C.P. No. 1390

C.P. No. 1390

PROCUREMENT EXECUTIVE, MINISTRY OF DEFENCE

AERONAUTICAL RESEARCH COUNCIL

CURRENT PAPERS

On The Application of Certain Statistical Methods to Wind - Tunnel Testing

by

J. McKie

Aerodynamics Dept., R.A.E., Farnborough Hants

LONDON: HER MAJESTY'S STATIONERY OFFICE

1978

£3-00 NET

UDC 533.6.071 : 519.272

*CP No.1390
January 1975

ON THE APPLICATION OF CERTAIN STATISTICAL METHODS TO WIND-TUNNEL TESTING

by

J. McKie

SUMMARY

The Report illustrates the use of some standard statistical techniques in wind-tunnel testing. The results of non-linear regression analysis are applied to the particular problem of comparing the data from experiments in two different tunnels on the same model. Residual variance is used as a measure of the repeatability of results and standard tests are applied to look for significant differences between the two tunnels. The accuracy of a measured aerodynamic coefficient is put in terms of confidence limits for a given probability level. A method is given for determining the minimum detectable effect of a model geometry change and also for finding the number of data points needed to measure a coefficient to a prescribed accuracy.

---

* Replaces RAE Technical Report 75018 - ARC 36049

CONTENTS

1 INTRODUCTION

During the planning of a recent wind-tunnel programme it was decided to group the runs into three separate series. The first two would be largely exploratory and would be carried out in different wind tunnels. It was arranged that some of the model configurations to be tested in these two series would be identical, so that a comparison of the results would enable a decision to be made as to which was the better tunnel to use for the third series of tests. It was realised at that stage that such a comparison would have to be made on a statistical basis, using such concepts as 'repeatability' and 'accuracy'. However, it was not clear how these ideas could be quantified when considering, for example, the model's pitching-moment behaviour with varying angle of incidence.

The experiments used a complete model that had externally-blown flaps and under-wing engines. The planned series of tests involved the investigation of the effects and sensitivities of a large number of parameters concerned with the model geometry and engine performance. Errors in setting up these parameters could be expected to be similar for the two tunnels, but other test parameters such as degree of constancy of wind speed, balance response, interferences from the air-supply connector, etc. would be peculiar to each tunnel, not forgetting the performances of the (different) tunnel crews! All these factors would have an influence on the accuracy of the results and it was decided to choose two configurations for the statistical investigation. For want of a way of calculating the required number of repetitions necessary to make meaningful comparisons, it was arbitrarily decided to repeat tests on each of these two configurations ten times in each tunnel. The twenty repeat runs were interspersed amongst the other runs, so that errors in setting up the configuration geometries would be roughly constant for each run, if it could be assumed that there was no 'learning-curve' behaviour.

When the time came to look at the results from the first two series of tests in order to decide which was the better wind tunnel for this application, those associated with the experiment found that their knowledge of statistical techniques was insufficient to enable them to assess the results properly. Moreover, there appeared to be little experience of using such techniques in the field of aerodynamics and wind-tunnel testing. They are widely used in experimental biology and production engineering for example, but the author was able to find very little information that directly relates to the present problem.

However, it proved possible to evolve a technique which enabled the idea of repeatability to be quantified and produced meaningful numbers on which the choice of tunnel could be made. The analysis of the results of the repeat runs enabled firm values to be placed on the levels of precision of the measured forces and moments, instead of the usual estimates based on the supposed performances of the wind-tunnel balance and data-recording systems. Some of these values were considerably different from those normally assumed.

The results of this work are of sufficient value to warrant their being brought to the attention of other wind-tunnel users. This is not a report about the merits and defects of the two wind tunnels, nor of the characteristics of the model used in the tests. It is about the use of certain statistical techniques which are considered standard and mandatory in other scientific areas, but which, up till now, appear to have been ignored in the field of experimental aerodynamics. Section 2 outlines the theoretical techniques required for the comparison of the two sets of results. The third section applies them to the experimental data and shows how conclusions can be made about repeatability, relative accuracies, etc. These statistical techniques theoretically, at least, enable an experimenter to calculate the number of runs required to measure a certain quantity to a required accuracy. This idea is developed and illustrated in section 4.

A brief glossary of the standard terminology used in the text for statistical quantities is included as an Appendix.

2 THEORETICAL TECHNIQUES

2.1 Initial ideas

Any particular performance characteristic or property Q of a wind-tunnel model is dependent on a number of parameters whose values can be set at levels independent of each other. For the particular model used in these experiments, they included angle of incidence; flap, tab and aileron angles; leading-edge boundary-layer-control blowing momentum; fin position; tail-plane presence, position and angle; engine-nacelle position and thrust. As well as these geometrical and power variables, other parameters such as wind speed and tunnel turbulence, were under the control of the tunnel crews to a greater or lesser degree. Most experiments are designed to determine how Q varies whilst only one of the independent variables α , say, is changed at a time: the others, hopefully, are kept constant. In practice, of course, small changes in the levels of the other parameters induce small changes in Q. The latter are lumped together with the errors incurred in the measurement of Q into a total error ε, so that for a given value of the independent variable, the measured Q equals the true value plus ε. In this Report it is assumed that any value of the independent variable is known exactly. This is an important qualification and should be borne in mind when considering the ultimate estimated accuracies (or errors). There exists a statistical theory, more complex than the one outlined in this Report, which does deal with the more realistic situation of errors in all the independent variables[1].

In the experiment under discussion, angle of incidence α was selected as the independent variable. It was measured by an accelerometer inside the model that transmitted a signal proportional to the sine of the angle. The properties Q were the customary six aerodynamic forces and moments. The experiment measured these quantities at approximately 4° intervals of α so that for both of the 'statistical runs' a group of ten observations was obtained at each nominal angle of incidence.

As it was impossible to set the model precisely at a specific angle of incidence, each group of results was scattered over a region in the (α,Q) plane. The centre of the region corresponded in some way to the mean value and the area to the variance. However, the concepts of measures of location (means) and of dispersion are normally applied to one-dimensional data, such as arise in sampling problems, for example. Thus an initial idea for analysing the data was to transform the observed values to values at the nominal angle of incidence, i.e.

$$Q(\alpha_{nom}) = Q(\alpha) - (\alpha - \alpha_{nom}) \left.\frac{\partial Q}{\partial \alpha}\right|_{\alpha}$$

where $\partial Q/\partial\alpha$ was expected to be weakly dependent on α and could be obtained graphically. Thus for each nominal α there would be a sample of ten observations. If it could be assumed that the errors in Q were random, then this set of ten results would be as good as any other set and could be used to estimate the mean value of Q at α_{nom}. The standard deviation of the points from the mean value would be a measure of the repeatability of the experiment for that particular angle of incidence. It would, of course, be possible to see if this repeatability bore a functional dependence on angle of incidence.

In a similar way, the behaviour of the mean values with α would be the sought dependence of Q on α . If very large samples were obtained, then it might reasonably be supposed that the behaviour of the means at the nominal angles of incidence would be a very good indication of the true dependence of Q on α . Unfortunately, with much smaller samples, the measured means are likely to be in error, which implies that so also would be the standard deviations; in fact they would always be optimistically low. The experimenter ameliorates this problem by fitting a curve through the observed values which is a compromise between one which fits the points and one which is 'smooth'. He uses his experience to strike the right balance, that is he gives the points varying weights according to his estimate of their reliability.

This is the point of failure of the statistical approach suggested so far, for the mean and standard deviation at any α_{nom} are derived in ignorance of the information provided by all the other points, particularly the immediately neighbouring ones. A second source of error is in the process of transforming the observed values of Q at $\alpha_{nom} + \delta\alpha$ to values at α_{nom} . It is clear that a better approach is to adopt curve-fitting techniques right from the start. The quality of the data will determine the goodness of fit of the resulting functional relationship, or the accuracy with which its characteristics, lift slope for example, can be determined.

2.2 Curvilinear regression

To a statistician, the variation of the mean value of Q for given values of α , when referred to all the observations, is called the regression of Q on α . This regression will only correspond to the true behaviour of Q with α if the sample is infinite. The customary method of estimating the regression is by the familiar method of least squares. A useful reference for the subject of least-squares analysis in the context of statistical methods is given in the book edited by Davies[1]. The results of using such methods are often looked upon with considerable suspicion. Davies takes pains to emphasise at all stages the care that has to be adopted in their application, if non-sensical statistics are to be avoided.

The nature of the expected (α,Q) behaviour is sometimes known beforehand; for example lift usually varies linearly with α over restricted ranges of α . In other situations the behaviour is not known in advance; such is often the case for the variation of pitching moment with α . In these latter situations, one would try various kinds of curves, transformations of the data, etc. in order to get a fit which looks 'reasonable'. In this Report, the (α, Q) curves will be restricted to those which can be represented as simple polynomials in α . As the experiments produced large numbers of data points (up to 200 in some cases), it is quite practicable to try and fit curves of relatively high degree. However, the least-squares normal equations tend to be ill-conditioned and difficulties often arise in fitting curves of degree six or more. For this reason, and for others concerned with simplifications to the statistical analysis, orthogonal polynomials are used. The theory outlined here largely follows that given by Forsythe[2].

The starting point is a set of n observations Q_i based on n values of the independent variable α_i . It is supposed that the Q_i can be written

$$Q_i = y(\alpha_i) + \varepsilon_i$$

and that the approximating function y is linearly dependent on the k functions $\alpha, \alpha^2, \alpha^3 \ldots \alpha^k$, so that $y(\alpha)$ can be written

$$y = c_0 + c_1\alpha + c_2\alpha^2 + \ldots c_k\alpha^k \quad . \tag{1}$$

It will be assumed that $k + 1$ is always less than n , so that it is generally impossible to make equation (1) fit all the observed points simultaneously. Equation (1) is said to be in 'standard' form and gives y the predicted value of Q for any α .

Suppose there exists a set of polynomials $\phi_i(\alpha)$ which are linear combinations of $\alpha, \alpha^2, \alpha^3$, etc. up to and including α^i . Then it is possible to rearrange equation (1) to read

$$y = b_0\phi_0 + b_1\phi_1 + \ldots b_k\phi_k \tag{2}$$

the so-called 'orthogonal' form, where the coefficients b_i and the multipliers in the ϕ_i are all functions of the coefficients c_i . In general, the observed value Q_i at α_i will not equal the predicted value y_i , but the desired objective is to make the error ε_i small for all $i = 1, \ldots n$ and to have zero mean. A measure of the goodness of fit is the amount of spread or scatter of the data points about the fitted curve. It is assumed in this analysis that the observations Q_i are distributed normally about the mean curve y with standard deviation σ. The true value of σ can only be determined by examining an infinite number of observations. In a real experiment, an estimate s of σ is made, based on n observations, where s approaches σ as $n \to \infty$.

The variance about regression is the square of σ and is equivalent to the mean of the squares of the errors. The total observed variation of Q with α is in part due to its functional relationship $y(\alpha)$ and in part due to random error. It is to be expected that, as the order k of the fitted curve (2) is increased, then s will decrease. This will continue until either (2) represents the true regression of Q on α and s then becomes practically constant, or when k becomes equal to $n - 1$ and the fitted curve joins all the (α, Q) points. This latter situation is not very meaningful and statistical theory takes it into account through the concept of degrees of freedom. Fitting a curve of degree k requires $k + 1$ degrees of freedom (or items of information), leaving only $n - (k + 1)$ on which to assess the variance, i.e.

$$s^2 = \frac{\text{total sum of squares about regression}}{n - k - 1} .$$

Clearly, when the fitted curve joins all the data points, $s^2 = 0/0$ so that no sensible statement can be made about the variance. s thus cannot be equated to rms error, which is defined using simply n in the denominator.

The method of least squares ensures that for a given value of k, the total sum of squares of the errors is a minimum

$$\sum_{i=1}^{n} (y(\alpha_i) - Q_i)^2 = \text{minimum} = S \qquad (3)$$

and the problem is to choose the b_i and ϕ_i so that this condition is achieved. The curve (2) will then be called the regression of Q on α. Equation (3) can be expanded in terms of the b_i:

$$S = \sum_{i=1}^{n} [Q_i - b_0\phi_0(\alpha_i) - b_1\phi_1(\alpha_i) - \ldots - b_k\phi_k(\alpha_i)]^2$$

and if S has a minimum for some set of the coefficients b_i , then

$$\frac{\partial S}{\partial b_j} = 0$$

for all $j = 0, 1 \ldots k$. Thus there result $k + 1$ equations

$$0 = \sum_{i=1}^{n} Q_i \phi_j(\alpha_i) - \sum_{i=1}^{n} b_0 \phi_0(\alpha_i)\phi_j(\alpha_i) - \sum_{i=1}^{n} b_1 \phi_1(\alpha_i)\phi_j(\alpha_i) - \ldots$$

$$- \sum_{i=1}^{n} b_k \phi_k(\alpha_i)\phi_j(\alpha_i)$$

for the unknown coefficients b_j . These are the normal equations of the least-squares method. It is convenient to introduce the notation:

$$(\phi_p, \phi_q) = \sum_{i=1}^{n} \phi_p(\alpha_i)\phi_q(\alpha_i)$$

so that the $k + 1$ normal equations can be put into matrix form

$$Y = CB \tag{4}$$

where Y is the column vector $\{(Q,\phi_0),(Q,\phi_1), \ldots (Q,\phi_k)\}$, C the symmetric square matrix

$$\begin{bmatrix} (\phi_0,\phi_0) & (\phi_0,\phi_1) & \ldots & (\phi_0,\phi_k) \\ \vdots & & & \\ (\phi_k,\phi_0) & (\phi_k,\phi_1) & \ldots & (\phi_k,\phi_k) \end{bmatrix}$$

and B the column vector $\{b_0, b_1 \ldots b_k\}$. Forsythe[2] shows that provided n exceeds k the determinant of C is not zero, so that C^{-1} exists and a unique vector B can be found by inverting C .

2.3 Use of orthogonal polynomials

The nature of the formation of C is to make it ill-conditioned and difficult to invert. This is especially so if k is larger than about six and if the data are clustered as in the current problem with groups of ten points all at the same nominal angle of incidence. However, if the ϕ_j can be chosen so as to make the off-diagonal terms in C all identically zero, then the inversion becomes trivial. This can be achieved by making the ϕ_j mutually orthogonal over the point set α_i , i.e.

$$(\phi_p,\phi_q) = \begin{cases} 0 & \text{if } p \neq q \\ M_p & \text{if } p = q . \end{cases} \tag{5}$$

The inverse of C is now simply

$$\begin{bmatrix} M_0^{-1} & 0 & 0 \ldots & 0 \\ 0 & M_1^{-1} & 0 \ldots & \\ \vdots & & & \\ 0 & 0 & & M_k^{-1} \end{bmatrix}$$

and the solution vector B has terms

$$b_j = \frac{\sum_{i=1}^{n} Q_i \phi_j(\alpha_i)}{\sum_{i=1}^{n} (\phi_j(\alpha_i))^2} . \tag{6}$$

A very important point to note about this relation is that it is independent of k , the degree of the curve fit. It was seen in the previous section that trial curve fits would have to be made, increasing the degree k of the polynomial, until the estimate s of the residual standard deviation tended to a roughly constant value. Normally, this would involve a new regression for each k resulting in a different set of coefficients b_i . However, using orthogonal polynomials going from order $j - 1$ to j does not change the values b_0 to b_{j-1} already calculated, so that only one regression need be done for, say, $j = 9$ which is likely to include the least value of k for the best fit.

The sum of squares about regression given by equation (3) can be written

$$S = (Q,Q) - 2\sum_{j=0}^{k} b_j(Q,\phi_j) + \sum_{j=0}^{k} b_j^2 M_j$$

or

$$S = (Q,Q) - \sum_{j=0}^{k} b_j^2 M_j \quad . \tag{7}$$

The first term on the right-hand side is the total sum of squares and the second term is the sum of squares due to regression. This latter is in a particularly simple form due to the use of orthogonal polynomials. The estimate of the residual variance is

$$s^2 = \frac{(Q,Q) - \sum_{j=0}^{k} b_j^2 M_j}{n - k - 1} \quad . \tag{8}$$

It is customary to analyse the total variance into its component parts and present the results in an analysis of variance table, in order to assess the significance of each of the terms in the regression. This is particularly simple when using orthogonal polynomials so that the table takes the form

Source	Sum of squares	DF	Mean square
Degree 0 term	$b_0^2 M_0$	1	$b_0^2 M_0$
Degree 1 term	$b_1^2 M_1$	1	$b_1^2 M_1$
...	...	...	⋮
Degree k term	$b_k^2 M_k$	1	$b_k^2 M_k$
Residual	$(Q,Q) - \sum_{j=0}^{k} b_j^2 M_j$	n - k - 1	s^2
Total	(Q,Q)	n - 1	

where DF stands for degrees of freedom. The values in the fourth column are obtained from those in the second through division by the appropriate numbers of degrees of freedom. For the analysis of the contributions due to each term in a curvilinear regression, this is redundant, as to each term is associated just one degree of freedom or constraint (so that, for example, a cubit fit provides a total of four constraints on the regression). However, the table takes the same standard form in more complicated analyses, in which each source of variation usually provides more than one degree of freedom. If the total variance is analysed in this way, then it is easy to see the individual contribution of each term in the curve fit.

