INTERNATIONAL STANDARD

ISO
7066-2

First edition
1988-07-01

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ

Assessment of uncertainty in the calibration and use of flow measurement devices —

Part 2: Non-linear calibration relationships

Évaluation de l'incertitude dans l'étalonnage et l'utilisation des appareils de mesure du débit —

Partie 2: Relations d'étalonnage non linéaires

Reference number
ISO 7066-2 : 1988 (E)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 7066-2 was prepared by Technical Committee ISO/TC 30, *Measurement of fluid flow in closed conduits.*

Users should note that all International Standards undergo revision from time to time and that any reference made herein to any other International Standard implies its latest edition, unless otherwise stated.

Contents

Page

Assessment of uncertainty in the calibration and use of flow measurement devices —

Part 2: Non-linear calibration relationships

0 Introduction

The method of fitting a straight line to flow measurement calibration data and of assessing the uncertainty in the calibration are dealt with in ISO 7066-1. ISO 7066-2 deals with the case where a straight line is inadequate for representing the calibration data.

1 Scope and field of application

This part of ISO 7066 describes the procedures for fitting a quadratic, cubic or higher degree polynomial expression to a non-linear[1)] set of calibration data, using the least-squares criterion, and of assessing the uncertainty associated with the resulting calibration curve. It considers only the use of polynomials with powers which are integers.

Because it is generally not practicable to carry out this type of curve fitting and assessment of uncertainty without using a computer, it is assumed in this part of ISO 7066 that the user has access to one. In many cases it will be possible to use standard routines available on most computers; as an alternative the FORTRAN program listed in annex C may be used.

Examples of the use of these methods are given in annex D.

Extrapolation beyond the range of the data is not permitted.

Annexes A, B, C, D and E do not form integral parts of this part of ISO 7066.

2 References

ISO 5168, *Measurement of fluid flow — Estimation of uncertainty of a flow-rate measurement.*[2)]

ISO 7066-1, *Assessment of uncertainty in the calibration and use of flow measurement devices — Part 1: Linear calibration relationships.*[3)]

3 Definitions

For the purposes of this part of ISO 7066, the following definitions apply.

3.1 method of least squares: Technique used to compute the coefficients of a particular form of an equation which is chosen for fitting a curve to data. The principle of least squares is the minimization of the sum of squares of deviations of the data from the curve.

3.2 polynomial (function): For a variable x, a series of terms with increasing integer powers of x.

3.3 regression analysis: The process of quantifying the dependence of one variable on one or more other variables.

NOTE — Many of the available computer programs suitable for curve fitting have the word "regression" in the title. For the purposes of this part of ISO 7066, the terms regression and least squares may be regarded as interchangeable.

3.4 standard deviation: The positive square root of the variance.

3.5 variance: A measure of dispersion based on the mean of the squares of deviations of values of a variable from its expected value.

4 Symbols and abbreviations

b_j coefficient of x_j

C_{jb} element of the inverse matrix

1) These procedures are also suitable for a linear set of calibration data.

2) At present at the stage of draft. (Revision of ISO 5168 : 1978.)

3) At present at the stage of draft.

$e_r(\ )$ random uncertainty of variable contained in parentheses[1)]

$e_s(\ )$ systematic uncertainty of variable contained in parentheses[1)]

$e(\hat{y}_c)$ total uncertainty of calibration coefficient[1)]

g_j coefficient of jth orthogonal polynomial

m degree of polynomial

n number of data values

$p_j(x)$ jth orthogonal polynomial

$s(\ )$ experimental standard deviation of variable contained in parentheses

s_r residual standard deviation of data values about the curve

t Student's t

x the independent variable

x^* arbitrary specified value of x

$\bar{x}$ arithmetic mean of the data values x_i

x_i value of x at the ith data point

x_j jth independent variable (in multiple linear regression)

x_{ji} value of x_j at the ith data point

y the dependent variable

$\bar{y}$ arithmetic mean of the data values y_i

$\hat{y}$ value of y predicted by the equation of the fitted curve

y_i value of y at the ith data point

$\hat{y}_i$ value of $\hat{y}$ at $x = x_i$

ν number of degrees of freedom

5 Curve fitting

5.1 General

Before attempting polynomial curve fitting, consideration should be given to whether a simple transformation of the x variable or the y variable or both may effectively linearize the data to enable the straight line methods described in ISO 7066-1 to be used. Some appropriate transformations are suggested in ISO 7066-1.

If it is not possible to establish a straight line, then the objective is to find the degree and coefficients of the polynomial function which best represents a set of n pairs of (x_i, y_i) data values obtained from calibration. If, for example, a quadratic expression is chosen, the curve will be of the form

$$\hat{y} = b_0 + b_1x + b_2x^2 \qquad \ldots (1)$$

The general polynomial expression is

$$\hat{y} = b_0 + b_1x + \ldots + b_jx^j + \ldots + b_mx^m$$

or

$$\hat{y} = \sum_{j=0}^{m} b_jx^j \qquad \ldots (2)$$

By applying the least-squares criterion, the coefficients b_j are computed to minimize the sum of squares of deviations of the data points from the curve:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where $\hat{y}_i$ is the value predicted by equation (2) at $x = x_i$.

In some cases, the degree m of the polynomial will be predetermined; for example, it may be known from experience that the calibration data will be satisfactorily represented by a cubic ($m = 3$) expression. Otherwise, the degree of fit is chosen by increasing the degree until an optimum is achieved (see 5.3).

If in increasing the degree of fit beyond a moderate degree significant improvements in the fit, as described in 5.3, continue to occur, then it is likely that the functional dependence is not suitable for representation by a polynomial; further, if the equation fitted has too many terms, the curve may display spurious oscillations. A not uncommon example is data which are virtually constant over most of the x range, but which vary strongly close to one end of the range.

In such cases, it is appropriate to divide the range into sections (see ISO 7066-1) which either are linear or can be fitted by a low-degree polynomial. Alternatively, transforming one or both variables may lead to a linear or low-degree polynomial function; transforming the independent variable to its reciprocal $1/x$ will in some cases result in adequate linearity.

The least-squares methods described in this part of ISO 7066 may not be appropriate if the effect of the random uncertainty $e_r(x)$ of the data values x_i is not negligible in comparison with that of the random uncertainty $e_r(y)$ of the y values. As in ISO 7066-1, if the magnitude of the slope[2)] of the calibration curve is always less than one-fifth of $e_r(y)/e_r(x)$, the methods may be regarded as appropriate; where this does not apply the

1) In some International Standards, the symbols U and E have been used instead of e.

2) "Slope" here means the derivative $d\hat{y}/dx = b_1 + 2b_2x + \ldots$.

mathematical treatment is outside the scope of this part of ISO 7066. If therefore the normal practice in calibrating any particular meter is to plot the variables in such a way that the above condition does not hold, then either the conventional choice of abscissa and ordinate is to be reversed or this part of ISO 7066 cannot be used.

If either variable is transformed before fitting, then the uncertainties referred to above, and later (clause 6), relate to the new transformed variables. If, as a result of transforming the dependent variable, the random uncertainty $e_r(y)$ cannot be regarded as constant over the range, then a weighted least-squares method should be used. The weighted least-squares method is not described in this part of ISO 7066 but many computer library routines allow the data to be weighted.

5.2 Computational methods

Standard library routines for least-squares curve fitting are available on most computers. The method for fitting a straight line described in ISO 7066-1 is commonly known as linear or simple linear regression: the equivalent method for fitting a polynomial may be described as polynomial or curvilinear regression, which is a special type of multiple linear regression. Annex A gives further information on regression methods and how to use them.

As an alternative to the standard regression routines, the orthogonal polynomial method described in annex B may be used: this method is particularly suitable when the degree of fit is not known beforehand. Annex C lists an appropriate orthogonal polynomial computer program.

When a computer is not available and the x values are uniformly spaced, a finite-difference method (see annex E) may be used to provide a quick indication of what degree of fit may be appropriate to represent the data. The coefficients of a polynomial representing the data may also be calculated, but this will not be the least-squares polynomial. The calculation of uncertainty using this method is beyond the scope of this part of ISO 7066.

5.3 Selecting the optimum degree of fit

The optimum fit is determined by trying increasing values of the degree m, either up to a specified maximum or until no further significant improvement occurs. The residual standard deviation s_r should be computed for each degree (s_r is the square root of the residual variance) using the equation

$$s_r^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n - m - 1) \qquad \ldots (3)$$

where $\hat{y}_i$ is the value predicted by the polynomial expression [equation (2)] at $x = x_i$.

NOTE — s_r^2 is equivalent to the term $s^2(y,x)$ used in ISO 7066-1.

The degree m should always be much less than the number n of data points.

If the data are well represented by a polynomial of degree m, then s_r will decrease significantly until the degree m is reached; thereafter s_r will remain approximately constant. In general, however, the degree at which the decrease in s_r ceases to be significant is not obvious, and an objective test of significance should be used as an aid to finding the optimum degree of fit.

