SPECIAL REPORT

A STATISTICAL APPROACH TO INSTRUMENT CALIBRATION

EM-710041

Robert R. Ziemer

Research Hydrologist

Pacific Southwest Forest and Range Experiment Station

David Strauss

Assistant Professor of Statistics

University of California, Riverside

JANUARY 1978

FOREST SERVICE

U.S. DEPARTMENT OF AGRICULTURE

WASHINGTON, D. C. 20013

CONTENTS

	Page
SUMMARY	v
TERMS AND SYMBOLS	vi
INTRODUCTION	1
DESCRIPTION OF THE PROBLEM	3
CALIBRATION PROCEDURE	7
Laboratory Calibration	7
Field Calibration	8
REGRESSION ANALYSIS OF FIELD DATA	11
Common Slope	14
Common Intercept	16
ESTIMATION OF ANGLE	19
CONCLUSION	27
REFERENCES	29

SUMMARY

It has been found that two instruments will yield different numerical values when used to measure identical points. A statistical approach is presented that can be used to approximate the error associated with the calibration of instruments. Included are standard statistical tests that can be used to determine if a number of successive calibrations of the same instrument can be considered to be the same curve, or if they are different. Essentially, these tests involve the determination of regression lines that have both a common slope and common intercept. The example was designed to monitor the deformation of mountain slopes. The statistical approach is general (standard), however, and does not depend upon the specific type or function of instrument being calibrated.

TERMS AND SYMBOLS

a_F, b_F, a_L, b_L the estimates of the population parameters;

k = number of years (number of field regression lines);

n_F, n_L = number of observations of (X, Y_F) and (θ, Y_L) respectively;

n_i = number of observations in the i^{th} year;

S_{xy} = $\sum_{i=1}^{k} (X_i - \bar{X})(Y_i - \bar{Y})$, the sum being taken over all in a particular year. Similarly define S_{xx}, S_{yy}, $S_{\theta\theta}$;

X = field strain measured in any particular year;

Y_F, Y_L = strain measured in 1972 in the field and laboratory respectively;

α_F, β_F = regression parameters of Y_F on X, the field calibration;

α_L, β_L = regression parameters of Y_L on θ, the laboratory calibration;

$\varepsilon_{F_2}, \varepsilon_{L_2}$ = errors of the dependent variables Y_F and Y_L, respectively;

σ_F, σ_L = variances of the ε_F and ε_L, respectively;

θ = angle set in the laboratory in 1972, in degrees.

INTRODUCTION

A problem that faces researchers and engineers who collect laboratory or field measurements is the calibration and subsequent stability of their instruments in relation to changes in time, temperature, pressure, or other external factors. These changes may cause the instrument readings to drift from the manufacturer's calibration. It is often necessary to recalibrate the instrument or use "secondary standards" to determine if the instrument is functioning properly.

A statistical approach has been developed to determine if new calibration lines are necessary for accurate use of the instrument. The example to be given outlines calibration of a borehole inclinometer designed to monitor the deformation of mountain slopes. The same statistical approach can be used with other instruments such as the neutron soil moisture probe, solar radiometer, suspended sediment sampler, or the precipitation gauge.

DESCRIPTION OF THE PROBLEM

Within a ten-year period, the Forest Service has studied the effects of land-management practices on acceleration of the rate of natural soil creep. The term *soil creep* refers to the slow, downslope movement of soil and rock materials. In mountainous regions of the Pacific Coast, natural soil creep processes play a major role in soil movement by directly contributing to sediment in streams, and indirectly contributing to landslides through progressive slope failure.

In 1964, the installation of a borehole access casings network was undertaken to study the character of natural soil creep. By 1966, approximately 150 access casings had been installed at 17 sites, representing a wide range of soil and vegetation types in northwestern California[1]. The casings were periodically surveyed with a modified strain-gauge inclinometer as described by Kallstenius and Bergau[2]. Strain readings were made with a *Balduin-Lima-Hamilton Model 120* strain indicator. The inclinometer is essentially a precise pendulum connected to a high-resolution strain gauge. The instrument is rotated in the borehole casing until the maximum strain reading is obtained. At this point, the orientation of the instrument is recorded in terms of strain reading and azimuth.