It is appropriate here to state how the orthogonal polynomials are chosen. The basis for their generation is a recurrence relation that gives ϕ_{j+1} in terms of ϕ_j and ϕ_{j-1} :

$$\phi_{j+1} = (\alpha - E_{j+1})\phi_j - D_{j+1}\phi_{j-1} \quad .$$

To start the sequence it is necessary to define

$$\phi_0 = 1$$

$$D_1 = 0$$

so that the second polynomial becomes

$$\phi_1 = \alpha - E_1$$

(in fact E_1 works out to be the mean α). Forsythe[2] shows how to choose the coefficients E_j and D_j so that the orthogonality relations (5) are fulfilled:

$$E_j = (\alpha\phi_{j-1}, \phi_{j-1})/M_{j-1}$$

$$D_j = M_{j-1}/M_{j-2}$$

where $(\alpha\phi_{j-1}, \phi_{j-1}) = \sum_{i=1}^{n} \alpha_i[\phi_{j-1}(\alpha_i)]^2$.

That this choice does ensure orthogonality can be verified by induction. It is clear from the forms given above for the first two polynomials that (ϕ_0,ϕ_1) is zero. Now suppose that $E_1, \ldots, E_i$ and $D_1, \ldots, D_i$ have been chosen so that $\phi_0, \ldots, \phi_i$ are all mutually orthogonal. It has to be demonstrated that (ϕ_{i+1},ϕ_i) also vanishes. The recurrence relation gives

$$\alpha\phi_{j-1} = \phi_j + E_j\phi_{j-1} + D_j\phi_{j-2}$$

so that for $j \leqslant i$

$$(\alpha\phi_{j-1},\phi_j) = M_j \quad .$$

It also follows from the recurrence relation that

$$(\phi_{i+1},\phi_j) = (\alpha\phi_i,\phi_j) - E_{i+1}(\phi_i,\phi_j) - D_{i+1}(\phi_{i-1},\phi_j) \quad .$$

For $j = i$, the last term on the right-hand side is zero and the choice of E_{i+1} given above ensures that the first two terms cancel. For $j = i - 1$, the second term is zero and the last becomes equal to M_j so that it cancels with the first. For $j < i - 1$, the last two terms both vanish. $\alpha\phi_j$ is a polynomial of degree less than i , so that it can be expressed in terms of $\phi_0, \ldots, \phi_{i-1}$. Hence the first term also is zero and orthogonality is proved.

2.4 Confidence limits

Although the analysis of variance table enables the amount of the total variance accounted for by each term in the regression to be easily seen, it is necessary to be more precise as to whether a term is important or not. Each of the coefficients is, of course, subject to error and the customary measure of this uncertainty is the standard error. If Q were independent of α , then the accuracy of the mean of n observations, that is the deviation of the observed mean from the true mean of all possible observations, would be $\sigma/\sqrt{n}$, called the standard error of the mean. Clearly this decreases if the number of observations is increased, as is to be expected. The extension to polynomial regression is more complicated and involves the inverse of matrix C (see equation (4)). The inverse is known as the variance-covariance matrix. Its diagonal terms give the variances or standard errors of the coefficients b_j and the off-diagonal terms the covariances or correlations between them. When using orthogonal polynomials these covariances are conveniently zero and the standard error of b_j is simply

$$SE(b_j) = \frac{\sigma}{\sqrt{M_j}} . \qquad (9)$$

The modulus M_j is directly related to the number of observations so that (9) is analogous to the standard error of the mean.

If a variable is distributed normally about its mean value with standard deviation σ , then it is well known that the probability of any one value lying within a distance of σ either side of the mean is 0.68 and within 2σ it is 0.95. That is, taking any group of 100 observations, on average 95 of them will lie within $\pm 2\sigma$ of the mean. This mean value is itself subject to error and the means of any two equal-sized samples from a population could not be expected to be the same. However, it is 95% probable that the true mean lies within a distance of two standard errors either side of the sample mean. These are the confidence limits for the true mean. It is an interesting fact that, even if the variable is not distributed normally, its sample mean rapidly tends to a normal distribution about the true mean as the sample size increases, with variance σ^2/n .

If the standard deviation is not known, but there exists for it an estimate s , which is itself subject to error, then the limits for any particular confidence level will generally be wider, depending on the number of degrees of freedom of s . The factor to apply is the quantity t_a , obtained from standard statistical tables[3], which is also a function of a , the level of probability being applied. Equation (9) expresses the variance of the regression coefficient b_j in terms of σ . An estimate of the standard error is obtained by substituting s for σ and the $(1 - 2a)$ limits (for the 95% level, $a = 0.025$) within which the true value of b_j can confidently be expected to lie are

$$b_j \pm \frac{t_a s}{\sqrt{M_j}} . \qquad (10)$$

For example, if s is based on only 5 degrees of freedom, t_a has the value 2.57 at the 95% confidence level, so that at this level the true b_j can be expected to lie within 2.57 standard errors of the calculated value. In fact, the probability of it lying within two standard errors has dropped to almost

90%. t_a does, of course, tend to 2 as the number of degrees of freedom increases, at the 95% level.

The accuracy of the estimated regression (2) or the variance of any prediction made from it, is clearly dependent on the variances of the individual terms and also on the value of the independent variable α . It is to be expected that the regression will be most accurate at the mean of the observed α values and least accurate at the extreme ends. A standard statistical rule is that the variance of a quantity is the sum of the variances of its constituent terms. Hence, if y is a prediction from equation (2) at a value α of the independent variable, its variance is given by

$$V(y) = \sum_{j=0}^{k} [\phi_j(\alpha)]^2 V(b_j)$$

$V(b_j)$ is the same as the square of the estimate of the standard error of b_j , given by equation (9) with s substituted for σ so that

$$V(y) = s^2 \sum_{j=0}^{k} M_j^{-1} [\phi_j(\alpha)]^2 \quad . \tag{11}$$

The independence of the coefficients b_j has already been mentioned. If a single term was being used in the regression, b_0 only, then this would correspond to the arithmetic mean of the observations and its variance would be the square of the standard error of the mean. In other words, $\phi_0 = 1$ and $M_0 = 1/\sqrt{n}$. These values are unchanged even if more than one term is used in equation (2), hence the variance of a prediction can be written.

$$V(y) = s^2 \left\{ \frac{1}{n} + \sum_{j=1}^{k} M_j^{-1} [\phi_j(\alpha)]^2 \right\} \quad . \tag{12}$$

Confidence limits on the prediction y at the $(1 - 2a)$ confidence level are given by

$$y \pm t_a \sqrt{V(y)} \quad .$$

In some practical situations, such as arose in the wind-tunnel tests to be discussed in the next section, it turns out that the first term in equation (12) is dominant, so that for practical purposes, the width of the confidence interval for the regression curve is $2t_a s/\sqrt{n}$. If s can be assessed (by a few trial runs of the experiment, say), then it is possible to calculate the number of data points n in order to achieve a curve fit of the required accuracy (see section 4).

Equation (12) expresses the variance of the best-fit mean curve. The scatter or spread of the data about this line can be equated to the variance of any individual observation. This is given by the sum of the variance of the regression and the mean variance about regression (i.e. s). Thus the variance of individual results is

$$V(Q) = s^2 \left\{ 1 + \frac{1}{n} + \sum_{j=1}^{k} M_j^{-1} [\phi_j(\alpha)]^2 \right\} \qquad (13)$$

and confidence limits are

$$Q \pm t_a \sqrt{V(Q)} \quad .$$

The t_a factor is based on (n - k - 1) degrees of freedom and if it is evaluated at a = 0.025 or 0.95 probability level say, then the limits can be interpreted by saying that for any 100 observations, on average only five will fall outside these limits. Equation (13) is much less dependent on the value α of the independent variable than is (12) and is fundamental to the experiment: it cannot be reduced significantly by increasing the number of observations. It is these limits which will be used to quantify the 'repeatability' of the experiment. For a large number of observations, the repeatability is practically directly proportional to the standard deviation of the experiment.

2.5 Significance

If the confidence limits for the true value of a regression coefficient (expression (10)) include the value zero, then it is said that the b_j is not statistically significant. This is using the word 'significant' in a specialised sense: physically, it may be quite a significant result that the coefficient is perhaps zero. The fact that the value of a parameter is statistically insignificant is not proof that the parameter is zero, merely that it could be. Changing the confidence level of the limits on the regression coefficient would alter the possibility that the coefficient is not significant. Hence the concept of statistical significance is coupled to levels of probability.

A fundamental idea in this area of statistics is the null hypothesis. The hypothesis is made that the experiment has indicated no true difference between two quantities or levels of a parameter. A test of significance is then applied to calculate the probability of the observed difference being due merely to chance. If this probability is small, then the null hypothesis is not true. More precisely, if the probability that the observed difference could have occurred through chance (i.e. the null hypothesis is true) is P and $P \leqslant a$, where a is a given probability level, then it is said that the result is significant at the level a. For the application to regression coefficients, the null hypothesis is made that b_j does not differ from zero. Chance could make it either positive or negative, so the test to be applied is called double-sided. A single-sided test would be used, for example, if it was required to know if an experiment in wind-tunnel A produced a result with a standard deviation which was worse than that for a similar experiment in tunnel B. (A double-sided test could be applied to see if there was a significant difference between the standard deviations.) For a double-sided test, the probability that the observed difference is due to chance is two-fold, so that the test level to be applied is 2a.

It is customary in statistical work to assess significance at certain standard levels of a : 0.1, 0.05, 0.01, 0.001. If the result could occur by chance at a probability greater than 0.1, then it is 'not significant' and there is no good reason for rejecting the null hypothesis. If $0.1 > P > 0.05$, then the result is 'probably significant'. For $0.05 > P > 0.01$, it is called 'significant' and if P is less than 0.01, the result is 'highly significant'. These are the jargon phrases corresponding to the various levels of a commonly in use, but clearly they should be used with caution as the context of a result is obviously of importance in establishing its significance.

The amount by which the regression coefficient differs from zero, if it is due purely to chance, clearly depends on the number of observations involved in its derivation. The appropriate test of significance is Student's t-test and in this situation involves computing a value t from the ratio of b_j to its standard error

$$t = b_j \sqrt{M_j} / s \quad . \tag{14}$$

This value is then referred to tables of the t-distribution for the standard a levels. The tables are entered at (n - k - 1) degrees of freedom and interpolated for P .

An alternative way of assessing the significance of a regression coefficient is to see by how much the total sum of squares (see section 2.3) is reduced through inclusion of this coefficient. If the mean sum of squares due to the jth term in the regression is of the same order as the residual mean square, then there is not much point in including it. This idea can be treated more precisely through application of Snedecor's F-test. This is generally to be applied when comparing two variances and the null hypothesis to be tested is that there is no true difference between the variances (or mean sums of squares in this case). The statistic used is the ratio of the variances and this is chosen always to be not less than one. The F-distribution is tabulated[3] for the standard probability levels and is entered at the point given by the degrees of freedom in the numerator and denominator. Reference to the analysis of variance table in section 2.3 shows that the value of F will be

$$F = b_j^2 M_j / s^2 \tag{15}$$

or its reciprocal, with one and (n - k - 1) degrees of freedom. In this example, F is the square of t given by equation (14). This simplification arises through the use of orthogonal polynomials, but is not generally true for polynomial regression.

In the tables, F is given for a single-sided test (does the variance due to the jth term in the regression exceed the residual variance?) and the a levels have to be doubled if the test is double-sided. The F-test is the appropriate one to use when assessing the significance of differing standard deviations for a repeat experiment in two wind tunnels.

3 RESULTS

3.1 Experimental background

In this section, the techniques outlined in the previous one will be demonstrated. The wind-tunnel data were obtained in the form of six-component overall forces and moments from the same model tested in two facilities, tunnels A and B. A data point for the statistical analysis consists of an uncorrected coefficient and its corresponding angle of incidence. For the purpose of obtaining data for the comparison of the two tunnels, two typical model arrangements were chosen: the high-lift and low-lift configurations. The former corresponds to the landing case with flaps set at 40^{o} and engine thrust on. In the latter configuration there was no engine thrust and the flaps were at the take-off setting of 20^{o}.

The two configurations were each repeated nine times during the courses of the two experiments. These both took about ten days of testing and the statistical runs accounted for approximately a third of the total number of runs. They were spread evenly through the courses of the experiments, during which a number of geometry changes occurred: fin position; tail-plane presence, angle and position on fin; flap, aileron and tab deflection angles. Engine pod spanwise position and blowing momentum coefficient were also varied. With engine thrust on, there was also boundary-layer control by part-span leading-edge blowing (at a fixed momentum coefficient). For the majority of the tests (including the statistical runs), the model was configured symmetrically relative to the lateral plane so that any non-zero measurement of side force, yawing or rolling moment represented an error of some kind. For a few runs, asymmetric thrust was deliberately generated by partial blockage of an internal air-supply duct in one of the engine pods. However, the difficulty of achieving perfectly symmetric thrust was probably the prime cause of non-zero lateral forces and moments.

Angle of incidence was selected as the independent variable. It was altered by an electro-hydraulic mechanism inside the model and measured by an accelerometer with an output in millivolts, which was calibrated against a clinometer. During the course of the runs in tunnel A, the incidence mechanism became defective so that it became increasingly difficult to set and maintain the model at a prescribed angle of incidence. Eventually, the system broke down completely and had to be repaired. Tunnel speed was controlled manually in both wind tunnels. Forces and moments were measured on identical virtual-centre balances with counter readings transferred directly to a teleprinter. One important difference between the two tunnels, however, was the provision for the blowing-air supply. The design of the six-component air connector was such as to make the possibility of constraints on one of the components unavoidable. For tunnel A interferences could be expected to occur with the measurement of pitching moment, and in tunnel B with the measurement of rolling moment. At the end of one of the low-lift statistical runs in tunnel A, the post-run pitching-moment balance zero was very different from the pre-run zero. The run was therefore repeated without any model changes or other interferences, simply the weighbeams rebalanced to obtain a new set of pre-run zeros and the airflow started again. This had to be done four times before satisfactory pitching-moment behaviour was achieved. These five consecutive runs thus provide a check on the errors due to model changes, etc, in the low-lift configuration. The last (successful) run was included in the set of ten repeat runs.

3.2 Curve fitting

It was shown in section 2.4 that the repeatability of an experiment is strongly dependent on the standard deviation of the regression. An estimate for σ can be obtained by the method of least squares. This is not a straightforward process, however, for a number of reasons.

Fig.1 shows a plot against angle of incidence of all the drag-coefficient data obtained in tunnel A for the high-lift configuration. It is clear that as the experiment progressed further into the stalled region, the scatter of the points increased. Most technical interest centres on the régime where the majority of the flow over the wing is attached. Consequently, it would be unfair to judge the repeatability of an experiment on the magnitude of a standard deviation that has been enlarged by the inclusion of data points from fully-separated flow. The regression line shown on the figure is drawn between the incidence limits of -4.5° and 21° and is based on all points included within these bounds. Where to draw these limits can only be decided by visual inspection of the complete set of results (for this particular coefficient).

For a small number of data points it is possible for the experimenter to assess their relative reliability by giving each point a 'weight' on a scale running from 0 to 1. However, this is a practical impossibility when the number of points becomes at all large (there are 191 points on Fig.1) and so this refinement has been omitted in the theory of section 2. The best that can be done in this direction is to use a binary weighting scheme: each point is either excluded or included in the least-squares analysis. Three of the points on Fig.1 are obviously incorrect, most probably because of malfunctions in the balance read-out equipment. During the course of each run the results were converted to uncorrected coefficient form and hand-plotted to check for bad points. Many angles of incidence were therefore repeated. If these latter points are included and the ones they replace excluded, then a reduced set of data points is obtained with a smaller scatter. This set of 130 points is shown on Fig.2. (In fact further points have been removed near the stall at odd angles of incidence to give groups of ten points at every 2^{o} between -4^{o} and 22^{o}.)

