NPL Report CMSC 22/03

Report to the National Measurement System Policy Unit, Department of Trade and Industry

From the Software Support for Metrology Programme

Extracting Features from Experimental Data

By
M G Cox, P M Harris, P D Kenward and I M Smith

May 2003

Extracting Features from Experimental Data

M G Cox, P M Harris, P D Kenward and I M Smith
Centre for Mathematics and Scientific Computing

May 2003

ABSTRACT

Univariate polynomial spline curves provide a flexible class of functions that are effective for modelling a wide variety of experimental data. However, the parameters defining such curves generally do not provide directly any physical information about the measurement system giving rise to the data. Instead such information is required to be *extracted* from the fitted model. The problem of extracting information from univariate polynomial spline curves is considered, where that information takes the form of *features* of the curve, including the positions of zero-crossing points, peaks, troughs and points of inflexion, and the width of peaks and troughs. The evaluation of the *uncertainties* associated with estimates of these features derived from a spline curve fitted to experimental data is addressed.

ISSN 1471–0005

Authorised by Dr Dave Rayner,
Head of the Centre for Mathematics and Scientific Computing

National Physical Laboratory,
Queens Road, Teddington, Middlesex, United Kingdom TW11 0LW

Contents

1 Introduction

Modelling experimental data is a key activity in metrology. Modelling comprises three main stages: model building, model solving and model prediction. *Model building* is concerned with developing a mathematical model of the measurement system in terms of mathematical equations involving parameters that describe all relevant aspects of the system. *Model solving* is concerned with determining estimates of the model parameters, together with the uncertainties associated with the estimates, from experimental data by solving the mathematical equations. Finally, *model prediction* involves making predictions about the measurement system using the model fitted to the experimental data, such as evaluating the model at points at which data is unavailable.

If a *physical* model of the measurement system exists, determined from a theoretical understanding of the system, the parameter estimates often convey directly information about the measurement system that gave rise to the experimental data. For example, the response of a damped oscillator when excited by a sinusoidal signal of frequency f_0 is

$$y(t) = A_0 \sin(2\pi f_0 t + \phi_0) + A_r e^{-d_r t} \sin(2\pi f_r t + \phi_r),$$

with parameters $A_0, \phi_0, A_r, d_r, f_r, \phi_r$. The first term in the model represents the steady-state response of the oscillator, and the second term its resonant behaviour. The parameters describe directly properties of the oscillator: A_0 is the steady-state amplitude (which may not be reached during the period of a measurement), f_r is the resonance frequency of the oscillator (which may be unknown), etc.

Empirical models are important in cases where knowledge of the underlying physics for a measurement system is insufficient to characterise it completely. For empirical models depending on one variable, polynomial and particularly *polynomial spline* curves, when used with care, are generally very satisfactory for representing data. A polynomial spline curve is composed of a sequence of polynomial curves joined together at points called knots and in such a way as to ensure appropriate smoothness of the complete curve.

Spline curves provide a flexible class of functions that are effective for representing a wide variety of shapes. However, the parameters defining such curves generally do *not* provide directly any physical information about the measurement system. Instead, such information is required to be *extracted* from the model fitted to the experimental data.

This report is concerned with extracting information from univariate polynomial spline curves where that information takes the form of *features* of the curve, including:

1. positions of zero-crossing points,

2. positions of peaks and troughs,
3. positions of points of inflexion, and
4. the width of peaks and troughs (at half the peak (trough) height relative to a defined baseline).

An important consideration that is addressed is the evaluation of the *uncertainties* associated with estimates of these features derived from a spline curve fitted to experimental data.

A software package[1] containing implementations of the procedures described is available through METROS[2] [2]. The software package extends the functionality of NPLFit[3], software developed by NPL for modelling experimental data using polynomial and polynomial spline curves that is also available through METROS.

The report is organised as follows. The modelling of experimental data by univariate polynomial spline curves is reviewed in Section 2. The aspects of the representation of spline curves in terms of B-splines, solving the least-squares fixed-knot spline approximation problem, and the evaluation of uncertainties associated with the solution to this problem are all considered. In Section 3 the extraction of features from a spline curve is considered. Section 4 considers the manner in which uncertainties associated with estimates of the features are evaluated. In Section 5 the application of the feature extraction procedures to the analysis of thermophysical data for the identification of temperature dependent material properties is presented. Conclusions are given in Section 6.

2 Spline approximation of experimental data

2.1 Univariate polynomial splines

Let $I := [x_{\min}, x_{\max}]$ be an interval of the x-axis partitioned into subintervals $\{I_j\}_0^N$, where

$$I_j = \begin{cases} [\lambda_j, \lambda_{j+1}), & j = 0, \ldots, N-1, \\ [\lambda_j, \lambda_{j+1}], & j = N, \end{cases}$$

and

$$x_{\min} = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1} \leq \lambda_N < \lambda_{N+1} = x_{\max}.$$

A spline $s(x)$ of order n (degree $n-1$) on I is a piecewise polynomial of order n on I_j, $j = 0, \ldots, N$. The spline $s(x)$ is C^{n-k-1} at λ_j if card$(\lambda_\ell = \lambda_j, \ell \in$

[1] See http://www.npl.co.uk/ssfm/metros/feature_extraction/
[2] See http://www.npl.co.uk/ssfm/metros/
[3] See http://www.npl.co.uk/ssfm/metros/nplfit/

$\{1,\ldots,N\}) = k$.[4] So, for example, a spline $s(x)$ of order 4 for which the points λ_j are distinct (cardinality one) is a piecewise cubic polynomial of continuity class C^2, i.e. continuous in value, first and second derivatives, at the points λ_j.

The partition points $\boldsymbol{\lambda} = \{\lambda_j\}_1^N$ are the (interior) *knots* of s. To specify the complete set of knots needed to define s on I in terms of B-splines, the knots $\{\lambda_j\}_1^N$ are augmented by end knots $\{\lambda_j\}_{1-n}^{-1}$ and $\{\lambda_j\}_{N+2}^{q}$, $q = N+n$, satisfying

$$\lambda_{1-n} \le \cdots \le \lambda_0, \qquad \lambda_{N+1} \le \cdots \le \lambda_q.$$

For many purposes, a good choice [9] of additional knots is

$$\lambda_{1-n} = \cdots = \lambda_0, \qquad \lambda_{N+1} = \cdots = \lambda_q.$$

It readily permits derivative boundary conditions to be incorporated in spline approximants [6].

On I, s has the *B-spline representation* [4]

$$s(x) := s(\mathbf{c}, \boldsymbol{\lambda}; x) = \sum_{j=1}^{q} c_j N_{n,j}(\boldsymbol{\lambda}; x), \tag{1}$$

where $N_{n,j}(\boldsymbol{\lambda}; x)$ is the *B-spline* [4, 12] of order n with knots $\{\lambda_k\}_{j-n}^{j}$ and $\mathbf{c} = (c_1,\ldots,c_q)^{\mathrm{T}}$ are the *B-spline coefficients* of s. Each $N_{n,j}(\boldsymbol{\lambda}; x)$ is a spline with knots $\boldsymbol{\lambda}$, is non-negative and has compact support.[5] Specifically,

$$N_{n,j}(\boldsymbol{\lambda}; x) > 0, \;\; x \in (\lambda_{j-n}, \lambda_j), \qquad \mathrm{supp}(N_{n,j}(\boldsymbol{\lambda}; x)) = [\lambda_{j-n}, \lambda_j]. \tag{2}$$