The access casings were installed in the field by drilling 8.9 cm boreholes, at least 1 m into bedrock whenever possible, to obtain a stable foundation. A 6.0 cm OD polyvinylchloride (PVC) tube was pressed into each borehole, then the entire site was allowed to stabilize for one winter before the initial inclinometer readings were taken.

Four measurements of orientation and inclination of the inclinometer in the casing were taken at 0.5 m depth increments from the soil surface to the bottom of the hole, and then repeated from the bottom to the surface. The eight pairs of readings at each measurement point were averaged to obtain the best estimate of strain and azimuth for that depth. These data were further reduced by computer to yield a projection of the casing on the plane of maximum displacement, and also on the horizontal plane by means of a polar projection. Comparison of the casing configuration between successive surveys reveals the direction and amount of movement of the casing with increasing depth in the soil profile.

By 1972 the accumulated data spanned eight years and revealed that the "movement" patterns of the borehole casings were extremely erratic. The movement of the borehole casing was found to be uniformly greater in the period from the initial survey to the second survey than the movement found in successive surveys. In addition, the direction of movement was neither progressive nor consistent.

Casings installed in boreholes about 1 m from each other showed markedly different patterns of movement, which could have been the result of three factors operating independently or in combination:

1. The method of installing the borehole casing was found to be a major cause of the problem. The initial assumption that the sides of the borehole would collapse within one year, resulting in an intimate contact of the casing with the surrounding soil materials, was found to be in

error. Alignment corrections in the borehole were still occurring eight years after installation[3].

2. The instrument error could have exceeded the measured rate of soil creep, and the indicated movement could be a reflection of random instrument error as well as the operator's ability to null the instrument. By using Monte Carlo techniques, the precision of the instrument for field measurements was found to be equivalent to about 2 mm of displacement for an 8 m deep borehole casing. Thus, the error of calculating displaced area is approximately 0.01 m^2 in soil 8 m deep[3].

3. The calibration of the instrument could reflect an inadequate expression of the instrument response, which would require a new calibration. The statistical calibration procedure, described in the following section, provides a satisfactory solution to this problem.

CALIBRATION PROCEDURE

LABORATORY CALIBRATION

During the course of the eight-year study, the inclinometer and strain indicator were periodically returned to the factory for adjustment and repair. Each spring, before the field measurement season, the instrument was calibrated in the laboratory by the following method: (a) the instrument was situated at "known" angles from the vertical; (b) using a high-precision granite plate, a sine plate, and gauge blocks, strain gauge readings associated with the angle were obtained; (c) a calibration curve of the strain gauge readings on the angle was then obtained by a "least-squares fit" to the data.

In the course of this investigation, it became evident that there were several serious errors in the laboratory calibration procedure. Although the angles set in the laboratory calibration covered a wide range, the angles of inclination of the hole casings encountered in actual field measurements were small, usually about 0.01 degrees. Therefore, almost all of the laboratory calibration data were collected at angles greater than the angles of interest.

In addition, under the incorrect assumption that a 0^o angle would yield a 0^o strain reading, no laboratory measurement was made at the vertical angle of 0^o. This procedure may be correct in theory; in practice, however, a 0^o angle is almost never found, because of the alignment of the strain gauge in the instrument and the practical difficulty of placing the instrument in a precise vertical position. The gauge blocks simply place the sine plate at precise angles from the initial set up. Therefore, not only does the annual laboratory calibration have error associated with the least-squares fit to the data, but the calibration line must be extrapolated beyond the actual data to the range where most of the field data are found.

This unsatisfactory situation was corrected when the 1972 laboratory calibration was made. The procedure was changed so that most of the laboratory data were collected within the range of angles encountered in the field measurements; however, field data that had been collected during the seven previous years by equipment without adequate laboratory calibrations, remained in the records of previous tests.