Another factor which is important in curve fitting is the choice of degree of the regression line (the value of k in equations (1) or (2)). As was shown in section 2.3, this is easy using orthogonal polynomials, as only one regression needs to be done, for a degree of fit which is likely to exceed the optimum. A degree 9 regression was performed on the data of Fig.2 (between the incidence limits of -4.5^{o} and 21^{o}) with the results as shown in the table.

Table 1

Regression of C_D on α ; tunnel A, high-lift configuration, selected points

Degree	Coefficients of regression line		Standard error	95% confidence limits		t
	Standard form	Orthogonal form		Upper	Lower	
0	-0.335	0.121	0.686×10^{-3}	0.122	0.119	176
1	0.191	0.545×10^{-1}	0.918×10^{-4}	0.547×10^{-1}	0.543×10^{-1}	593
2	-0.132×10^{-1}	0.570×10^{-3}	0.136×10^{-4}	0.597×10^{-3}	0.543×10^{-3}	42
3	-0.312×10^{-1}	-0.240×10^{-4}	0.205×10^{-5}	-0.199×10^{-4}	-0.281×10^{-4}	12
4	0.716×10^{-2}	-0.154×10^{-5}	0.322×10^{-6}	-0.899×10^{-6}	-0.218×10^{-5}	4.8
5	0.483×10^{-3}	-0.240×10^{-6}	0.524×10^{-7}	-0.136×10^{-6}	-0.344×10^{-6}	4.6
6	-0.315×10^{-3}	-0.136×10^{-7}	0.876×10^{-8}	0.386×10^{-8}	-0.310×10^{-7}	1.55
7	0.412×10^{-4}	-0.292×10^{-8}	0.152×10^{-8}	0.976×10^{-10}	-0.594×10^{-8}	1.93
8	-0.254×10^{-5}	-0.254×10^{-9}	0.270×10^{-9}	0.283×10^{-9}	-0.792×10^{-9}	0.94
9	0.769×10^{-7}	0.405×10^{-11}	0.438×10^{-10}	0.912×10^{-10}	-0.831×10^{-10}	0.09

The second column shows the values of the coefficients c_j in the standard-form regression, equation (1). This set would take different values if a regression of different degree had been calculated. The third column gives the best estimates of the coefficients b_j in the orthogonal-form regression, equation (2). This set is invariant with respect to the degree of the regression. The fact that the coefficients become extremely small as the degree of fit increases is not to say that they are insignificant. This is a result of using unscaled values of the independent variable. At $\alpha = 20^{o}$, the term 20^{9} would arise in an expansion of the orthogonal polynomials, so that $\phi_9 b_9$ could be a relatively sizable quantity. The significance of the coefficients is assessed by the ratio of their magnitudes to their standard errors (see equation (14)). The values in the fourth column are derived from relation (9), where the estimate s of the standard deviation is the square root of the mean residual variance after fitting the degree 9 regression. In other words, the standard errors and the values of the t-factor (last column as calcualted from (14)) are dependent on the degree of fit.

Table 2

Analysis of variance for the regression of Table 1

Degree	Sum of squares	Degrees of freedom	Mean square
0	1.8719	1	1.8719
1	21.3798	1	21.3798
2	0.1068	1	0.1068
3	0.835×10^{-2}	1	0.835×10^{-2}
4	0.139×10^{-2}	1	0.139×10^{-2}
5	0.127×10^{-2}	1	0.127×10^{-2}
6	0.146×10^{-3}	1	0.146×10^{-3}
7	0.225×10^{-3}	1	0.255×10^{-3}
8	0.539×10^{-4}	1	0.539×10^{-4}
9	0.520×10^{-6}	1	0.520×10^{-6}
Residual	0.723×10^{-2}	119	0.608×10^{-4}
Total	23.3772	128	

An analysis of the variance for the degree 9 regression on the data of Fig.2 is shown in Table 2. 129 data points are included in the calculation so that the number of degrees of freedom of the residual variance is 119. The estimate of the standard deviation is thus $(0.00723/119)^{\frac{1}{2}}$, i.e. s = 0.00780. If a degree 5 regression had been fitted, then the residual sum of squares would have equalled the value shown in the table plus the sums of squares accounted for by the sixth, seventh, eighth and ninth terms, and would be based on 123 degrees of freedom. s would have been $(0.00766/123)^{\frac{1}{2}}$, i.e. 0.00789 which is hardly different from the result using the degree 9 regression, so that the inclusion of these terms is unnecessary.

The significance of the terms is assessed from their corresponding t-values. Table 3 gives the probability points of the single-sided t-distribution for 120 degrees of freedom (extracted from Ref.3). Insignificant terms can be dropped from the regression by comparing their t-values from

Table 3

Probability points of the t-distribution (single-sided), 120 degrees of freedom

a	t
0.1	1.29
0.05	1.66
0.025	1.98
0.010	2.36
0.005	2.62

Table 1 with those in Table 3 and, as has been demonstrated above, this does not materially alter s or the t-values themselves. For example, if terms are to be ignored if they do not differ significantly from zero at the 95% confidence level, then the critical value of t is 1.98 (this is a double-sided test). Hence terms of degree higher than five are unnecessary, and this agrees with the observation that the upper and lower limits to the probable values of b_j change sign for these coefficients.

The degree-5 curve (equation (2)) has been drawn through the data points on Fig.2. It is of interest that, to the same level of significance, a degree-3 regression suffices for the full set of data, Fig.1. This is because the larger standard deviation ($s = 0.0278$) puts less constraint on the curve at higher angles of incidence.

It should be emphasised, however, that the t-value ought not to be used as the sole factor in assessing the number of terms to be included in the curve fit. The statistical theory of section 2 assumes that the errors in the observations are distributed normally about the true values. If the errors do not belong to a normal distribution, are not random but are systematic, then the theory is not strictly applicable. Furthermore, it works best if the data points are distributed evenly over the range of the independent variable. Fig.3 shows some lift-coefficient data for the five consecutive runs in the low-lift configuration. The t-values for the regression are listed in Table 4 (for the range of angle of incidence of -4.5° to 15°).

Table 4

Significance of the coefficients in the regression of C_L on α ; tunnel A consecutive runs, selected points

Degree	t
0	2810
1	888
2	61
3	0.10
4	16
5	4
6	11
7	2.28
8	8
9	0.97

A satisfactory degree-6 regression curve has been drawn on Fig.3. However, it appears from the tabulated t-values that terms up to degree 8 should be included. Unfortunately, if this is done, then, as shown on Fig.4, an undesirable waviness results at the ends of the curve. This is due to excessive 'clumping' of the data, so that the effective number of degrees of freedom is reduced and the situation becomes very close to the exact fitting of a degree-n curve to $n + 1$ data points.

A further instance of the need to beware of making curve-fit judgements solely on the basis of the t-value is illustrated on Fig.5. This shows yawing-moment coefficients obtained in tunnel A. Since the model was configured symmetrically about the lateral plane, all these results represent errors. The polynomial regression of C_n on α produces t-values greater than two for the first five coefficients. However, it is unlikely that there is any physical basis for a quartic relationship, so a constant fit (i.e. mean value) has been drawn on Fig.5, as the zero-degree term had by far the largest t-value. The repeat experiment in tunnel B produced the data shown on Fig.6. In this instance, a linear dependence seemed the most sensible form of curve fit.

3.3 Comparisons

The accuracy of the experiment in determining a relationship between angle of incidence and an aerodynamic coefficient is indicated by the amount of variance of the curve fit. This is given by expression (12) in which the mean residual variance s^2 results from the inclusion of all significant terms in the regression. The first term in (12) is equivalent to the square of the estimated standard error of the mean of the measured coefficients. The other terms are dependent on angle of incidence and their total value is a minimum at the mean angle. The variance of the curve fit can be visualised by plotting confidence limits for the regression. These are shown at the 95% level on all the figures by the inner pair of dashed lines. If another, similar experiment was carried out, with the same number of data points, then it is 95% probable that the regression line would lie between the two confidence limits derived for the first experiment.

On Fig.5, these lines are parallel to the (constant) regression; on the next figure, however, the extra degree of freedom of the regression line (the uncertainty about its slope) makes its end values less reliable than those nearer the mean angle of incidence, so that the confidence limits are curved.

The two figures are drawn to the same scale so that the differing amounts of scatter are obvious. The outer pair of dashed lines represent 95% confidence limits for an individual result, given by expression (13). They indicate the accuracy of a prediction and it is to be expected that, on average, 95 out of 100 data points will be contained by these limits. Fig.2, for example, shows 130 points, so that six or seven might be expected to lie outside the lines and indeed this is the case.

Comparison of the vertical separations of the confidence limits on Figs.5 and 6 gives a direct indication of the relative accuracies of the two tunnels in determining C_n. Clearly, tunnel B is inferior to A in this respect. The scatter is inherent in the experiment and cannot be altered unless the cause is identifiable and can be corrected. However, a prime reason for curve fitting is to make use of the information provided by all the data points, and the more of these there are, the more accurate will be the regression. This idea is developed further in section 4.

The comparisons of the measurements of side force and rolling moment in the two tunnels are similar to Figs.5 and 6. The longitudinal coefficients are compared on Figs.7 to 12. The coefficients are uncorrected values so that the differences in mean levels can be attributed to different blockage factors, incidence constraints, etc. for the two tunnels. The question now arises of whether or not there is a significant difference between the capabilities of the two tunnels to measure a certain quantity. This could be done visually, for example the vertical separations of the confidence limits on individual results for Figs.7 and 8 look about the same. Turning to the next pair of figures, is the apparent smaller scatter for tunnel B statistically significant? More precisely, is the difference between the two estimated standard deviations larger than can reasonably be explained by errors in the experiment? Clearly the number of observations used to obtain the estimates is important, and the question is best quantified by applying the F-test (see section 2.5). Table 5

Table 5

Comparison of estimated experimental standard deviation for the low-lift configuration

	Estimated σ		F	P
	Tunnel A	Tunnel B		
C_L	0.01337	0.01422	1.13	0.5
C_D	0.00453	0.00193	5	<0.001
C_m	0.01211	0.00829	2.13	<0.001
C_ℓ	0.00078	0.00390	25	<0.001
C_Y	0.00375	0.00759	4	<0.001
C_n	0.00038	0.00113	9	<0.001

compares the estimated standard deviations for all six components for the low-lift configuration and shows in the fourth column the square of the ratio of the two estimates (with the larger as numerator). The null hypothesis to be applied is that any apparent difference in estimated σ (F larger than one) is merely due to random errors of measurement. Tables of F are entered at the appropriate point to determine the value of P, the probability that the difference is due to chance. This is a double-sided test and the probability points appropriate to 120 degrees of freedom in both denominator and numerator[3] have been plotted on Fig.13 (P is on a logarithmic scale). Column 5 of Table 5 shows the relevant values of P, extracted from this figure. Regarding the value of P = 0.5 for lift coefficient, this is clearly not significant and there is no reason to reject the null hypothesis. For all the other coefficients, the differences are very highly significant. Tunnel A is clearly superior to B for the lateral coefficients, inferior for the measurement of drag and pitching moment and just as good for lift.

The same statistical test can be applied in other contexts. Figs.3 and 7 show the effect on the measurement of lift in tunnel A of model changes. For the consecutive runs, the model remained untouched between runs, the wind was merely turned on and off and the balance rezeroed. Table 6 compares the standard deviations and shows the probability P that the differences are not

Table 6

Comparison of standard deviations for separate and consecutive runs in tunnel A

	Estimated σ		F	P
	Separate runs	Consecutive runs		
C_L	0.01337	0.00759	3	<0.001
C_D	0.00453	0.00122	14	<0.001
C_m	0.01211	0.01236	1.04	>0.2
C_ℓ	0.00078	0.00045	3	<0.001
C_Y	0.00375	0.00154	6	<0.001
C_n	0.00038	0.00026	2.1	<0.01

systematic, but due to chance. (The values of P were not extracted from Fig.13 since the estimates of σ for the consecutive runs are based on 49, not 120 degrees of freedom.) There is a highly significant improvement in all coefficients except for pitching moment.

It is ironic that the reason for performing the consecutive runs in the first place was the poor results for pitching moment. What has happened is that firstly the method of assuming linear drift of the balance zero throughout the run was successful in correcting for bad post-run zeros. Secondly, the consecutive runs were completed after the breakdown of the incidence-changing equipment in tunnel A and its subsequent repair. This effect could be checked statistically by applying the F-test to the data from the four pre-breakdown runs and the six post-breakdown runs. For tunnel B, the consecutive runs produced a decrease in s for lift coefficient from 0.01422 to 0.01334, which is not significant (P = 0.7) and for drag, a decrease from 0.00193 to 0.00152, which is just significant (P = 0.05).

3.4 Accuracies

In statistical terms the accuracy of an experiment in determining, say, the dependence of lift on angle of incidence, is equivalent to the variance of the regression of C_L on α. This is given by expression (12) and it is clear from the various examples shown on the figures that when the variance is visualized by means of confidence limits, it is usually a weak function of incidence. In other words, for the assessment of accuracy consideration of the mean residual variance and the number of data points is sufficient. Thus the accuracy of a curve fit can be expressed at a certain confidence level by

$$\pm t_a s/\sqrt{n} \qquad (16)$$

where t_a is the appropriate Student's t-factor, s the estimated standard deviation appropriate to a regression of sufficient degree to include all significant terms and n the number of data points used. For the comparisons given in Table 5 between the two wind tunnels with the model in the low-lift configuration, n is the order of 100, so that at the 95% confidence level the appropriate value of t_a is 1.98. Table 7 gives the estimated accuracies of these results; the values should all be preceded by the $\pm$ symbol.

Table 7

Comparison of estimated accuracies at the 95% level for the low-lift configuration

	Estimated accuracies	
	Tunnel A	Tunnel B
C_L	0.00266	0.00283
C_D	0.00090	0.00038
C_m	0.00241	0.00165
C_ℓ	0.00016	0.00078
C_Y	0.00071	0.00144
C_n	0.00008	0.00023

This table shows, for example, that after ten repeat runs, drag in the low-lift configuration could be measured to within nine 'counts' in tunnel A and four 'counts' in tunnel B. However, after five consecutive runs in tunnel A, drag could be measured to an accuracy of $\pm 2.003 \times 0.00122/\sqrt{55}$ or three 'counts' at the 95% confidence level. These consecutive runs were performed after the breakdown and repair of the incidence mechanism, so that the comparison may indicate an improvement due to the repair, as well as the benefit of minimising model-configuration changes. It is of interest to compare the results quoted in Table 7 with those obtained for the high-lift configuration. On Fig.2, the estimated standard deviation of the curve-fit through 129 points is 0.00789, which gives an accuracy at the 95% level of ±14 drag 'counts'. It is possible that at the higher lifts, the defects in the incidence mechanism gave rise to proportionately larger errors in drag. However, the prime cause was propably errors in setting up and maintaining engine thrust in both pods at the values required for constant blowing-momentum coefficient.

Usually an aerodynamic coefficient is a function of more than one independent variable, but tests are conducted varying only one of these at a time. The effect of a second variable is found by subtraction of one set of results in which this variable takes one value, from another set in which it takes a second value. It is most appropriate that the actual subtractions are performed on least-squares curve fits and it has to be assumed that the errors in the two sets of results are completely independent of each other (i.e. uncorrelated). It is a general principle in statistics that the variance of a combination is the sum of the variances of its component parts. Thus if s_1 and s_2 are the estimated standard deviations of the two curve fits based on n_1 and n_2 data points respectively, the standard error of the difference between the two curves is

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad . \tag{17}$$

This expression is approximate since the first term only of the variance given by (12) has been used: the dependence on angle of incidence has been ignored.

In practice it is likely that comparable numbers of data points will be used in each configuration and that the variances of both sets of results will be similar. Thus the standard error of the difference between the two sets is $s\sqrt{2/n}$. Table 5 gives some values of s found after ten runs of the model in both wind tunnels. If these values can be taken to be typical of similar runs with small changes in model geometry, then they can be used to obtain confidence limits on the effects of these changes on the aerodynamic coefficients. At the 95% level, the results will be approximately $\sqrt{2}$ times the values shown on Table 7 since t_a is hardly changed going from 100 to 200 degrees of freedom. In tunnel A, for example, 95% confidence limits on an increment in rolling moment are ±0.00023. Hence at this level of confidence it is not possible to measure increments in rolling moment which fall within these limits, i.e. less than 0.00046. In tunnel B, on the other hand, the minimum increment in rolling moment that can reliably be detected at the 95% level is 0.00221.

