

**An Algorithm for the Removal of Noise and Jitter
in Signals and its Application to Picosecond
Electrical Measurement**

M G Cox, P M Harris and D A Humphreys

June 1993

An Algorithm for the Removal of Noise and Jitter in Signals and its Application to Picosecond Electrical Measurement*

M G Cox, P M Harris and D A Humphreys[†]
Division of Information Technology and Computing
National Physical Laboratory
Teddington
Middlesex
United Kingdom
TW11 0LW

ABSTRACT

The accuracy of data recorded by many measurement systems is limited both by uncertainty in the measured value and by uncertainty in the trigger input to the system which controls when a measurement is taken. The former effect, which appears as noise on the underlying signal, is due, in part, to the sampling process and can often be reduced to an acceptable level by averaging many measurements. Noise on the trigger input gives rise to uncertainty in time between the trigger and the measurement points. This effect, known as jitter, causes a distortion of the signal which cannot be removed by averaging.

We describe and analyse the effects of noise and jitter on a waveform, and an algorithm for removing, or reducing, these effects is presented. The work is motivated by an application in picosecond electrical and optoelectronic metrology where a laser pulse is measured by a system consisting of a photodiode and a sampling oscilloscope. Here, since the length of the pulse is so short, perhaps only tens of picoseconds in duration, the effect of jitter is as pronounced as that of measurement noise. Results obtained by applying the algorithm to simulated data obtained from this application are presented.

*The work described was presented at the NATO Advanced Research Workshop on Algorithms for Approximation held at Lady Margaret Hall, Oxford, UK, 27-31 July 1992.

[†]Division of Electrical Science, National Physical Laboratory.

© Crown copyright 1993

ISSN 0262-5369

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

Extracts from this report may be reproduced
provided that the source is acknowledged

Approved on behalf of Chief Executive, NPL,
by Mr A J Marks, Head, Division of Information Technology and Computing

CONTENTS

	Page
1 INTRODUCTION	1
2 SIMULATING THE MEASUREMENT PROCESS	2
3 NOTATION	2
4 ANALYSIS OF THE MEASUREMENT PROCESS	6
4.1 THE MEAN OF THE MEASURED SIGNAL	6
4.2 THE VARIANCE OF THE MEASURED SIGNAL	7
4.3 SUMMARY	8
5 AN ALGORITHM FOR REMOVING NOISE AND JITTER	8
5.1 WEIGHTING OF THE MEASURED VARIANCE VALUES	9
6 EXAMPLES	10
6.1 EXAMPLE 1	10
6.2 EXAMPLE 2	13
7 CONCLUSION	13
8 ACKNOWLEDGEMENT	15
9 REFERENCES	15

ILLUSTRATIONS

Figure 1	Data generated by adding measurement noise to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable. The units for the vertical scale are millivolts; the horizontal scale, nominally 0-300 picoseconds, has been normalised to the interval [0, 1].	3
Figure 2	Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 1. The true response is shown as a dotted curve.	3
Figure 3	Data generated by adding jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.	4
Figure 4	Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 3. The true response is shown as a dotted curve.	4
Figure 5	Data generated by adding measurement noise and jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.	5

Figure 6	Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 5. The true response is shown as a dotted curve.	5
Figure 7	A second set of data generated by adding measurement noise and jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.	11
Figure 8	Initial weighted least-squares spline fit (Step 1) to the data shown in Figure 7.	12
Figure 9	Final fit (after 3 iterations) to the data shown in Figure 7; this takes into account the bias and correct weighting of the measurements using estimates of ω and τ	12
Figure 10	Initial weighted least-squares spline fit (Step 1) to the data shown in Figure 5.	13
Figure 11	Final fit (after 6 iterations) to the data shown in Figure 5; this takes into account the bias and correct weighting of the measurements using estimates of ω and τ	14

TABLES

Table 1	Results produced by the algorithm for the data shown in Figure 7. For this data $\omega = 2.5 \times 10^{-3}$ and $\tau = 8.0 \times 10^{-3}$. The model chosen is a fourth order polynomial spline function with twenty interior knots.	11
Table 2	Results produced by the algorithm for the data shown in Figure 5. For this data $\omega = 5.0 \times 10^{-3}$ and $\tau = 1.6 \times 10^{-2}$. The model chosen is a fourth order polynomial spline function with twenty interior knots.	13

1 INTRODUCTION

Metrology, or measurement science, is concerned with measuring physical processes. For many areas of metrology this involves two stages. Firstly, a physical measurement system appropriate to the application is used to give a set of data which represents the process. Secondly, mathematical modelling and numerical analysis are used to derive from the data a more useful, and hopefully more accurate, representation of the process than is provided by the data. It is the latter stage that we address in this paper.

We suppose the physical process of interest can be summarised by an underlying functional relationship between a control (or independent) variable, t , and a response (or dependent) variable, v . The data returned by any real system used to measure the process is usually inexact, being subject to errors arising from a number of sources. Having collected the data, we wish to determine from it the true response or at least a good approximation to the response. This requires a model (physical or empirical) for the underlying process, as well as knowledge of the nature of the errors in the data. The latter may be provided by a model of the measurement process itself.

A widely used model for the measurement process, suitable for many problems, assumes that the values of the control variable at which measurements are taken are known exactly, but the corresponding measured response values are contaminated by random error, known as *noise*. It is assumed that the errors in the measured values are statistically independent, and are samples of a Normal distribution whose mean is zero. Under these circumstances it is appropriate to use a *least-squares* analysis to give an approximation to the underlying function. Unequal *weighting* of the measurements may be used to model any dependence of the variance of the Normal distribution on the control variable.

Since the technology used in measurement science has progressed, and its application has become more demanding, this model is no longer satisfactory and it is necessary to consider more complete models of the measurement process used. Humphreys ([9]; Part II, Chapter 3) is concerned with using a measurement system, consisting of a photodiode and sampling oscilloscope combination, to measure a short laser pulse whose duration may be only a few tens of picoseconds. A trigger signal to the sampling oscilloscope is used to control when a measurement is taken. However, noise on the trigger input gives rise to an uncertainty in time between the trigger and measurement points. This effect is known as *jitter*, and is in addition to noise on the measured signal brought about by the sampling process. It is the purpose of this work to take proper account of the presence of both noise and jitter when analysing data collected by such measurement systems.