FIELD CALIBRATION

One of the criteria established for drilling the borehole was to extend the hole at least 1 m into bedrock to obtain a stable foundation for the inclinometer casing. If this was actually accomplished for each access casing, as is indicated in the drilling log, the inclination of the access casing at the bottom two measurement positions should remain constant from year to year. It then follows that the measured strain gauge readings would also remain constant for these depths, with two sources of variation: (a) the random instrument and operator error, previously found to be equivalent to an average displacement of about 2 mm; (b) the shift in instrument calibration from survey to survey. For each survey, then, the strain gauge measurement for a given angle would either be the same or would have a consistent shift relative to previous surveys, with variation due to the random instrument and operator error.

Significant variation in strain measurement from survey to survey for a given access casing (outside of those expected to be associated with the random error and the calibration shift) would lead one to conclude that the bottom of the access casing was not established in stable bedrock. In checking the data, several of the access casings produced readings that were consistently and obviously deviant ($>5\sigma$) from the readings obtained from the other casings. These deviant casings were rejected from further analysis.

The ultimate objective is to estimate the vertical angle of the bore-hole casing at each measurement point for each year in which measurements were made; however, the only year for which an adequate laboratory calibration exists is 1972. Thus, there is a problem in determining a relationship between the field strain measurements made in 1972 and those made at the same stable field locations in previous years.

The 1972 field strain data can be used as a "bridge", since they are related both to angle (through the 1972 laboratory calibration) and to the field strain gauge measurements made in previous years. The first step is to calculate regression lines relating the 1972 field strain gauge data to the strain data collected in the surveys of 1965, 1966, 1968, and 1970. The next step is to establish if the relationship has remained constant with time--that is, if the five separate regression lines have a common slope and intercept. If this can not be established, the instrument calibration has then shifted and a different calibration formula is required for each year. The third step is to estimate the relatiohship between the strain gauge measurement and the angle precisely set in the 1972 laboratory tests. The two analyses are then combined to estimate the angle of the borehole casing from the field strain gauge measurements, and give confidence intervals for these angles.

REGRESSION ANALYSIS OF FIELD DATA

There are a number of preliminary questions to be answered before it can be determined whether the five regression lines of field strain measurement made in 1972 (Y) for each of the five previous years (X) may be regarded as the same line.

1. Why fit regression lines? In a strict sense, it is not correct to regard the strain measured in 1965, for instance, as being an independent variable (X) measured without error, and the 1972 strain (Y) as the dependent variable with an error term attached. In fact, the theoretically correct approach (Kendall and Stuart[4,5]) supposes a structural relationship between X and Y. Such an approach is not necessary in practice for this case, because the fit to a straight line is good, as indicated by the high R^2 values (Table 1). This is fortunate, since the theoretical developments for the structural case would be much more difficult than for the regression case.

2. The range in X varies from year to year; also, not all of the plots were measured in each of the five years. Given the conditions, is it justifiable to assume that the regression line can validly be extrapolated beyond the range of the data for those years where the range of X is rather short? In general, the answer is that extrapolations cannot be justifiably made because they can lead to dangerously misleading conclusions. In this study, however, the ranges of X for five years of data

vary by no more than a factor of 2, and a careful examination of the scatter of points on the X and Y plots for each year give no indication of departure from a straight line. Therefore, in this case, a *limited* extrapolation is justified.

TABLE 1-- Summary of regression statistics for the five calibration lines.

Year	Number of points $= n$	Error Mean Sq.	Regression Sum sq.	Slope b	P^2
1965	18	163.75	671,195	1.296	0.99611
1966	21	81.84	4,738,688	1.004	0.99967
1967	56	407.35	164,110,312	0.961	0.99866
1968	15	49.53	2,248,886	1.025	0.99971
1970	39	139.13	10,647,179	0.998	0.99952

S_{xx} - summed over the years 1966 through 1970 is 35,307,342.