The B-spline basis $\{N_{n,j}(\boldsymbol{\lambda}; x)\}_{j=1}^{q}$ for splines of order n with knots $\boldsymbol{\lambda}$ is generally very well-conditioned [9]. Moreover, the basis functions (and their derivatives) for any $x \in [x_{\min}, x_{\max}]$ can be formed in an unconditionally stable manner using the three-term recurrence relations [8]

$$N_{n,j}(\boldsymbol{\lambda}; x) = \frac{x - \lambda_{j-n}}{\lambda_{j-1} - \lambda_{j-n}} N_{n-1,j-1}(\boldsymbol{\lambda}; x) + \frac{\lambda_j - x}{\lambda_j - \lambda_{j-n+1}} N_{n-1,j}(\boldsymbol{\lambda}; x), \; n > 1,$$

and

$$N_{n,j}^{(\ell)}(\boldsymbol{\lambda}; x) = (n-1)\left(\frac{N_{n-1,j-1}^{(\ell-1)}(\boldsymbol{\lambda}; x)}{\lambda_{j-1} - \lambda_{j-n}} - \frac{N_{n-1,j}^{(\ell-1)}(\boldsymbol{\lambda}; x)}{\lambda_j - \lambda_{j-n+1}} \right),$$

with

$$N_{1,j}(\boldsymbol{\lambda}; x) = \begin{cases} 1, & x \in I_{j-1}, \\ 0, & x \notin I_{j-1}. \end{cases}$$

Figure 1 shows the B-spline basis for splines of order 3 with interior knots at $x = (1, 2, 5)^{\mathrm{T}}$ and coincident end knots at $x = 0$ and 10. The derivatives of the spline

[4] card(A) is used to denote the *cardinality* of the set A, i.e., the number of elements in the set.

[5] A function f of a real variable x has *compact support* if it is zero outside an interval, denoted by supp(f), that is closed and bounded.

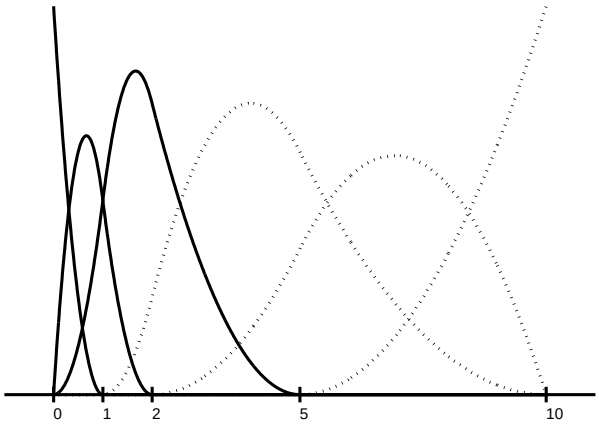

Figure 1: The B-spline basis for splines of order 3 for some non-uniformly spaced knots. The first three B-spline basis functions are shown as solid lines and the remaining three as dotted lines.

---

s are themselves splines with the same (interior) knots but of lower order, and consequently have the B-spline representation [8]

$$s^{(\ell)}(x) := \frac{\partial^\ell s(\mathbf{c}, \boldsymbol{\lambda}; x)}{\partial x^\ell} = \sum_{j=1}^{q} c_j N_{n,j}^{(\ell)}(\boldsymbol{\lambda}; x) = \sum_{j=1}^{q-l} c_j^{(\ell)} N_{n-\ell,j}(\boldsymbol{\lambda}; x),$$

where[6]

$$c_j^{(\ell)} = \begin{cases} c_j, & \ell = 0, \\ (n-\ell)\left(\frac{c_{j+1}^{(\ell-1)} - c_j^{(\ell-1)}}{\lambda_j - \lambda_{j-n+\ell}}\right), & \ell = 1, \ldots, n-1. \end{cases} \tag{3}$$

Figure 2 illustrates a spline s of order 4 (piecewise cubic) having interior knots at $x = (1, 2, 5)^{\mathrm{T}}$, coincident end knots at $x = 0$ and 10, and B-spline coefficients

$$\mathbf{c} = (0.00, 0.20, 0.60, 0.22, 0.18, 0.14, 0.12)^{\mathrm{T}}.$$

The spline has a "non-polynomial" shape, and demonstrates the advantage of using splines to represent complicated behaviour compared with (simple) polynomials. To reproduce this shape to visual accuracy with a polynomial would require a high degree and hence many more defining coefficients. Figures 3 and 4 show the first and second derivatives of the spline shown in Figure 2. The derivatives are, respectively, splines of order 3 (piecewise quadratic) and 2 (piecewise linear) with the same knots at $x = (1, 2, 5)^{\mathrm{T}}$. The coefficients in the B-spline representation

[6]Note that if $\lambda_j = \lambda_{j-n+\ell}$, $c_j^{(\ell)}$ cannot be computed from (3). However, since $N_{n-\ell,j}(\boldsymbol{\lambda}; x)$ is defined to be the zero function when this condition obtains, an arbitrary value can be assigned to $c_j^{(\ell)}$, since its value will not influence that of $s^{(\ell)}(x)$ [8].

of the derivatives are obtained by applying the recurrence (3). The graphs of the spline and its derivatives are produced by evaluating their B-spline representations at many values of x in the interval $[0, 10]$.

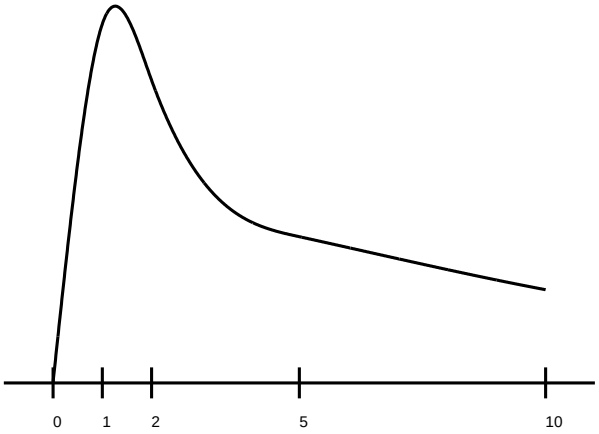

Figure 2: A spline with "non-polynomial" shape.

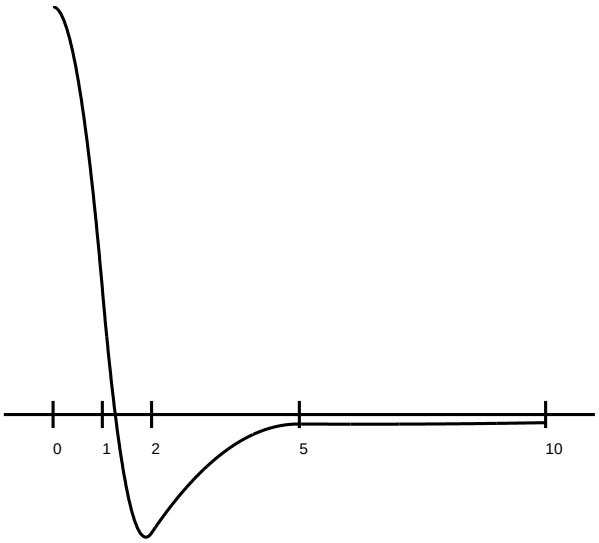

Figure 3: First derivative of the spline shown in Figure 2.