The paper is organised as follows. In Section 2 we present a numerical simulation of these two sources of error which illustrates their effect on measured data. In Section 3 we provide the notation to be used throughout the work, and state any assumptions required in the analysis of the measurement process. This analysis is given in Section 4, and it leads to a description in Section 5 of an algorithm for finding an approximation to the true response given data recorded by such a measurement system. In addition, the algorithm returns estimates of the standard deviations of the noise and jitter probability distributions. Results obtained by applying the algorithm to simulated data obtained from an application in picosecond electrical metrology are presented in Section 6. A conclusion is made in Section 7, where suggestions for further work are presented.

2 SIMULATING THE MEASUREMENT PROCESS

The following numerical simulation of data contaminated by noise and jitter illustrates the effect of these sources of error on measured data. In this description we denote the control variable, time, by t and the response variable, voltage, by v . The exact data (t_i^*, v_i^*) , $i = 1, \dots, m$, for the example is obtained by sampling at $m = 300$ uniformly spaced points a signal typical of that measured by the system considered in [9]. In this application, the signal lasts for about 300 picoseconds (ps), and the measured voltage varies from about -2×10^{-2} millivolts (mv) to about 7×10^{-2} mv . Moreover, the noise and jitter applied to the exact data in this simulation are typical for the application considered. For convenience, we have normalised the control variable to lie in the interval $[0, 1]$.

Firstly, and in order to illustrate measurement noise alone, we add to each value v_i^* a random sample taken from a Normal distribution with mean zero and standard deviation 5.0×10^{-3} mv . We carry out this procedure thirty-two times in order to generate repeat measurements at each of the values t_i^* of the control variable. The resulting data set, consisting of 9600 points, is plotted in Figure 1. As we might expect, by averaging the replicate values corresponding to each time-point, we can reduce the effect of measurement noise to a more acceptable level. This is illustrated in Figure 2 where we show the mean value (solid curve) and standard deviation (broken curve) of the data at each temporal point together with a plot of the original signal (dotted curve). The mean values provide a reasonable approximation to the true response. Moreover, the standard deviation values allow us to estimate the amount of measurement error in the data, expressed in terms of the standard deviation of the probability distribution from which the noise is sampled.

To simulate the effect of jitter alone, we add to each value t_i^* a random sample taken from a Normal distribution with mean zero and standard deviation 1.6×10^{-2} ps to give a value t_i for the control variable at which the actual measurement is taken. The collected data point consists of t_i^* paired with the true value of the underlying signal at t_i . The procedure is again carried out thirty-two times in order to generate repeat measurements for each of the values t_i^* . The complete set of data is shown in Figure 3, and in Figure 4 we have plotted the mean value and standard deviation of the data at each time-point. It is clear that the presence of jitter has produced a distortion of the signal, notably where the signal exhibits peaks and troughs, that cannot be removed by averaging. Moreover, even though the jitter added to each point comes from the same probability distribution, the standard deviation curve for the measured data varies with time.

Finally, Figures 5 and 6 show data generated by the addition of both noise and jitter. Using the notation introduced above, a collected data point consists of t_i^* paired with the value of the underlying signal, contaminated by noise, at t_i . The mean and standard deviation functions plotted in Figure 6 exhibit many of the features of those illustrated in Figure 4; averaging of the data produces a similar distortion in the mean function, and the standard deviation curve varies with the control variable.

3 NOTATION

We suppose that $v = g(t)$ describes the underlying signal which we wish to model, and we require that g is twice continuously differentiable and that the derivative functions $g^{(1)}(t)$ and $g^{(2)}(t)$ are not identically zero. Let t_i^* , $i = 1, \dots, m$, denote the nominal values for the time points at which the signal is sampled, and suppose that for each sampling point R measurements

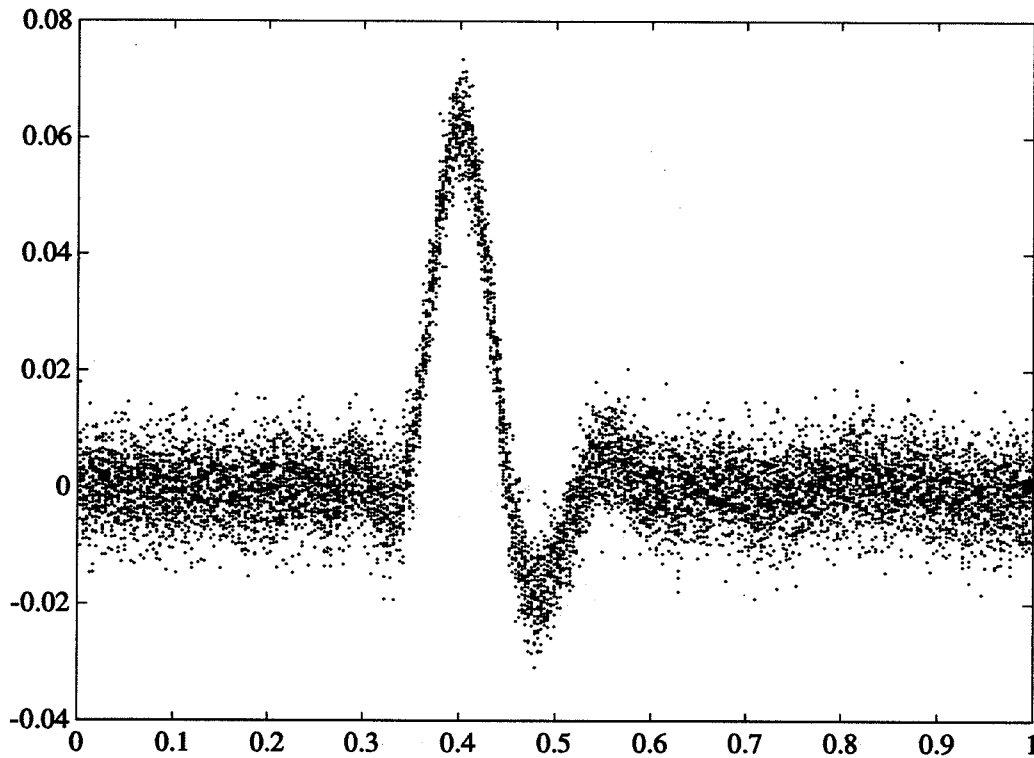


Figure 1 Data generated by adding measurement noise to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable. The units for the vertical scale are millivolts; the horizontal scale, nominally 0-300 picoseconds, has been normalised to the interval $[0, 1]$.