S_{xy} - summed over the years 1966 through 1970 is 34,662,401.

S_{yy} - summed over the years 1966 through 1970 is 34,075,193.

3. Are simple linear regressions adequate, or should quadratic or cubic regressions be fitted to the data? The data for each of the five years were analyzed, using combinatorial screening for multivariate regression[6]. Tests were made of X, X^2, X^3, and X^4 alone and in combination, This analysis confirmed that the higher order transformations did not significantly add predictive power to the linear fit.

4. Is the error variance the same for each regression line? The standard analysis of regression assumes this property of homoscedasticity.

It is known, however, that the analysis is relatively insensitive to small departures from the assumption. In Table 1, the five error mean squares show the variances to be roughly comparable--although the one for 1967 looks suspiciously large.

The equality of variance can be determined by Bartlett's test[7], in which the quantity

$$2 \log \lambda \left\{ 1+\frac{1}{3(k-1)} \sum_{i=1}^{k} \frac{1}{n_i} - \frac{1}{\Sigma n_i} \right\}^{-1} \quad \ldots\ldots\ldots\ldots\ (1)$$

will have the χ^2 distribution with $(k-1)$ d.f., if the population variances are indeed equal. Here, if $\hat{\sigma}_i^2$ is the error mean square in the i^{th} year, based on n_i data points $\tilde{\sigma}^2$ is the polled estimate of variance based on k different $\hat{\sigma}_i^2$, then λ is calculated from

$$\lambda = \left[\prod_{i=1}^{k} (\hat{\sigma}_i^2)^{(n_i-1)/2} \right] \left[(\tilde{\sigma}^2)^{(\Sigma n_i - k)/2} \right]^{-1} \quad \ldots\ldots\ldots\ (2)$$

Using the compiled data from all five years, a λ of 32.14 is obtained which is highly significant as χ^2 with 4 d.f. It is easily shown that the variance for 1967 is significantly larger than the combined variance for the other four years. The discrepancy in 1967, however, was probably not great enough to invalidate the subsequent analysis, so the 1967 data are retained. However, it must be noted for future reference that the variance for the 1967 data is anomalous.

A comparison of the five separate regression lines of *Y* (1972 field strain gauge measurement) on each *X* (field strain gauge measurement made in a particular year--1965, 1966, 1967, 1968 or 1970) was made. From the comparison, the following questions were found pertinent:

1. Do the lines have a common slope?
2. If the lines have a common slope, are the parallel lines actually the same line (i.e., do they have a common intercept)?
3. Do the various lines pass through the origin?
4. Do the lines have a slope of 1?

The last two questions are of little importance here, and will not be investigated in detail. However, it is important to realize that they are not "obviously" true, even though they seem plausible on physical grounds; indeed, with the present data, the questions would be irrelevant in the majority of cases. For example, it would be quite wrong not to include a constant term in our regression model for Y on X.

The theory is outlined as needed here; for a more detailed account refer to Williams[8].

COMMON SLOPE

<u>Theory</u>

For a particular year, let S_{xy} be the sum of $(X_j-\bar{X})(Y_j-\bar{Y})$ over all the data points for that year. Similarly, S_{xx} is the sum of $(X_j-\bar{X})^2$, and S_{yy} is the sum of $(Y_j-\bar{Y})^2$.

Then, the total sum of squares for regression is

$$\sum_{1}^{k} (S_{xy}{}^{2}/S_{xx}),$$

with k degrees of freedom, the sum being taken over all the k years.

The next requirement is the regression sum of squares when all lines are assumed to have a common slope. It is easiest to imagine that the means ($\bar{X}$, $\bar{Y}$) for the k sets of data are moved to a common point. The procedure is then continued as if fitting a single line to all k sets of points taken together. Thus, the total sum of squares for X is

$$\sum_{1}^{k} S_{xx}.$$

The total sum of products for XY is

$$\sum_{1}^{k} S_{xy},$$

and the total sum of squares for Y is

$$\sum_{1}^{k} S_{yy}.$$

Hence, the regression sum of squares for the k lines with a common slope is

$$\left(\sum_{1}^{k} S_{xy}\right)^2 \Big/ \left(\sum_{1}^{k} S_{xx}\right)$$

with one degree of freedom.