Increasing the degree from $m-1$ to m is regarded as providing a statistically significant improvement in the fit if the new coefficient b_m differs significantly from zero, i.e. if $b_m + t_{95}\, s(b_m)$ and $b_m - t_{95}\, s(b_m)$ (the 95 % confidence limits of b_m) do not include zero.

This condition may be expressed as

$$\left| \frac{b_m}{s(b_m)} \right| > t_{95}$$

where t_{95} is the Student's t value for the 95 % confidence level with $\nu = n - m - 1$.

The value of t_{95} as a function of the number of degrees of freedom ν can be computed from the following empirical equation:

$$t_{95} = 1{,}96 + 2{,}36/\nu + 3{,}2/\nu^2 + 5{,}2/\nu^{3{,}84} \qquad \ldots (4)$$

For the orthogonal polynomial coefficient g_m (see annex B), the condition is

$$\left| \frac{g_m}{s(g_m)} \right| > t_{95}$$

Expressions for the variances of the coefficients $s^2(b_m)$ and $s^2(g_m)$ are given in annex A and annex B respectively.

It is important to test the effect of increasing the degree at least one degree beyond that which first shows no significant improvement, since it is often the case that either only the odd terms or only the even terms produce a significant improvement.

From a statistical point of view, the highest degree which produces an improvement in the fit which is significant at the 95 % confidence level may be regarded as the optimum degree. However, before this degree is selected as providing the most suitable expression to represent the data, other factors should be considered. These factors include any knowledge of the expected shape of the curve, the desirability of having a functional form which is not too complex, the range which it is necessary to represent, and the accuracy which is sought.

In assessing these factors, it is always advisable to plot graphs showing the data and the possible curves; these graphs will also highlight other possible problems. For example, if the degree is too low, then the curve will fail to represent a real trend in the data, and the predicted value $\hat{y}$ may have a bias over some of the range. If the degree is too high, the curve may be fitting the scatter of the data rather than the underlying trend.

The examples given in annex D illustrate the application of some of these principles.

6 Uncertainty

The random component of the uncertainty, at the 95 % confidence level, of a predicted value $\hat{y}$, is given by

$$e_r(\hat{y}) = t_{95}\, s(\hat{y})$$

where $s(\hat{y})$ is the square root of the variance $s^2(\hat{y})$ of $\hat{y}$. Expressions for $s^2(\hat{y})$ are given in annexes A and B; in general, $s^2(\hat{y})$ may be expressed as a polynomial function of x of degree $2m$. It is important to ensure that enough significant figures are used in the computation of $s^2(\hat{y})$ to avoid large rounding errors which result from subtraction.

It should be noted that the estimate of uncertainty provided by $e_r(\hat{y})$ will only be valid to the extent that the polynomial expression chosen is a good approximation to the true functional relationship between y and x.

The 95 % random confidence limits for the true value of y are

$$y \pm e_r(\hat{y})$$

As in ISO 7066-1, the uncertainty in the calibration coefficient is given by

$$e(\hat{y}_c) = \left[e_r^2(\hat{y}) + e_s^2(\hat{y}) \right]^{1/2}$$

where $e_s(\hat{y})$ is the systematic component of the uncertainty in $\hat{y}$.

NOTE — In the revised version of ISO 5168, in preparation, guidelines are provided for using either the linear addition or the root-sum-square combination of random and systematic errors.

If the dependent variable has been transformed, then all the above uncertainties refer to the transformed variable.

Annex A

Regression methods

(This annex does not form an integral part of the standard.)

A.1 Introduction

Regression methods for curve fitting are widely available under various names as standard routines in computer libraries. The documentation provided with these routines tends to assume a certain level of knowledge of regression analysis. The purpose of this annex is to provide a general description of the methods and terminology of regression curve fitting as a background to the documentation of the library routines.

The most widely available regression technique, apart from simple linear regression, is multiple linear regression; curve fitting can be carried out using a special type of multiple linear regression known as polynomial or curvilinear regression. If a polynomial regression routine is not available, then a multiple linear regression method can be used, although it is less convenient. "Stepwise" and "backwards elimination" or "back solution" are special types of multiple linear regression methods which may be used.

A.2 Multiple linear regression

In the following, the summation sign Σ is used to represent $\sum_{i=1}^{n}$ unless otherwise noted.

A dependent variable y is assumed to be related linearly to m independent variables $x_1, x_2, \ldots, x_m$ by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + U \qquad \ldots (5)$$

where

β_0 to β_m are the unknown regression coefficients;

U is a measure of the random effects which cause the dependence of y on the m independent variables to depart from exact linearity.

From the n sets of observations

$$(y_i, x_{1i}, x_{2i}, \ldots, x_{mi}), \quad i = 1, 2, \ldots, n$$

the estimates of the regression coefficents are

$$b_0, b_1, \ldots, b_m$$

so that the estimate $\hat{y}$ of the true value corresponding to the ith set of observations of the independent variables is

$$\hat{y}_i = b_0 + b_1 x_{1i} + \ldots + b_m x_{mi} \qquad \ldots (6)$$

The application of the least-squares procedure to minimize $\Sigma(y_i - \hat{y}_i)^2$ leads to a set of $m + 1$ simultaneous equations, commonly known as the "normal equations":

$$\begin{aligned} nb_0 + \Sigma(x_{1i})b_1 + \Sigma(x_{2i})b_2 + \ldots + \Sigma(x_{mi})b_m &= \Sigma y_i \\ \Sigma(x_{1i})b_0 + \Sigma(x_{1i})^2 b_1 + \ldots + \Sigma(x_{1i}x_{mi})b_m &= \Sigma(x_{1i}y_i) \\ \Sigma(x_{mi})b_0 + \Sigma(x_{mi}x_{1i})b_1 + \ldots + \Sigma(x_{mi})^2 b_m &= \Sigma(x_{mi}y_i) \end{aligned} \qquad \ldots (7)$$

These can then be solved for the $m + 1$ unknowns $b_0, b_1, \ldots, b_m$.

A.3 Polynomial (curvilinear) regression

When a relationship between two variables is not linear, but may be fitted by a polynomial function

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \dots + b_m x^m$$

there is said to be a polynomial or curvilinear regression of y on x. This can be treated as a multiple linear regression with the independent variables $x_1, x_2, \dots, x_m$ replaced by $x, x^2, \dots, x^m$.

In clauses A.4 and A.5, any of the multiple linear regression expressions may be transformed to the equivalent polynomial regression expressions by replacing the jth independent variable x_j by x^j, and the corresponding data values x_{ji} by x_i^j.

A.4 Computation of coefficients and variances

Consider the multiple linear regression equation with $m = 2$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \qquad \dots (8)$$

which is equivalent to

$$\hat{y} = b_0 + b_1 x + b_2 x^2 \qquad \dots (9)$$

in the polynomial regression case.

When the least-squares criterion is applied, the normal equations are

$$n b_0 + \Sigma(x_{1i}) b_1 + \Sigma(x_{2i}) b_2 = \Sigma(y_i) \qquad \dots (10)$$

$$\Sigma(x_{1i}) b_0 + \Sigma(x_{1i})^2 b_1 + \Sigma(x_{1i} x_{2i}) b_2 = \Sigma(x_{1i} y_i) \qquad \dots (11)$$

$$\Sigma(x_{2i}) b_0 + \Sigma(x_{2i} x_{1i}) b_1 + \Sigma(x_{2i})^2 b_2 = \Sigma(x_{2i} y_i) \qquad \dots (12)$$

The traditional method for solving the normal equations involves computing the inverse of the 3×3 matrix of coefficients of b_0, b_1 and b_2. If the elements of this inverse matrix are

$$\begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

then

$$\begin{aligned} b_0 &= C_{00} \Sigma y_i + C_{01} \Sigma(x_{1i} y_i) + C_{02} \Sigma(x_{2i} y_i) \\ b_1 &= C_{10} \Sigma y_i + C_{11} \Sigma(x_{1i} y_i) + C_{12} \Sigma(x_{2i} y_i) \qquad \dots (13) \\ b_2 &= C_{20} \Sigma y_i + C_{21} \Sigma(x_{1i} y_i) + C_{22} \Sigma(x_{2i} y_i) \end{aligned}$$

or, in generalized form,

$$b_j = \sum_{k=0}^{m} \left[C_{jk} \Sigma(x_{ki} y_i) \right]$$

where $x_{ki} = 1$ for $k = 0$.

Note that since the matrix from the normal equations is symmetric, the inverse matrix is also symmetric.