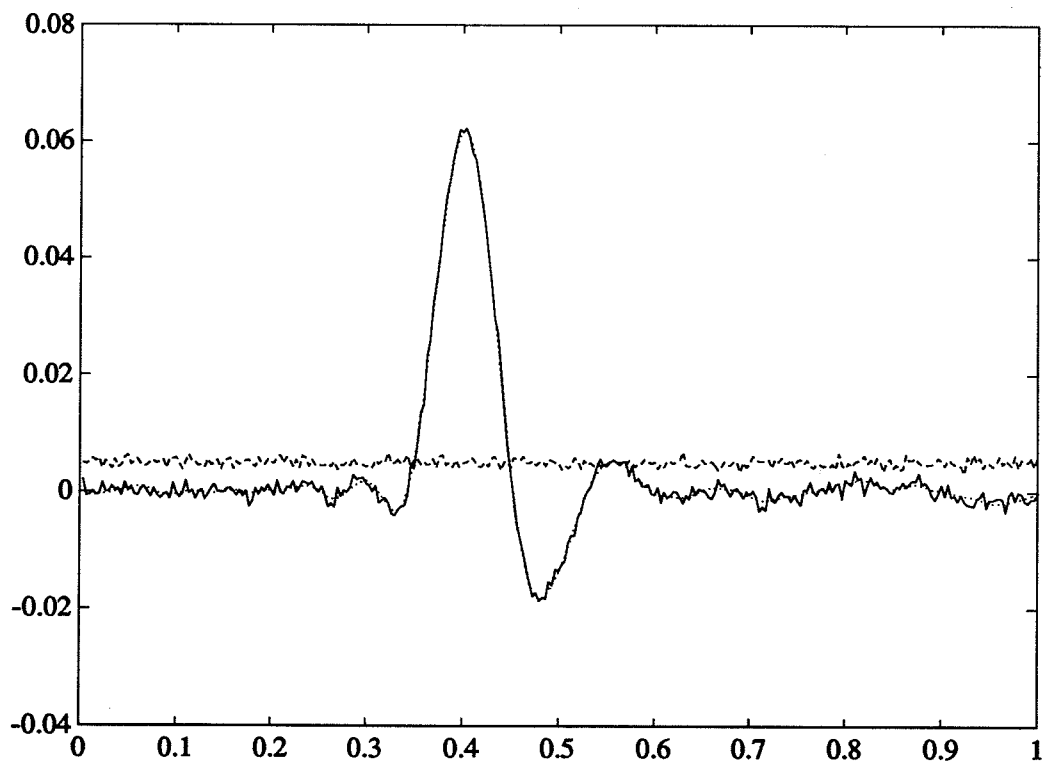


Figure 2 Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 1. The true response is shown as a dotted curve.

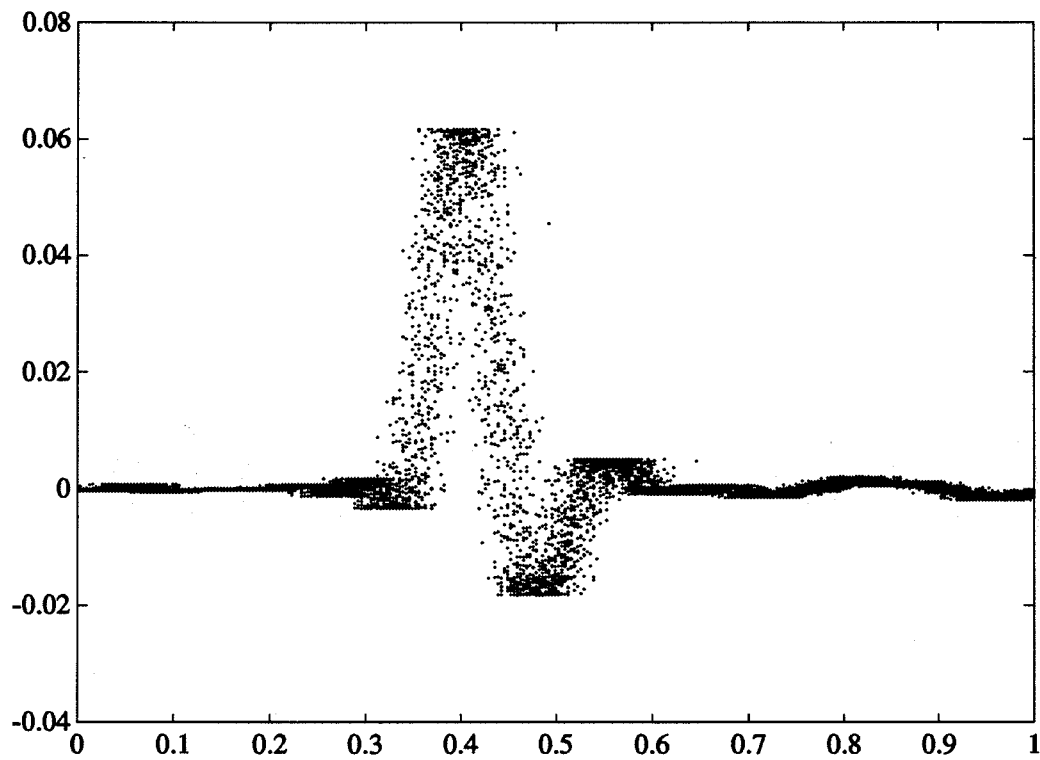


Figure 3 Data generated by adding jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.