2.2 Least-squares fixed-knot spline approximation

The least-squares data approximation problem for splines with fixed knots can be posed as follows. Given are data points $\{(x_i, y_i)\}_1^m$, with $x_1 \leq \cdots \leq x_m$, and corresponding *weights* $\{w_i\}_1^m$ or *standard uncertainties* $\{u_i\}_1^m$. The w_i reflect the

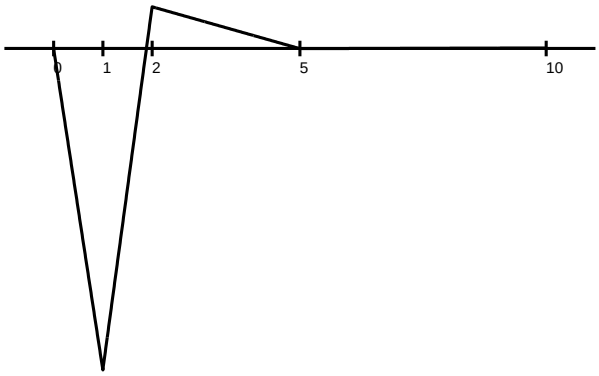

Figure 4: Second derivative of the spline shown in Figure 2.

---

relative quality of the y_i,[7] u_i is the standard uncertainty associated with y_i and corresponds to the standard deviation of possible "measurements" at $x = x_i$ of the function underlying the data, y_i being one realisation. Given also are the N knots $\boldsymbol{\lambda} = \{\lambda_j\}_1^N$ and the order n of the spline s.

Define *residuals* $\{e_i\}_1^m$ by

$$e_i = y_i - s(\mathbf{c}, \boldsymbol{\lambda}; x_i), \qquad i = 1, \ldots, m.$$

When *weights* are specified, the problem is to determine the spline s of order n, with knots $\boldsymbol{\lambda}$, such that the sum of squares of the elements $\{w_i e_i\}_1^m$ is minimised with respect to $\mathbf{c}$. When *standard uncertainties* are specified, the sum of squares of the elements $\{u_i^{-1} e_i\}_1^m$ is minimised with respect to $\mathbf{c}$. If $w_i = u_i^{-1}$, $i = 1, \ldots, m$, the two formulations are identical in terms of the spline produced. When weights are specified, s is referred to as a *spline approximant*. When uncertainties are prescribed, s is known as a *spline model*.

The use of a formulation in terms of standard uncertainties, together with the B-spline representation (1) of s, gives the linear algebraic formulation[8]

$$\min_{\mathbf{c}} \mathbf{e}^{\mathrm{T}} V_{\mathbf{y}}^{-1} \mathbf{e}, \qquad \mathbf{e} = \mathbf{y} - A\mathbf{c}, \tag{4}$$

where $\mathbf{y} = (y_1, \ldots, y_m)^{\mathrm{T}}$, A is an $m \times q$ matrix with $a_{i,j} = N_{n,j}(x_i)$, and

$$V_{\mathbf{y}} = \mathrm{diag}(u_1^2, \ldots, u_m^2).$$

The linear algebraic solution to (4) can be effected using Givens rotations to triangularise the system, back-solution then yielding the coefficients $\mathbf{c}$ [5]. Specifically,

[7] The x_i are taken as exact for the treatment here. A generalised treatment is possible, in which the x_i are also regarded as inexact. The problem becomes nonlinear (in $\mathbf{c}$).

[8] A further generalisation is possible in which mutual dependencies are permitted among the measurement errors. In this case, $V_{\mathbf{y}}$ is non-diagonal.

let the *QR*-decomposition of A be given by

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{5}$$

where Q is an orthonormal matrix ($Q^{\mathrm{T}}Q = I$, the identity matrix) and R is upper-triangular. Then, the coefficients $\mathbf{c}$ are obtained by back-substitution as the solution to

$$R\mathbf{c} = \mathbf{t},$$

where $\mathbf{t}$ contains the first n elements of $Q^{\mathrm{T}}\mathbf{y}$. As a consequence of property (2) of the B-splines, A is a rectangular banded matrix of bandwidth n [7].

Although splines provide a flexible class of functions for representing a wide variety of shapes, the effectiveness of a spline approximant or model of experimental data can depend critically on the number and locations of the interior knots. Various strategies for knot placement have been proposed [10]. In addition, the distribution of the interior knots in relation to the data abscissa values $\{x_i\}_1^m$ affects whether the coefficients $\mathbf{c}$ are uniquely determined [11]. The Schoenberg–Whitney conditions [14] provide criteria for testing whether $\mathbf{c}$ is uniquely determined by the data prior to effecting the solution to (4).

2.3 Uncertainty evaluation for fitted model

The key entity is the covariance matrix $V_{\mathbf{c}}$ of the spline coefficients $\mathbf{c}$. Using recognised procedures of linear algebra,

$$V_{\mathbf{c}} = (A^{\mathrm{T}} V_{\mathbf{y}}^{-1} A)^{-1}.$$

In the case that the measurements $y_i,\ i = 1, \ldots, m$, are independent and identically distributed,

$$V_{\mathbf{y}} = \sigma^2 I,$$

where I is the identity matrix. If σ is unknown, it may be estimated by the root-mean-square residual value associated with the fitted model,

$$\hat{\sigma} = \sqrt{\frac{\sum_i (y_i - s(x_i))^2}{m - q}},$$

so that

$$V_{\mathbf{c}} = \hat{\sigma}^2 (A^{\mathrm{T}} A)^{-1}.$$

In terms of the upper triangular factor R in the *QR*-decomposition of A (Section 2.2), we may also write

$$V_{\mathbf{c}} = \hat{\sigma}^2 (R^{\mathrm{T}} R)^{-1} = \hat{\sigma}^2 U^{\mathrm{T}} U, \tag{6}$$

where

$$R^{\mathrm{T}} U = I.$$

3 Extracting features from a spline curve

3.1 Positions of zero-crossing points

The position z_0 of a zero-crossing point of the spline curve s satisfies

$$s(z_0) := s(\mathbf{c}, \boldsymbol{\lambda}; x)|_{x=z_0} = 0. \tag{7}$$

Provided the interior knots $\boldsymbol{\lambda}$ of s satisfy

$$\text{card}(\lambda_\ell = \lambda_j, \ell \in 1, \ldots, N) < n, \qquad j = 1, \ldots, N, \tag{8}$$

solutions to equation (7) may be obtained by applying the bisection method [13] as follows.

1. Let $\alpha < \beta$ be such that $s(\alpha)s(\beta) < 0$.[9]
2. Evaluate s at the midpoint of the interval.
3. If the values of the spline at α and the midpoint have the same sign, replace α by the midpoint. Otherwise, replace β by the midpoint.
4. Repeat from step 2 until (the floating-point representations of) the endpoints are equal.

After each iteration of the procedure the length of the interval containing the zero halves and the procedure is consequently guaranteed to terminate. Candidate intervals to start the procedure in step 1 may be obtained by evaluating the spline at a large number of points equally-spaced in the interval $[x_{\min}, x_{\max}]$ and identifying consecutive points at which s changes sign.

3.2 Positions of peaks, troughs and points of inflexion

The zeros of the first derivative of the spline curve with respect to x define the positions of the peaks and troughs of the curve. Denoting by z_1 the position of a peak (trough), z_1 satisfies the equation

$$s^{(1)}(z_1) := \left.\frac{\partial s(\mathbf{c}, \boldsymbol{\lambda}; x)}{\partial x}\right|_{x=z_1} = 0. \tag{9}$$

Provided the interior knots $\boldsymbol{\lambda}$ of s satisfy

$$\text{card}(\lambda_\ell = \lambda_j, \ell \in 1, \ldots, N) < n - 1, \qquad j = 1, \ldots, N,$$

[9] This condition says that the values of the spline at the endpoints of the interval $[\alpha, \beta]$ are non-zero and have opposite signs. The condition (8) ensures that s is a continuous function and, consequently, the interval (α, β) must contain a zero of s.

the bisection method, as described in Section 3.1, can be used to find solutions to (9).

