EVALUATING SPC TECHNIQUES
AND
COMPUTING THE UNCERTAINTY OF FORCE CALIBRATIONS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1989

Johnson Space Center

Prepared by:	Sharon E. Navard, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Virginia Commonwealth University Department of Mathematical Sciences Richmond, Virginia 23284-2014

ABSTRACT

In recent years there has been a push within NASA to use statistical techniques to improve the quality of production. Two areas where statistics is used is in establishing product and process quality control of flight hardware and in evaluating the uncertainty of calibration of instruments. The Flight Systems Quality Engineering branch is responsible for developing and assuring the quality of all flight hardware; the statistical process control methods employed are reviewed and evaluated in the first section of this report. The Measurement Standards and Calibration Laboratory performs the calibration of all instruments used on-site at JSC as well as those used by all off-site contractors. These calibrations must be performed in such a way as to be traceable to national standards maintained by the National Institute of Standards and Technology, and they must meet a four-to-one ratio of the instrument specifications to calibrating standard uncertainty. In some instances this ratio is not met, and in these cases it is desirable to compute the exact uncertainty of the calibration and determine ways of reducing it. A particular example where this problem is encountered is with a machine which does automatic calibrations of force. The second section of this report describes in detail the process of force calibration using the United Force Machine, identifies the sources of error and quantifies them when possible, and makes suggestions for improvement.

INTRODUCTION

Historically, quality assurance at NASA has not relied heavily on statistical techniques because of the nature of the work done. The equipment developed for space flight was so specialized that very few like items were manufactured and thus one hundred percent inspection was possible. However, since the advent of the Space Transportation System with its reusable orbiter, and particularly since the 51-L accident, there has been a push within NASA to apply trend analysis techniques to maintain tighter control on the quality of flight hardware. The first section of this report covers our examination of the statistical process control (SPC) techniques which were being used by the Quality Assurance and Engineering Division and our suggestions for improvement.

A particular area of Quality Assurance which is prone to a variety of statistical problems is the Measurement Standards and Calibration Lab (MSCL). The MSCL is responsible for calibrating all of the instruments used on-site as well as those of all off-site contractors. These calibrations must be performed in such a way as to be traceable back to national standards, to control the measurement process error, and to meet a four-to-one ratio of the uncertainties of calibrated item to calibrating standard. In some cases, the laboratory has been unable to achieve the last two criteria. The second section of this report presents our efforts to calculate the uncertainty associated with calibrations done by a United Force Machine.

STATISTICAL PROCESS CONTROL

Quality control is maintained on each of ten divisions by the use of separate control charts for each division. The data is obtained as follows. When any type of test is to be performed, a Test Preparation Sheet (TPS) must be filed. One TPS can cover an item as simple as a single screw or as complex as an entire satellite. After the test, any irregularities must be reported with a Discrepancy Report (DR). These DR's fall into three categories--Type I (damage or potential damage to hardware, persons, or both), Type II (deterioration of performance), and Type III (incidental). It is possible to have more than one DR for one TPS. Data consists of the number of TPS's and DR's per division since January 1986, with the DR's being classified as Type I, Type II, or Type III since January 1989.

The type of control chart which was being used was the p-chart with constant control limits, as described in Feigenbaum (1984), pp. 432 - 435. The p-chart is used for plotting the fraction of defective units out of n total units; constant control limits are used when the sample size n remains the same for each inspection period. This is not appropriate for this situation. The p-chart, which is based on the

normal approximation to the binomial distribution, assumes that the same number n of independent, identical units is being observed each period, and each unit is being classified as either defective or nondefective. In fact, the number of TPS's filed per month changes, the TPS's are not identical, and it is possible to get more than one defective per unit.

A more appropriate type of chart to use for this situation is the u-chart with variable control limits, as described in Duncan (1965), pp. 376 - 378. Using variable control limits will account for the fact that the number of TPS's filed per month does not remain constant. Furthermore, the u-chart, which is based on the Poisson rather than the binomial distribution, plots the fraction of defectives per (identical) unit rather than the proportion of defective (identical) units. U-charts with variable control limits are now being employed.