The difference between the two regression sums of squares is the sum of squares for departure from parallelism, on (k-1) d.f., and it can be tested for significance against the pooled estimate of residual sum of squares from all the lines. The latter is the "combined residual" term in Table 2.

Analysis

Some summary statistics for the regression analysis are given in Table 1. It is immediately obvious that the Slope b for 1965 is widely divergent from the others. Obviously, predicting angles from the 1965 field strain gauge readings would require an equation different from those for the other year. Therefore, data for 1965 are omitted from the analysis to see whether the other four lines appear to differ. The analysis of variance for the remaining four lines is shown in Table 2. Clearly, the *F* value of 23.1 provides evidence that the four lines do not have a common slope.

TABLE 2 -- Analysis of variance for common slope test.

Item	D.F.	S.C.	MS.	F
Regression with common slope	1	34,029,328.5		
Difference of slopes[1]	3	16,546.2	5,515.4	23.1[4]
Sum of 4 Regression S.S.	4	34,045,784.7		
Combined residuals[2]	123[3]	29,342.9	238.6	

[1] By subtraction

[2] Sum of four separate residual S.S.

[3] $(n_{1966}-2)+(n_{1967}-2)+(n_{1968}-2)+(n_{1970}-2) = (21-2)+(56-2)+(15-2)+(39-2)$

[4] Statistically significant at the 1% level

COMMON INTERCEPT

This case is presented for completeness only, since it has been already shown that the lines are not parallel and so can hardly be identical.

Theory

The sum of squares corresponding to the difference in positions of the regression lines, as opposed to the difference in slopes sum of squares, must be obtained. If differences in years are disregarded, all the data can be combined, giving a grand total sum of squares on

$$\sum_{i=1}^{k} n_i - 1 \text{ d.f.}$$

This can be separated into three components:

1. Combined residual sum of squares, as in Table 2 (123 d.f.).
2. Sum of squares for the four separate slopes of the lines, as in row 3 of Table 2 (4 d.f.).
3. *Sum* of squares for difference in positions (intercepts) of the lines, obtained by subtraction (3 d.f.).

Analysis

The analysis of variance for the four lines is shown in Table 3. For compatibility with Table 2, the 1965 data have been omitted. It is seen that the difference in positions is significantly large, having an *F* of 64.7. As noted above, however, this is of little interest, once the question of parallelism has been rejected.

TABLE 3 -- Analysis of variance for common intercept test.

Item	D.F.	S.S.	M.S.	F.
Sum of 4 Regression S.S.[1]	4	34,045,784.7		
Difference of intercept (by subtraction)	3	46,295.6	15,431.9	64.7[2]
Combined residual[1]	123	29,342.9	238.6	
TOTAL S.S.	130	34,127,423.2		

1 See Table 2.
2 Statistically significant at the 1% level.

ESTIMATION OF ANGLE

Once the comparability of the field strain gauge data from year to year is examined and a relationship between the years by regression is established, the main problem is calibrating the borehole inclinometer so as to estimate the vertical angle which corresponds to the strain gauge reading. In the standard calibration procedure for this type of instrument, the angle (8) expressed in degrees is precisely set in the laboratory, and the strain (Y_L) is measured with the inclinometer and strain gauge. Here and subsequently, the subscript L refers to the laboratory calibration; F refers to the field calibration.

For the 1972 data, the regression line is

$$Y_L = -6.66 \quad + \quad 11{,}714.65 \ \theta \qquad (3)$$

The residual mean square error for the laboratory calibration (σ_L^2) is 281.3 with 156 d.f.