The variances of the coefficients are

$$s^2(b_0) = s_r^2 C_{00}$$
$$s^2(b_1) = s_r^2 C_{11}$$
$$s^2(b_2) = s_r^2 C_{22}$$

where the residual variance, s_r^2, is given as in 5.3 by

$$s_r^2 = \frac{\Sigma(y_i - \hat{y}_i)^2}{n - m - 1}$$

Because the inverse matrix is symmetric,

$$\begin{aligned} C_{01} &= C_{10} \\ C_{02} &= C_{20} \\ C_{12} &= C_{21} \end{aligned} \qquad \ldots (14)$$

These non-diagonal terms are used to calculate the covariances[1] between the coefficients b_j; using COV to denote covariance,

$$\begin{aligned} \mathrm{COV}(b_0, b_1) &= s_r^2 C_{01} \\ \mathrm{COV}(b_0, b_2) &= s_r^2 C_{02} \\ \mathrm{COV}(b_1, b_2) &= s_r^2 C_{12} \end{aligned} \qquad \ldots (15)$$

At specified values $x_1 = x_1^*$ and $x_2 = x_2^*$, the value predicted by the regression equation is

$$\hat{y} = b_0 + b_1 x_1^* + b_2 x_2^* \qquad \ldots (16)$$

The variance of this value of $\hat{y}$ is given by

$$s^2(\hat{y}) = s_r^2 \left[C_{00} + C_{11}(x_1^*)^2 + C_{22}(x_2^*)^2 + 2C_{01}x_1^* + 2C_{02}x_2^* + 2C_{12}x_1^*x_2^* \right] \qquad \ldots (17)$$

The factor of 2 arises because $C_{jk} = C_{kj}$ for each j, k.

The general formula is

$$s^2(\hat{y}) = s_r^2 \sum_{j=0}^{m} \sum_{k=0}^{m} (C_{jk} x_j^* x_k^*) \qquad \ldots (18)$$

where x_j^*, $x_k^* = 1$ for j, $k = 0$.

For polynomial regression, $x_j^* = (x^*)^j$ and $x_k^* = (x^*)^k$, and so

$$s^2(\hat{y}) = s_r^2 \sum_{j=0}^{m} \left[\sum_{k=0}^{m} C_{jk}(x^*)^{j+k} \right]$$

Adapting this expression to the form of a polynomial of degree $2m$ gives

$$s^2(\hat{y}) = s_r^2 \sum_{j=0}^{m} \left[\left(\sum_{k=0}^{j} C_{k,j-k} \right) (x^*)^j \right] + s_r^2 \sum_{j=m+1}^{2m} \left[\left(\sum_{k=j-m}^{m} C_{k,j-k} \right) (x^*)^j \right] \qquad \ldots (19)$$

1) The covariance of two coefficients indicates the effect of a change in one on the magnitude of the other. The inverse matrix multiplied by the scalar s_r^2 is known as the variance, covariance, or variance-covariance matrix.

A.5 Centred formulation

The least-squares or regression analysis is sometimes expressed in "centred" form, in which each variable is replaced by its deviation from its mean. In this form, equation (8) is replaced by

$$\hat{y} - \bar{y} = b_1 (x_1 - \bar{x}_1) + b_2 (x_2 - \bar{x}_2) \qquad \dots (20)$$

where the bar over a symbol is used to denote the mean value of the quantity represented by the symbol for the n measurements.

A.6 Numerical techniques used in computer libraries

For the least-squares or regression computations discussed in this annex, a computer library routine may make use of one of a variety of numerical techniques. The main numerical techniques used by computers for regression and least-squares matrix manipulations are

a) Gauss or Gauss-Jordan elimination,

b) Cholesky decomposition, and

c) orthogonal decompositions (usually Householder or modified Gram-Schmidt).

The particular technique used is in general not of importance to the user. However, it should be noted that elimination methods are susceptible to the build-up of rounding error, so that the computed coefficients b_j may be significantly in error for a high-degree polynomial; for a moderate degree, up to $m = 3$ or 4, this should not be a problem.

Annex B

Orthogonal polynomial curve fitting

(This annex does not form an integral part of the standard.)

This annex describes the main features of orthogonal polynomial curve fitting in relation to the regression methods discussed in annex A. Orthogonal polynomial curve fitting is particularly efficient when the degree of fit is unknown, and it is not subject to the rapid build-up of rounding error which can occur with elimination methods (see annex A, clause A.6).

The results of orthogonal polynomial curve fitting will be identical, apart from rounding error, to those produced by the regression methods described in annex A.

Computer library routines using orthogonal polynomials do not in general provide enough information to allow uncertainty to be easily computed: the program listed in annex C, however, provides full information on uncertainty.

In the following, the summation sign Σ is used to represent $\sum_{i=1}^{n}$ unless otherwise noted.

With the orthogonal polynomial method, the polynomial

$$\hat{y} = b_0 + b_1x + b_2x^2 + \dots + b_mx^m$$

is replaced by an equivalent form

$$\hat{y} = g_0p_0(x) + g_1p_1(x) + g_2p_2(x) + \dots + g_mp_m(x) \qquad \dots (21)$$

where

$p_j(x)$ are polynomials of degree j which obey for all $j \neq k$ the orthogonality condition

$$\Sigma\left[p_j(x_i)p_k(x_i)\right] = 0 \qquad \dots (22)$$

$$p_0(x) = 1$$

These polynomials are described as orthogonal over the data points x_i; the coefficients which define them are derived using a three-term recurrence relation, given by Forsythe[1].

Because of the orthogonality condition, all the elements in the matrix and inverse matrix derived from the normal equations (see clause A.4), except for those on the diagonal ($j = k$), are zero, and the coefficients g_j are obtained directly from the normal equations as

$$g_j = \frac{\Sigma[y_ip_j(x_i)]}{\Sigma[p_j(x_i)]^2} \qquad \dots (23)$$

The variances of the coefficients are obtained from the inverse matrix elements, as in annex A:

$$s^2(g_j) = s_r^2C_{jj} = \frac{s_r^2}{\Sigma[p_j(x_i)]^2} \qquad \dots (24)$$

At a specified value $x = x^*$, since the covariances are zero,

$$s^2(\hat{y}) = s^2(g_0) + [p_1(x^*)]^2s^2(g_1) + \dots + [p_m(x^*)]^2s^2(g_m)$$

$$= \frac{s_r^2}{n} + s_r^2\sum_{j=1}^{m}\frac{[p_j(x^*)]^2}{\Sigma[p_j(x_i)]^2} \qquad \dots (25)$$

Because the coefficients g_j are computed simply from equation (23), as the degree of fit is increased, the previous coefficients are unchanged: it is this feature that makes orthogonal polynomial curve fitting particularly convenient when the degree of fit is not known beforehand. When the optimum degree has been finally chosen, however, it is necessary to convert the orthogonal polynomial form for $\hat{y}$ [equation (21)]

$$\hat{y} = g_0 p_0(x) + g_1 p_1(x) + g_2 p_2(x) + \dots + g_m p_m(x)$$

to the more convenient simple power series

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \dots + b_m x^m$$

using the coefficients defining the orthogonal polynomials derived using the original recurrence relation.

Bibliography

[1] FORSYTHE, G.E. Generation and use of orthogonal polynomials for data fitting with a digital computer, *J. Soc. Ind. Appl. Maths*, *5* (2) (1957), pp. 14-88.

Annex C

Computer program using orthogonal polynomials

(This annex does not form an integral part of the standard.)

C.1 Input and output

The program operates interactively, requesting input as required. The initial input is the maximum degree of fit (which should be at least two degrees higher than is expected to be required), the number of data values, and the data, which are entered as one pair of (x_i, y_i) values per line.

The residual standard deviation s_r and the percentage significance are then printed for each degree up to the maximum, and the highest degree for which the coefficient is significant at the 95 % confidence level is suggested as the optimum degree of fit.

The user then selects and enters a degree of fit, which may be different from the suggested optimum degree, and the coefficients of the least-squares polynomial and the coefficients of the polynomial for the square of the random uncertainty are then printed. Finally, a table of data values (x_i, y_i), predicted values $\hat{y}_i$, deviations $y_i - \hat{y}_i$, and random uncertainties $t_{95}s(\hat{y})$ is printed. Another degree may then be entered, or -1 may be entered to terminate the execution.

C.2 Program description

After the input has been entered, the data are fitted using subroutine ORFIT up to a maximum degree MAXD1; the percentage significance of each coefficient is obtained from function PCTSQ. With a specified degree JDEG1, subroutine POWSER is used to compute the coefficients POLCO of the least-squares polynomial and the coefficients UVCO of the polynomial for the un-normalized variance $[s^2(\hat{y})/s_r^2$ from equation (25)]. With JDEG = JDEG1 + 1, t_{95} is computed at N − JDEG degrees of freedom using equation (4), and s_r^2 from D(JDEG) / (N − JDEG). The coefficients UVCO are then multiplied by $t^2 s_r^2$ to obtain USQCO, which are the coefficients of the polynomial representing the square of the random uncertainty.

The code used is that of standard FORTRAN IV except for the use of the arc cosine function ACOS in PCTSQ.

C.3 Possible modifications

The program can be used as listed, but in general it will be more convenient to make some modifications, particularly to the input and output.

Most implementations of FORTRAN allow data to be input in free format; this is more convenient than the fixed format required by standard FORTRAN.

The output provided by the listed program is for illustration purposes only; the most useful way of presenting the output will depend on what output devices are to be used. If a plotting or graphics device is available, then a plot which includes the curve, the data values and the 95 % confidence limits $\hat{y} \pm e_r(\hat{y})$, as illustrated in annex D, can be produced. If no such device is available, a printer can be used to give an approximate plot of, for example, the deviations $y_i - \hat{y}_i$ of the data from the curve.