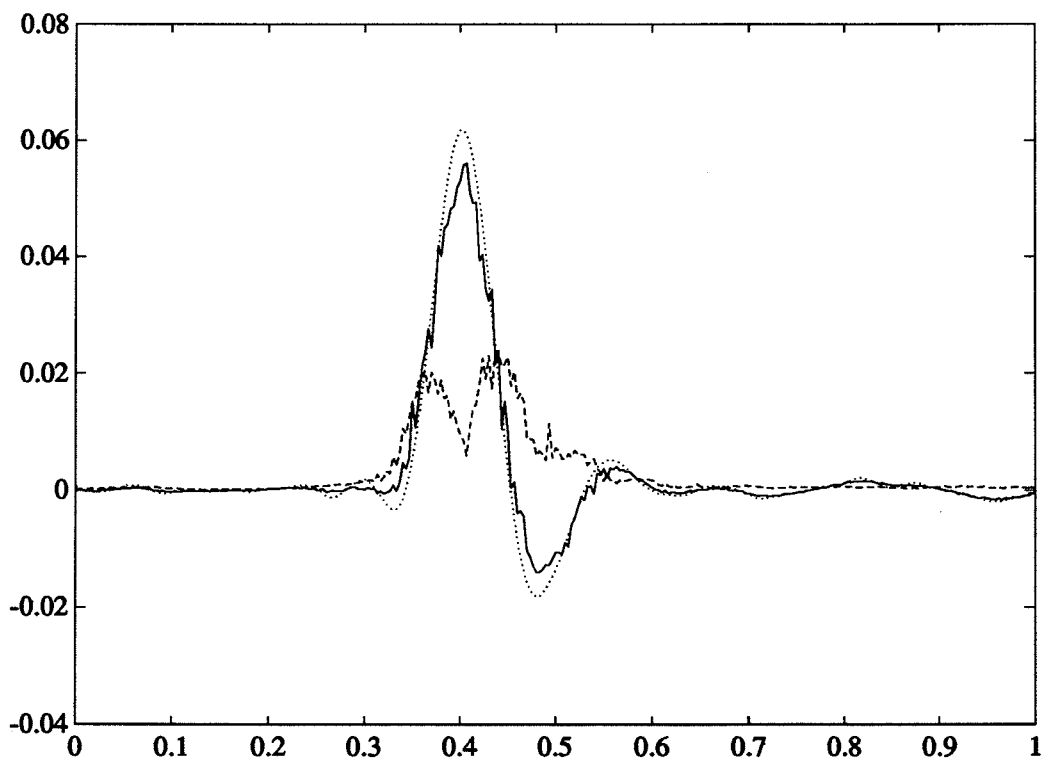


Figure 4 Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 3. The true response is shown as a dotted curve.

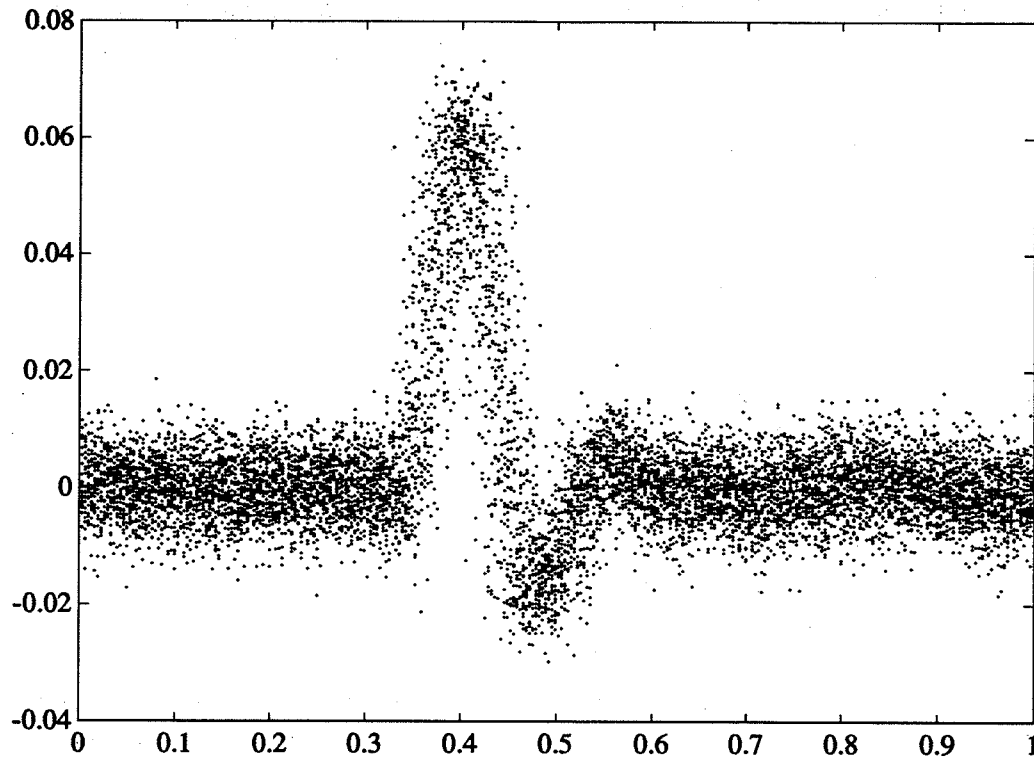


Figure 5 Data generated by adding measurement noise and jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.

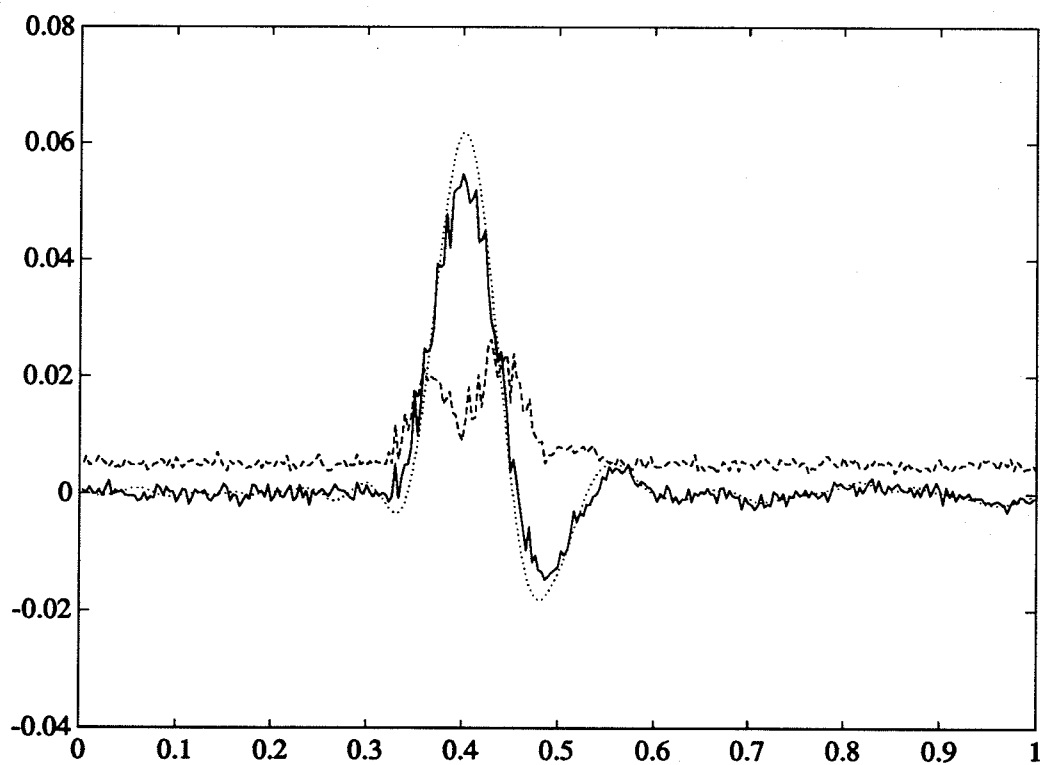


Figure 6 Mean value (solid curve) and standard deviation (broken curve) of the data shown in Figure 5. The true response is shown as a dotted curve.