The zeros of the second derivative of the spline curve with respect to x define the points of inflexion of the curve. Denoting by z_2 the position of a point of inflexion, z_2 satisfies the equation

$$s^{(2)}(z_2) := \left.\frac{\partial^2 s(\mathbf{c}, \boldsymbol{\lambda}; x)}{\partial x^2}\right|_{x=z_2} = 0. \tag{10}$$

Provided the interior knots $\boldsymbol{\lambda}$ of s satisfy

$$\text{card}(\lambda_\ell = \lambda_j, \ell \in 1, \ldots, N) < n - 2, \qquad j = 1, \ldots, N,$$

the bisection method, as described in Section 3.1, can be used to find solutions to (10).

3.3 Peak width

Let z_1 denote the position of a peak (trough) evaluated as in Section 3.2. Given a baseline value γ, the positions z_a and z_b $(> z_a)$ of half-height with respect to the baseline satisfy

$$s(z_a, \mathbf{c}) = s(z_b, \mathbf{c}) = \frac{s(z_1, \mathbf{c}) + \gamma}{2}.$$

The *full-width at half-height* d is then given by

$$d = z_b - z_a,$$

and is used as a measure of the *width* of the peak (trough). See Figure 5.

4 Uncertainty evaluation for features

4.1 Positions of zero-crossing points

Let $\mathbf{C}$ denote the vector of coefficients of the spline curve s. We use the "upper-case" notation to emphasise that, due to the "stochastic" nature of the data, $\mathbf{C}$ is a vector of random variables with expectation $\mathbf{c}$, the calculated coefficients of the least-squares best-fit spline curve to the given experimental data, and covariance matrix $V_{\mathbf{c}}$.

Consider the estimation of the positions of the zeros-crossing points of the spline curve s. The position Z_0 of a zero-crossing point for s satisfies

$$s(Z_0, \mathbf{C}) = 0. \tag{11}$$

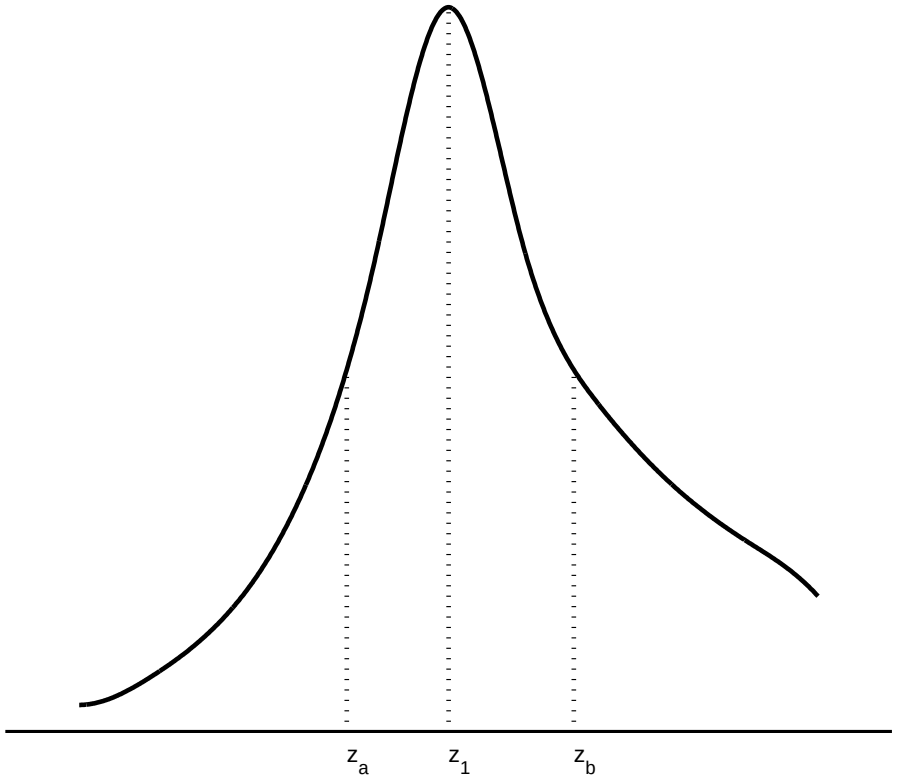

Figure 5: Position z_1 of a peak and the positions z_a and z_b at half-peak-height relative to a baseline (shown here as the horizontal line). The peak width d is evaluated as $z_b - z_a$.

In the terminology of the Guide to the Expression of Uncertainty in Measurement (GUM) [3], (11) is the *measurement model* and relates the *measurand* (or output quantity) Z_0 to the *input quantities* $\mathbf{C}$. A best estimate z_0 of the zero-crossing point is obtained by solving the (generally non-linear) equation

$$s(z_0, \mathbf{c}) = 0$$

(see Section 3.1).

From a first order Taylor series expansion of (11), we obtain

$$s(z_0, \mathbf{c}) + J_z\, \delta Z_0 + J_c\, \delta \mathbf{C} = 0, \tag{12}$$

where

$$\begin{aligned} \delta Z_0 &= Z_0 - z_0, \\ \delta \mathbf{C} &= \mathbf{C} - \mathbf{c}, \end{aligned}$$

$$\begin{aligned} J_z &= \frac{\partial s}{\partial Z}, \\ J_c &= \left(\frac{\partial s}{\partial C_1}, \ldots, \frac{\partial s}{\partial C_q} \right), \end{aligned}$$

with all derivatives evaluated at $Z_0 = z_0$ and $\mathbf{C} = \mathbf{c}$. In particular,

$$J_z = s^{(1)}(z_0),$$

and

$$J_c = \left(N_{n,1}(\boldsymbol{\lambda}; z_0), \ldots, N_{n,q}(\boldsymbol{\lambda}; z_0) \right).$$

Now, $\delta\mathbf{C}$ is a vector of random variables with expectation

$$E(\delta\mathbf{C}) = E(\mathbf{C}) - \mathbf{c} = \mathbf{0},$$

and covariance matrix

$$\text{var}(\delta\mathbf{C}) = \text{var}(\mathbf{C}) = V_{\mathbf{c}}.$$

From (12),

$$\delta Z_0 = -J_z^{-1} J_c \, \delta\mathbf{C}, \tag{13}$$

and also

$$\delta Z_0^2 = J_z^{-1} J_c \, \delta\mathbf{C} \, \delta\mathbf{C}^{\mathrm{T}} J_c^{\mathrm{T}} J_z^{-1}. \tag{14}$$

Taking expectations of (13), we obtain

$$E(\delta Z_0) = -J_z^{-1} J_c E(\delta\mathbf{C}) = 0.$$

It follows that

$$\text{var}(Z_0) = \text{var}(\delta Z_0) = E(\delta Z_0^2),$$

and so, from (14),

$$\text{var}(Z_0) = J_z^{-1} J_c E(\delta\mathbf{C}\delta\mathbf{C}^{\mathrm{T}}) J_c^{\mathrm{T}} J_z^{-1} = J_z^{-1} J_c V_{\mathbf{c}} J_c^{\mathrm{T}} J_z^{-1}.$$

In terms of the upper triangular factor R in the *QR*-decomposition of A (Section 2.2), we may write

$$\text{var}(Z_0) = \hat{\sigma}^2 J_z^{-1} J_c (R^{\mathrm{T}} R)^{-1} J_c^{\mathrm{T}} J_z^{-1}, \tag{15}$$

or, equivalently,

$$\text{var}(Z_0) = \hat{\sigma}^2 U^{\mathrm{T}} U,$$

where U solves the lower-triangular system of equations

$$R^{\mathrm{T}} U = J_c^{\mathrm{T}} J_z^{-1},$$

and may be determined by forward substitution.