4 NUMBER OF POINTS FOR A GIVEN ACCURACY

In many practical situations a wind-tunnel programme is designed to test as many parameters or configurations as possible, within the allowable time. This is often dictated by pressure of other tests or by the available funds. Accuracy of results is estimated on the basis of supposedly known performances of the wind-tunnel and its data read-out equipment.

However, it is conceivable that a programme could set out to measure certain quantities to a predetermined accuracy. For example, it may be pointless to do the experiment if drag cannot be measured to within plus or minus so many 'counts'. Such results could only be guaranteed after an infinite number of observations had been made, when there would then be complete certainty about σ, their standard deviation. This is impossible of course and the desired accuracy has to be set against a confidence level. If an estimate s of the standard deviation is known in advance, then the variance of the curve fit or standard errors of the regression coefficients can be made as small as is desired, by increasing the number of data points. If this proves to be too expensive, or time consuming, a lower confidence level has to be adopted.

For tests on a well-tried model, in a wind tunnel with known characteristics and using a regular tunnel crew, the standard deviation for a particular coefficient may well be known in advance. On the other hand, this is unlikely to be the case for a research model and the values of s have to be determined during the course of the experiment. The question that has to be asked in this situation, therefore, is how many data points are required in order to obtain a good estimate of σ for any particular configuration and coefficient?

It might be expected that as more and more observations are made, their variance for a given angle of incidence will settle down to a constant value, which will give a good estimate of σ. That is, the width of the outer pair of confidence limits on the figures will remain much the same although the density of points increases. In the experiments in tunnels A and B, an arbitrary number of ten repeat runs were carried out for the two configurations and five for the consecutive runs. For the consecutive runs in tunnel A, a regression analysis has been carried out for each coefficient, using results from the first run only, the first plus second, first plus second plus third, etc. Fig.14 shows the effect of the number of runs included in the regression of lift coefficient on angle of incidence, as well as the degree of regression. If a degree 6 regression is considered sufficient (*cf*. Fig.3), then in this case only two runs are required to establish a value of 0.0077 for s .

The next figure shows some results for the regression of C_L on α for the ten runs in the high-lift configuration. It is apparent that the inclusion of points from run 4 increased the residual variance for all degrees of regression. It would appear that this was a 'bad' run and its scatter was atypical in some way. The effect of including points from the seventh run is similar and it takes another two or three runs for the standard deviation to settle down again. The very poor results for yawing moment in the low-lift configuration as measured in tunnel B (see Fig.6) are reflected in Fig.16. Even after ten runs, no consistent estimate of σ has been achieved.

Figs.14 to 16 make it clear that the residual variance is very sensitive to experimental error. It is customary to spot gross errors during the course of the experiment by plotting the results as the run progresses, and repeating incidences as required. If on-line computing facilities are available, it would be possible to calculate and display s as each data point is acquired. Bad points could then be easily detected and deleted. Once s is reliably known, it is a simple matter to employ expression (16) to calculate the number of data points required to achieve the desired accuracy of curve fit.

5 CONCLUSIONS

This Report has described the use of some standard statistical techniques in the comparison of results obtained from a wind-tunnel model tested in two tunnels. The basis of the comparisons was least-squares curve fits using orthogonal polynomials. Excluded from the regressions were obviously 'bad' points and those for angles of incidence above the angle where the scatter was noticeably worse due to the commencement of the stall. The degree of regression was primarily selected by including all the coefficients of the polynomial curve which were statistically significant. However, if the data points were too strongly grouped or clumped together, then occasionally unwarranted oscillations appeared in the regression curve. This is a situation which often arises when fitting a degree k curve through k - 1 data points. For this reason, if it is intended to apply the techniques of this Report to other tunnel tests, it is best to distribute the data points uniformly over the range of incidence of interest, rather than to try and duplicate readings at nominal angles.

The idea of 'repeatability' of a result has been identified with the degree of scatter or variance of an individual data point for a given angle of incidence. This scatter can be expressed in terms of limits within which a repeat point can confidently be expected to lie, for a prescribed probability level. For the model and tunnels used in these tests, this variance is practically independent of angle of incidence and is equal to the residual variance of the curve fit. It is thus fundamental to the nature and characteristics of the experiment and, if all the data points are equally reliable and sufficient of them have been obtained (30 say), it cannot be improved by further tests.

The overall reliability of the regression for a particular aerodynamic coefficient can be visualized graphically by drawing the upper and lower confidence limits (at, say, the 95% probability level) on either side of the regression curve. If the results from the two tunnels are plotted using similar scales, they can be compared at once. If there is a difference between the two estimates for the residual variance, then this could be a fundamental, statistically significant difference, or it could arise naturally through the limited numbers of data points used. This question can be answered by application of the ratio test which gives the probability that the observed difference is merely due to chance.

Clearly, the reliability of the regression, or the precision with which it can be used to predict a coefficient for a given angle of incidence, is dependent on the number of data points used, as well as the residual variance. If the latter remains roughly constant, then the more data points employed, the more precise is the curve fit. Confidence limits at a prescribed probability level can be deduced for the true value of a coefficient at a given incidence. The widths of these limits are generally very weakly dependent on α but almost directly proportional to the standard deviation of the curve fit and inversely proportional to the square root of the number of data points. Thus the customary way of reporting that a result is such and such, plus or minus an error has been refined by relating this error to a confidence level. For example, $C_L = 2.0 \pm 0.015$ at the 95% level (at the 98% level, the error might be ±0.022). These errors can be deduced from the actual experimental results, rather than estimating them from supposedly known accuracies of the tunnel balance, speed control, etc.

Very often the effect of a change to the geometry of the model is required to be found from the difference between two sets of results. The minimum effect that can confidently be detected at any specified angle of incidence will be equal to the width of the confidence limits on the difference curve at that angle. Thus the minimum detectable effect can be made smaller if the degree of confidence required in it is relaxed. If the basic results have similar standard deviations and are based on roughly equal numbers of data points, then this minimum detectable effect will be approximately $\sqrt{2}$ of the confidence width of a basic curve.

An analysis of the effect of number of runs on the residual variance of best-fit curves has shown that they are very sensitive to the quality of the data. If these are all of similar worth, then in one example two or three runs of 13 points each were sufficient to establish a reliable value for s ; further runs had no significant effect. In other examples, however, the inclusion of bad points from one or two runs considerably worsened the overall scatter and a consistent value for the experimental standard deviation was not obtained even after ten runs. It is suggested that the standard deviations of curve fits for each coefficient should, if possible, be computed after the acquisition of each data point during a tunnel run. It will then be possible to detect and reject atypical points. Once a reliable estimate of s has been obtained, it is then possible to calculate the number of data points necessary to measure a coefficient to a given accuracy at a prescribed confidence level.

Acknowledgments

The author expresses his appreciation to the members of the Wind Tunnel Department of Hawker Siddeley Aviation Limited (Hatfield) for their assistance with the experiments and also to Mr. J. Withrington of the Research and Future Projects Department for discussions and help with the analysis.

Appendix

GLOSSARY OF STATISTICAL TERMS USED

Some of these terms are illustrated diagrammatically on Fig.17.

Confidence limits - Limits defining the confidence interval within which, for a stated probability, the true value lies.

Degree of freedom - A statistic has degrees of freedom equal to the number of observations involved, less the number of constraints employed in its derivation. For example, calculating the mean of n points 'uses up' one degree of freedom, so that the mean, estimated variance, etc. have $n - 1$ degrees of freedom. The coefficients of a straight-line fit would have $n - 2$ degrees of freedom.

Mean - In this Report, the arithmetic mean or average is implied. If the data points are from a sample of a population distributed normally, their average value is the best possible estimate of the population mean.

Normal distribution (or Gaussian distribution) - A particular kind of frequency distribution that has convenient statistical properties. Fortunately, it commonly arises naturally or in experimental work. If α is distributed normally with mean $\bar{\alpha}$ and standard deviation σ, and is transformed to a new variable μ by $\mu = (\alpha - \bar{\alpha})/\sigma$, then the probability density of μ can be visualised as the normal curve in standard form, illustrated on Fig.17. μ has unit standard deviation and zero mean; the area under the curve is unity. On the figure, the area to the left of the line at μ' represents the probability of μ being less than μ'. The two hatched areas each cover 2.5% of the total area and have bounds (the confidence limits) at $\mu = \pm 1.96$. The probability of μ being within 1.96 standard deviations from the mean is thus (1 - 2a) or 95%.

Null hypothesis - A belief that two quantities are, in fact, really equal and that the apparent difference is due to errors of observation. The hypothesis is put to test and is rejected if the probability of it being true is sufficiently small.

Population or universe - The complete set of all possible observations of the kind being made.

Regression - The relationship between the mean value of an observed quantity Q and the independent variable α is called the regression of Q on α. If a one-one correspondence exists between Q and α, then the regression is equivalent to the functional relationship $Q(\alpha)$. Usually the regression must be estimated from a sample of observations. The estimated regression $y(\alpha)$ would tend closer to the true regression $Q(\alpha)$ as the size of the sample was increased.

Residual variance - The complete summation of the squares of the observed values ('total sum of squares') is in part due to the nature of the regression $Q(\alpha)$ ('sum of squares due to regression'). The remainder is called the 'sum of squares about regression' and, when divided by the appropriate number of degrees of freedom, gives an estimate of the residual variance.

Sample - A subset, or finite set if the distribution is continuous, of a population.

Significance - A quantity is statistically significant if its confidence interval does not contain zero.

Standard deviation - A measure of the average scatter or deviation of a sample of observations from their mean value.

Standard error - The standard deviation of a mean or averaged quantity. Whatever the nature of the distribution from which a sample of n observations is taken, if its population standard deviation is σ , the standard error of the mean is $\sigma/\sqrt{n}$.

Test of significance - Such tests are used to investigate the reasonableness of a hypothesis and to see, for example, if there is a significant difference between two quantities at a given level of probability.

Variance - The square of the standard deviation; a measure of the variability of the observations.

SYMBOLS

a	probability level used in t-test
b_i	coefficients of regression curve in orthogonal form
B	column vector of b_i in normal equations
c_i	coefficients of regression curve in normal form
C	square matrix of the sums (ϕ_p, ϕ_q) in normal equations
C_D	drag coefficient
C_ℓ	rolling-moment coefficient
C_L	lift coefficient
C_m	pitching-moment coefficient
C_n	yawing-moment coefficient
C_Y	side-force coefficient
D_j, E_j	coefficients in the recurrence relation for ϕ_i
F	ratio of two variances
k	degree of regression
M_j	modulus of ϕ_j over the point set α_i
n	number of data points
P	probability
Q	observed quantity, dependent variable
s	estimate of σ
S	sum of squares of errors
SE	standard error
t	statistic used to test the significance of a quantity
V	variance
y	value of Q predicted by the regression curve
Y	column vector in normal equations
α	independent variable, angle of incidence
ε	error
ϕ_i	orthogonal polynomial of proper degree i in α
σ	population standard deviation

REFERENCES

No.	Author	Title, etc.
1	O.L. Davies	*Statistical Methods in Research and Production.* Third edition, Edinburgh, Oliver and Boyd (1957)
2	G.E. Forsythe	Generation and use of orthogonal polynomials for data-fitting with a digital computer. *J. Soc. Ind. Appl. Math.*, 5, 74 (1957)
3	C.S. Pearson H.O. Hartley	*Biometrika Tables for Statisticians.* Vol.1, Cambridge University Press (1958)