of the signal are taken. Let $(t_i^{(j)}, v_i^{(j)})$, $i = 1, \dots, m$, $j = 1, \dots, R$, be the j^{th} replicate value measured at the i^{th} sampling point, corresponding to the true point $(t_i^*, g(t_i^*))$ lying on the underlying curve.

Now, the error between the measured and true points may be decomposed into a “horizontal” component

$$\delta t_i^{(j)} = t_i^{(j)} - t_i^*$$

due to jitter, and a “vertical” component

$$\delta v_i^{(j)} = v_i^{(j)} - g(t_i^{(j)})$$

due to measurement error. These errors are samples of probability distribution functions P_J and P_N , respectively, that are assumed to be independent of time and response and of each other. The true and measured values are thus related by the equations

$$v_i^{(j)} = g(t_i^* + \delta t_i^{(j)}) + \delta v_i^{(j)}, \quad i = 1, \dots, m, \quad j = 1, \dots, R, \quad (1)$$

or, regarded as a continuous function of time t , by the equation

$$v(t) = g(t + \xi) + \psi, \quad (2)$$

where ξ and ψ are independent random variables with probability distribution functions P_J and P_N , respectively. We assume further that P_J and P_N have zero means and, respectively, nonzero variances τ^2 and ω^2 .

4 ANALYSIS OF THE MEASUREMENT PROCESS

In this section we derive expressions for the mean and variance of the measured signal $v(t)$ as functions of t and the parameters τ and ω . These expressions form the basis of an algorithm for computing an approximation to the true underlying signal $g(t)$.

4.1 THE MEAN OF THE MEASURED SIGNAL

We denote by $E(f(t))$ the mean of a function f at time t . From (2) we may write

$$E(v(t)) = E(g(t + \xi)) + E(\psi),$$

and since $E(\psi) = 0$, so

$$E(v(t)) = E(g(t + \xi)) = \int_{-\infty}^{\infty} g(t + \xi) P_J(\xi) d\xi.$$

Applying Taylor's Theorem,

$$E(v(t)) = \int_{-\infty}^{\infty} \{g(t) + \xi g^{(1)}(t) + \frac{1}{2} \xi^2 g^{(2)}(t) + O(\xi^3)\} P_J(\xi) d\xi, \quad (3)$$

and, since

$$\int_{-\infty}^{\infty} P_J(\xi) d\xi = 1, \quad (4)$$

$$\int_{-\infty}^{\infty} \xi P_J(\xi) d\xi = 0, \quad (5)$$

$$\int_{-\infty}^{\infty} \xi^2 P_J(\xi) d\xi = \tau^2, \quad (6)$$

we have, assuming higher order terms are negligible,

$$E(v(t)) = g(t) + \frac{1}{2}\tau^2 g^{(2)}(t). \quad (7)$$

Thus, the mean or expected value of $v(t)$ differs from $g(t)$ by an amount $\frac{1}{2}\tau^2 g^{(2)}(t)$. This quantity represents a non-constant *bias* or *systematic* error in the measured response. We notice that the bias is (a) independent of any measurement noise in the data, and (b) greatest where the magnitude of the second derivative of the underlying function is greatest. Consequently, as Figures 4 and 6 illustrate, the departure between the true signal and the computed mean signal is greatest where there are peaks and troughs in the underlying signal.

4.2 THE VARIANCE OF THE MEASURED SIGNAL

We denote by $\text{var}(v(t))$ the variance of the measured signal where

$$\begin{aligned} \text{var}(v(t)) &= E\{v(t) - E(v(t))\}^2 \\ &= E\{(v(t))^2\} - \{E(v(t))\}^2. \end{aligned}$$

Now,

$$\begin{aligned} E\{(v(t))^2\} &= E\{(g(t + \xi) + \psi)^2\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{g(t + \xi) + \psi\}^2 P_N(\psi) P_J(\xi) d\psi d\xi \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{(g(t + \xi))^2 + 2g(t + \xi)\psi + \psi^2\} P_N(\psi) d\psi P_J(\xi) d\xi \\ &= \int_{-\infty}^{\infty} \{(g(t + \xi))^2 + \omega^2\} P_J(\xi) d\xi, \end{aligned}$$

using the counterparts of (4)–(6) for $P_N(\psi)$. Thus,

$$E\{(v(t))^2\} = \int_{-\infty}^{\infty} (g(t + \xi))^2 P_J(\xi) d\xi + \omega^2.$$

Again, ignoring higher order terms,

$$\begin{aligned} E\{(v(t))^2\} &= \int_{-\infty}^{\infty} \{g(t) + g^{(1)}(t)\xi + \frac{1}{2}g^{(2)}(t)\xi^2\}^2 P_J(\xi) d\xi + \omega^2 \\ &\simeq \int_{-\infty}^{\infty} [(g(t))^2 + 2g(t)g^{(1)}(t)\xi + \{g(t)g^{(2)}(t) + (g^{(1)}(t))^2\}\xi^2] P_J(\xi) d\xi + \omega^2 \\ &= (g(t))^2 + \tau^2\{g(t)g^{(2)}(t) + (g^{(1)}(t))^2\} + \omega^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{var}(v(t)) &= (g(t))^2 + \tau^2\{g(t)g^{(2)}(t) + (g^{(1)}(t))^2\} + \omega^2 - \{g(t) + \frac{1}{2}\tau^2 g^{(2)}(t)\}^2 \\ &\simeq \tau^2(g^{(1)}(t))^2 + \omega^2. \end{aligned} \quad (8)$$

Here, we notice that (a) where $g^{(1)}(t)$ is zero, the jitter makes no contribution to the standard deviation of the measured data, and (b) the standard deviation is greatest where $|g^{(1)}(t)|$ is greatest; for example, at points of inflexion between peaks and troughs in the underlying signal. Both these properties are exhibited in Figures 4 and 6 for the numerical example described in Section 1.