Finally, the *standard uncertainty* associated with the estimate z_0 of the position Z_0 of the zero-crossing point is given by

$$u(z_0) = \sqrt{\text{var}(Z_0)} = \hat{\sigma} \sqrt{U^{\mathrm{T}} U}.$$

A similiar analysis may be applied to obtain the standard uncertainties associated with estimates of the positions of peaks, troughs and points of inflexion (Appendix A), and associated with an estimate of full-width at half-height (Appendix B).

4.2 Identifying significant features

Consider the positions of the zero-crossing points of the spline curve s.[10] Suppose the (ordered) positions of the zero-crossing points of s are denoted by $z_1, z_2, \ldots$, having been identified using the bisection method as described in Section 3.1.

Evaluate s at a large number M of uniformly-spaced points in the interval $[x_{\min}, x_{\max}]$, and construct the standard uncertainty $u(x)$ by evaluating the standard uncertainty associated with $s(x)$ at the same M points. For a given coverage probability p (e.g., 0.95), construct the *lower expanded uncertainty envelope* $l(x) = s(x) - k_p u(x)$ and the *upper expanded uncertainty envelope* $h(x) = s(x) + k_p u(x)$, where k_p is the coverage factor corresponding to p [3].[11]

The region bounded by these envelopes defines an "uncertainty swathe". A zero of s is regarded as *significant* if it corresponds to the *complete* uncertainty swathe crossing the x-axis. Thus, in Figure 6, z_1 and z_2 are not significant zeros, whereas z_3 is a significant zero. It is observed that a significant zero is bracketed by one zero of l and one zero of h.

Determine the zeros $l_1, l_2, \ldots$ of l and those $h_1, h_2, \ldots$ of h.[12] Order the complete set of zeros $z_1, z_2, \ldots, l_1, l_2, \ldots, h_1, h_2, \ldots$, labelling each one as to its type: "lower", "middle" or "upper". Scan the list for "middle" zeros. Accept a zero of s as a significant zero if it is bracketed as above, i.e., a "lower" immediately preceding it and an "upper" immediately following it, or *vice versa*. Thus, in Figure 6, z_3 is accepted as a significant zero, since it is preceded by l_3 and followed by h_1. However, z_1 is followed by z_2 and therefore not accepted. Similarly, z_2 is not accepted.

5 Application to the analysis of thermophysical data

The example considered here concerns the application of thermal analysis techniques, including differential scanning calorimetry (DSC), dynamic mechanical analysis (DMA) and deflection temperature under load (DTUL), for the assessment of the processing and performance of polymer composites and adhesives. DSC [1] is a thermal analysis technique in which the difference between the heat flux (power) into a test specimen and that into a reference specimen is measured as a function of temperature (and/or time) while the test specimen and the reference specimen are subjected to a controlled temperature programme.

[10] The positions of peaks, troughs and points of inflexion, corresponding to the zero-crossing points of the first and second derivatives of s are handled similarly.

[11] Under the usual assumption that the distribution for the value of s at x is Gaussian, $k_{0.95} \approx 2$ [3].

[12] The method of bisection may also be used for this purpose.

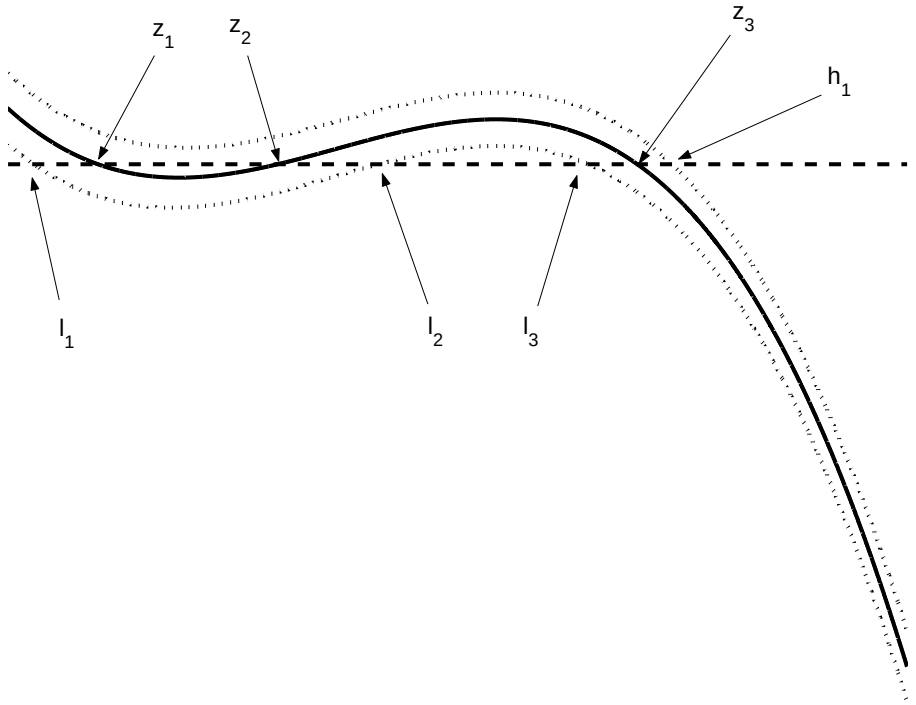

Figure 6: Using an "uncertainty swathe" to determine significant features. A spline curve with its lower and upper uncertainty envelopes are shown. The relative positions of the zeros of the spline and its uncertainty envelopes are used to decide whether a zero of the spline is significant.

An example of data from the application of DSC is shown in Figure 7. In this figure the x-axis represents temperature (°C) and the y-axis heat flow (mW). The data defines a DSC curve, features of which correspond to characteristic temperatures of the test specimen. For example, the positions of the main peak and main trough of the curve defined by the data in Figure 7 correspond to crystallization and melting transitions of the material.

It is for the purpose of estimating such characteristic temperatures, together with the standard uncertainties associated with these estimates, that the procedures described in Sections 3 and 4 are applied. For this data set, the positions of the peaks and troughs of the underlying curve have practical importance. The procedures to estimate other features, including the positions of zero-crossing points and inflexion points and the width of peaks and troughs, are also applied for illustration purposes, although these features have less practical value in this example.

Figure 8 shows a spline approximant to the DSC data of order 4 with 70 uniformly-spaced (interior) knots in the interval spanned by the data. The figure shows the

Position of peak/trough (°C)	Standard uncertainty (°C)
114.477	11.620
115.233	12.639
118.838	1.095
138.236	0.004

Table 1: Estimates of the positions of the peaks and troughs identified from the spline curve in the range 110 °C to 175 °C, together with the standard uncertainties associated with these estimates.

spline curve superimposed on the DSC data, and it is therefore evident that the spline curve represents adequately the "shape" described by the data.[13]

In the following we restrict our attention, for ease of presentation, to the temperature range (on the x-axis) from 110 °C to 175 °C. Figures 9, 10 and 11 each show the spline curve restricted to this range together with the positions of, respectively, the zero-crossing points, the peaks and troughs, and the points of inflexion within this range. Two zero-crossing points, four peaks and troughs and six inflexion points are identified in the three figures.