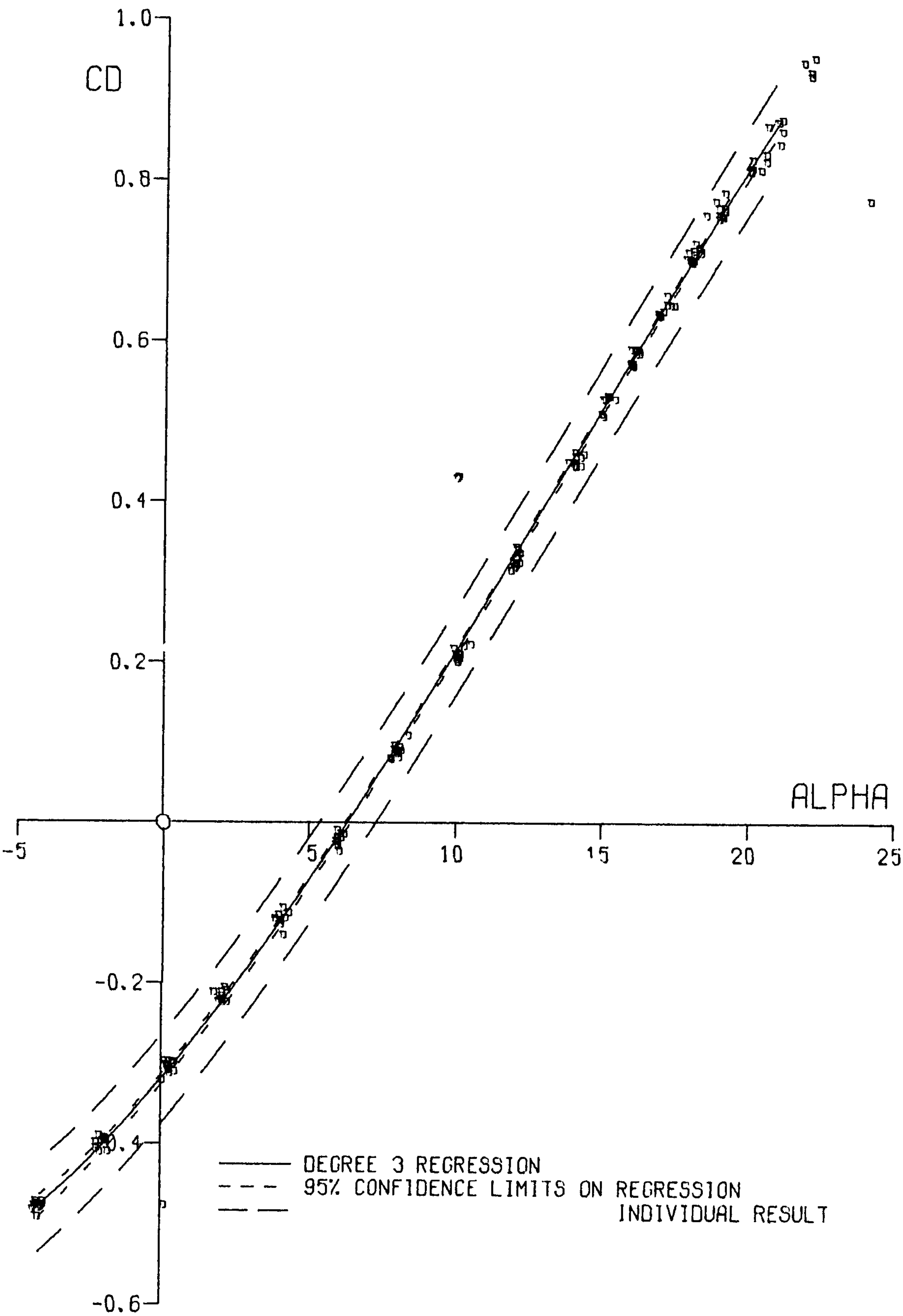


Fig.1 Drag coefficient (Tunnel A, high-lift configuration) all data points

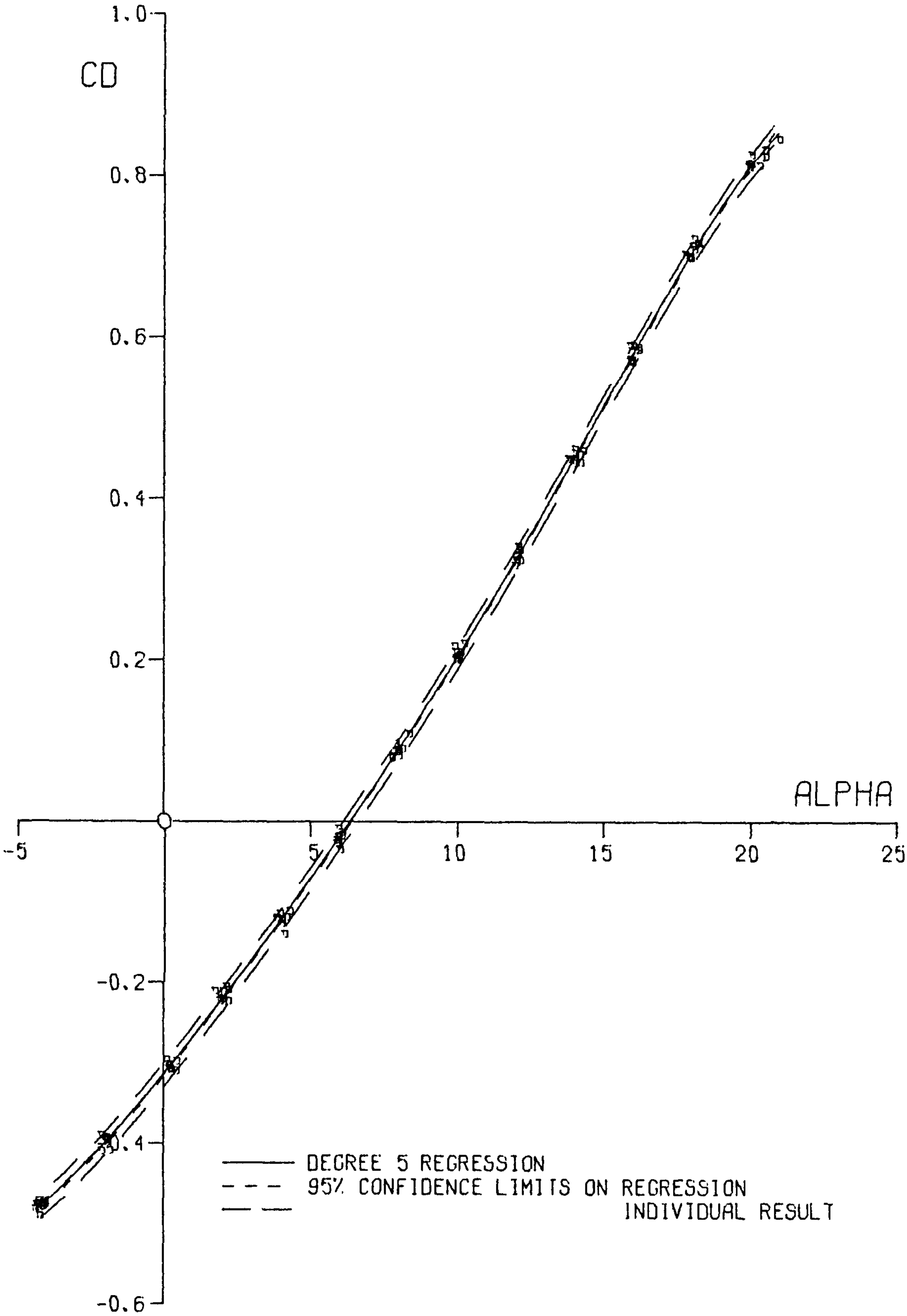


Fig.2 Drag coefficient (Tunnel A, high-lift configuration) selected points

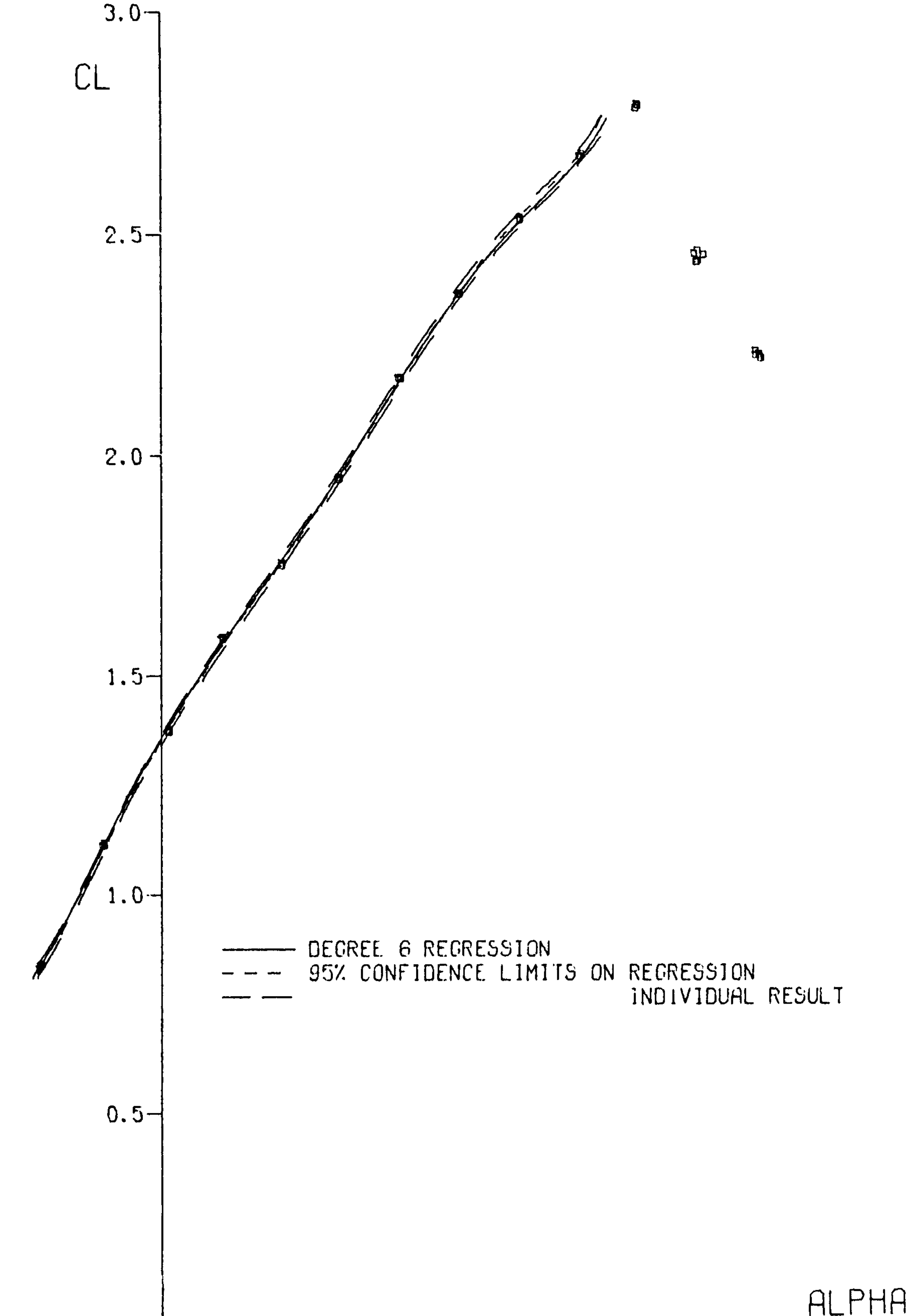


Fig.3 Lift coefficient (Tunnel A, consecutive runs) selected points, degree 6

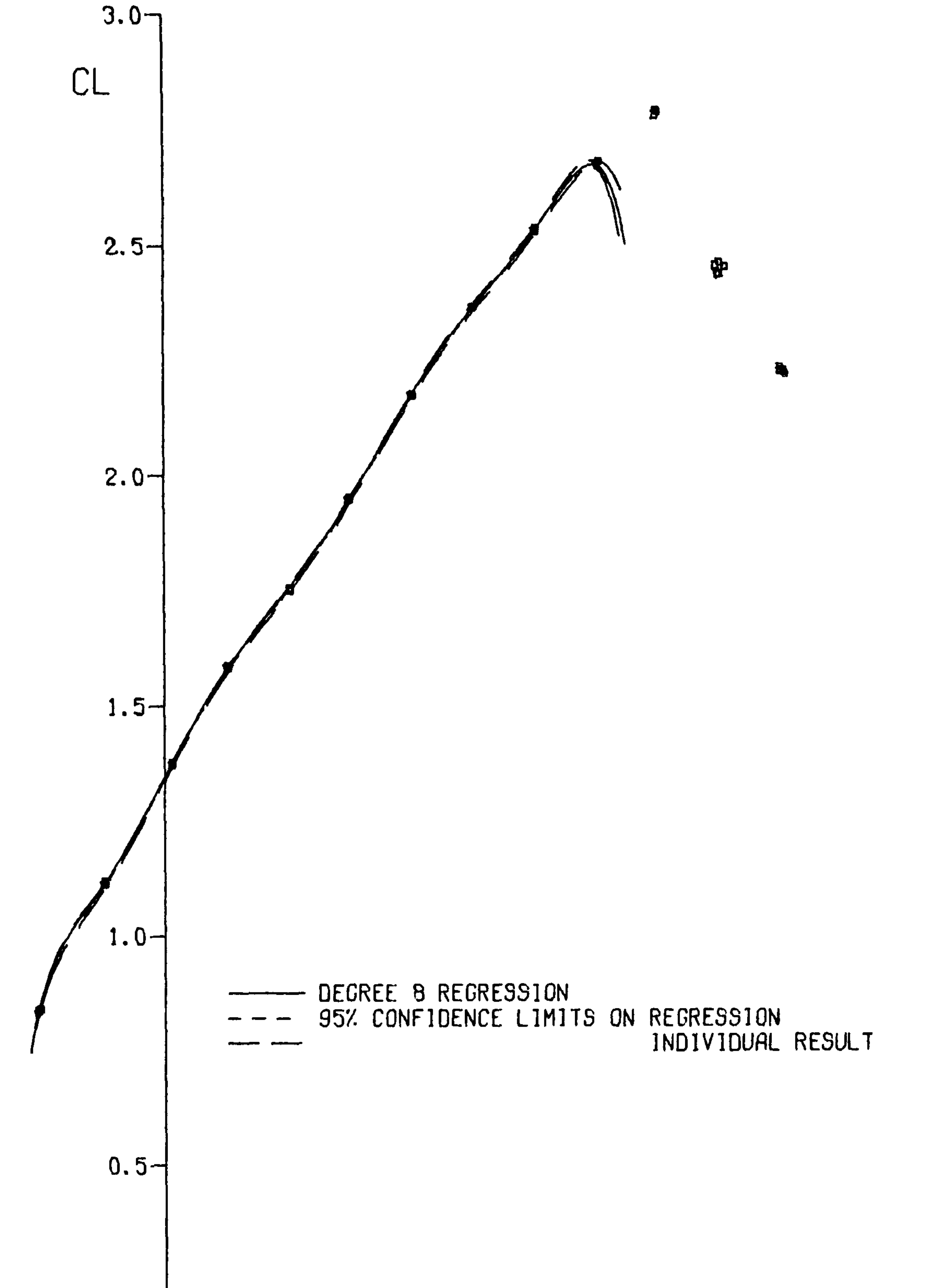


Fig.4 Lift coefficient (Tunnel A, consecutive runs) selected points, degree 8

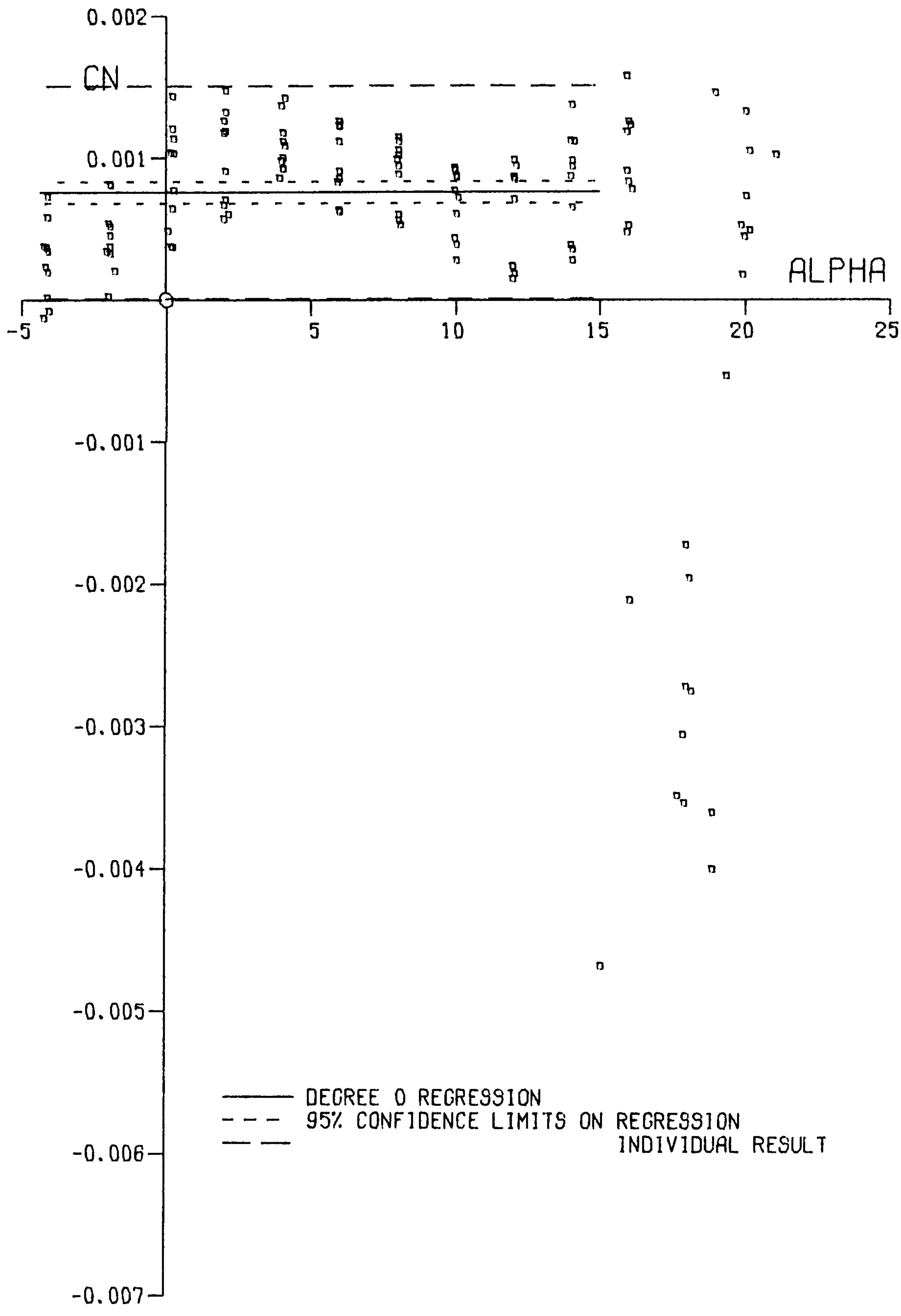


Fig.5 Yawing moment (Tunnel A, low-lift configuration) selected points

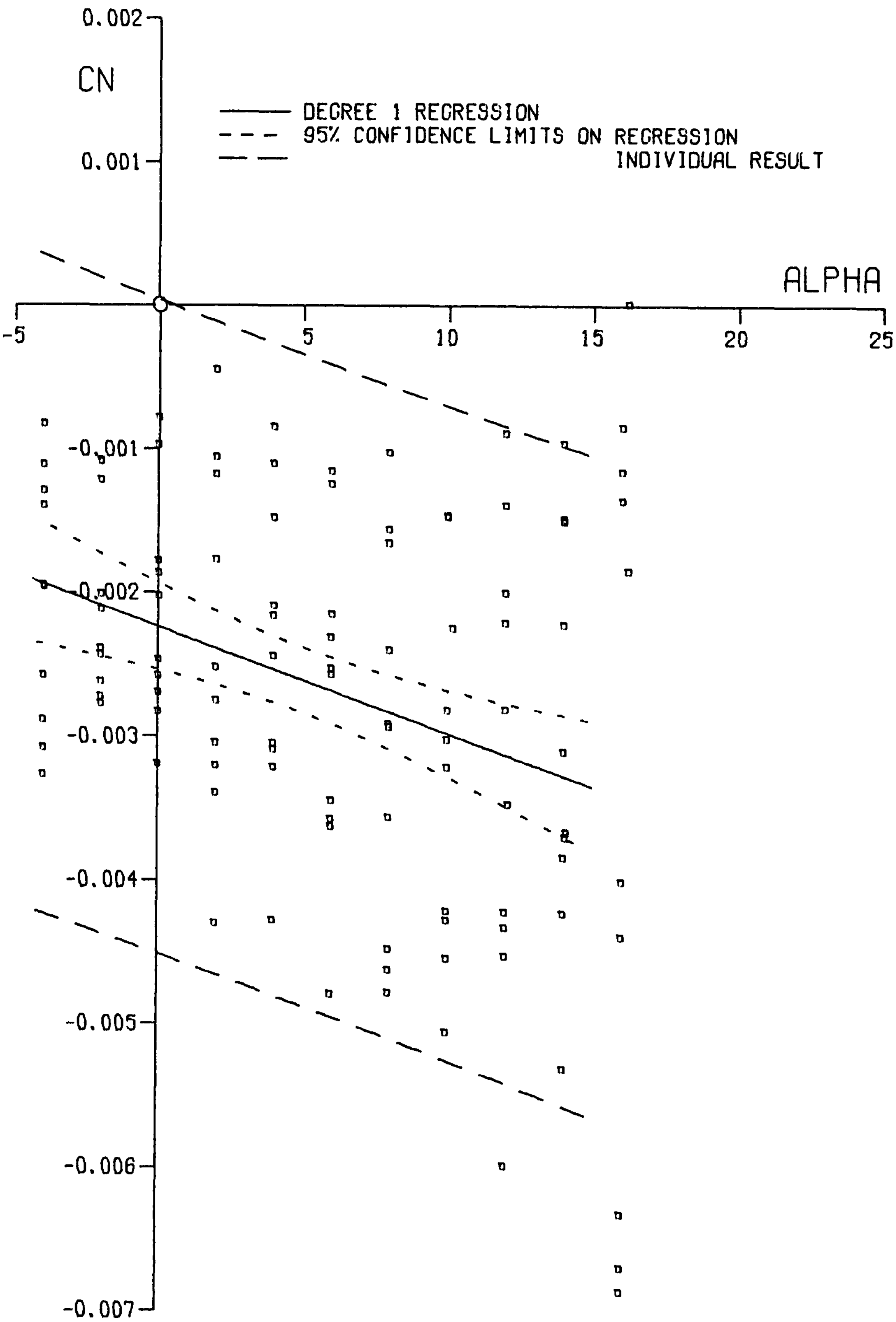


Fig.6 **Yawing moment (Tunnel B, low-lift configuration) selected points**

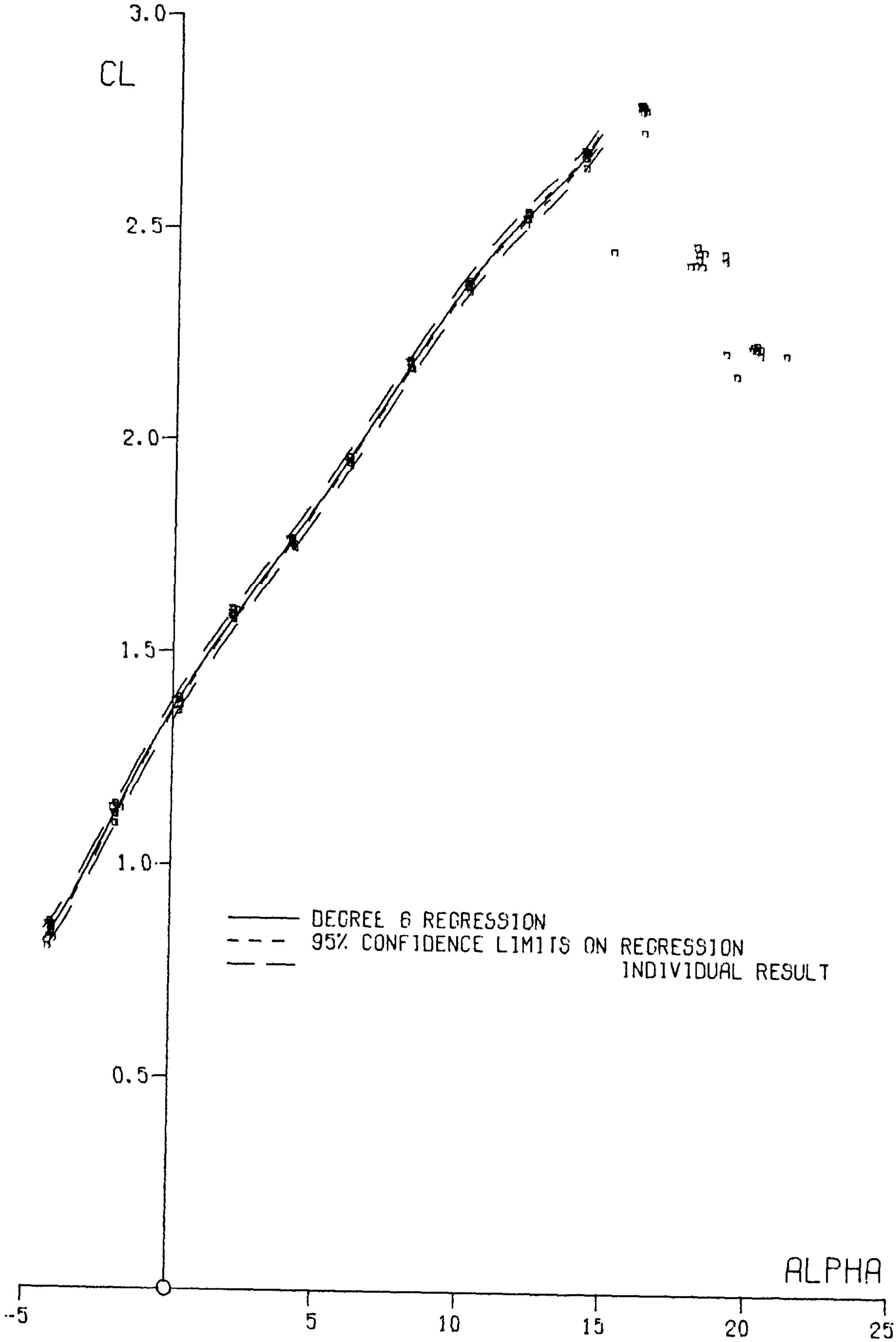


Fig.7 Lift coefficient (Tunnel A, low-lift configuration) selected points

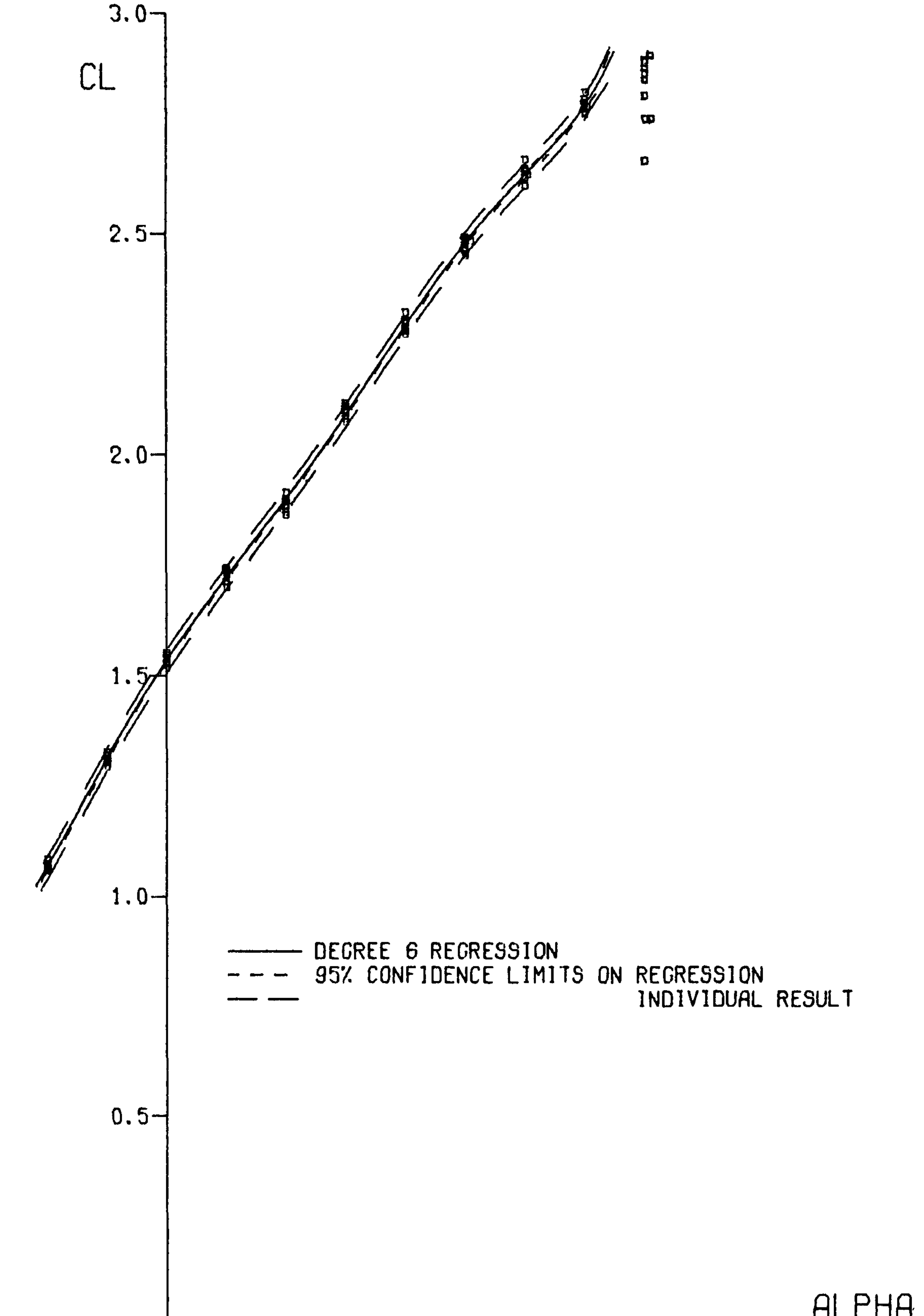


Fig.8 Lift coefficient (Tunnel B, low-lift configuration) selected points

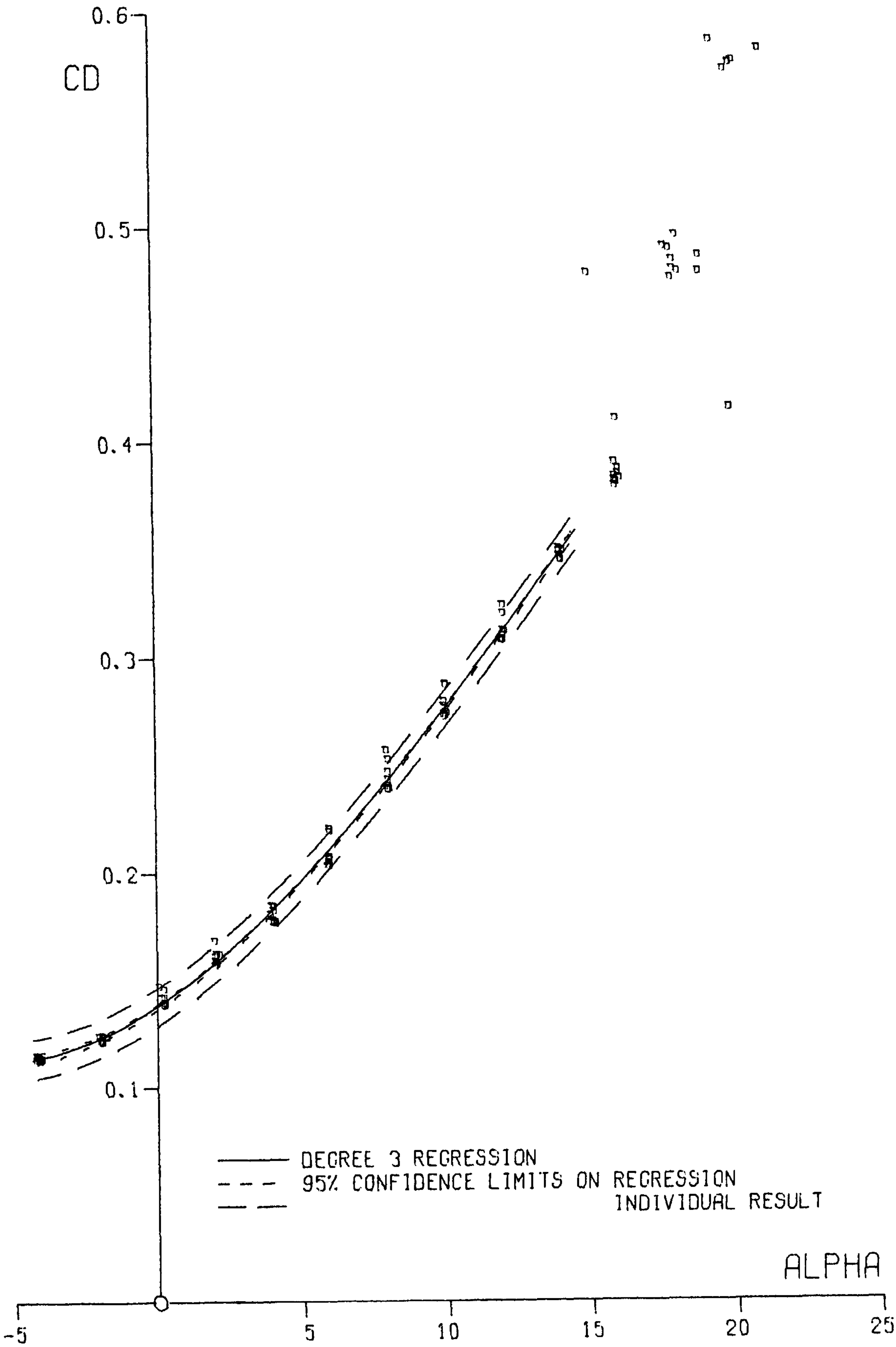


Fig.9 Drag coefficient (Tunnel A, low-lift configuration) selected points

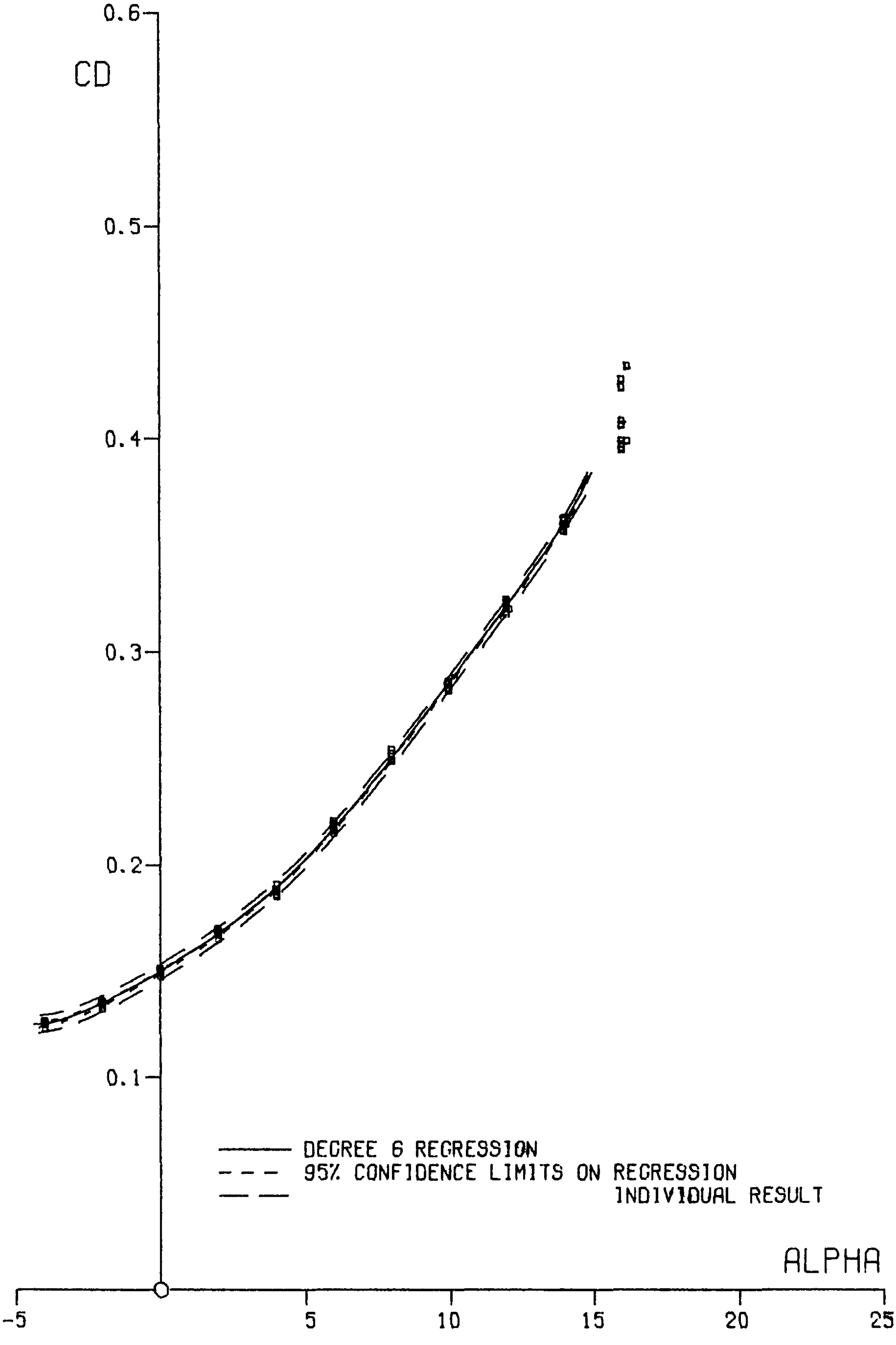


Fig.10 Drag coefficient (Tunnel B, low-lift configuration) selected points

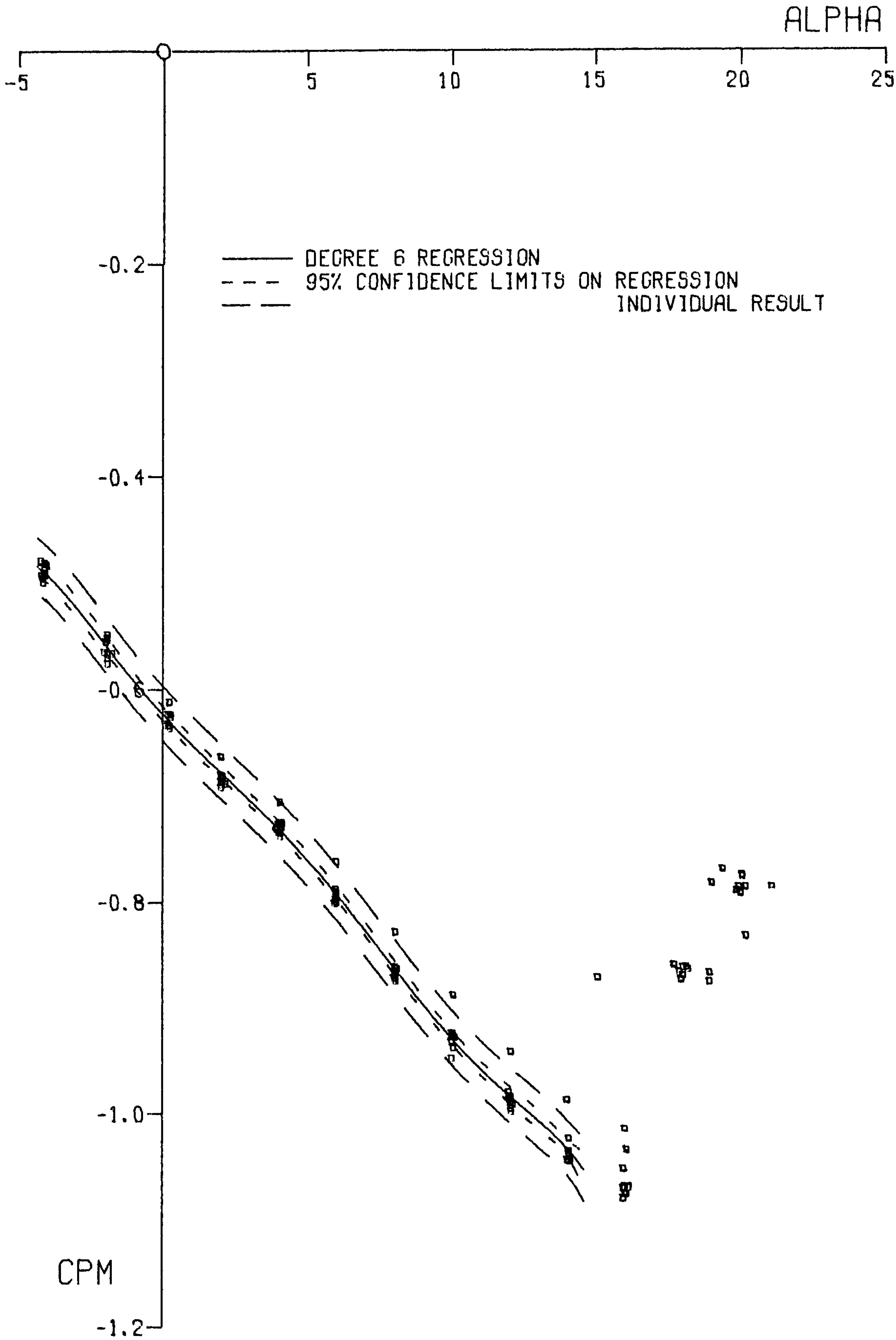


Fig.11 Pitching moment (Tunnel A, low-lift configuration) selected points

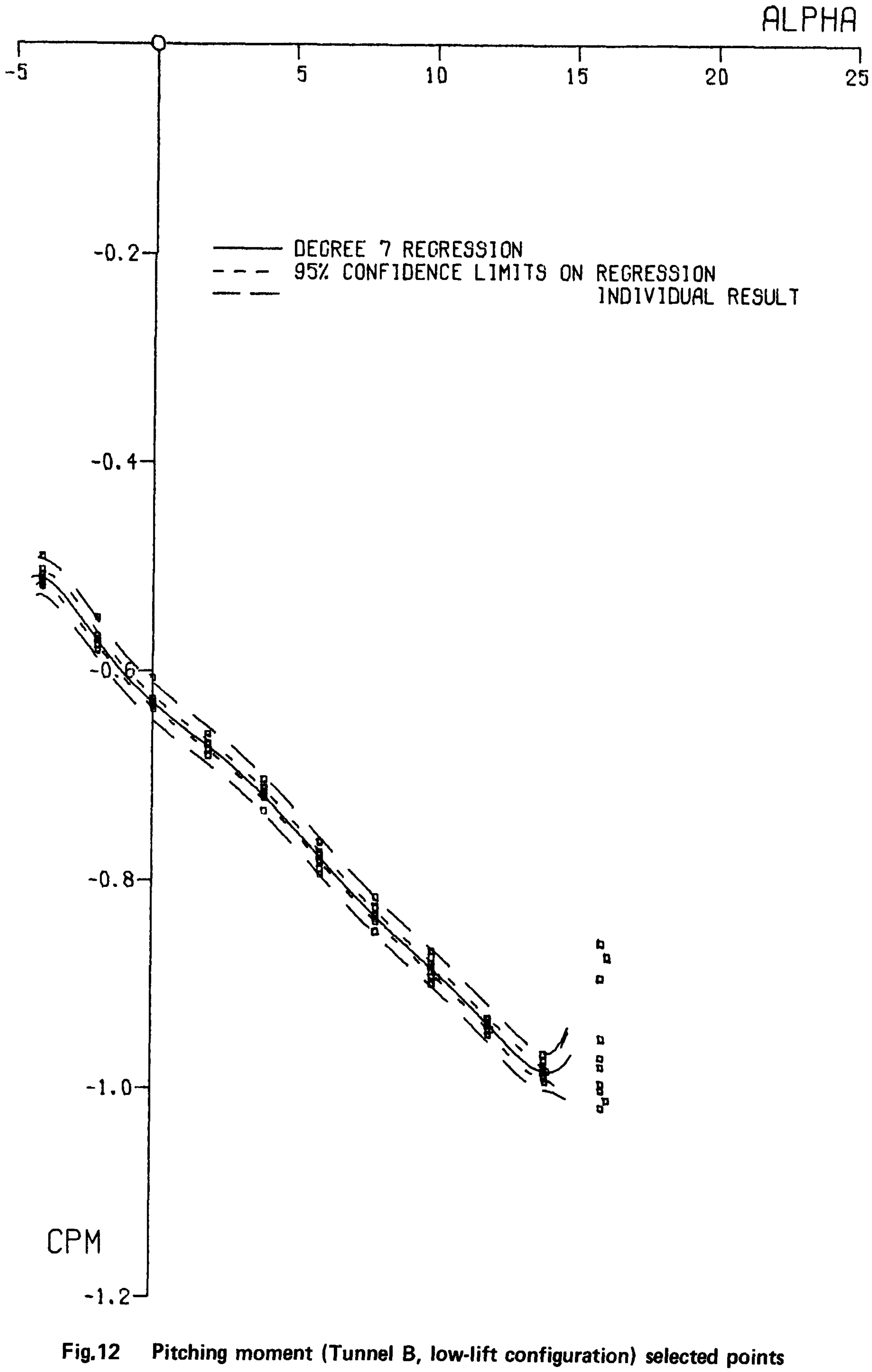


Fig.12 Pitching moment (Tunnel B, low-lift configuration) selected points

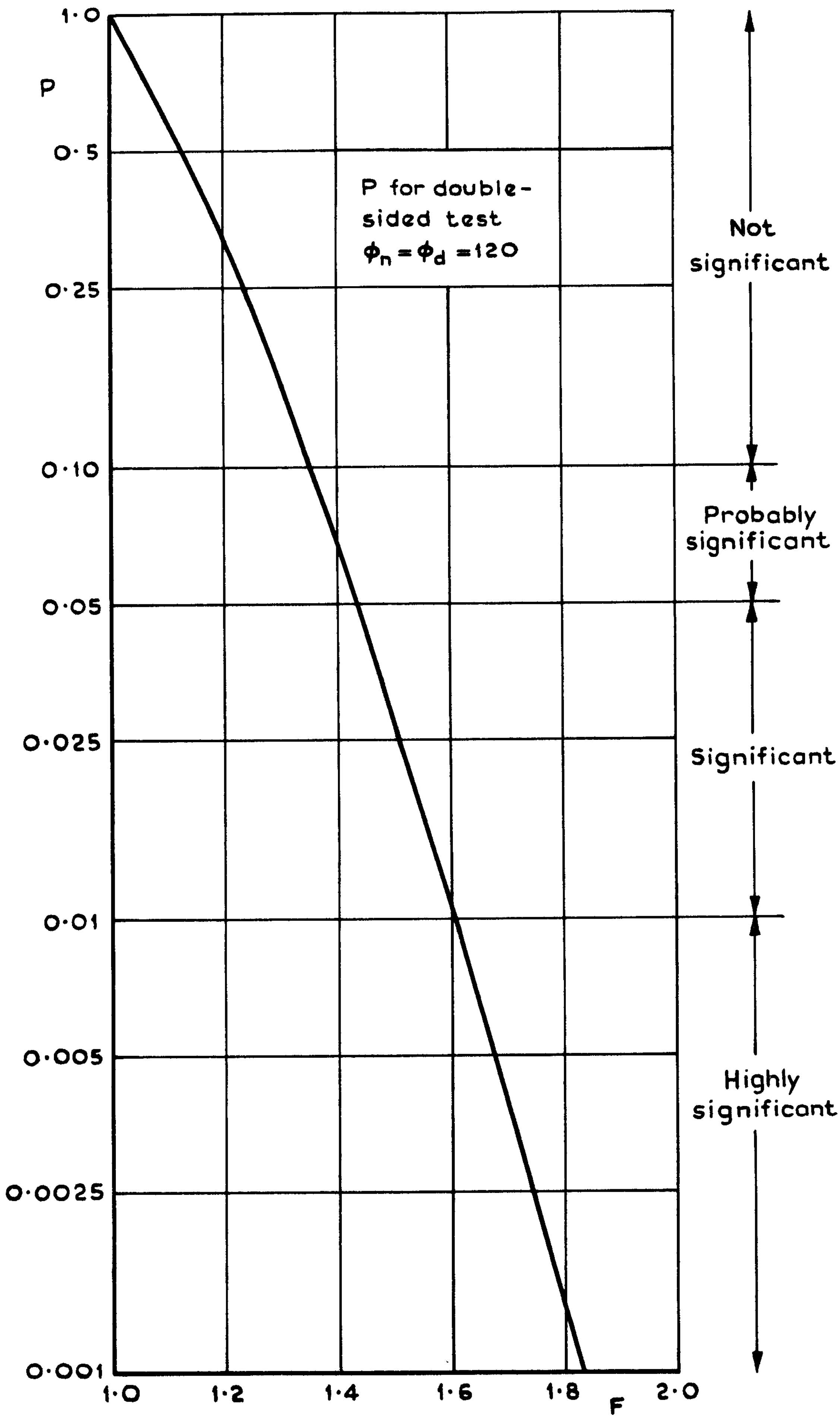


Fig.13 Probability curve of the variance ratio

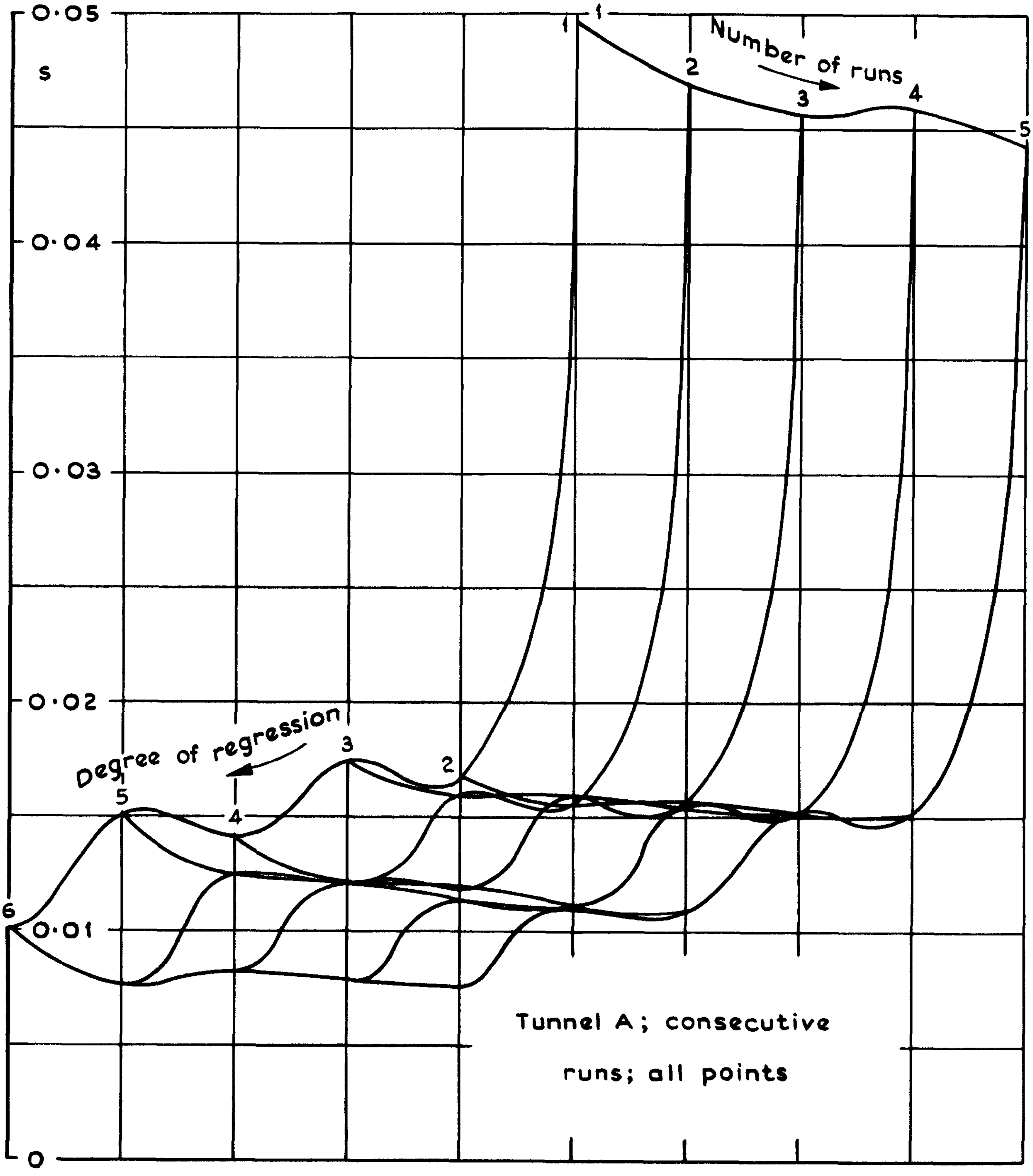