The fact that the TPS's are not identical means that even u-charts are not correct for this situation. The information gleaned from these charts could be improved by weighting the TPS's according to their complexity--for example, by the number of inspections actually done for each TPS. Unfortunately, this information has not been entered into the data base. The charts could also be improved by weighting the DR's according to their criticality; this type of weighting is discussed in Besterfield (1987), pp. 171 - 174. At present, not enough historical data is available to set control limits based on weighted DR's, but these weights will be implemented when sufficient data has been recorded.

COMPUTING THE UNCERTAINTY OF FORCE CALIBRATIONS

The Measurement Standards and Calibration Laboratory

The Measurement Standards and Calibration Laboratory is composed of three departments: Reference Standards Laboratory/Metrology Engineering, which maintains the standards; Instrument Calibration & Repair Services, which does the customer instrument calibrations and repairs, and Metrology Information & Management Services, which supplies technical and information support. The MSCL is responsible for calibrating and repairing all of the instruments used on-site at JSC, as well as all instruments used by off-site contractors. It services over eighty customers, performing an average of 18,500 calibrations and 1500 instrument repairs per year. The calibrations require and average of 2.26 manhours per item, with an average turn around time of 5.4 working days per instrument.

Each of these calibrations must be performed using a standard which is traceable back to the national standards maintained by the National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards. Furthermore, a check standard should

be maintained--that is, repeated observations on an artifact, or differences in artifacts, under all environmental and operational conditions under which calibrations will be performed. The purpose of the check standard is twofold. It gives an estimate of the random error of the measurement process itself, which is a component of the uncertainty, and it also allows one to use control charts to monitor the stability of the measurement process. For more information on check standards, see Croarkin (1984). Finally, the ratio of the specifications of the customer test item to the uncertainty of the calibrating standard must be at least four-to-one. For example, if the specification of an instrument is accuracy to within 1%, the uncertainty of the standard used to calibrate it cannot exceed 0.25%.

The MSCL uses standards calibrated by the NIST to assure traceability. However, for many of the calibrations done, no check standard is maintained, often due to time and financial constraints. When no check standard is available, it becomes impossible to use control charts to monitor the measurement process stability. It is also not possible to estimate the random error and, therefore, the uncertainty of the calibrated item. In cases where it is impossible to compute an exact value, the uncertainty of a calibrated item is arbitrarily assigned a value four times the uncertainty of the standard used to calibrate it. However, when this is done, in some instances the MSCL's standards no longer meet the four-to-one ratio. Thus it is desirable to have a means of computing the exact uncertainty.

Calibration of Force Measuring Devices

Considered here are two types of devices which can be used to measure force: proving rings and load cells. A proving ring is a metal ring equipped with a micrometer which measures the actual deflection of the ring when a force, either tension or compression, is applied. A load cell is a device which outputs an electric current when a force is applied. The particular situation examined here is the use of a United Force Machine to do automatic calibrations of customers' load cells.

In order to maintain traceability to national standards, all calibrations begin with a primary standard which was calibrated by the NIST. For force calibrations, the primary standard is a proving ring which the NIST calibrates using dead weights of known mass. Thirty data points are used--two repetitions each of fifteen known loads or three repetitions of ten known loads--and ordinary least-squares is used to fit the deflection (D) as a quadratic function of the load (L):

$$D = A + BL + CL^2.$$

Ordinary least-squares is appropriate for this situation because it can be assumed that the exact values of the load are known without error.

These calibrations are done at 23° C. The uncertainty associated with the calibration is taken to be 2.4s, where s is the estimate of the standard deviation of the residuals, σ_ε^2.

The laboratory does not use the primary standards to calibrate customers' instruments. Rather, they are used to calibrate the laboratory's secondary standards, which are then used for customer calibrations. For the MSCL force calibrations, the secondary standard is a load cell. This standard load cell is calibrated by the United Force Machine, using the primary standard to determine the values of the loads applied. Because the ambient temperature may not be 23° C, a temperature correction must be made to the deflection values of the proving ring. The loads (L) as measured by the proving ring and corresponding responses (R) of the load cell (in MV/V) are used to obtain the least-squares fit

$$L = A + BR + CR^2.$$

Least-squares is not entirely appropriate at this step because the responses as well as the loads are measured with error. The uncertainty of this calibration is taken to be four times the uncertainty reported by the NIST for the proving ring.