The estimates of the positions of the four peaks and troughs, together with the standard uncertainties associated with the estimates, are listed in Table 1. From the considerations of the "uncertainty swathe" for the first derivative of the spline function (Section 4.2) only the fourth peak (trough), at approximately 138 °C, is accepted as "significant". The identification of this feature is particularly important to the application giving rise to the data. The other three peaks and troughs within the temperature range that have been identified are either *real*, i.e., a feature in the underlying DSC curve, but not physically important, or *spurious*, i.e., introduced as a consequence of modelling the measured data by an (empirical) spline function. By considering the *width* of the "uncertainty swathe" for the first derivative centred on zero corresponding to the significant peak, we find that an estimate of the standard uncertainty associated with the estimated position of this feature is 0.004 °C.[14] We note that there is excellent agreement between the estimates of the standard uncertainty evaluated in the two ways.

Figure 12 shows the spline curve restricted to the temperature range from 125 °C to 150 °C, that contains the *single* peak at approximately 138 °C, and the positions

[13] Model validation, including an examination of the residual deviations between the data and the fitted model, should be undertaken to confirm that the spline curve represents the data to a degree that is consistent with the measurement uncertainty associated with the data.

[14] The half-width of the "uncertainty swathe" is equal to twice this estimate of the standard uncertainty.

of the points corresponding to the peak position and half-height peak position used in the calculation of the full-width at half-height parameter. For this calculation, the height of the baseline (from which the peak height is measured) is set at -2.5 mW. The width of the peak is estimated to be 8.51 °C with an associated standard uncertainty of 0.009 °C.

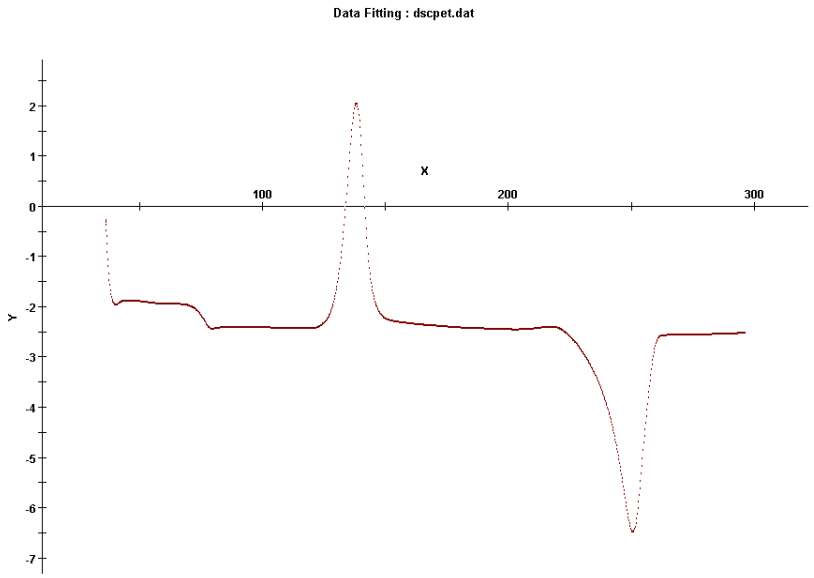

Figure 7: Data from the application of the thermal analysis technique of differential scanning calorimetry.

6 Conclusions

The purpose of modelling experimental data is usually to infer information about the measurement system giving rise to the data. If a physical model of the measurement system exists, the parameters defining the model often convey directly information about the measurement system. In the absence of a physical model, it is necessary to use instead an empirical function. Univariate polynomial spline curves consitute an important class of empirical functions that are generally very satisfactory for representing data. However, the parameters defining such curves generally do not provide directly any physical information about the measurement system.

This report has been concerned with the problem of extracting information from a univariate polynomial spline curve fitted to experimental data. The information is assumed to take the form of *features* of the curve, including positions of zero-

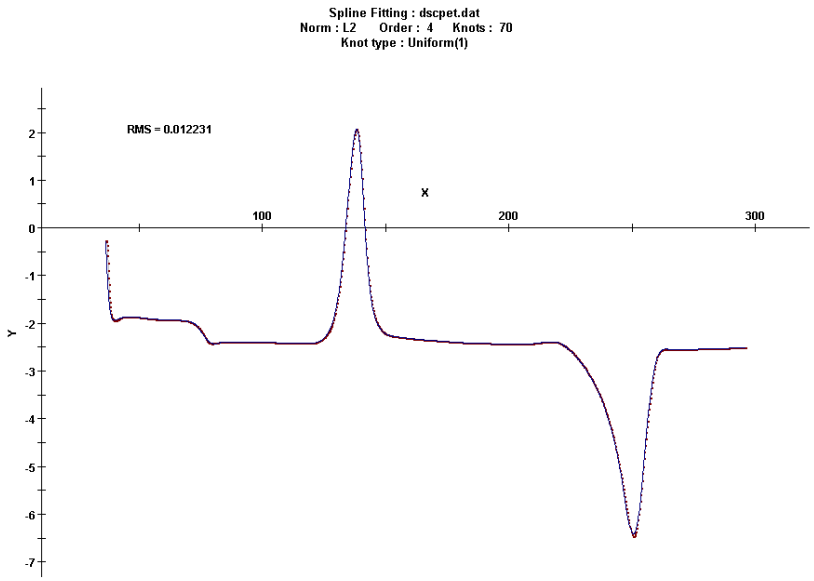

Figure 8: Spline curve of order 4 with 70 uniformly-spaced knots for the DSC data shown in Figure 7. The spline curve is superimposed on the DSC data.

---

crossing points, peaks, troughs and points of inflexion, and the width of peaks and troughs (full-width at half-height relative to a specified baseline).

Consideration has been given to (a) providing unambiguous definitions of the features, (b) the evaluation of the *uncertainty* associated with an estimate of a feature derived from a spline curve fitted to experimental data, and (c) guidance on accepting an estimate of a feature as "significant".

The procedures described in this report have been implemented in software that is available through METROS. The procedures have been illustrated using an example concerned with the application of thermal analysis techniques to estimating characteristic temperatures of a test specimen.

7 Acknowledgements

The work described here was supported by the National Measurement System Directorate of the UK Department of Trade and Industry as part of its NMS Software Support for Metrology programme.

We thank our colleagues Graham Sims and David Mulligan for providing information on the application of thermal analysis techniques.

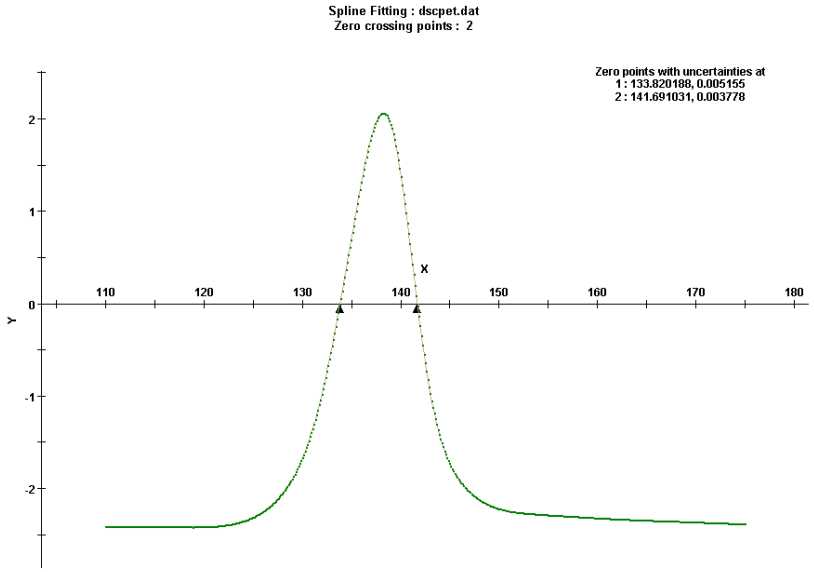

Figure 9: Positions (marked by triangles) between 110 °C and 175 °C of the zero-crossing points for the spline curve shown in Figure 8.

