

CHROM. 4857

STATISTICAL PROCESSING OF CALIBRATION DATA IN
QUANTITATIVE ANALYSIS BY GAS CHROMATOGRAPHY

P. BOČEK AND J. NOVÁK

Institute of Instrumental Analytical Chemistry, Czechoslovak Academy of Science, Brno (Czechoslovakia)

(Received June 2nd, 1970)

SUMMARY

The requirements that have to be met in the processing of calibration data by the linear regression method in quantitative gas chromatography are discussed and documented on four typical variants of the absolute calibration technique.

A mean slope method has been suggested for the precise processing of the data in cases where the statistically dependent variable displays a constant relative error. The effect of the systematic error on the course of the calibration line has been shown.

INTRODUCTION

Methods involving the work with a calibration curve are among the very popular procedures in quantitative analysis by gas chromatography. The most used methods are variants of the absolute calibration and internal standard techniques¹. The calibration curve is often constructed by merely interlacing the experimental points by rule of thumb; however, when more precise work is required, it is necessary to process the calibration data with the aid of statistics. While in the first case mentioned there are no special limitations, in the second case, it is necessary to consider carefully the character of the problem in hand and to match the procedure of obtaining the most plausible curve with the requirements dictated by statistics².

THEORETICAL

The plotting of a calibration line by statistically processing the experimental data may be characterized as the determination of the most plausible linear relation between a number of the values of a random variable y and a nonrandom variable x . Then, the relation obtained provides for the best estimation of a value Y of the variable y for a given precise value X of the variable x . From the statistical viewpoint, the above procedure is characterized as linear regression, and the calibration line obtained corresponds to the regression straight line. Most of the experimental techniques of quantitative gas chromatography (the standard addition technique is an exception¹) merely produce calibration lines that pass through the origin of the coordinate system.

To follow the correct procedure in the linear regression, it is necessary to decide

which of the variables involved should be considered as a dependent one and to ascertain the character of the variance of the above variable in the calibration range. From the statistical point of view, the variable which is looked upon as dependent is that one the values of which are substantially less precise than those of the other variable, which is considered as independent. However, this statistical concept has nothing in common with the actual relationship between the variables and, furthermore, is in no way connected with the order of obtaining the individual values, e.g. in the premise "given-found".

Hence, there are two situations which may occur in reading out the results of an analysis from the calibration line:

(1) *The result corresponds to the dependent variable y*

In this case, the value of the result required (Y) for a given precise value X is read out from the calibration line with a slope K according to the relation $Y = KX$. Thus, only the variance of the slope, S_K^2 , has any effect on the variance of the result, S_Y^2 , the latter being given by $S_Y^2 = X^2 S_K^2$.

(2) *The result corresponds to the independent variable x*

This case represents a situation when the value X is read out for a measured value Y from the calibration line corresponding to the relation $X = Y/K$, so that the variance of the final result of analysis is given by the relation $S_X^2 = (Y^2 S_K^2 / K^4) + S_Y^2 / K^2$. Hence it follows that the variance of the result comprises both the variance of the slope of calibration line and the variance of the measured value Y .

When considering the properties of the variance of the dependent variable, it is important to know whether the variance is constant within the calibration limits and whether its values are dependent on the value corresponding to the quantity measured. In quantitative gas chromatographic analysis, one is concerned most frequently with the following situations:

(a) *The dependent variable has a constant variance*

In this case, the calculation of the most plausible slope of the calibration line may be carried out by the least squares method. In order to use this the constancy of the variance of the dependent variable is a necessary condition (cf. ref. 2). The estimation of the calibration line slope is given by $K = \frac{\sum [(Y - \bar{Y})(X - \bar{X})]}{\sum (X - \bar{X})^2}$ where \bar{Y} and \bar{X} are the arithmetic averages of the respective values measured. The variance of the slope is given by $S_K^2 = (\sum Y^2 - K \sum XY) / (n - 2) \sum (X - \bar{X})^2$ where n is the number of experimental points. The expression $(\sum Y^2 - K \sum XY) / (n - 2)$ gives the variance of the dependent variable, S_Y^2 . This variance may be determined independently of the construction of the calibration line (cf. ref. 3), the variance of the slope being given by $S_K^2 = S_Y^2 / \sum (X - \bar{X})^2$.