4.3 SUMMARY

From the results of the previous two sections, it follows that an unbiased estimate s of g is given by approximating $v(t)$ by $s(t) + \frac{1}{2}\tau^2 g^{(2)}(t)$ using a (squared) weighting function inversely proportional to the variance function $\tau^2(g^{(1)}(t))^2 + \omega^2$. It is this observation that forms the basis of the algorithm presented in the next section.

In the algorithm we require a sensible choice S for the class of functions from which the approximation s is to be taken, as well as estimates of τ , ω , $g^{(1)}(t)$ and $g^{(2)}(t)$. Assuming these estimates can be provided (this is discussed in the next section), S could usefully be taken as a set of (cubic) polynomial splines having suitably selected knots. An interesting consequence of this particular choice is that if we replace g by s in the analysis, (3) takes the form

$$\begin{aligned} E(v(t)) &= \int_{-\infty}^{\infty} \sum_{k=0}^3 \frac{1}{k!} s^{(k)}(t) \xi^k P_J(\xi) d\xi \\ &= \sum_{k=0}^3 \frac{1}{k!} s^{(k)}(t) \int_{-\infty}^{\infty} \xi^k P_J(\xi) d\xi. \end{aligned}$$

If we further assume that the probability distribution function P_J is symmetric, then the counterpart of (7),

$$E(v(t)) = s(t) + \frac{1}{2}\tau^2 s^{(2)}(t),$$

is *exact*. We conclude that the only essential limitation of our proposed approach is the ability of s to represent adequately the underlying function. In practice, this limitation can be minimized by paying careful attention to the distribution and number of knots used to define S .

5 AN ALGORITHM FOR REMOVING NOISE AND JITTER

Given the data $(t_i^*, v_i^{(j)})$, $i = 1, \dots, m$, $j = 1, \dots, R$, we describe in this section an algorithm for determining a cubic polynomial spline approximation to the true response function represented by the data. The algorithm returns also estimates of the variances τ^2 and ω^2 . We assume that interior knots defining the spline function are given. The proposed algorithm consists of the following steps.

Step 1 Compute a weighted least-squares spline fit to the data $(t_i^*, v_i^{(j)})$, $i = 1, \dots, m$, $j = 1, \dots, R$, to give an *initial* approximation s to g . At this stage we ignore the bias in the data caused by the effect of jitter, and assign to each measurement a weight $w_i^{(j)}$ derived by a simple analysis of the replicate data. Thus,

$$w_i^{(j)} = \sigma_i^{-1}, \quad i = 1, \dots, m, \quad j = 1, \dots, R,$$

where σ_i is the standard deviation of the values $v_i^{(j)}$, $j = 1, \dots, R$.

Step 2 Evaluate $s^{(1)}$ at each measurement point t_i^* , $i = 1, \dots, m$.

Step 3 Using the values computed at Step 2, determine estimates for τ^2 and ω^2 by fitting in a *weighted* least-squares sense the linear model

$$\tau^2(s^{(1)}(t))^2 + \omega^2$$

to the measured variance values σ_i^2 , $i = 1, \dots, m$. The particular weighting of the measured variance values that is used is discussed in Section 5.1.

Step 4 Using the values $s^{(1)}(t_i^*)$, $i = 1, \dots, m$, τ and ω obtained at Steps 2 and 3, compute a new approximation $\bar{s}$ (to g) which takes into account the bias and correct weighting of the measurements. We write $\bar{s}$ in terms of the cubic B-spline basis functions $N_{4,k}(t)$ ([4], [5], [6], [7]) defined on the chosen knot set:

$$\bar{s}(t) = \sum_k c_k N_{4,k}(t).$$

The coefficients c_k defining $\bar{s}$ are given by fitting the data in a weighted least-squares sense by the function

$$\sum_k c_k \{N_{4,k}(t) + \frac{1}{2}\tau^2 N_{4,k}^{(2)}(t)\}$$

using weights

$$\bar{w}_i^{(j)} = \{\tau^2 (s^{(1)}(t_i^*))^2 + \omega^2\}^{-1/2}, \quad i = 1, \dots, m, \quad j = 1, \dots, R.$$

Step 5 Setting $s = \bar{s}$, repeat from Step 2 until the process has stabilized. The termination criterion used is to require that

$$\max_{i=1, \dots, m} |s(t_i^*) - \bar{s}(t_i^*)| \leq tol$$

for some user-prescribed tolerance tol .

Routines from DASL, the National Physical Laboratory's Data Approximation Subroutine Library [1], may be used to evaluate the B-spline basis and the spline s and its derivatives, and to solve the weighted least-squares problems indicated in the algorithm.

It should be noticed that no proof of "convergence" of our algorithm is available. In practice, the calculations stabilize after fewer than ten iterations (see Section 6). In this regard, our process bears some resemblance to de Boor's knot-placement technique [7], in which the spline s and its derivatives at each stage are used to obtain an improved spline at the next.

5.1 WEIGHTING OF THE MEASURED VARIANCE VALUES

In an original version of the algorithm described above, estimates of τ^2 and ω^2 were obtained at Step 3 using an *unweighted* least-squares fit to the measured variance values σ_i^2 , $i = 1, \dots, m$. This is appropriate if the random uncertainties associated with the values σ_i^2 are assumed to be equal. However, when applied to simulated data, the approach gave estimates of τ^2 and ω^2 that were poor and also exhibited a bias. Consequently, the assumption of equal weighting may not be valid.

In order to remove these problems, and thus improve the performance of the algorithm, the uncertainties in the values σ_i^2 are estimated using the *jackknife* [8], a technique used in non-parametric statistical analysis. This shows that the assumption of equal weighting is indeed inappropriate, and provides a means of assigning weights in Step 3 to the values σ_i^2 that gives much improved results. The application of the jackknife is described below.