Once the standard load cell is calibrated, it is then used to determine the values of the loads to calibrate customers' load cells. The process is similar to the previous step except that no temperature adjustment is made, and the reported calibration equation is linear, that is,

$$L = A + BR.$$

The residuals, or deviations of each data point from the line, are computed, and if none of the residuals exceeds the specified accuracy of the load cell, it is said to be calibrated. However, if the specified accuracy of the calibrated load cell is not a least four times as great as the uncertainty of the standard load cell, the calibration does not meet the ratio. When this occurs, the MSCL is required to give a full report of the calibration procedure to the customer and inform him that the ratio was not met.

Because of the increasing precision of customer load cells, it is becoming more frequent for the MSCL to do force calibrations which do not meet the four-to-one ratio. For this reason, it is of interest to compute the actual uncertainty associated with the standard load cell, rather than using four times the uncertainty of the proving ring. Unfortunately, it will be impossible to compute the random error component of the uncertainty because at present no check standard is being maintained on the United Force Machine. However, other sources of error can be identified and possibly reduced. The following

sections describe in more detail the sources of systematic error inherent in the calibration of the standard load cell.

Inverse Regression versus Classical Calibration

The first problem encountered in using the primary standard to calibrate the standard load cell is obtaining values of the load from the deflection of the proving ring. The equation supplied by the NIST gives the deflection of the ring as a function of the load, yet the ring must be used to obtain the values of the load as a function of the deflection. This is known as the calibration problem, and there are two approaches to solving it.

The classical method of calibration is to use the given equation, $D = A + BL + CL^2$, and simply solve it for L as a function of D. Since it is a quadratic equation, there will be two solutions; the correct one is the one which lies in the calibration range of the proving ring. Denote the classical estimator of load by L_C. Then

$$L_C = [-B + \sqrt{B^2 - 4C(A - D)}] / 2C.$$

The standard error of the classical estimator of L is different from σ_e^2, the quantity currently being used by the MSCL; it will, in fact, be larger because the least-squares equation is being used in the opposite direction than the one in which the errors were minimized. Not much work has been done on determining estimates of σ_C, the standard error of the classical calibration estimator, when the response is quadratic.

The second approach to the calibration problem, known as inverse regression, is to take the original data points supplied by the NIST and use least-squares to fit the load as a quadratic function of the deflection. Denote the inverse regression estimator of load by L_{IR}. Then

$$L_{IR} = \alpha + \beta D + \gamma D^2,$$

where α, β, and γ are the parameter estimates obtained using ordinary least-squares. As in the case of the classical estimator, little work has been done in obtaining estimates of σ_{IR}, the standard error of the inverse regression estimator, when the response is quadratic. It will not be simply the standard error of the residuals because one of the major assumptions of least-squares has been violated: it is the dependent variable, rather than the independent variable, which has been measured without error. Inverse regression is the technique currently being employed by the MSCL.

A serious question at this point is to ask which of the two methods is "better." This is a question which has been addressed

extensively in the literature, although generally only a linear response is considered. The first author to address the issue was Eisenhart (1939), who concluded that the classical method is the only reasonable one to use for two reasons. First, the classical method minimizes the actual observed errors, while the inverse regression method minimizes the "errors" in a variable that was measured without error. Second, the classical method is asymptotically unbiased, whereas the inverse regression estimator is asymptotically biased.

This seemed to be the final word on the subject until Krutchkoff (1967) reopened the question with a simulation paper which indicated that the inverse regression estimator has uniformly smaller mean squared error (MSE) than the classical estimator, and is therefore superior based on that criterion. In an ensuing letter, Krutchkoff (1968) pointed out that the classical estimator will in fact have infinite MSE if the linear slope term (or in our case, the quadratic term, C) is allowed to be zero, and that truncating the estimator will introduce a bias. Then, in Krutchkoff (1969), he demonstrated that as the number of observations at each L value increases, the MSE of the classical estimator decreases faster than that of the inverse estimator and can in fact get smaller. The classical estimator also gave a smaller MSE when there was an ignored quadratic term.