The number of data values allowed in the listed program is 100, and the maximum permitted degree of fit m_{max} is 7. At degrees of fit above about 7, the computation of random uncertainty from the polynomial coefficients USQCO may be subject to large rounding errors. Note that the arrays A, B, G, D, E and POLCO are dimensioned $m_{max} + 1$, and UVCO and USQCO are dimensioned $2m_{max} + 1$.

ORTHOGONAL POLYNOMIAL COMPUTER PROGRAM

```
C     ORTHOGONAL POLYNOMIAL CURVE-FITTING  -  MAIN PROGRAM;
C     SUBROUTINES ORFIT AND POWSER, AND FUNCTION PCTSQ, ARE REQUIRED.
C
      DOUBLE PRECISION A,B,G,D,E,FAC,POLCO,USQ,USQCO,UVCO,X,Y,YPOL
      DIMENSION A(8), B(8), G(8), D(8), E(8), POLCO(8)
      DIMENSION X(100), Y(100), UVCO(15), USQCO(15)
C
C ARRAYS:
C
C  A  ALPHA COEFFICIENTS IN ORTHOGONAL POLYNOMIAL RECURRENCE RELATION
C  B  BETA COEFFICIENTS IN ORTHOGONAL POLYNOMIAL RECURRENCE RELATION
C  G  COEFFICIENTS OF ORTHOGONAL POLYNOMIAL SERIES
C  D  RESIDUAL SUM OF SQUARES
C  E  SQUARE OF COEFFICIENT G/VARIANCE OF G, FOR SIGNIFICANCE TESTING
C
C  POLCO  COEFFICIENTS OF SIMPLE POLYNOMIAL FOR Y
C  UVCO   COEFFICIENTS OF POLYNOMIAL FOR UNNORMALISED VARIANCE OF Y
C  USQCO  COEFFICIENTS OF POLYNOMIAL FOR SQUARE OF RANDOM UNCERTAINTY
C
C                    ***** INITIAL INPUT *****
C
      WRITE (6,120)
      READ (5,130) MAXD1
      IF (MAXD1.GT.7) MAXD1=7
      WRITE (6,140)
      READ (5,150) N
      IF (N.LE.100) GO TO 10
      WRITE (6,160)
      GO TO 110
   10 WRITE (6,170) N
      MAXD=MAXD1+1
      IF (MAXD.GT.N) MAXD=N
      DO 20 I=1,N
   20   READ (5,180) X(I),Y(I)
C
C                 ***** PRELIMINARY FITTING *****
C
      CALL ORFIT (X,Y,A,B,G,D,E,N,MAXD)
      WRITE (6,190)
      JOPT=0
      DO 30 J=1,MAXD
        IF (J.GE.N) GO TO 40
        J1=J-1
        SD=DSQRT(D(J)/FLOAT(N-J))
        SE=E(J)
        PC=PCTSQ(SE,N-J)
C     PC IS PERCENTAGE SIGNIFICANCE OF COEFFICIENT
        IF (PC.GE.95.) JOPT=J1
   30   WRITE (6,200) J1,SD,PC
   40 WRITE (6,210) JOPT
C
```

ORTHOGONAL POLYNOMIAL COMPUTER PROGRAM (CONTINUED)

```
C      ***** INFORMATION FOR A SPECIFIED DEGREE OF FIT *****
C
C     ENTER DEGREE
C
   50 WRITE (6,220)
      READ (5,130) JDEG1
      IF (JDEG1.LT.0) GO TO 110
      JDEG=JDEG1+1
      IF (JDEG.LE.MAXD) GO TO 60
      WRITE (6,230)
      GO TO 50
C
C     COMPUTE POWER SERIES (SIMPLE POLYNOMIAL) COEFFICIENTS
C
   60 CALL POWSER (A,B,G,JDEG,N,POLCO,UVCO)
C
      WRITE (6,240)
      WRITE (6,250) (POLCO(J),J=1,JDEG)
C
C     COMPUTE NORMALISING FACTOR FOR SQUARE OF RANDOM UNCERTAINTY
C     FROM RECIPROCAL OF DEGREES OF FREEDOM RDF, RESIDUAL SUM OF SQUARES
C     IN D, AND EMPIRICAL EQUATION FOR STUDENT T
C
      RDF=1./FLOAT(N-JDEG)
      FAC=D(JDEG)*RDF*(1.96+2.36*RDF+3.2*RDF**2+5.2*RDF**3.84)**2
      MDEG=2*JDEG-1
      DO 70 J=1,MDEG
   70   USQCO(J)=UVCO(J)*FAC
      WRITE (6,260)
      WRITE (6,250) (USQCO(J),J=1,MDEG)
C
C     TABULATE DATA VALUES, DEVIATIONS AND UNCERTAINTY
C
      WRITE (6,270)
C
      DO 100 I=1,N
        YPOL=0.0D0
        DO 80 J=1,JDEG
          JJ=JDEG+1-J
   80     YPOL=YPOL*X(I)+POLCO(JJ)
        USQ=0.0D0
        DO 90 J=1,MDEG
          JJ=MDEG+1-J
   90     USQ=USQ*X(I)+USQCO(JJ)
        YDEV=Y(I)-YPOL
        RUNC=0.0
        IF (USQ.GT.0.0D0) RUNC=DSQRT(USQ)
        XX=X(I)
        YY=Y(I)
        YP=YPOL
  100   WRITE (6,280) XX,YY,YP,YDEV,RUNC
      GO TO 50
  110 WRITE (6,290)
      STOP
```

ORTHOGONAL POLYNOMIAL COMPUTER PROGRAM (CONTINUED)

```
C
  120 FORMAT (1H0,32HENTER MAXIMUM DEGREE OF FIT .(I2))
  130 FORMAT (I2)
  140 FORMAT (33H ENTER NUMBER OF DATA VALUES (I3))
  150 FORMAT (I3)
  160 FORMAT (21H TOO MANY DATA POINTS)
  170 FORMAT (7H ENTER ,I3,29H PAIRS OF (X,Y) VALUES (2F10))
  180 FORMAT (2F10.5)
  190 FORMAT (1H0,37HDEGREE  RESIDUAL STANDARD  PERCENTAGE/12X,27HDEVIAT
     1ION      SIGNIFICANCE)
  200 FORMAT (I5,G18.6,F13.2)
  210 FORMAT (18H SUGGESTED DEGREE-,I2)
  220 FORMAT (1H0,32HENTER DEGREE (I2), OR -1 TO EXIT)
  230 FORMAT (16H DEGREE TOO HIGH)
  240 FORMAT (59H POLYNOMIAL COEFFICIENTS, LISTED IN INCREASING POWERS O
     1F X-)
  250 FORMAT (4G16.8)
  260 FORMAT (47H COEFFICIENTS FOR SQUARE OF RANDOM UNCERTAINTY-)
  270 FORMAT (1H0,10X,4HDATA,10X,36HPOLYNOMIAL    RESIDUAL   RANDOM UNC
     1/60H     X           Y           Y        Y - Y(POL)   OF Y(POL))
  280 FORMAT (4G12.5,G12.4)
  290 FORMAT (1H0,17H END OF EXECUTION)
C
      END
```

```
      REAL FUNCTION PCTSQ (TSQ,NU)
C
C     TSQ CONTAINS THE RATIO OF THE SQUARE OF A COEFFICIENT TO ITS
C     VARIANCE (CORRESPONDING TO THE SQUARE OF THE STUDENT T): PCTSQ
C     IS THE PERCENTAGE LEVEL AT WHICH THE COEFFICIENT CAN BE SAID TO
C     DIFFER SIGNIFICANTLY FROM ZERO.
C
      ANU=FLOAT(NU)
      X=ANU/(TSQ+ANU)
      RTX=SQRT(X)
      NUODD=NU-NU/2*2
      SUM=0.
      IF (NU.EQ.1) GO TO 30
      TERM=1.
      DO 10 J=2,NU,2
        IF (SUM.GT.TERM*1.E10) GO TO 20
        SUM=SUM+TERM
   10   TERM=TERM*X*(1.-1./FLOAT(J+NUODD))
   20 SUM=SUM*SQRT(1.-X)
   30 IF (NUODD.GT.0) SUM=0.636619772*(ACOS(RTX)+SUM*RTX)
      PCTSQ=100.*SUM
      RETURN
C
      END
```