Recall that σ_i^2 is the variance of the values $v_i^{(1)}, \dots, v_i^{(R)}$;

$$\sigma_i^2 = \text{var}\{v_i^{(1)}, \dots, v_i^{(R)}\}.$$

To estimate the uncertainty in σ_i^2 we compute for each j the variance of the values $v_i^{(1)}, \dots, v_i^{(R)}$ with $v_i^{(j)}$ removed, and denote this by $\sigma_{i,j}^2$:

$$\sigma_{i,j}^2 = \text{var}\{v_i^{(1)}, \dots, v_i^{(R)} \setminus v_i^{(j)}\}.$$

Each of the values $\sigma_{i,j}^2$, $j = 1, \dots, R$, provides an estimate of σ_i^2 , and the variance of these values gives a measure of the uncertainty in σ_i^2 . Thus,

$$\text{var}\{\sigma_i^2\} = \text{var}\{\sigma_{i,1}^2, \dots, \sigma_{i,R}^2\},$$

and in the least-squares analysis carried out at Step 3 of the algorithm, the weight associated with σ_i^2 is taken to be inversely proportional to the square root of this uncertainty.

6 EXAMPLES

In this section we present results obtained by applying the algorithm described in Section 5 to simulated data obtained from an application in picosecond electrical metrology. We hope in the future to report on the results obtained for *real* data arising from this application. However, for the purpose of assessing the behaviour of the algorithm, it is more useful to consider data generated from a known underlying function. The procedure for simulating data contaminated by noise and jitter given an underlying function and values for ω and τ is discussed in Section 1. For the two examples considered, the true response underlying the data sets is the same as that used in Section 1. This function is illustrated by the dotted curve in, for example, Figure 6. Moreover, the data sets are constructed to have thirty-two repeat measurements for each of the values t_i^* of the control variable.

6.1 EXAMPLE 1

In our first example, the data is generated by setting $\omega = 2.5 \times 10^{-3}$ and $\tau = 8.0 \times 10^{-3}$. The complete set of data is illustrated in Figure 7. The model used to approximate the true response is a fourth order polynomial spline function with twenty interior knots. The knots, marked as asterisks on the independent variable axes in Figures 8 and 9, are chosen in such a way that the underlying function may be adequately represented by a fourth order spline with these knots. In Section 7 we make some further comments about (automatic) knot placement when choosing a suitable model. The tolerance *tol* used to control the convergence of the algorithm (see Step 5) is set as 1.0×10^{-4} .

In Figure 8 we show the weighted least-squares fit to the complete set of data, as generated at Step 1 of the algorithm, together with the true response (dotted curve). This fit is used to initialize the algorithm. An “iteration” of the algorithm consists of using the current spline approximation s to provide estimates for ω and τ (Steps 2 and 3), and then to determine a new spline approximation $\bar{s}$ (Step 4). In Table 1 we give the results produced at each iteration of the algorithm which proceeds until a measure of the change between s and $\bar{s}$ is judged to be small (Step 5). In Figure 9 we illustrate the final computed fit to the data together with the true response that we wish to approximate. We observe that we have realised reasonable approximations to ω and τ , as well as an acceptable approximation to the true response. In particular, in the region of the main peak and trough the approximation is much improved, although this is less so in the regions of the secondary peaks and troughs. The use of well-chosen knots in these regions would, we expect, give improvement.

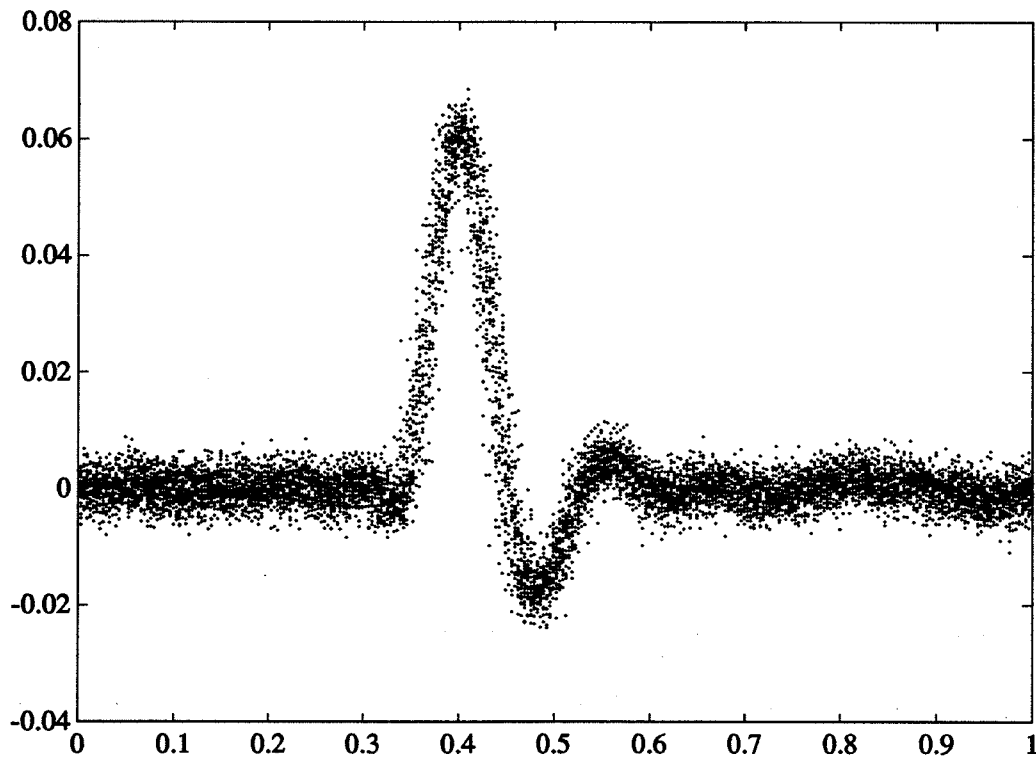


Figure 7 A second set of data generated by adding measurement noise and jitter to samples of a true signal. Thirty-two repeat measurements are made at each value of the control variable.

Table 1 Results produced by the algorithm for the data shown in Figure 7. For this data $\omega = 2.5 \times 10^{-3}$ and $\tau = 8.0 \times 10^{-3}$. The model chosen is a fourth order polynomial spline function with twenty interior knots.