At this point more authors began to address the question. Williams (1969) demonstrated that not only the classical estimator, but *any* unbiased estimator will have infinite MSE and therefore MSE is not a good criterion to use to compare the estimators. Berkson (1969) claimed that classical estimator is consistent, whereas the inverse estimator is inconsistent, and the inverse estimator should thus never be used. Martinelle (1970) derived analytic results which basically agreed with Krutchkoff's simulation results. Halperin (1970) compared the two estimators using Pitman closeness, concluding that the classical estimator is generally better. Hoadley (1970) approached the problem from a Bayesian point of view and demonstrated that the inverse estimator is actually Bayes with respect to a particularly informative prior. For additional information on the calibration problem, consult the bibliography.

Further research is required to determine which of the two methods is better for use in force calibrations. Currently, inverse regression is being used. Table 1. demonstrates the differences in the estimates of load obtained using the two methods for determining compression loads for the 100,000 pound proving ring. It can be seen that the percent difference in the two estimators is not very large, especially for the larger loads.

TABLE 1.- ESTIMATES OF LOAD USING THE CLASSICAL AND INVERSE REGRESSION ESTIMATORS

	Load	
Deflection	Classical Method	Inverse Regression
0	-10.695	0.624
100	9919.771	9926.369
200	19826.557	19829.628
300	29709.660	29710.400
400	39569.465	39568.686
500	49405.969	49404.485
600	59219.176	59217.798
700	69009.461	69008.624
800	78777.219	78776.963
900	88522.055	88522.816
1000	98244.352	98246.182

Propagation of Temperature Errors

Another problem encountered when using the primary standard to calibrate the standard load cell is the effect of differences in temperature. The proving ring was calibrated at 23° C, and if it is used at a different temperature, then the deflection readings must be adjusted according to the equation

$$d_{23} = d_t[1 + k(t - 23)],$$

where d_{23} is the deflection at 23° C, t is the current temperature, d_t is the observed deflection, and k is the expansion coefficient of the ring. For the proving rings used by the MSCL, k = -.00027.

Unfortunately, the thermometer being used to determine the temperature of the proving ring is only accurate to within ± .95° F ≅ .53 ° C. For a temperature error of .53° C, the percent error in the deflection, denoted $\%_D$, is computed as

$$\%_D = |\text{Actual } d_{23} - \text{Computed } d_{23}| / \text{Actual } d_{23},$$

and is approximately .00015, or .015% for temperatures near 23° C.

To determine the effect of this deflection error on the values of the load, the error must be propagated through the expression

$$L = F(D) = \alpha + \beta D + \gamma D^2.$$

The general propagation of errors formula as given in Ku (1966) is
$$\mathrm{Var}\ (L) = [dF/dD]^2\ \mathrm{Var}(D).$$

An estimate of Var(D) is required in order to use this formula. One may be obtained by noting that the percent error in deflection is the ratio of the error in the deflection to the deflection; but the error in deflection can be expressed as the uncertainty, which is generally a multiple r of the standard deviation of deflection, σ_D. (The NIST uses r = 2.4.) Thus
$$\%_D = (\text{error in D})/D \cong \text{uncertainty}/D \cong r\sigma_D/D,$$
and
$$\mathrm{Var}(D) \cong [D\%_D/r]^2.$$

Using this estimate of Var(D) in the propagation of error formula yields an estimate of the standard deviation of load values due only to the .53° C error in temperature, $\sigma_{L(t)}$:

$$\sigma_{L(t)} \cong D\%_D[\beta + 2\gamma D]/r.$$

Thus the percent uncertainty of load values due to this error in temperature, $\%_{L(t)}$, is

$$\%_{L(t)} \cong r\sigma_{L(t)}/(\text{maximum load}) \times 100\%.$$

For the 1000 and 100,000 pound proving rings, this error ranged from .015% to .017%. These values are approximately equal to the percent uncertainties associated with the calibration of the proving rings themselves. Therefore the error due to inaccuracy of the temperature readings is a significant contributor to the total uncertainty of the standard load cell.