With constant variance, the half width of the interval of confidence (i) is also constant, and the experimental points cover, with a probability of $1 - \alpha$, an area defined by the parallels $Y = K'X \pm i$ around the calibration line, i being given by $i = t_\alpha(v) S_Y$ where $t_\alpha(v)$ is the Student coefficient for the confidence level α and the number of degrees of freedom corresponding to the number of measurements used

for calculating the independent variance S_Y^2 . The value of K' is the correct value of the slope and may be substituted by the value of K , obtained from the above mentioned relation, when dealing with a large number of experimental data.

(b) *The dependent variable has a nonconstant variance*

In this case, the least square method is not appropriate to the determination of the slope of the regression line. However, if the dependent variable displays a constant relative error, which very frequently occurs in quantitative gas chromatography, it is possible to process the experimental data precisely by the method of mean slope. In this method, the slope of the calibration line plays the role of a random (dependent) variable with a constant variance, and the best estimation of it is the arithmetic mean. Hence, $K = (1/n) \sum'' (Y/X)$, and the variance of the slope is given by $S_K^2 = \frac{\sum'' [(Y/X) - K]^2}{n(n-1)}$.

If the relative error of the dependent variable is known, *e.g.*, if this error has been obtained independently of the calibration as the coefficient of variation I_Y , given by $I_Y = 100 S_Y/Y$ where S_Y is the standard deviation of the value Y , it is possible to write for the variance of the slope $S_K^2 = (1/n) (I_Y/100)^2 K^2$. In this case, the experimental points cover, with a probability of $1 - \alpha$, a divergent area defined by the straight lines $Y = K'(1 \pm i)X$ where $i = t_{\alpha}(v)I_Y/100$. The correct value of K' may again be substituted by the value of K calculated from the respective relationship.

When evaluating calibration lines by the methods mentioned in connection with situations (a) and (b), cases frequently occur in which the statistically dependent variable suffers from a significant systematic error, d . Though this error may manifest itself in different ways, two cases are the most frequent.

In the first case, the error manifests itself as a constant absolute error, *i.e.*, a measured value Y and the correct value Y' are related to each other by $Y = Y' + d$. The calibration line is most often obtained by the procedure described under (a) in this case. When following the above procedure, the value of the slope is not modified by the error d and, therefore, the calculated slope, K_1 , corresponds to the correct value, K' ; the error d will manifest itself as an intercept on the y co-ordinate, *i.e.*, the calibration line does not pass through the origin of the co-ordinate system. The d value may be calculated from the data necessary for constructing the calibration line by using the relation $d = \bar{Y} - K_1 \bar{X}$.

In the second case, the error d presents itself as a constant relative error, so that the relation between the measured and correct values may be expressed in the form $Y = Y'(1 + d)$. In this case, the calibration line is most often constructed by the method mentioned in connection with the situation under (b). It is characteristic for this procedure that the error d manifests itself by a modification of the slope of the calibration line; the relation between the calculated slope K_2 and the correct slope K' is as follows: $K_2 = K'(1 + d)$.

EXPERIMENTAL

The validity of the above relations was confirmed by comparing the spread of the experimental points about the regression line with the interval of confidence

calculated on the basis of the independently determined variance of the statistically dependent variable. The procedure was based on measuring the independent variances of all the variables occurring in the calibration; by comparing the relative errors of the given variables, it was ascertained which of the variables had associated with it a greater error and, consequently, would be considered as statistically dependent. This variable was always assigned the perpendicular co-ordinate (y).

In addition, an adequate number (40) of points was measured within the calibration limits investigated and, allowing for the character of the variance of the dependent variable, a suitable method (situations (a) or (b)) was employed to obtain the respective calibration line. By virtue of the independent variance, the respective interval of confidence was constructed around the calibration line and compared with the lay-out of the experimental points. The effect of the systematic error (d) on the calibration line was tested by comparing two calibration procedures in which the systematic error manifested itself either as the intercept on the y co-ordinate or as the change in the slope.