ORTHOGONAL POLYNOMIAL COMPUTER PROGRAM (CONTINUED)

```
      SUBROUTINE ORFIT (X,Y,A,B,G,D,E,N,MAX)
C
C     METHOD FROM G.E.FORSYTHE, 'GENERATION AND USE OF ORTHOGONAL
C     POLYNOMIALS FOR DATA FITTING WITH A DIGITAL COMPUTER',
C     J.S.I.A.M., VOL 5, 2, JUNE 1957, PP 74-88
C
      DOUBLE PRECISION X,Y,P,A,B,G,D,E,Q,R,S,SA,SB,SG,SD,RN
      DIMENSION X(N), Y(N), P(100), Q(100), R(100)
      DIMENSION A(MAX), B(MAX), G(MAX), D(MAX), E(MAX)
C
      SA=0.0D0
      SG=0.0D0
      SD=0.0D0
      RN=1.0D0/N
      DO 10 I=1,N
        P(I)=1.0D0
        Q(I)=0.0D0
        SA=SA+X(I)
        SG=SG+Y(I)
   10   SD=SD+Y(I)*Y(I)
C
      A(1)=SA*RN
      B(1)=0.0D0
      G(1)=SG*RN
      D(1)=SD-G(1)*SG
      E(1)=1.0D20
      IF (D(1).GT.0.0D0) E(1)=G(1)*SG*(N-1)/D(1)
      SD=N
      J=1
C
   20 IF (J.GE.MAX) RETURN
      SA=0.0D0
      SB=0.0D0
      SG=0.0D0
      DO 30 I=1,N
        R(I)=Q(I)
        Q(I)=P(I)
        P(I)=(X(I)-A(J))*Q(I)-B(J)*R(I)
        S=P(I)*P(I)
        SA=SA+X(I)*S
        SB=SB+S
   30   SG=SG+Y(I)*P(I)
C
      J=J+1
      A(J)=SA/SB
      B(J)=SB/SD
      G(J)=SG/SB
      D(J)=D(J-1)-G(J)*G(J)*SB
      E(J)=1.0D20
      IF (D(J).GT.0.0D0) E(J)=G(J)*SG*(N-J)/D(J)
      SD=SB
      GO TO 20
C
      END
```

ORTHOGONAL POLYNOMIAL COMPUTER PROGRAM (CONTINUED)

```
      SUBROUTINE POWSER (A,B,G,MAX,N,COEF,UVCO)
      DOUBLE PRECISION A,B,G,COEF,F,H,UVCO,SCO,VCO
      DIMENSION A(MAX), B(MAX), G(MAX), COEF(MAX), UVCO(15), F(8,8)
C
C     INITIALISE
C
      DO 10 J=1,MAX
        DO 10 L=J,MAX
   10     F(L,J)=0.0D0
      F(1,1)=1.0D0
      F(1,2)=-A(1)
      F(2,2)=1.0D0
      K=MAX-1
      IF (K.LT.2) GO TO 30
C
C     USING THE RECURRENCE RELATION, COMPUTE THE COEFFICIENTS
C     F(L,J) OF THE J-TH ORTHOGONAL POLYNOMIAL
C
      DO 20 J=2,K
        H=0.0D0
        JJ=J+1
        DO 20 L=1,JJ
          F(L,JJ)=H-F(L,J)*A(J)-F(L,J-1)*B(J)
   20     H=F(L,J)
C
C     POLYNOMIAL COEFFICIENTS FOR Y
C
   30 DO 40 L=1,MAX
        COEF(L)=0.0D0
        DO 40 J=L,MAX
   40     COEF(L)=COEF(L)+F(L,J)*G(J)
C
C     POLYNOMIAL COEFFICIENTS FOR UNNORMALISED VARIANCE OF Y
C
      MU=2*MAX-1
      DO 50 L=1,MU
   50   UVCO(L)=0.0D0
      VCO=1.0D0/FLOAT(N)
      UVCO(1)=VCO
C
      IF (MAX.LE.1) RETURN
      DO 70 J=2,MAX
        VCO=VCO/B(J)
        M=2*J-1
        DO 70 L=1,M
          SCO=0.0D0
          K1=1
          IF (L.GT.MAX) K1=1+L-MAX
          K2=L+1-K1
          DO 60 K=K1,K2
   60       SCO=SCO+F(K,J)*F(L-K+1,J)
   70     UVCO(L)=UVCO(L)+SCO*VCO
C
      RETURN
C
      END
```

Annex D

Examples

(This annex does not form an integral part of the standard.)

This annex contains three examples — two on the calibration of flow-meters for use in pipes and one on the calibration of a river gauging station using current-meters.

The data given in tables 1, 2 or 3 may be used to test the operation of the program listed in annex C, or any other appropriate program. The precision with which the results are computed will depend on the method used and on the accuracy of the computer used: the results in this annex were obtained using double-precision arithmetic equivalent to 18 decimal digits.

D.1 Example 1: Calibration of a differential pressure flow-meter

Table 1 lists 12 pairs of data values obtained from the calibration of a differential pressure device. Figure 1 shows the data plotted with y as the discharge coefficient and x as the pipe Reynolds number divided by 10^6.

To check first that the least-squares methods described in this part of ISO 7066 are appropriate for approximating the functional relationship between y and x, it is necessary first to show that the random error in x can be neglected. In this case the methods of ISO 5168 give values of approximately 0,001 3 and 0,005 for $e_r(y)$ and $e_r(x)$ respectively, so that $e_r(y) / e_r(x)$ is 0,26. By inspection of figure 1, it can be seen that the magnitude of the slope of any fitted curve will not exceed about 0,015, which is less than one-fifth of $e_r(y) / e_r(x)$, and so the least-squares methods are appropriate.

Any method described in annex A or annex B may be used to fit the data: they will all give results which are identical apart from the rounding error. Using the orthogonal polynomial computer program listed in annex C to fit the data in table 1 up to a maximum degree of 5 gives the following output.

```
DEGREE   RESIDUAL STANDARD   PERCENTAGE
              DEVIATION.     SIGNIFICANCE
   0          .150309-02        100.00
   1          .126028-02         96.11
   2          .643462-03         99.96
   3          .641446-03         66.60
   4          .673798-03         36.77
   5          .727772-03          1.14
SUGGESTED DEGREE- 2

ENTER DEGREE (I2), OR -1 TO EXIT
>
```

NOTE — Some numbers are output by the computer in "scientific notation"; for example, the first residual standard deviation is printed as ".150309-02" which is equivalent to 0,150 309 × 10^{-2} or 0,001 503 09.

Instead of testing whether or not each new coefficient, as the degree of fit is increased, differs significantly from zero at the 95 % confidence level, as described in 5.3, this program prints the percentage significance level at which the coefficient differs from zero.

In this example, the improvement obtained between degrees 1 and 2 is highly significant (99,96 %), whereas at higher degrees, there is no significant improvement, and so the suggested degree 2 is appropriate. Entering the degree 2 to obtain details of the fit gives

```
> 2
 POLYNOMIAL COEFFICIENTS, LISTED IN INCREASING POWERS OF X-
   .97273964+000  -.11222161-001   .85781873-002
 COEFFICIENTS FOR SQUARE OF RANDOM UNCERTAINTY-
   .38979504-005  -.21527711-004   .45708054-004   -.40537128-004
   .12833299-004

             DATA               POLYNOMIAL     RESIDUAL      RANDOM
      X               Y             Y        Y - Y(POL)   UNCERTAINTY
   .22000          .97046        .97069      -.22595-03     .9862-03
   .30800          .97031        .97010       .21303-03     .7311-03
   .35500          .96945        .96984      -.38684-03     .6373-03
   .45000          .96989        .96943       .46325-03     .5465-03
   .56200          .96927        .96914       .12785-03     .5663-03
   .65700          .96841        .96907      -.65944-03     .6157-03
   .76800          .97042        .96918       .12394-02     .6529-03
   .88800          .96954        .96954       .13637-05     .6471-03
   .99800          .96911        .97008      -.97383-03     .6126-03
   1.1480          .97131        .97116       .14818-03     .6180-03
   1.2490          .97174        .97211      -.36514-03     .7493-03
   1.3850          .97407        .97365       .41816-03     .1134-02

 ENTER DEGREE (I2), OR -1 TO EXIT
>-1

  END OF EXECUTION
>
```

In the print-out, the polynomial coefficients are listed in sequence, and so the expression for the curve is

$$\hat{y} = 0{,}972\,74 - 0{,}011\,22\,x + 0{,}008\,578\,x^2$$

The five "coefficients for square of random uncertainty" listed in the fifth and sixth lines define the fourth-degree polynomial which represents the square of the random uncertainty $e_r(\hat{y})$ as a function of x. It can be seen in the print-out and in figure 1, that the random uncertainty varies between 0,000 55 and 0,000 65 for most of the range, reaching up to 0,001 13 at the extremes. If the range of the calibration data is wider than the range over which the calibration is required, then the increase in random uncertainty at the extremes will not be important; for example, it can be seen from the print-out that the random uncertainty is within 0,000 75 over the range of x values from 0,30 to 1,25.

D.2 Example 2: Calibration of a turbine meter

In the previous example, the choice of the best degree of fit was straightforward since the significance of the coefficients fell abruptly from 99,96 % to values much less than 95 %. In general, however, the situation is less clear-cut. Table 2 lists calibration data for a turbine meter; x is the frequency (in hertz) and y is the meter coefficient (in pulses per cubic metre).