The value for σ_L^2 is close to σ_F^2, the combined residual mean square error of 238.6 obtained from the four regression lines developed from the field calibration with 123 d.f. (Table 2). There was no reason to expect this in advance, because σ_F^2 is the residual mean square error for the 1972 field strain gauge measurement, and σ_L^2 is the 1972 laboratory experiment error. Conceivably, the error in reading the instrument might be the same in each case, but there are, of course, other sources of variation about a regression line. The implication is that the problem cannot be simplified by assuming that one error variance is negligible compared to the other.

The standard error of the slope for the laboratory calibration, calculated by the usual formula

$$\sigma_L \Big/ \sqrt{S_{\theta\theta}}$$

is 14.16. The standard error of the intercept, based on a sample of size n_L, as calculated by

$$\sigma_L \sqrt{\frac{1}{n_L} + \frac{\Sigma\theta^2}{S_{\theta\theta}n_L}} \qquad (4)$$

is 2.00. Thus, the answer to the question of a line through the origin is negative.

Theoretically, it is possible to obtain 0 angle results in a strain reading different from 0, and this would imply that there was some error associated with the installation of the pendulum in the instrument when it was fabricated. It must be remembered that the *common practice of assuming that the calibration line passes through the origin is invalid.*

Theory

Again let X denote the field strain gauge measurement in some particular year and let Y_L be the 1972 laboratory strain gauge measurement, Y_F the 1972 field strain gauge measurement, and θ the vertical angle in degrees. Thus, the 1972 laboratory strain measurement on angle becomes

$$Y_L = \alpha_L + \beta_L\theta + \varepsilon_L \qquad (5)$$

and the regression of the 1972 field strain gauge measurement (Y_F) upon strain in a particular year (X) becomes

$$Y_F = \alpha_F + \beta_F X + \varepsilon_F \quad \text{. . . . .(6)}$$

In this case, the ε_L's are independently distributed as normal with mean 0 and variance σ_L^2, and the ε_F's are independently normal with mean 0 and variance σ_F^2. All of the parameters $\alpha_L, \beta_L, \alpha_F, \beta_F, \sigma_L^2$ and σ_F^2 have been estimated, the denotation of the first four by a_L, b_L, a_F and b_F.

The objective is to estimate the vertical angle (θ) from the strain measurement made in the field (X). It must be assumed, however, that $Y_L = Y_F$, which is a basic assumption of nearly every instrument calibration experiment. That is, the laboratory measurements and the field measurements are equal, and are subjected to the same errors. We have already seen that the error variances are similar. Thus, the angle (6) can be estimated from

$$\hat{\theta} = \frac{a_F + b_F X - a_L}{b_L} \quad \text{. . . . .(7)}$$

Next, the problem of finding confidence intervals must be solved for the estimated 8. Reviewing the ordinary regression case, where a line is fitted

$$Y = a + bX \quad \text{. . . . .(8)}$$

Two cases arise in determing the variability of the prediction at $X = X_o$.

1. To know the accuracy of the estimate of Y on the *regression line* at $X = X_o$, the standard formula for the variance of the prediction is

$$\text{Var } \hat{Y} = \sigma^2 \left[\frac{1}{n} + \frac{(X_o - X)^2}{S_{xx}} \right] \quad \ldots\ldots\ldots\ldots \quad (9)$$

where σ^2 is the residual mean square error of Y.

2. To know the accuracy of the estimate of Y at an *additional* observation will have its own error variance σ^2. It is now possible to calculate *tolerance limits* for the prediction based on the variance

$$\sigma^2 \left[1 + \frac{1}{n} + \frac{(X_o - \bar{X})^2}{S_{xx}} \right] \quad \ldots\ldots\ldots\ldots \quad (10)$$

These analogies should clarify the following discussion of the problem.

(1) Predicting the angle (θ) at a known value strain (X_o).