The above procedure has been carried out with the absolute calibration technique, in which the calibration line is obtained by plotting the peak areas or peak heights against the absolute amounts injected of the substance under analysis. The concentrations have been expressed in molarities (M), the volumes injected (v) in μl , the peak areas (A) in cm^2 , and the peak heights (h) in cm . The peak area was determined as the product of the peak height and the peak width at its half height (r). The experimental design covers four relevant cases, *i.e.*, four combinations of the possibilities that the final result is, from the statistical viewpoint, either a dependent (1) or an independent (2) variable, having either a constant (a) or a nonconstant (b) variance. The final results are expressed in the absolute amounts (Mv) of the test substance introduced. The effect of the systematic error has been tested in the case of the systematic error of the volume injected, which is the most frequent error in the method employed; the most usual source of the above error is the evaporation of part of the sample from the needle of the syringe used.

The measurement was carried out on a Becker Multigraph F, Model 410 (Becker Delft, N.V., Delft, The Netherlands) equipped with a Servogor RE 511 recorder (Goerz Electro, G.m.b.H., Austria). 1 m long aluminum columns, 4 mm I.D., with 4 g of 20 wt. % squalane on Celite 545 30–60 mesh were used, and kept at 60° . The N_2 , H_2 , and air flow rates were 0.80, 1.25, and 10 ml/sec, respectively, as measured at the detector jet nose (24° , 746 mm Hg). The column inlet excess pressure was 0.2 atm. The injection port was kept at 140° . The samples were introduced by a Hamilton 701-N (10 μl) syringe (Hamilton Co., Whittier, U.S.A.). The model mixtures were prepared by weighing chromatographically pure chloroform, benzene, and toluene (Lachema, N.E., Brno, Czechoslovakia) on a Meopta A3/100 analytical balance (Meopta, N.E., Prague, Czechoslovakia) with a precision better than 10^{-2} % of the value weighed.

The principle of the method used makes it necessary to take into account the variables M , h , r , A , and v as well as the respective variances, S_M^2 , S_h^2 , S_r^2 , S_A^2 , and S_v^2 (*cf.* ref. 3). The concentration (M) was determined by weighing and, therefore, the respective variance (S_M^2) may be neglected in comparison with the error of the other variables. When measuring the lengths h and r by the same gauge, the variances S_h^2 and S_r^2 are identical and will be designated by S_l^2 from now on. The measurement

of lengths was carried out by a rule; the respective variance (S_l^2) was determined by measuring 15 standard lengths and amounted to $4.0 \times 10^{-4} \text{ cm}^2$. The variance of the area calculated by $A = hr$ is given by $S_A^2 = (r^2 + h^2)S_l^2$ and its value has to be determined for particular values of h and r . The variance of measuring the volumes injected, S_v^2 , was determined, for the given 10 μl Hamilton syringe, by weighing 15 measured volumes of tetrabromoethane, and amounted to $16 \times 10^{-4} \mu\text{l}^2$.

RESULTS AND DISCUSSION

Variant 1(a)

The above variant corresponds to the case when the dependent variable represents the final result of analysis and displays a constant variance. In the absolute calibration technique, this situation occurs when the measurement of the sample charge suffers from the larger error and when the calibration points have been obtained by injecting different volumes of the same sample mixture. In our experiments,

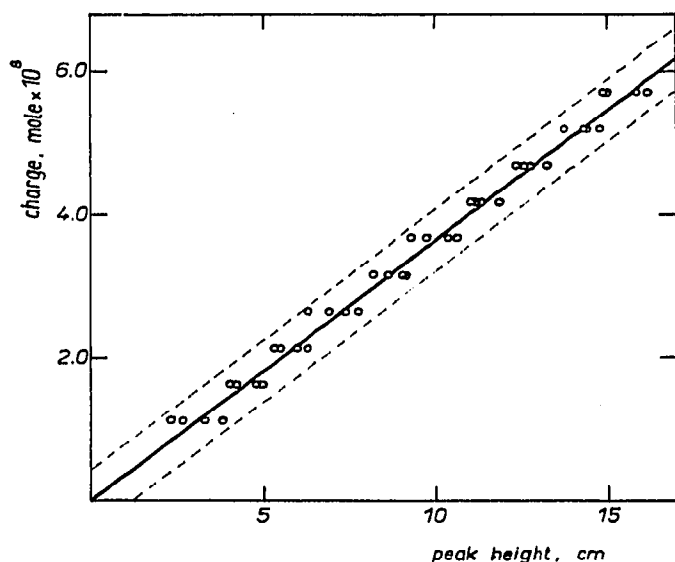


Fig. 1. Variant 1(a). Main source of error — measuring of the sample charge; calibration data obtained by injecting different volumes of a sample of a constant concentration of the substance analyzed; $K = 0.3634 \times 10^{-8} \text{ mole cm}^{-1}$.