When the orthogonal polynomial computer program is used to process these data, the preliminary fitting process gives

```
DEGREE   RESIDUAL STANDARD   PERCENTAGE
              DEVIATION      SIGNIFICANCE
   0          1.05171           100.00
   1          .929832            98.58
   2          .532487           100.00
   3          .448948            99.30
   4          .455227            50.25
   5          .416441            95.13
   6          .428974            11.37
SUGGESTED DEGREE- 5
```

The degree 5 fit is just significant at the 95 % confidence level, and so this degree is suggested. If the data had been slightly different, then the percentage significance might well have been less than 95 % for degree 5, and degree 3 would have been suggested. In this situation, it is more difficult to choose the optimum degree. The fifth-degree polynomial gives a better fit to the data, but it is not certain that such a high-degree polynomial will provide a better approximation to the true underlying functional relationship between y and x.

In the end, the choice of degree is a matter of judgement. It is easiest to apply judgement if each curve is plotted out, together with its confidence limits and the data points. Figure 2 and figure 3 show the effect of fitting the data with a degree 3 curve and a degree 5 curve respectively.

From experience, it is known that the turbine meter coefficient tends to decrease fairly steeply below a certain flow-rate: higher in the range, the trend is level. The degree 3 curve follows this pattern better, and it is simpler, so it is the better choice.

D.3 Example 3: Calibration of a stream flow station

Table 3 lists 44 pairs of data from a stream flow station giving stage values and corresponding current-meter discharge values.

The computer program in annex C gives the following output when the data in table 3 are fitted up to a maximum degree of 5.

```
DEGREE   RESIDUAL STANDARD   PERCENTAGE
              DEVIATION      SIGNIFICANCE
   0          15107.8           100.00
   1          5927.44           100.00
   2          1539.71           100.00
   3          534.002           100.00
   4          503.890            98.04
   5          499.663            79.50
SUGGESTED DEGREE- 4

ENTER DEGREE (12), OR -1 TO EXIT
>
```

Entering the degree 4 to obtain details of the fit gives the following output:

```
> 4
 POLYNOMIAL COEFFICIENTS, LISTED IN INCREASING POWERS OF X-
   .48004925+004  -.37421273+004   .10730031+004  -.12228391+003
   .60793445+001
 COEFFICIENTS FOR SQUARE OF RANDOM UNCERTAINTY-
   .89518922+009  -.86129348+009   .35689887+009  -.83190923+008
   .11933195+008  -.10790425+007   .60091427+005  -.18852659+004
   .25524470+002
```

DATA X	Y	POLYNOMIAL Y	RESIDUAL Y - Y(POL)	RANDOM UNC OF Y(POL)
4.9200	1390.0	1361.5	28.502	481.8
4.9500	1450.0	1386.6	63.402	459.1
5.0500	1500.0	1472.2	27.774	392.1
5.1500	1600.0	1560.9	39.132	338.6
5.2100	1650.0	1615.5	34.496	313.1
5.3000	1750.0	1699.5	50.491	284.2
5.4700	1820.0	1865.0	-44.990	256.9
5.5000	1890.0	1895.1	-5.1311	255.1
5.5800	2000.0	1976.9	23.091	253.8
5.6100	2010.0	2008.1	1.8932	254.3
5.7300	2100.0	2135.9	-35.856	259.8
5.8100	2160.0	2223.7	-63.710	265.2
5.9000	2270.0	2325.2	-55.194	271.6
6.1000	2500.0	2561.2	-61.201	283.2
6.2500	2750.0	2748.3	1.7459	287.6
6.5000	2950.0	3080.8	-130.84	286.2
6.7000	3300.0	3367.4	-67.434	278.5
6.9000	3410.0	3674.3	-264.25	267.5
7.1000	3800.0	4003.4	-203.35	255.8
7.2000	3810.0	4177.0	-366.97	250.6
7.3000	4800.0	4357.0	442.95	246.3
7.5000	4500.0	4737.9	-237.86	241.3
7.6000	5100.0	4939.3	160.70	241.0
7.7000	5300.0	5148.6	151.43	242.3
7.8000	5220.0	5366.1	-146.06	245.1
7.9000	5400.0	5592.2	-192.17	249.4
7.9000	6100.0	5592.2	507.83	249.4
8.0000	6500.0	5827.3	672.69	254.8
8.1000	6100.0	6071.9	28.104	261.2
8.4000	6900.0	6866.9	33.126	284.3
8.6000	7350.0	7452.6	-102.60	300.4
9.0000	8900.0	8776.2	123.79	328.8
9.5000	10100.	10762.	-662.29	351.0
9.6000	12200.	11210.	990.24	353.8
10.100	14000.	13735.	265.28	363.9
10.500	14600.	16143.	-1542.6	375.4
11.400	22500.	23096.	-596.44	435.5
11.900	28700.	28061.	639.21	468.2
12.100	31500.	30302.	1198.1	473.1
12.600	36000.	36614.	-614.14	452.3
13.200	45000.	45682.	-681.72	410.1
13.500	52000.	50898.	1101.9	483.3
13.500	51000.	50898.	101.88	483.3
13.800	56000.	56613.	-612.90	694.9

```
 ENTER DEGREE (I2), OR -1 TO EXIT
>-1

  END OF EXECUTION
>
```

NOTE — The number of zeros does not reflect the precision of the test data.

The expression for the curve is

$$\hat{y} = 4\,800 - 3\,742x + 1\,073{,}0x^2 - 122{,}28x^3 + 6{,}079x^4$$

The fitted curve, together with its random uncertainty limits at the 95 % confidence level, is shown in figure 4.

NOTE — Normal plotting practice requires dependent variables to be on the vertical axis but it is normal practice in hydrology to produce the plot as shown in figure 4.

Table 1 — Calibration data for a differential pressure flow-meter

Reynolds number ($\times 10^{-6}$)	Discharge coefficient
0,220	0,970 46
0,308	0,970 31
0,355	0,969 45
0,450	0,969 89
0,562	0,969 27
0,657	0,968 41
0,768	0,970 42
0,888	0,969 54
0,998	0,969 11
1,148	0,971 31
1,249	0,971 74
1,385	0,974 07

Table 2 — Calibration data for a turbine meter

Frequency Hz	Meter coefficient pulse/m^3
28,24	573,76
32,12	574,71
35,58	575,14
40,16	574,84
43,48	575,74
48,82	576,20
52,06	576,50
54,36	576,44
54,86	575,61
56,48	576,40
58,18	575,54
58,38	576,67
60,92	575,94
64,72	575,41
67,74	575,01
71,72	574,51
76,52	574,88
82,64	574,42
83,06	574,05
88,40	574,88
91,94	573,69
96,90	573,25
99,58	573,07

Table 3 — Calibration data for a stream flow station

Stage m	Discharge m^3/s
4,92	1 390
4,95	1 450
5,05	1 500
5,15	1 600
5,21	1 650
5,30	1 750
5,47	1 820
5,50	1 890
5,58	2 000
5,61	2 010
5,73	2 100
5,81	2 160
5,90	2 270
6,10	2 500
6,25	2 750
6,50	2 950
6,70	3 300
6,90	3 410
7,10	3 800
7,20	3 810
7,30	4 800
7,50	4 500
7,60	5 100
7,70	5 300
7,80	5 220
7,90	5 400
7,90	6 100
8,00	6 500
8,10	6 100
8,40	6 900
8,60	7 350
9,00	8 900
9,50	10 100
9,60	12 200
10,10	14 000
10,50	14 600
11,40	22 500
11,90	28 700
12,10	31 500
12,60	36 000
13,20	45 000
13,50	52 000
13,50	51 000
13,80	56 000