X_o can now be regarded as a constant. From eq. (7), define

$$u = a_F + b_F X_o - a_L, \text{ thence } Var (\hat{\theta}) = Var_L \left(\frac{u}{b}\right) \ldots\ldots \quad (11)$$

This quantity can be approximated by the delta method (see, for example, Kendall and Stuart', Chapter 10), since all of a_F, b_F, a_L, b_L have variances inversely proportional to sample size. Now, approximately,

$$Var (\hat{\theta}) = \left(\frac{E(u)}{E(b_L)}\right)^2 \left(\frac{Var (u)}{E^2 (u)} - \frac{2 \text{ cov } (u, b_L)}{E(u)E(b_L)} + \frac{Var (b_L)}{E^2 (b_L)}\right) \ldots \quad (12)$$

All these quantities can be estimated; clearly

$$E(u) = a_F + \beta_F X_o - \alpha_L \quad \ldots\ldots\ldots\ldots \quad (13)$$

$$E(b_L) = \beta_L \quad \ldots\ldots\ldots\ldots \quad (14)$$

Further,

$$Var (b_L) = \sigma_L^2 / S_{\theta\theta} \text{ as usual} \quad \ldots\ldots\ldots\ldots \quad (15)$$

$$Cov\,(u, b_L) = -\,Cov\,(a_L, b_L) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (16)$$

Since (a_L, b_L) is dependent of (a_F, b_F)

$$Cov\,(u, b_L) = +\,\bar{\theta}\,\sigma_L^2 / S_{\theta\theta} \quad \ldots\ldots\ldots\ldots\ldots\ldots (17)$$

$$Var\,(u) = \text{Var}\,(a_F + b_F X_O) + Var\,(a_L) \ldots\ldots\ldots\ldots \quad (18)$$

$$Var\,(u) = \sigma_F^2 \left(\frac{1}{n_F} + \frac{(X_O - X)^2}{Sxx} \right) + \sigma_L^2\,\frac{\Sigma\theta^2}{n_L S\theta\theta} \quad \ldots\ldots\ldots\ldots (19)$$

Hence, confidence intervals for $\hat{\theta}$ are available (although somewhat complicated to compute!).

(2) Predicting θ for an additional measurement $X = X_O$.

Here X_O must be assumed to have its own error variance σ_F^2. If the calibrations are based on a fairly large sample of data (a sample size of 20 or more), then the error in X_O will dominate those of the regression parameters. Put slightly differently, the uncertainty about the additional observation will far outweigh the uncertainty about the rather accurately-determined regression lines. In this case, the variation in a_F, b_F, a_L, b_L can be neglected, and approximately

$$Var\,(\hat{\theta}) = \frac{b_F^2}{b_L^2}\,\sigma_F^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

In practical applications, case (2) above seems to be the more commonly appropriate; it has been used in making the calculations below.

Before the numerical values for $\hat{\theta}$ and $Var\,(\hat{\theta})$ are given for the problem, a further question arises: *Is it really necessary to use a different formula for the angle $\hat{\theta}$ for each year?*

It may be objected that this is wasteful of data, for if there were not 5, but 500 regression lines, and if it were known that they were not the same, then none of the information on the first 499 lines could be used for the five-hundredth line. This may be considered a disadvantage, but the difficulty is inherent in the particular problem, rather than in the analysis. If information from previous years is to be used to improve the prediction for this year, some strong assumptions are necessary about the fluctuations in the regression lines from year to year. Such an approach would be possible using Bayesian statistics, where the slopes (b_F) for different years might be regarded as random samples, from some distribution (perhaps normal). In such a case, the estimate of b_F for a particular year would be a compromise between the estimate based on the data for that particular year, and the overall mean of the b_F's for every year.

Such an approach has not been adopted here because the assumption seems unrealistic in this context, and because one can hardly estimate an underlying distribution from only four values (four years). But from the point of view of classical--as opposed to Bayesian--statistics, there is nothing to be said about using previous years' b_F values to estimate this year's,

For the following five years of data given (Table 1), the equations for estimating angle (θ) over the range found in the field measurements from measured strain in a given year (X), and the standard error for the angle are given in Table 4.