a 0.05105 M solution of benzene in toluene was injected, the sample size ranging from 0.20 to 1.10 μl . The respective peak heights varied from 3 to 15 cm. The calibration line determined by the least square method is in Fig. 1. The dashed lines depict the interval of confidence calculated from the independent variance by $S_{Mv}^2 = M^2 S_v^2$. Hence, for the half width of the interval of confidence, i , $i = t_{0.05}(14)MS_v$, and the numerical value is $0.44 \times 10^{-8} \text{ mole}$ for $t_{0.05}(14) = 2.14$; the respective interval of confidence is within the parallels $Mv = Kh \pm 0.44 \times 10^{-8}$.

Variant 1(b)

This variant is partly analogous to the preceding one. The dependent variable again expresses the final result of the analysis, but the variance of the former is not constant. In this case, the relative error, *i.e.*, the coefficient of variation of the de-

pendent variable, is constant. In the absolute calibration technique, this variant again represents a case when it is the measurement of the volume injected that suffers from the larger error, but the calibration line is constructed from points obtained by injecting equal volumes of sample solutions of different concentrations. In the respective experiments, 1.1 μl charges of different solutions of benzene in toluene were

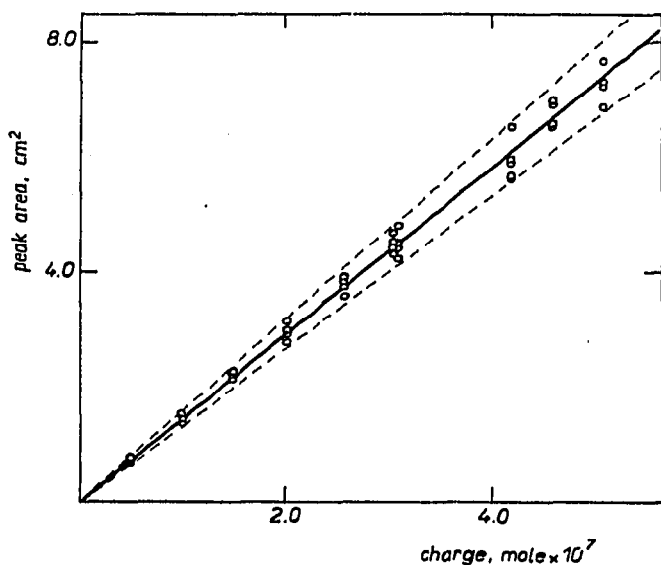


Fig. 2. Variant 1(b). Main source of error — measuring of the sample charge; calibration data obtained by injecting equal volumes of samples of different concentrations of the substance analyzed; $K = 0.3655 \times 10^{-8} \text{ mole cm}^{-1}$.