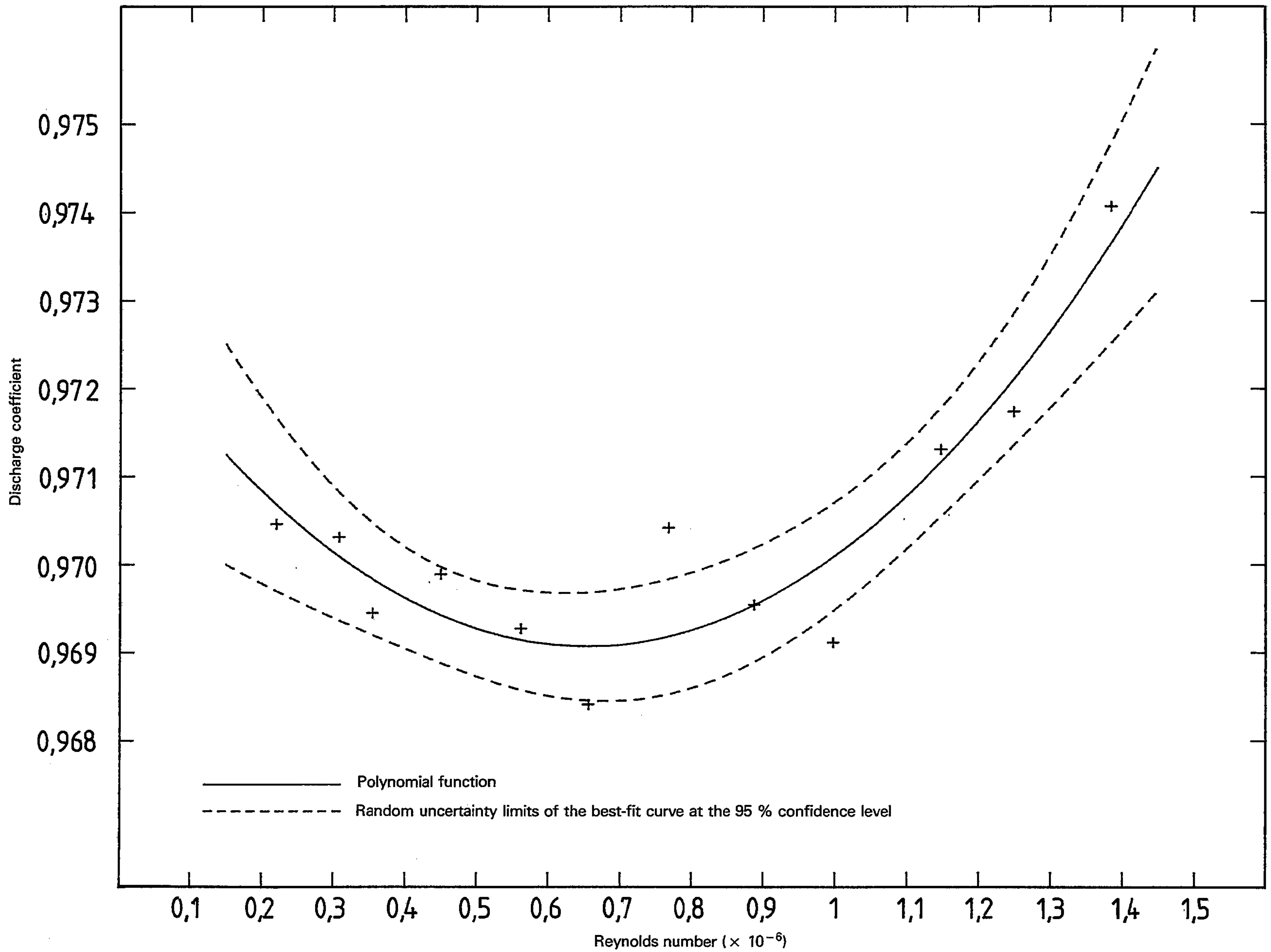

Figure 1 — Calibration data fitted using a second-degree polynomial

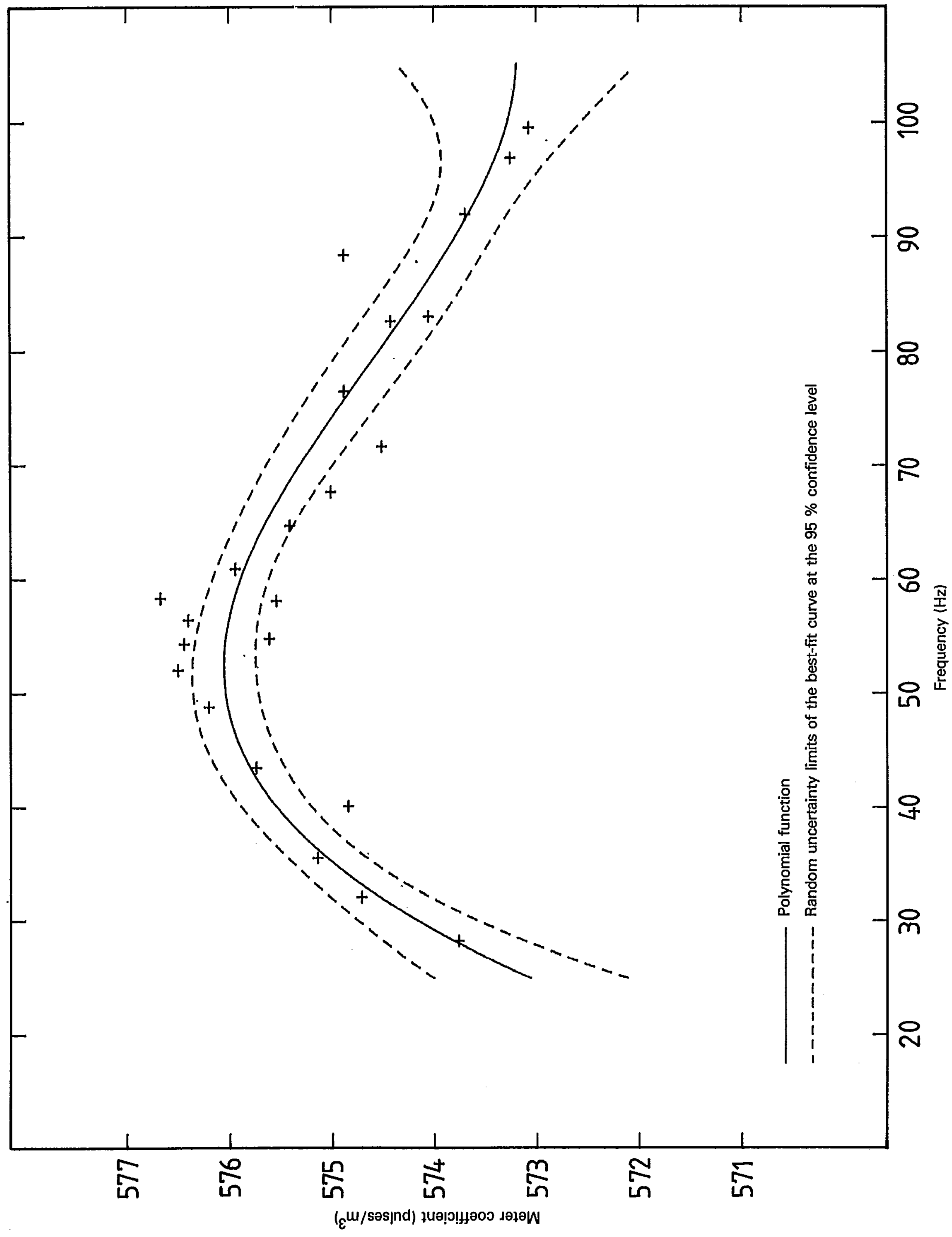

Figure 2 — Calibration data fitted using a third-degree polynomial

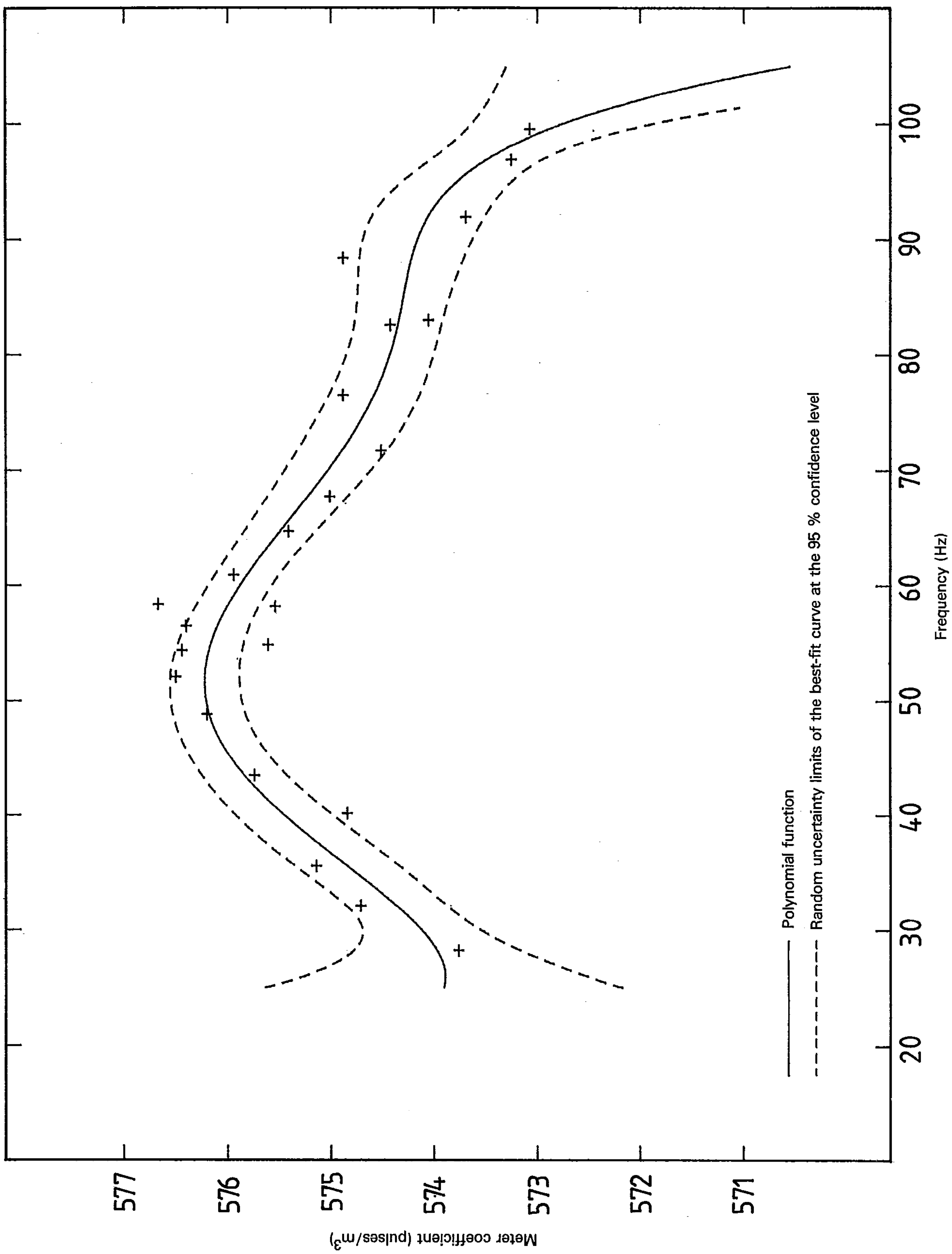

Figure 3 — Calibration data fitted using a fifth-degree polynomial

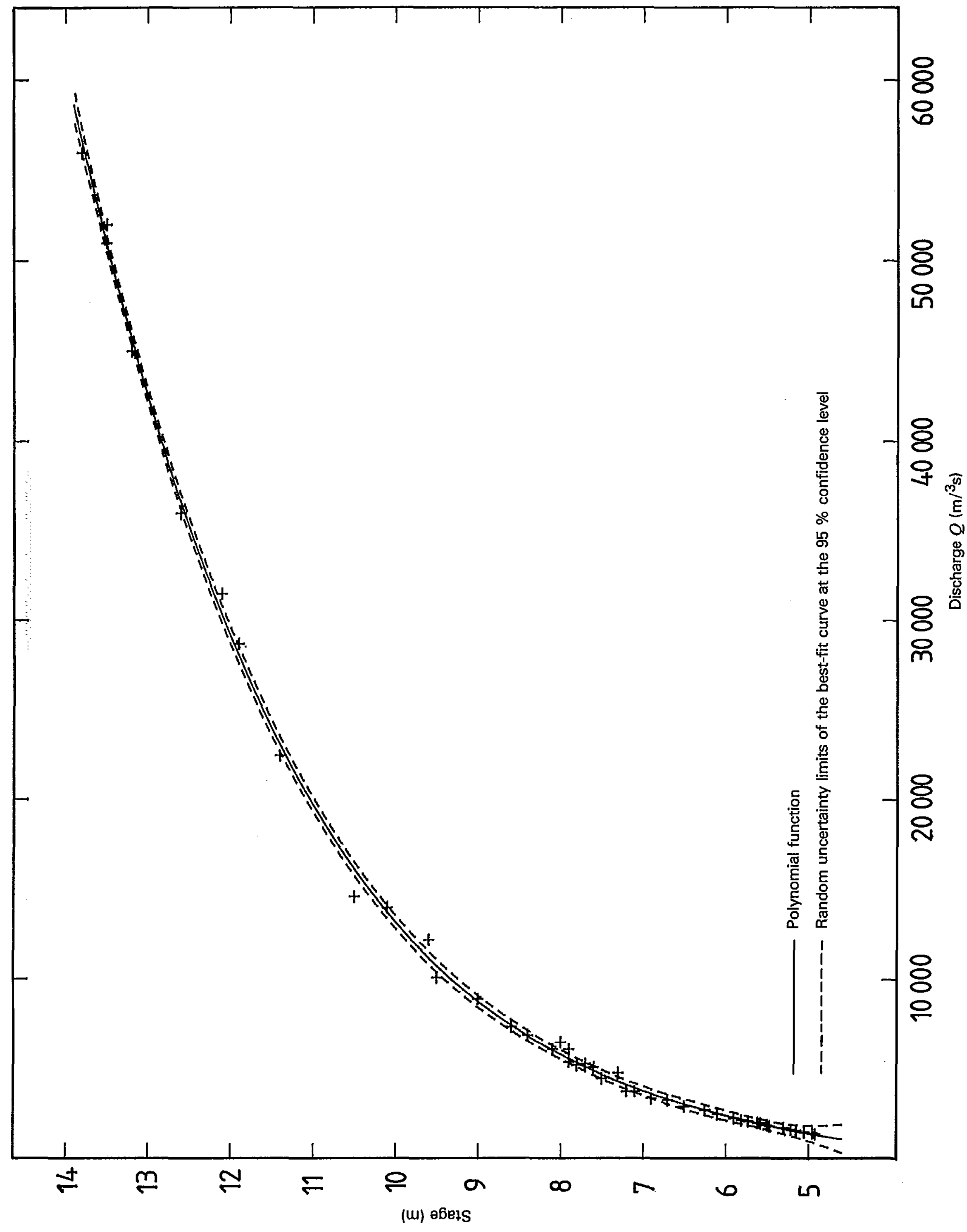

Figure 4 — Stage-discharge curve fitted using a fourth-degree polynomial

Annex E

Finite-difference method

(This annex does not form an integral part of the standard.)

When a computer is not available and the x values are uniformly spaced, a finite-difference table may be used to provide a quick indication of what degree of fit may be appropriate to represent the data. The coefficients of a polynomial representing the data may also be calculated, although this will not be the least-squares polynomial. Calculation of uncertainty using this method is beyond the scope of this part of ISO 7066.

The finite-difference table is constructed as shown in the example below. First, second and third differences, $\Delta^{(1)}$, $\Delta^{(2)}$ and $\Delta^{(3)}$, for a set of n values (x_i, y_i), are given by

$$\Delta_i^{(1)} = y_{i+1} - y_i \qquad \text{for } i = 1 \text{ to } n - 1$$

$$\Delta_i^{(2)} = \Delta_{i+1}^{(1)} - \Delta_i^{(1)} \qquad \text{for } i = 1 \text{ to } n - 2$$

$$\Delta_i^{(3)} = \Delta_{i+1}^{(2)} - \Delta_i^{(2)} \qquad \text{for } i = 1 \text{ to } n - 3$$

The values obtained are given in table 4.

The arithmetical mean value for each column of figures ($\overline{x}$, $\overline{y}$, $\overline{\Delta}^{(1)}$, $\overline{\Delta}^{(2)}$ and $\overline{\Delta}^{(3)}$) is given at the foot of table 4.

Table 4 — Finite-difference table

x	y	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$
0,06	3 622			
		40		
0,07	3 662		−15	
		25		29
0,08	3 687		14	
		39		−28
0,09	3 726		−14	
		25		4
0,10	3 751		−10	
		15		20
0,11	3 766		10	
		25		−25
0,12	3 791		−15	
		10		18
0,13	3 801		3	
		13		−19
0,14	3 814		−16	
		− 3		23
0,15	3 811		7	
		4		−16
0,16	3 815		− 9	
