and in the near future dead-weight calibration capacity will be extended to 1,000,000 lb, one solution is to fit multiple ring calibrations by the same method as that for the dead-weight calibrations. The following discussion is based on the results of calibrations of rings fitted by this method.

Examination of these results showed no evidence of bias in the sense that residuals of the fit at the two adjoining increments of load, i.e. the last dead-weight and the first multiple ring load, are not unusually large or consistently of opposite sign. For this to remain true it is necessary that the calibrations of the rings used to determine the load in a multiple ring setup be unbiased. To insure that this condition is maintained the rings owned by the Bureau are usually reconditioned yearly and are calibrated frequently.

For dead-weight calibration the errors of the applied loads were assumed to be negligible in fitting the data; for multiple ring calibration, errors are introduced in the determination of the loads applied. Thus a non-linear functional relationship is to be estimated between deflection and load where the measurements of both are subject to error. There is no simple solution to this problem except that experience in this laboratory has shown that the least square fitting procedure still gives satisfactory estimates provided the errors are small compared with the range covered. This requirement is satisfied since each increment of load is more than 700 times the magnitude of the error involved.

Considering the above, and from a study of numerous past calibrations, it was decided the deflection should be fitted as a function of load since the former is believed to have larger errors than the latter.

Since the dead-weight calibration is presumably more precise than the multiple ring calibration, the question of weighting the observations prior to least square fitting was considered. Results of the rings studied indicated that the standard deviation of an average deflection obtained from multiple ring calibrations was not significantly larger than that for the dead weights. Thus the inflation of this deviation due to the errors in the loads does not increase the total imprecision by any appreciable amount. The use of weighting factors is therefore not of practical importance.

Examination of the plot of residuals resulting from fitting deflections to the loads, both in dead weights and in multiple ring setups, indicates that the deviations of the data points from the fitted curve contribute a large part of the total error. In view of this it appears reasonable that the averages of the deflection readings should be used for fitting, similar to the procedure used for dead-weight calibration. Thus, the standard error includes the imprecision components of the calibration error for both the ring being calibrated and the rings being used to mea.ure the applied load.

Bounds for systematic error of a multiple ring setup can be estimated by summing (1) the systematic error due to the dead weights (2) the systematic difference due to change with time in the calibrated values of the load measuring devices, and (3) other sources of error due to the inherent difficulties in using and reading the devices simultaneously. For example, such an estimate can be given an percent error of applied load for loads in excess of the dead weights.

CONCLUSION

In the above we have presented a procedure for the determination of limits of uncertainty for the calibration of proving rings. The method of analysis includes; the fitting of this type of data to an appropriate curve by the method of least squares, the use of confidence intervals and bands as limits of imprecision, and the estimate of bounds for systematic error.

Since many types of devices and instruments are calibrated similiarly at selected points along their ranges, it is believed that the procedures outlined above may be useful, when properly modified, in yielding a realistic evaluation of the uncertainties associated with their calibration.

ACKNOWLEDGMENT

Acknowledgment is made to Churchill Eisenhart of the National Bureau of Standards for the many helpful discussions and suggestions.

REFERENCES

(1) B. L. Wilson, D. R. Tate, and G. Borkowski, "Proving Rings for Calibrating Testing Machines", NBS Circular C 454, U. S. Government Printing Office, Washington 25, D. C.

(2) M. G. Natrella, "Experimental Statistics," NBS Handbook 91, U. S. Government Printing Office, Washington 25, D. C.

(3) M. G. Kendall and Alan Stuart, Advanced Theories of Statistics, Hafner Publishing Co., New York, 1961.

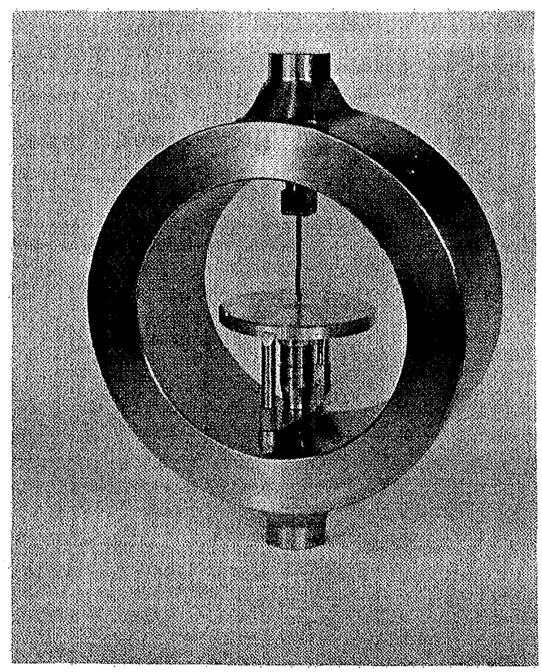

Figure 1 Proving Ring

Table 1 A Calibration of Proving Ring A

Applied load lb	Deflection Run 1 div	Deflection Run 2 div	Deflection Run 3 div
10,000	68.32	68.35	68.30
20,000	136.78	136.68	136.80
30,000	204.98	205.02	204.98
40,000	273.85	273.85	273.80
50,000	342.70	342.63	342.63
60,000	411.30	411.35	411.28
70,000	480.65	480.60	480.63
80,000	549.85	549.85	549.83
90,000	619.00	619.02	619.10
100,000	688.70	688.62	688.58