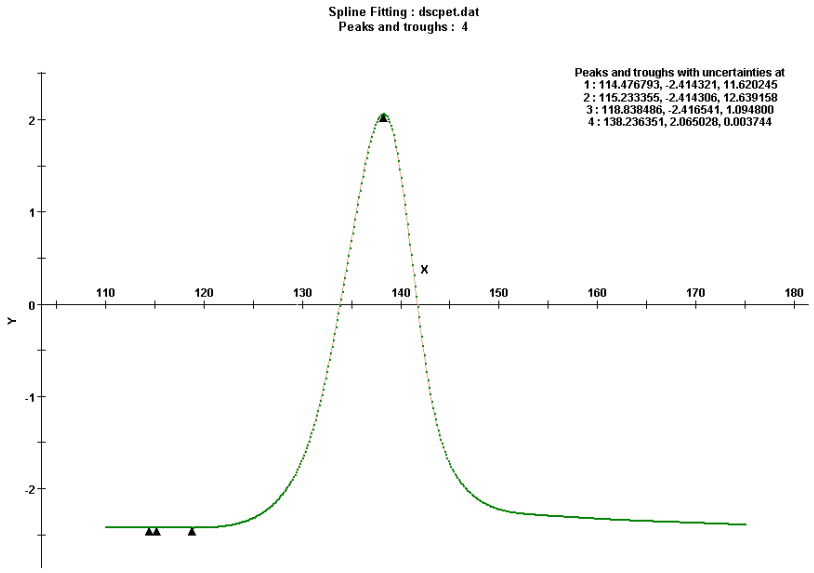

Figure 10: Positions (marked as triangles) between 110 °C and 175 °C of the peaks and troughs for the spline curve shown in Figure 8.

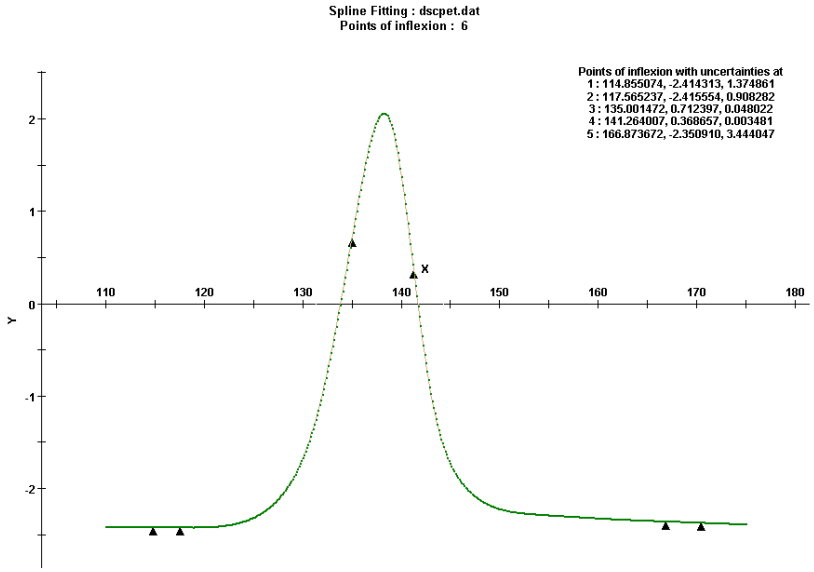

Figure 11: Positions (marked as triangles) between 110 °C and 175 °C of the points of inflexion for the spline curve shown in Figure 8.

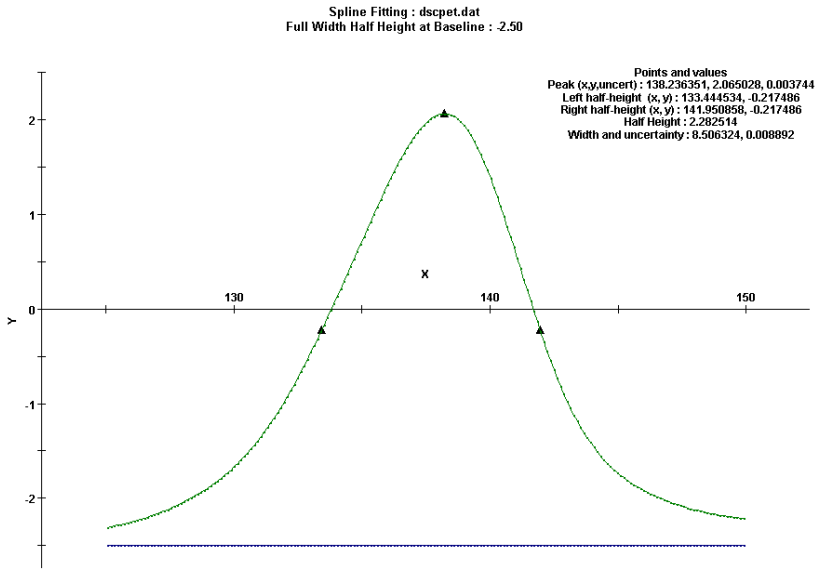

Figure 12: Points (marked as triangles) in the calculation of the width of the (main) peak in the range 125 °C to 150 °C for the spline curve shown in Figure 8.

References

[1] ISO 11357. Plastics – Differential scanning calorimetry (DSC) – Part 1: General principles, 1997.

[2] R. M. Barker. Best Practice Guide No. 5. Software Re-use: Guide to METROS. Technical report, National Physical Laboratory, Teddington, UK, 2000.

[3] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement, 1995. ISBN 92-67-10188-9, Second Edition.

[4] M. G. Cox. The numerical evaluation of B-splines. *J. Inst. Math. Appl.*, 10:134–149, 1972.

[5] M. G. Cox. A survey of numerical methods for data and function approximation. In D. A. H. Jacobs, editor, *The State of the Art in Numerical Analysis*, pages 627–668, London, 1977. Academic Press.

[6] M. G. Cox. The incorporation of boundary conditions in spline approximation problems. In G. A. Watson, editor, *Lecture Notes in Numerical Analysis 630: Numerical Analysis*, pages 51–63, Berlin, 1978. Springer-Verlag.

[7] M. G. Cox. The least squares solution of overdetermined linear equations having band or augmented band structure. *IMA J. Numer. Anal.*, 1:3–22, 1981.

[8] M. G. Cox. Practical spline approximation. In P. R. Turner, editor, *Notes in Mathematics 965: Topics in Numerical Analysis*, pages 79–112, Berlin, 1982. Springer-Verlag.

[9] M. G. Cox. Algorithms for spline curves and surfaces. In L. Piegl, editor, *Fundamental Developments of Computer-Aided Geometric Modelling*, pages 51–76, London, 1993. Academic Press.

[10] M. G. Cox, P. M. Harris, and P. D. Kenward. Fixed- and free-knot univariate least-squares data approximation by polynomial splines. Technical Report CMSC 13/02, National Physical Laboratory, Teddington, UK, 2002.

[11] M. G. Cox and J. G. Hayes. Curve fitting: a guide and suite of algorithms for the non-specialist user. Technical Report NAC 26, National Physical Laboratory, Teddington, UK, 1973.

[12] C. de Boor. On calculating with B-splines. *J. Approx. Theory*, 6:50–62, 1972.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran: The Art of Scientific Computing*. Cambridge University Press, New York, second edition, 1992.

[14] I. J. Schoenberg and Anne Whitney. On Pólya frequency functions III. *Trans. Am. Math.*, 74:246–259, 1953.