		− 5		8
0,17	3 810		− 1	
		− 6		− 4
0,18	3 804		− 5	
		−11		4
0,19	3 793		− 1	
		−12		−10
0,20	3 781		−11	
		−23		7
0,21	3 758		− 4	
		−27		6
0,22	3 731		2	
		−25		
0,23	3 706			
$\overline{x}$ = 0,145	$\overline{y}$ = 3 757,17	$\overline{\Delta}^{(1)}$ = 4,941 2	$\overline{\Delta}^{(2)}$ = −4,062 5	$\overline{\Delta}^{(3)}$ = 1,133 3

In table 4, the $\Delta^{(1)}$ column shows a clear trend, from positive to negative. In the $\Delta^{(2)}$ column, although there are significant fluctuations, the average of any three or four consecutive values is never very different from the mean value for the column of $-4{,}062\,5$. In the $\Delta^{(3)}$ column, the fluctuations are larger, but negative and positive numbers are generally balanced, and no clear trend away from zero is discernible. In this case, because the third differences are fluctuating about a value close to zero, a polynomial of degree 2 is appropriate.

The coefficients of the degree 2 polynomial can be calculated from

$$b_0 = \bar{y} + \frac{(n^2 - 1)\,\overline{\Delta}^{(2)}}{24} - \frac{\overline{\Delta}^{(1)}\,\bar{x}}{d_x} + \frac{\overline{\Delta}_2\bar{x}^2}{2d_x^2}$$

$$b_1 = \frac{\overline{\Delta}^{(1)}}{d_x} - \frac{\overline{\Delta}^{(2)}\bar{x}}{d_x^2}$$

$$b_2 = \frac{\overline{\Delta}^{(2)}}{2d_x^2}$$

where d_x is the difference between consecutive x values.

The polynomial is then

$$\hat{y} = 3\,313{,}12 + 6\,384{,}74x - 20\,312{,}5x^2$$

For comparison, the least-squares polynomial for the same set of data is

$$\hat{y} = 3\,306{,}97 + 6\,484{,}63x - 20\,663{,}7x^2$$

In the case of data for which the second differences fluctuate randomly about zero, a linear expression may be used; the coefficients are then

$$b_0 = \bar{y} - \frac{\overline{\Delta}^{(1)}\bar{x}}{d_x}$$

$$b_1 = \frac{\overline{\Delta}^{(1)}}{d_x}$$

The finite-difference method works best with data in which the random scatter is relatively small.

UDC 681.121.87 : 532.575

Descriptors: flow measurement, liquid flow, calibration, flow rate, error analysis.

Price based on 28 pages