Table 2 - Computed Load Table in lb for 70 Degrees F for Proving Ring A

Deflection Div	0	1	2	3	4	5	6	7	8	9
60.	-	-	-	-	-	-	-	-	9952.	10099.
70.	10245.	10392.	10538.	10685.	10831.	10978.	11124.	11270.	11417.	11564.
80.	11710.	11856.	12003.	12149.	12295.	12442.	12588.	12735.	12881.	13027.
90.	13174.	13320.	13467.	13613.	13759.	13906.	14052.	14199.	14345.	14491.
100.	14638.	14784.	14930.	15077.	15223.	15369.	15516.	15662.	15808.	15954.
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
640.	93007.	93151.	93295.	93439.	93582.	93727.	93871.	94014.	94158.	94302.
650.	94446.	94590.	94734.	94878.	95021.	95165.	95309.	95453.	95597.	95741.
660.	95885.	96029.	96173.	96316.	96460.	96604.	96748.	96892.	97035.	97179.
670.	97323.	97467.	97611.	97754.	97898.	98042.	98186.	98330.	98473.	98617.
680.	98761.	98905.	99048.	99192.	99336.	99480.	99623.	99767.	99911.	100054.

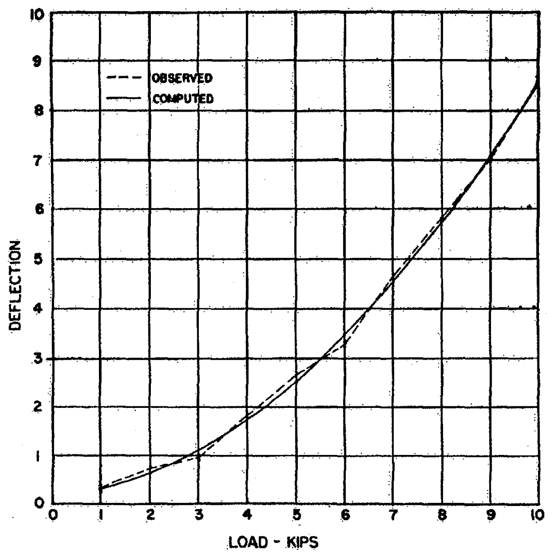

Figure 2 Observed and Fitted Calibration Curves for Proving Ring A

Note: Deflection = Deflection value minus 68.00 x (the number of the load increment)

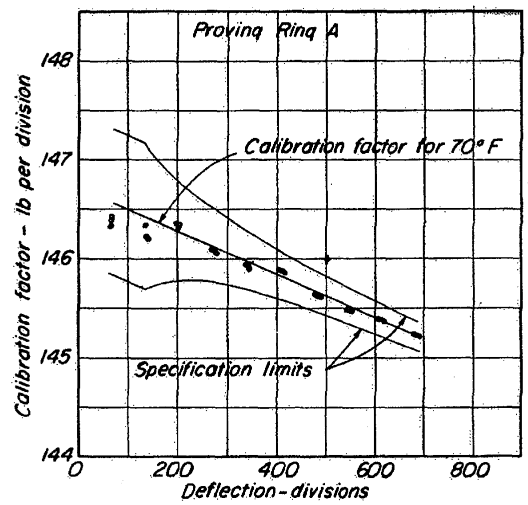

Figure 3 Calibration Graph for Proving Ring A

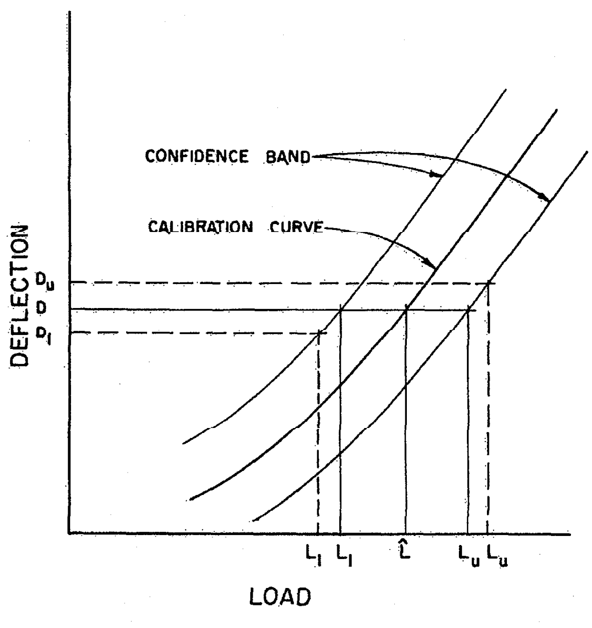

Figure 4 Determination of Confidence Limits for a Load given a Deflection Value

Table 3 - Sample Results of Experiment Designed by W. J. Youden of NBS for Proving Ring A

A Load applied to ring lb	B Computed load using second degree fitting method lb	Column A minus Column B lb	C Computed load using "best fit" method lb	Column A minus Column C lb	Ring reader
10,070	10,077	- 7	10,090	-20	1
30,000	29,987	+13	29,994	+ 6	1
40,050	40,059	- 9	40,064	-14	1
80,020	80,032	-12	80,036	-16	1
10,020	10,033	-13	10,046	-26	2
30,050	30,047	+ 3	30,053	- 3	2
40,000	40,011	-11	40,016	-16	2
80,070	80,081	-11	80,082	-12	2
10,000	9,993	+ 7	10,007	- 7	3
30,070	30,063	+ 7	30,069	+ 1	3
40,020	40,029	- 9	40,034	-14	3
80,050	80,061	-11	80,062	-12	3
10,050	10,058	- 8	10,068	-18	4
30,020	30,010	+10	30,013	+ 7	4
40,070	40,087	-17	40,096	-26	4
80,000	80,025	-25	80,030	-30	4