TABLE 4 -- The standard error for the angle.

1965:	θ =	**0.000046**	-	**0.000111** X	s.e. of θ =	**0.001**
1966:	θ =	**0.000945**	+	**0.000086** X		**0.001**
1967:	θ =	**-0.001700**	+	**0.000082** X		**0.002**
1968:	θ =	**-0.001380**	+	**0.000875** X		**0.001**
1970:	θ =	**0.000502**	+	**0.000085** X		**0.001**

CONCLUSION

Nearly all instruments commonly used for collection of laboratory or field data require some type of calibration to relate the instrument energy output (be it millivolts, ohms, electronic pulses, etc.) to the desired physical measurement (be it langleys of solar energy, soil moisture or density, inches of rainfall, height of streamflow, etc.). Once the basic instrument is calibrated in general terms, a second calibration is necessary for the reading to be useful in the specific case under study.

For example, once a streamflow recorder is calibrated so that a change in instrument voltage shows the depth of a river, a second calibration is required to relate river stage to volume of flow, which is based on the cross sectional area and velocity of the river.

Another example is the neutron soil moisture meter, in which the instrument is calibrated carefully by collecting repetitive neutron counts in tanks containing soil of known and uniform moisture content. Because it is very time-consuming and difficult to prepare tanks of uniform soil, a secondary calibration procedure is consistently used. In this case, tanks containing various concentrations of boric acid replace the tanks of soil. A relationship is then established between count rate in the soil tanks and the count rate in the boric acid tanks.

Numerous other examples can be found in which rather elaborate and complicated procedures are used to calibrate instruments. Too often, such calibrations are taken at face value. Seldom is any indication of the error associated with the calibration provided for the user--indeed, seldom is the calibration curve supplied with data points so that the user may assess the scatter of data about his calibration. As a result, many investigators have been perplexed to find that two instruments yield different numerical values when used to measure identical points.

The example of a statistical approach can be used to approximate the error associated with the calibration of instruments. The example further provides the standard statistical tests, which are used to determine whether a number of successive calibrations of the same instrument can be considered to be the same curve or whether they are different. Essentially, these tests involve determining if the lines have a common slope and a common intercept.

Field calibrations were based on regressions of strain measurements in 1972 on strain measurements in a previous year. A laboratory calibration was based on a regression of strain measurements in 1972 on angle. The procedure shows how to combine field calibrations with the laboratory calibration, thereby providing statistical inferences about the angle associated with the strain measured in previous years.

REFERENCES

1. Kojan, E. *Mechanics and rates of natural soil creep.* Proc. 5th Ann. Eng. Geology and Soils Engineers Symp., Pocatello, Idaho, April 1967, pp. 233-253.
2. Kallstenius, T. and W. Bergau. *In situ determination of horizontal ground movements,* 5th Int. Conf. on Soil Mech. and Found. Eng., 1961, pp. 481-485.
3. Ziemer, R. R. *Measurement of soil creep by inclinometer,* ETR 7100-4 Forest Service, U.S. Department of Agriculture, Washington, D. C., 1977, 10 pp.
4. Kendall, M. G. and A. Stuart. *Advanced Theory Of Statistics, Vol. I: Distribution Theory,* Hafner Pub. Co., New York, N. Y., 1969.
5. Kendall, M. G. and A. Stuart. *Advanced* Theory *of Statistics, Vol. II: Inference and Relationship,* Hafner Pub. Co., New York, N. Y., 1967.
6. Grosenbaugh, L. R. *REX--Fortran-4 System for combinatorial screening or conventional analysis of multivariate regressions.* U.S. Forest Service Res. Paper PSW-44, Pacific Southwest Forest and Range Expt. Stn., Berkeley, California, 1967, 47 pp.
7. Mood, A. M. *Introduction to the Theory of Statistics,* McGraw-hill, New York, N. Y., 1950.
8. Williams, E. J. *Regression Analysis,* Wiley, New York, N. Y., 1959.