Fig.14 Effect of number of runs on s for the regression of C_L on α

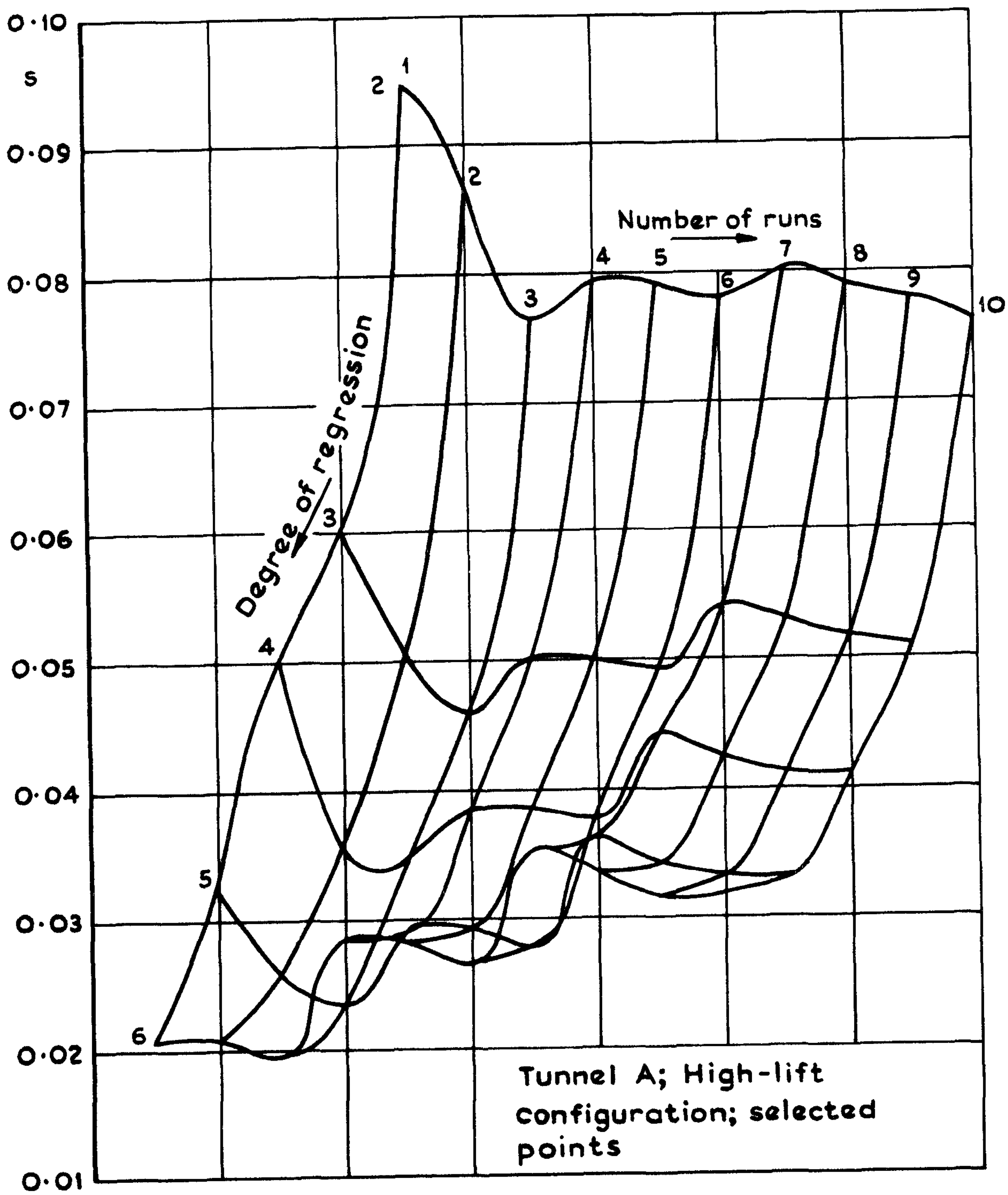


Fig.15 Effect of number of runs on s for the regression of C_L on α

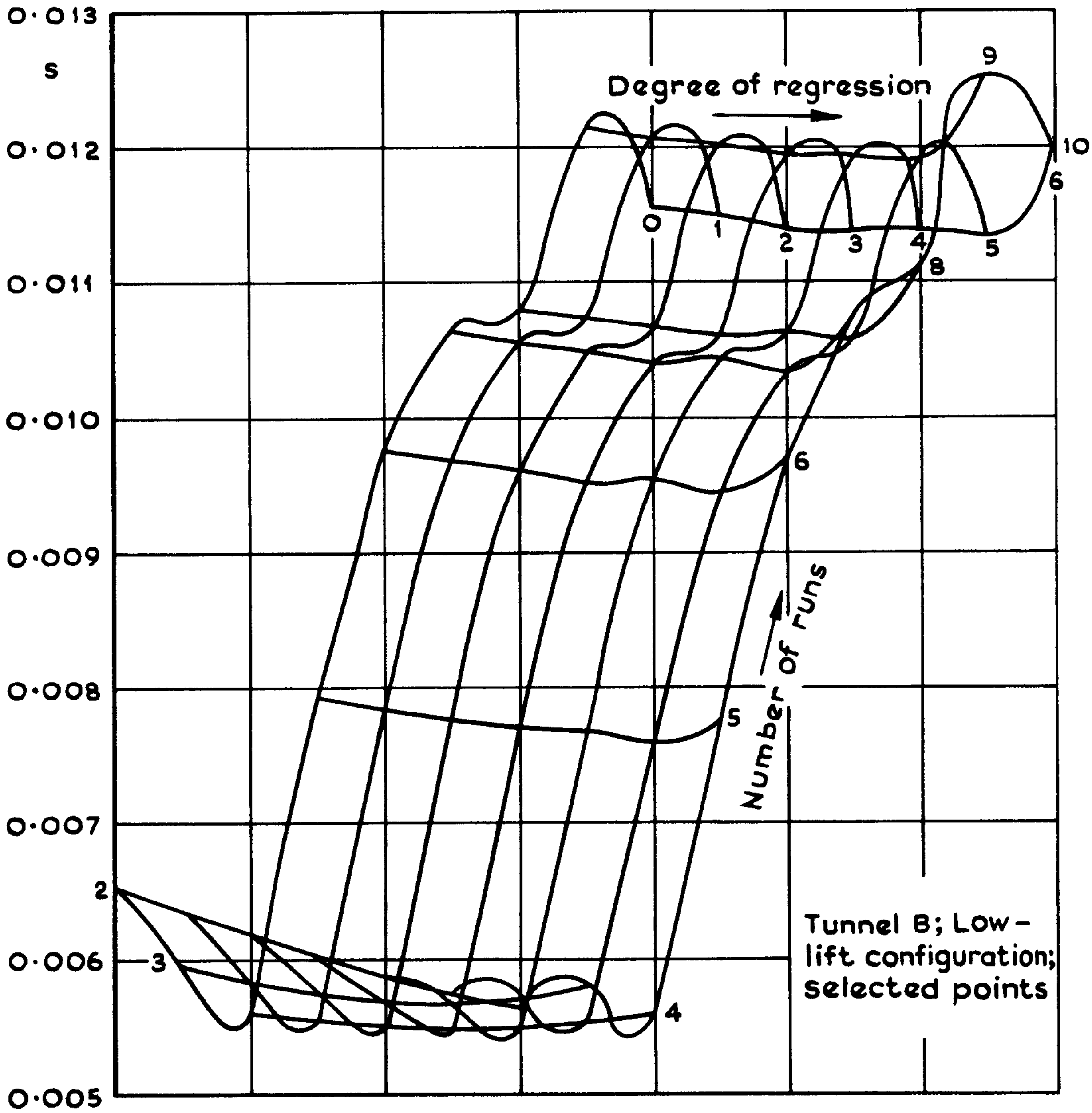
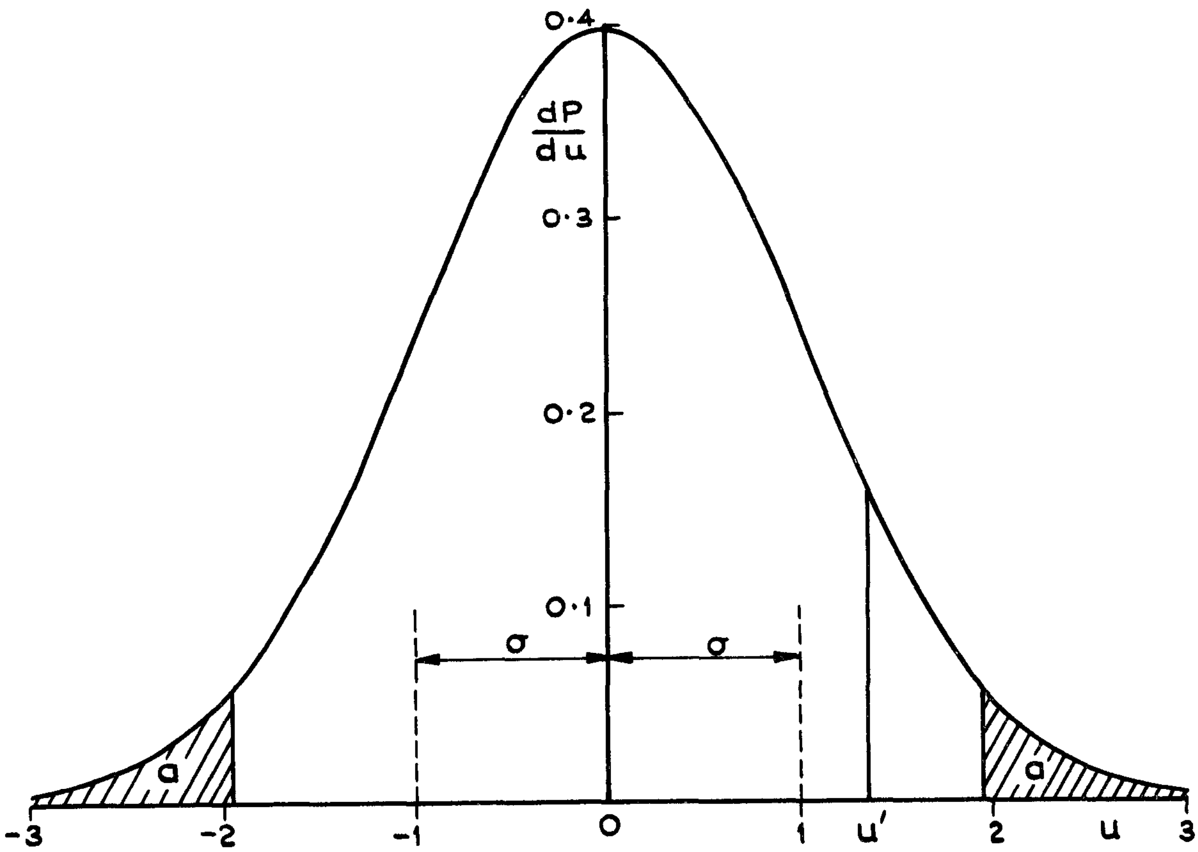
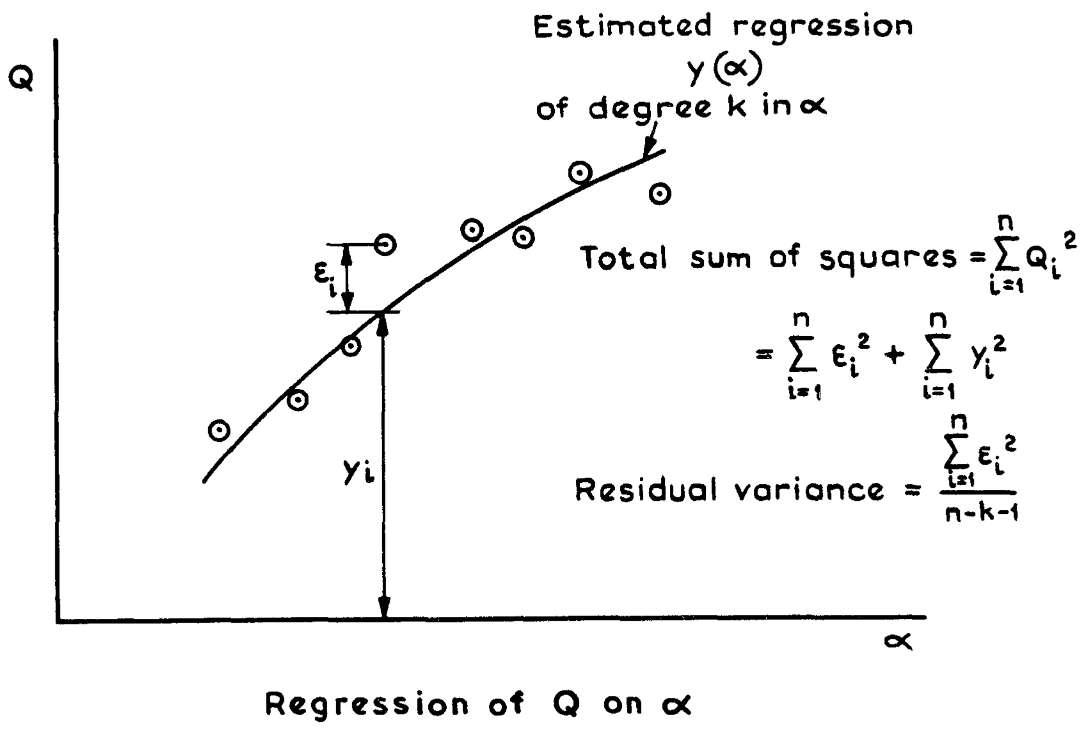


Fig.16 Effect of number of runs on s for the regression of C_n on α



Probability density of a normal distribution, or normal curve, in standard form



Regression of Q on α

Fig.17 Illustrations of some statistical terms

Printed in England for Her Majesty's Stationery Office by the Royal Aircraft Establishment, Farnborough Dd 587483 K5 1/78

ARC CP No.1390 — 533.6.071 : 519.272
January 1975

J. McKie

ON THE APPLICATION OF CERTAIN STATISTICAL METHODS TO WIND-TUNNEL TESTING

The Report illustrates the use of some standard statistical techniques in wind-tunnel testing. The results of non-linear regression analysis are applied to the particular problem of comparing the data from experiments in two different tunnels on the same model. Residual variance is used as a measure of the repeatability of results and standard tests are applied to look for significant differences between the two tunnels. The accuracy of a measured aerodynamic coefficient is put in terms of confidence limits for a given probability level. A method is given for determining the minimum detectable effect of a model geometry change and also for finding the number of data points needed to measure a coefficient to a prescribed accuracy.

ARC CP No.1390 — 533.6.071 : 519.272