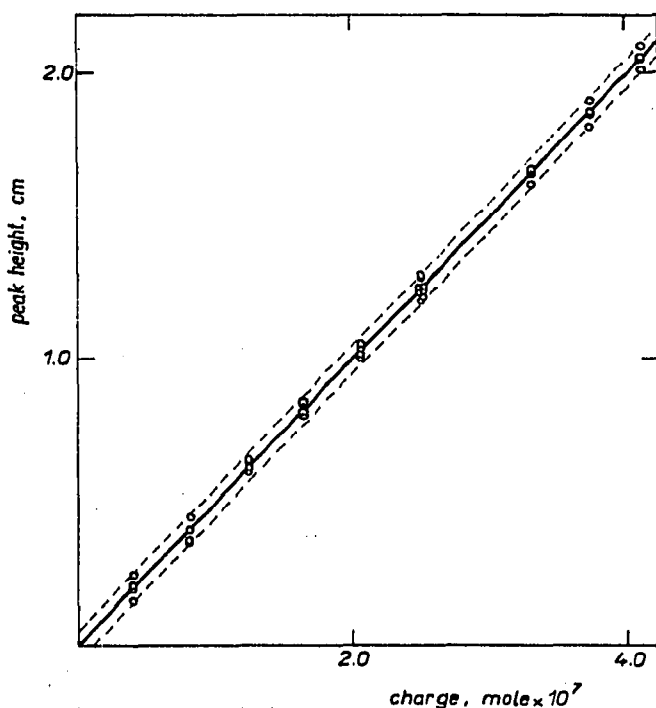


Fig. 3. Variant 2(a). Main source of error — measurement of the chromatogram; peak height is the quantitative parameter; $K = 0.4975 \times 10^7 \text{ cm mole}^{-1}$.

injected, the benzene peak heights varying from 3 to 15 cm. The calibration line obtained by the mean slope method is given in Fig. 2. The interval of confidence, given by the lines $Mv = K(1 \pm i)h$ where $i = t_{0.05}(14)I_{Mv}/100$ and I_{Mv} stands for the coefficient of variation of the amount injected, is designated by the dashed lines. The numerical value of I_{Mv} , calculated from the independent variance S_v^2 and for the volume injected, is 3.6%. The lines defining the interval of confidence are given by $Mv = K(1 \pm 0.076)h$.

Variant 2(a)

This variant represents a case when the final result of analysis, Mv , is the independent variable, and the dependent one has a constant variance. The corresponding variant of the absolute calibration technique is that in which the measurement of the chromatogram is associated with the larger error and has a constant variance. In our case, 10 μ l charges of solutions of chloroform in toluene were injected, the corresponding peak heights varying from 0.2 to 2 cm. The respective calibration line, obtained by the least squares method, is shown in Fig. 3. The interval of confidence is given by the parallels $h = KMv \pm i$ where $i = t_{0.05}(14)S_i$; substitution of the numerical values gives $h = KMv \pm 0.043$.

Variant 2(b)

In this case, the result of the analysis is the independent variable. The dependent variable has a nonconstant variance, but displays a constant relative error within the entire calibration range. In the absolute calibration technique, this variant corresponds to a case when the peak area is determined as the product hr and the measurement of the peak width suffers from the largest error ($r^2 \ll h^2$). Then, the variance of the dependent variable, S_A^2 , may be expressed by $S_A^2 = h^2S_i^2$, and the coefficients of variation of the variables A and r are constant and equal to each other. 10 μ l sample

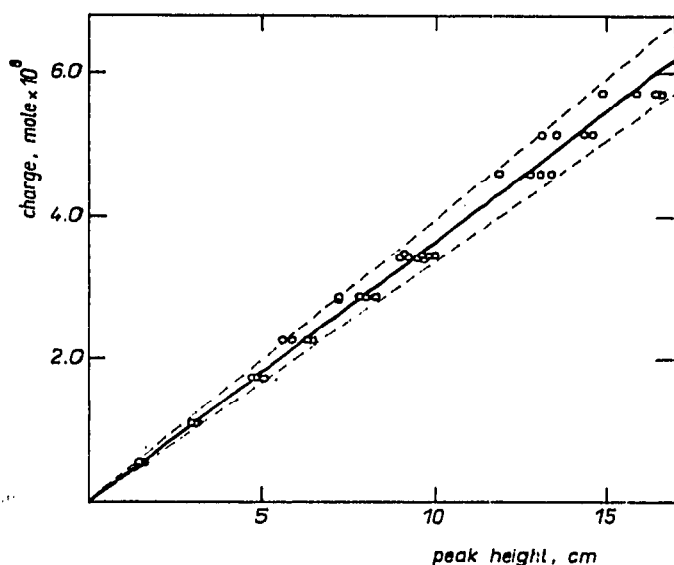


Fig. 4. Variant 2(b). Main source of error — measurement of the chromatogram; the product of the peak height and the peak width at the half height is the quantitative parameter; $K = 1.460 \times 10^7 \text{ cm}^2 \text{ mole}^{-1}$.

charges of different benzene-in-toluene solutions were injected. The benzene peak widths amounted to about 0.5 cm and the peak heights varied from 1.5 to 16 cm. The respective calibration line is shown in Fig. 4. The interval of confidence is given by the lines $A = Kmv(1 \pm i)$ where $i = t_{0.05}(14)I_r/100$. The value of I_r , expressed by means of S_i^2 for $r = 0.5$ cm, is 4%, so that the equations to the lines indicating the interval of confidence are given by $A = KMv(1 \pm 0.086)$.