The Errors-in-Variables Model

When the proving ring is used to determine values of the load in order to calibrate the standard load cell, there is error associated with these load values as well as with the response values of the load cell. Therefore, when the MSCL uses least-squares to fit the equation $L = A + BR + CR^2$, this is not an example of inverse regression, but is rather what is known as an errors-in-variables (EV) model. EV models result when both of the variables are measured with error, and in this situation even determining consistent estimators of the regression parameters when one of the error variances or their ratio is unknown

is an open problem. See Gunst and Lakshminarayanan (1984) for more details on the EV model.

More research is required in order to determine the best estimates of the parameters of the EV model and their associated standard errors. Also, since the MSCL is currently using least-squares to estimate the regression parameters, an estimate of the standard error of load estimates thus obtained, σ_{EV}, is also a quantity which needs to be determined in order to compute the uncertainty of the standard load cell.

To compute the uncertainty associated with the standard load cell, it is necessary to combine the systematic and random errors. While it is still a source of contention as to how exactly this should be done, the general consensus seems to be that independent systematic errors can be added in quadrature, while correlated systematic errors and random errors should be added linearly. Thus the uncertainty of the standard load cell, uncertainty$_{SLC}$, can be estimated as

$$\text{uncertainty}_{SLC} = 2.4[\sqrt{\sigma_{IR}^2 + \sigma_{L(t)}^2} + \sigma_{EV} + \sigma_{RE}],$$

where σ_{RE} is the random error of the measurement process, as determined by the check standard.

Calibrating the Customer's Load Cell

The problems encountered when calibrating customers' load cells are similar to those of calibrating the standard load cell. Once again, there is error in the values of the load determined by the standard load cell, as well as in the response of the customer's load cell, so this is also a case of an EV model. There is also an uncompensated systematic error due to a drift in the response of load cells at different temperatures. Thus the uncertainty associated with the calibration will be a sum of the uncertainties of the standard load cell, the temperature drift, the EV model, and the random error of the measurement process.

DISCUSSION

Improving Use of the United Force Machine

Under current practices, a customer's load cell can be properly calibrated only if its specifications are more than sixteen times the uncertainty of the proving ring used to calibrate it. This is because the MSCL takes the uncertainty of the standard load cell to be four times that of the proving ring calibrated by the NIST, and the uncertainty of the customer's load cell to be four times that of the standard load cell.

If the actual uncertainty were computed, it may or may not be smaller. However, it will not be possible to compute the uncertainty until some of its component values are determined.

The most important step for the MSCL to take at this time is to institute a check standard for the United Force Machine. This will supply an estimate of the random error of the measurement process, which is actually added twice into the overall uncertainty. It will also enable the stability of the measurement process itself to be monitored through the use of control charts, thus allowing a faster identification of any problems with the process which may arise.

The uncertainty of the calibrations can be reduced by obtaining more accurate measurements of the temperature of the proving rings during calibration of the standard load cells. When considering more precise measurements of the temperature, however, one is faced with the problem of the temperature varying on different parts of the ring, due to stress on the ring or changes in the ambient temperature. This problem could be reduced if the proving rings could be stored and used in a temperature-controlled environment. Using proving rings with a smaller expansion coefficient would also reduce the error due to inaccurate temperature readings.

The systematic error due to the temperature drift of the load cells could be reduced by obtaining an accurate determination of the drift from the manufacturer, if such is available, and incorporating the temperature correction into the software. If this is done, the error in temperature readings will become a factor, but this error should be smaller than the full drift error.

Applications to Other Areas of the MSCL

The discussion here has centered on automatic force calibrations using the United Force Machine. However, some of the same principles can be applied to other calibration processes. In particular, the use of check standards should be made on as many of the processes as possible, especially those which are subject to large variations and those where mistakes are most likely to be made. Not only will this provide an estimate of the random error, but it will also facilitate the identification of operator errors, changes in the values of the reference standards, effects due to unusual environmental fluctuations, etc.

A tighter control on the environment is also something which should be sought. Many of the calibrations done in the MSCL are affected by changes in temperature, pressure, humidity, etc. In some cases these effects are compensated for, but even in these instances errors in the readings can have significant effects on the final results. The more stable the environment is, the more accurate the calibrations will be.