January 1975

J. McKie

ON THE APPLICATION OF CERTAIN STATISTICAL METHODS TO WIND-TUNNEL TESTING

The Report illustrates the use of some standard statistical techniques in wind-tunnel testing. The results of non-linear regression analysis are applied to the particular problem of comparing the data from experiments in two different tunnels on the same model. Residual variance is used as a measure of the repeatability of results and standard tests are applied to look for significant differences between the two tunnels. The accuracy of a measured aerodynamic coefficient is put in terms of confidence limits for a given probability level. A method is given for determining the minimum detectable effect of a model geometry change and also for finding the number of data points needed to measure a coefficient to a prescribed accuracy.

Cut here

ARC CP No.1390 — 533.6.071 : 519.272
January 1975

J. McKie

ON THE APPLICATION OF CERTAIN STATISTICAL METHODS TO WIND-TUNNEL TESTING

The Report illustrates the use of some standard statistical techniques in wind-tunnel testing. The results of non-linear regression analysis are applied to the particular problem of comparing the data from experiments in two different tunnels on the same model. Residual variance is used as a measure of the repeatability of results and standard tests are applied to look for significant differences between the two tunnels. The accuracy of a measured aerodynamic coefficient is put in terms of confidence limits for a given probability level. A method is given for determining the minimum detectable effect of a model geometry change and also for finding the number of data points needed to measure a coefficient to a prescribed accuracy.