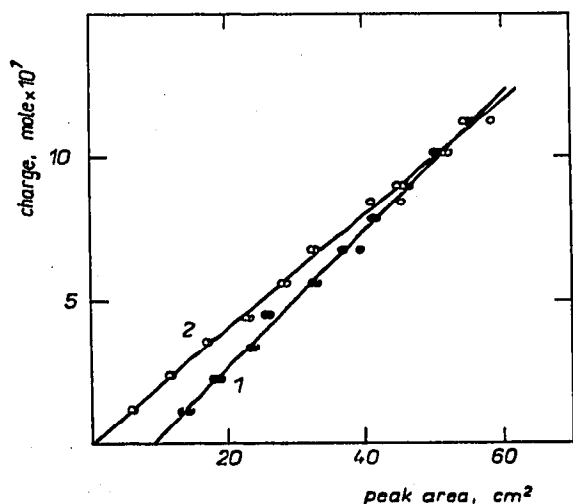


Fig. 5. Effect of the systematic error on the calibration line. Line 1: injection of different volumes of a sample of constant concentration, $K_1 = 1.990 \times 10^{-9}$ mole cm^{-2} , the intercept corresponds to 0.196 μl ; line 2: injection of equal volumes of samples of different concentrations, $K_2 = 2.399 \times 10^{-9}$ mole cm^{-2} .

Effect of the systematic error of sample introduction

To demonstrate the above effect, two procedures were employed for constructing the calibration line. In the first case, a 0.1119 M solution of isooctane in toluene was injected, the sample charges varying from 0.10 to 1.0 μl . The statistically dependent variable was the quantity Mv , the independent variable was the isooctane peak height, and the calculation was carried out by the procedure mentioned in connection with variant 1(a). In this case, the error d manifests itself as a constant systematic absolute error of the magnitude Md . The respective calibration line (No. 1) is shown in Fig. 5. The slope of the line is $K_1 = 1.990 \times 10^{-9}$ mole cm^{-2} and the intercept on the x co-ordinate corresponds to 2.20×10^{-8} mole. With the given sample solution, the systematic error incidental to the above intercept amounts to 0.196 μl .

In the second case, solutions of isooctane of different concentrations in toluene were injected, the sample volumes being 1.0 μl . In this case, the error d exerts itself as a constant systematic relative error. The calibration line, constructed according to the above 1(b) variant, is designated by numeral 2 in Fig. 5. The line passes through the origin of the co-ordinate system and its slope is $K_2 = 2.399 \times 10^{-9}$ mole cm^{-2} . The value of the slope K_2 may be calculated using the error d and the correct value of K_1 (equal to K') known from the preceding case; the value calculated ($K_2 = 2.380 \times 10^{-9}$ mole cm^{-2}) is in good agreement with the actual value (the difference is about 0.8%).

CONCLUSION

In processing calibration data by linear regression, it is necessary to proceed with regard to which of the variables involved is associated with the larger error; this variable has to be regarded, from the statistical viewpoint, as the dependent one. Linear regression may be carried out by the method of least squares only in the cases where the dependent variable displays a constant variance within the calibration limits. If the dependent variable with a nonconstant variance has a constant relative error, precise processing of the calibration data is possible by means of the mean slope method. The presence of a systematic error is manifested according to the working procedure used, either as a systematic constant absolute error or as a systematic constant relative error. In the first case, an intercept on the respective co-ordinate is incidental to the error whereas, in the second case, the error is responsible for a deviation in the slope of the calibration line.

REFERENCES

- 1 J. NOVÁK, in J. C. GIDDINGS AND R. A. KELLER (Editors), *Advances in Chromatography*, Marcel Dekker, New York, in press.
- 2 W. J. YOUDEN, *Statistical Methods for Chemists*, Wiley, New York, 1951.
- 3 P. BOČEK, J. NOVÁK AND J. JANÁK, *J. Chromatog.*, 42 (1969) 1.

J. Chromatog., 51 (1970) 375-383