REFERENCES

Berkson, Joseph: Estimation of a Linear Function for a Calibration Line; Consideration of a Recent Proposal. *Technometrics*, vol. 11, no. 4, Nov. 1969, pp. 649-660.

Besterfield, Dale H.: *Quality Control*, 2nd Ed. Prentice-Hall, 1987.

Croarkin, Carroll: Measurement Assurance Programs Part II: Development and Implementation. National Bureau of Standards Special Publication 676-II, Apr. 1984.

Duncan, Acheson J.: *Quality Control and Industrial Statistics*, 3rd Ed. Richard D. Irwin, Inc., 1965.

Eisenhart, C.: The interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research. *Annals of Mathematical Statistics*, vol. 10, no. 2, June 1939, pp. 162-186.

Feigenbaum, Armand V.: *Total Quality Control*, 3rd Ed. McGraw-Hill, Inc., 1984.

Gunst, Richard F., and Mani Y. Lakshminarayanan: Estimation of Parameters in Linear Structural Relationships: Sensitivity to the Choice of the Ratio of Error Variances. *Biometrika*, vol. 71, 1984, pp. 569-573.

Halperin, Max: On Inverse Estimation in Linear Regression. *Technometrics*, vol. 12, no. 4, Nov. 1970, pp. 727-736.

Hoadley, Bruce: A Bayesian Look at Inverse Linear Regression. *Journal of the American Statistical Association*, vol. 12, no. 4, Nov. 1970, pp. 356-369.

Krutchkoff, R. G.: Classical and Inverse Regression Methods of Calibration. *Technometrics*, vol. 9, no. 3, Aug. 1967, pp. 425-439.

Krutchkoff, R. G.: Letter to the Editor. *Technometrics*, vol. 10, no. 2, May 1968, pp. 430-431.

Krutchkoff, R. G.: Classical and Inverse Regression Methods of Calibration in Extrapolation. *Technometrics*, vol. 11, no. 3, Aug. 1969, pp. 605-608.

Ku, Harry H.: Notes on the Use of Propagation of Error Formulas. *Journal of Research of the National Bureau of Standards--C. Engineering and Instrumentation,* vol. 70C, no. 4, Oct. - Dec. 1966.

Martinelle, Sten: On the Choice of Regression in Linear Calibration. Comments on a paper by R. G. Krutchkoff. *Technometrics,* vol. 12, no. 1, Feb 1970, pp. 157-161.

Williams, E. J.: A Note on Regression Methods in Calibration. *Technometrics,* vol. 11, no. 1, Feb. 1969, pp. 189-192.

BIBLIOGRAPHY

Brown, G. H.: An Optimization Criterion for Linear Inverse Estimation. *Technometrics*, vol. 21, no. 4, Nov. 1979, pp. 575-579.

Hunter, William G., and Lamboy, Warren F.: A Bayesian Analysis of the Linear Calibration Problem. *Technometrics*, vol. 23, no. 4, Nov. 1981, pp. 323-328.

Kalotay, A. J.: Structural Solution to the Linear Calibration Problem. *Technometrics*, vol. 13, no. 4, Nov. 1971, pp. 761-769.

Krutchkoff, R. G.: Letter to the Editor. *Technometrics*, vol. 12, no. 2, May 1970, pp. 433-434.

Ku, Harry H., ed.: *Precision Measurement and Calibration--Selected Papers on Statistical Concepts and Procedures*. National Bureau of Standards Special Publication 300, vol. 1, Feb. 1969.

McKeon, James John: Statistical Calibration Theory. Ph. D. Thesis, Old Dominion University, 1985.

Minder, Ch. E., and Whitney, J. B.: A Likelihood Analysis of the Linear Calibration Problem. *Technometrics*, vol. 17, no. 24, Nov. 1975, pp. 463-471.

Perng, S. K., and Tong, Y. L.: Optimal Allocation of Observations in Inverse Linear Regression. *Annals of Statistics*, vol. 5, no.1, 1977, pp. 191-196.

Scheffe, Henry: A Statistical Theory of Calibration. *Annals of Statistics*, vol. 1, no.1, 1973, pp. 1-37.

Shukla, G. K.: On the Problem of Calibration. *Technometrics*, vol. 14, no. 3, Aug. 1972, pp. 547-553.