ARC CP No.1390 — 533.6 071 : 519.272
January 1975

J. McKie

ON THE APPLICATION OF CERTAIN STATISTICAL METHODS TO WIND-TUNNEL TESTING

The Report illustrates the use of some standard statistical techniques in wind-tunnel testing. The results of non-linear regression analysis are applied to the particular problem of comparing the data from experiments in two different tunnels on the same model. Residual variance is used as a measure of the repeatability of results and standard tests are applied to look for significant differences between the two tunnels. The accuracy of a measured aerodynamic coefficient is put in terms of confidence limits for a given probability level. A method is given for determining the minimum detectable effect of a model geometry change and also for finding the number of data points needed to measure a coefficient to a prescribed accuracy.

Cut here

DETACHABLE ABSTRACT CARDS

DETACHABLE ABSTRACT CARDS

C.P. No. 1390

1978
Published by
HER MAJESTY'S STATIONERY OFFICE

Government Bookshops
49 High Holborn, London WC1V 6HB
13a Castle Street, Edinburgh EH2 3AR
41 The Hayes, Cardiff CF1 1JW
Brazennose Street, Manchester M60 8AS
Southey House, Wine Street, Bristol BS1 2BQ
258 Broad Street, Birmingham B1 2HE
80 Chichester Street, Belfast BT1 4JY

Government Publications are also available through booksellers

C.P. No. 1390

ISBN 011 471136 4