A Uncertainty evaluation for positions of peaks, troughs and points of inflexion

The position Z_1 of a peak or trough of the spline curve s satisfies

$$s^{(1)}(Z_1, \mathbf{C}) = 0.$$

Based on a first order Taylor series expansion of this model, we obtain

$$\text{var}(Z_1) = \hat{\sigma}^2 U^{\mathrm{T}} U,$$

where U solves the lower-triangular system of equations

$$R^{\mathrm{T}} U = K_c^{\mathrm{T}} K_z^{-1},$$

and

$$\begin{aligned} K_z &= \frac{\partial s^{(1)}}{\partial Z}, \\ K_c &= \left(\frac{\partial s^{(1)}}{\partial C_1}, \ldots, \frac{\partial s^{(1)}}{\partial C_q} \right), \end{aligned}$$

with all derivatives evaluated at $Z_1 = z_1$ and $\mathbf{C} = \mathbf{c}$. In particular,

$$K_z = s^{(2)}(z_1),$$

and

$$K_c = \left(N_{n,1}^{(1)}(\boldsymbol{\lambda}; z_1), \ldots, N_{n,q}^{(1)}(\boldsymbol{\lambda}; z_1) \right).$$

A point of inflexion Z_2 for the spline curve satisfies

$$s^{(2)}(Z_2, \mathbf{C}) = 0.$$

Based on a first order Taylor series expansion of this model, we obtain

$$\text{var}(Z_2) = \hat{\sigma}^2 U^{\mathrm{T}} U,$$

where U solves the lower-triangular system of equations

$$R^{\mathrm{T}} U = L_c^{\mathrm{T}} L_z^{-1},$$

and

$$\begin{aligned} L_z &= \frac{\partial s^{(2)}}{\partial Z}, \\ L_c &= \left(\frac{\partial s^{(2)}}{\partial C_1}, \ldots, \frac{\partial s^{(2)}}{\partial C_q} \right), \end{aligned}$$

with all derivatives evaluated at $Z_2 = z_2$ and $\mathbf{C} = \mathbf{c}$. In particular,

$$L_z = s^{(3)}(z_2),$$

and

$$L_c = \left(N_{n,1}^{(2)}(\boldsymbol{\lambda}; z_2), \ldots, N_{n,q}^{(2)}(\boldsymbol{\lambda}; z_2) \right).$$

B Uncertainty evaluation for peak width

The left-hand position Z_a of half-height satisfies

$$2s(Z_a, \mathbf{C}) = s(Z_1, \mathbf{C}) + \gamma, \tag{16}$$

where Z_1 is the position of the peak (trough) and γ denotes the value of the baseline. A first order Taylor series expansion of (16) gives

$$2s(z_a, \mathbf{c}) + 2J_{z,a}\delta Z_a + 2J_{c,a}\delta\mathbf{C} = s(z_1, \mathbf{c}) + J_{z,1}\delta Z_1 + J_{c,1}\delta\mathbf{C} + \gamma, \tag{17}$$

where

$$\begin{aligned}
J_{z,1} &= \frac{\partial s}{\partial Z} \text{ evaluated at } Z = z_1 \text{ and } \mathbf{C} = \mathbf{c},\\
J_{c,1} &= \left(\frac{\partial s}{\partial C_1}, \ldots, \frac{\partial s}{\partial C_q}\right) \text{ evaluated at } Z = z_1 \text{ and } \mathbf{C} = \mathbf{c},\\
J_{z,a} &= \frac{\partial s}{\partial Z} \text{ evaluated at } Z = z_a \text{ and } \mathbf{C} = \mathbf{c},\\
J_{c,a} &= \left(\frac{\partial s}{\partial C_1}, \ldots, \frac{\partial s}{\partial C_q}\right) \text{ evaluated at } Z = z_a \text{ and } \mathbf{C} = \mathbf{c}.
\end{aligned}$$

From (16) and (17),

$$\begin{aligned}
\delta Z_a &= \frac{1}{2}J_{z,a}^{-1}\left(J_{z,1}\delta Z_1 + J_{c,1}\delta\mathbf{C} - 2J_{c,a}\delta\mathbf{C}\right)\\
&= \frac{1}{2}J_{z,a}^{-1}\left(J_{c,1}\delta\mathbf{C} - 2J_{c,a}\delta\mathbf{C}\right),
\end{aligned}$$

since $J_{z,1} = 0$.

Similarly for Z_b, the right-hand position of half-height, define

$$\begin{aligned}
J_{z,b} &= \frac{\partial s}{\partial Z} \text{ evaluated at } Z = z_b \text{ and } \mathbf{C} = \mathbf{c},\\
J_{c,b} &= \left(\frac{\partial s}{\partial C_1}, \ldots, \frac{\partial s}{\partial C_q}\right) \text{ evaluated at } Z = z_b \text{ and } \mathbf{C} = \mathbf{c},
\end{aligned}$$

to obtain

$$\delta Z_b = \frac{1}{2}J_{z,b}^{-1}\left(J_{c,1}\delta\mathbf{C} - 2J_{c,b}\delta\mathbf{C}\right).$$

Now define

$$\mathbf{Z} = \begin{pmatrix} Z_a \\ Z_b \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} z_a \\ z_b \end{pmatrix},$$

and

$$\delta\mathbf{Z} = \begin{pmatrix} \delta Z_a \\ \delta Z_b \end{pmatrix} = \begin{pmatrix} \frac{1}{2}J_{z,a}^{-1}\left(J_{c,1}\delta\mathbf{C} - 2J_{c,a}\delta\mathbf{C}\right) \\ \frac{1}{2}J_{z,b}^{-1}\left(J_{c,1}\delta\mathbf{C} - 2J_{c,b}\delta\mathbf{C}\right) \end{pmatrix} = Q\delta\mathbf{C}, \qquad (18)$$

where

$$Q = \begin{pmatrix} \frac{1}{2}J_{z,a}^{-1}\left(J_{c,1} - 2J_{c,a}\right) \\ \frac{1}{2}J_{z,b}^{-1}\left(J_{c,1} - 2J_{c,b}\right) \end{pmatrix}.$$

Then,

$$\delta\mathbf{Z}\delta\mathbf{Z}^{\mathrm{T}} = Q\delta\mathbf{C}\delta\mathbf{C}^{\mathrm{T}}Q^{\mathrm{T}}. \qquad (19)$$

Taking expectations of (18), we obtain

$$E(\delta\mathbf{Z}) = QE(\delta\mathbf{C}) = 0.$$

It follows that

$$\text{var}(\mathbf{Z}) = \text{var}(\delta\mathbf{Z}) = E(\delta\mathbf{Z}\delta\mathbf{Z}^{\mathrm{T}}),$$

and so, from (19),

$$\text{var}(\mathbf{Z}) = QV_{\mathbf{c}}Q^{\mathrm{T}}.$$

The full-width at half-height satisfies

$$D = Z_b - Z_a = \begin{pmatrix} -1 & 1 \end{pmatrix}\mathbf{Z}.$$

We can therefore write

$$\begin{aligned} \text{var}(D) &= \begin{pmatrix} -1 & 1 \end{pmatrix}\text{var}(\mathbf{Z})\begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 1 \end{pmatrix}QV_{\mathbf{c}}Q^{\mathrm{T}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

In terms of the upper triangular factor R in the *QR*-decomposition of A (Section 2.2), we may write

$$\text{var}(D) = \begin{pmatrix} -1 & 1 \end{pmatrix}\hat{\sigma}^2U^{\mathrm{T}}U\begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

where U solves the lower-triangular system of equations

$$R^{\mathrm{T}}U = Q^{\mathrm{T}},$$

and may be determined by forward substitution.