Iteration	Estimate of ω	Estimate of τ	Error $\ s - \bar{s}\ _{\ell_\infty}$
1	2.394605×10^{-3}	7.236869×10^{-3}	2.050455×10^{-3}
2	2.394876×10^{-3}	6.883045×10^{-3}	2.078842×10^{-4}
3	2.394838×10^{-3}	6.927162×10^{-3}	2.603884×10^{-5}

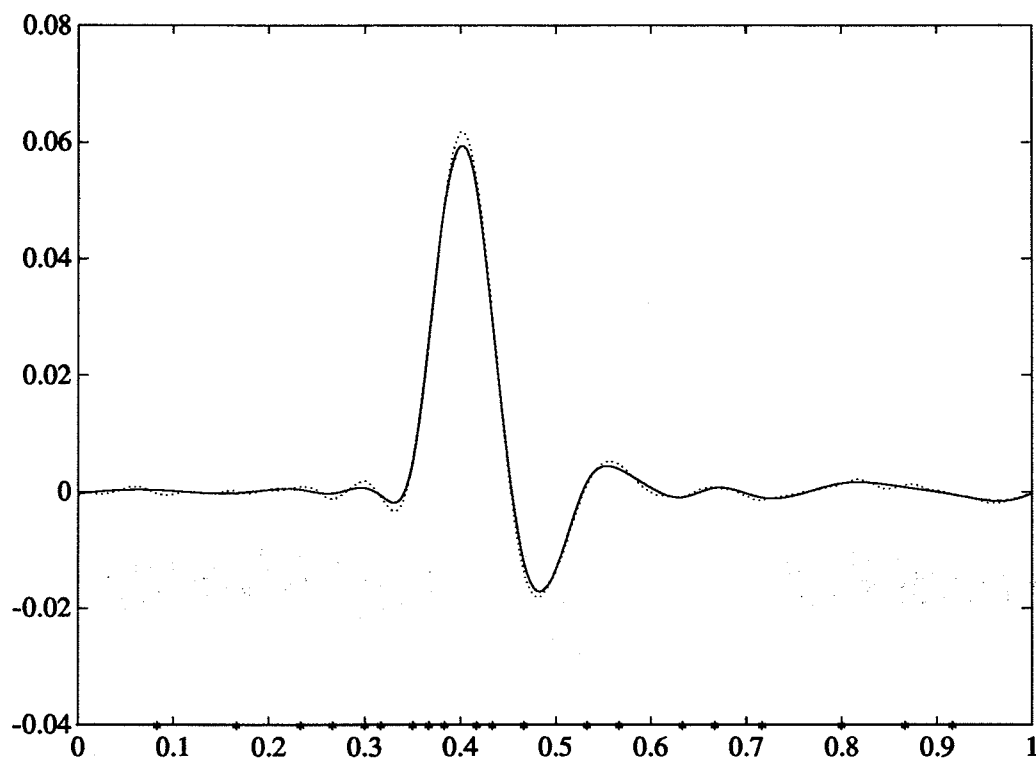


Figure 8 Initial weighted least-squares spline fit (Step 1) to the data shown in Figure 7.

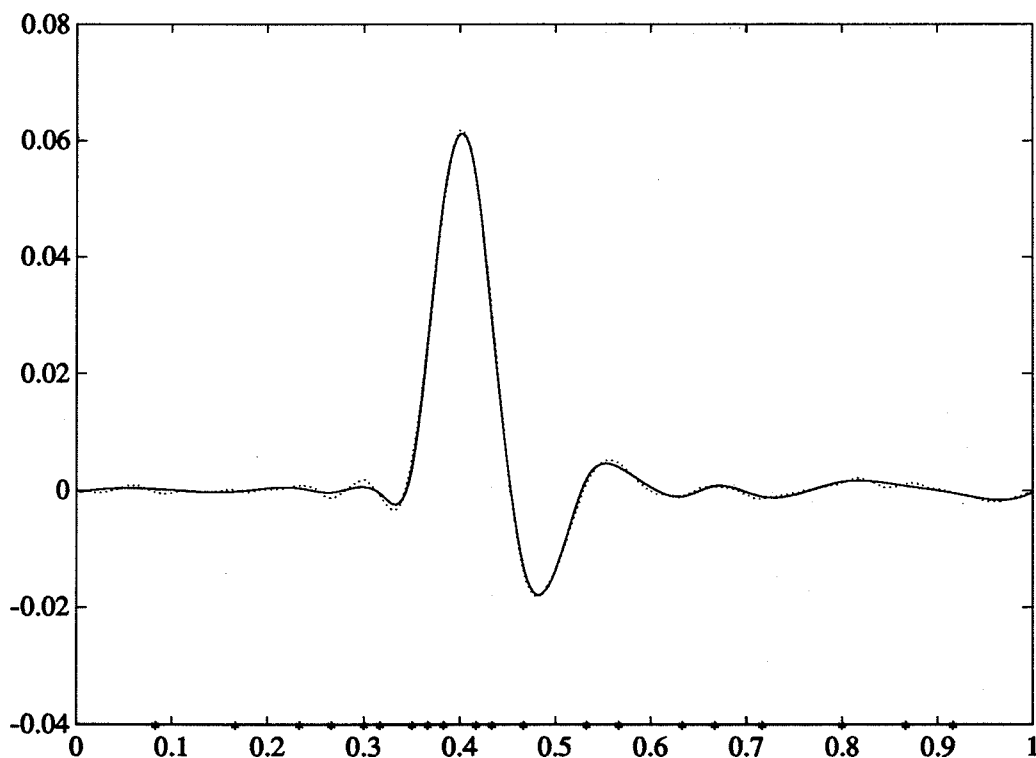


Figure 9 Final fit (after 3 iterations) to the data shown in Figure 7; this takes into account the bias and correct weighting of the measurements using estimates of ω and τ .

Table 2 Results produced by the algorithm for the data shown in Figure 5. For this data $\omega = 5.0 \times 10^{-3}$ and $\tau = 1.6 \times 10^{-2}$. The model chosen is a fourth order polynomial spline function with twenty interior knots.

Iteration	Estimate of ω	Estimate of τ	Error $\ s - \bar{s}\ _{\ell_\infty}$
1	4.800556×10^{-3}	1.601684×10^{-2}	8.597203×10^{-3}
2	4.813575×10^{-3}	1.231002×10^{-2}	7.978492×10^{-3}
3	4.819147×10^{-3}	1.287071×10^{-2}	3.165307×10^{-3}
4	4.816172×10^{-3}	1.324244×10^{-2}	5.728471×10^{-4}
5	4.815111×10^{-3}	1.316436×10^{-2}	1.679683×10^{-4}
6	4.815509×10^{-3}	1.315809×10^{-2}	1.174168×10^{-5}

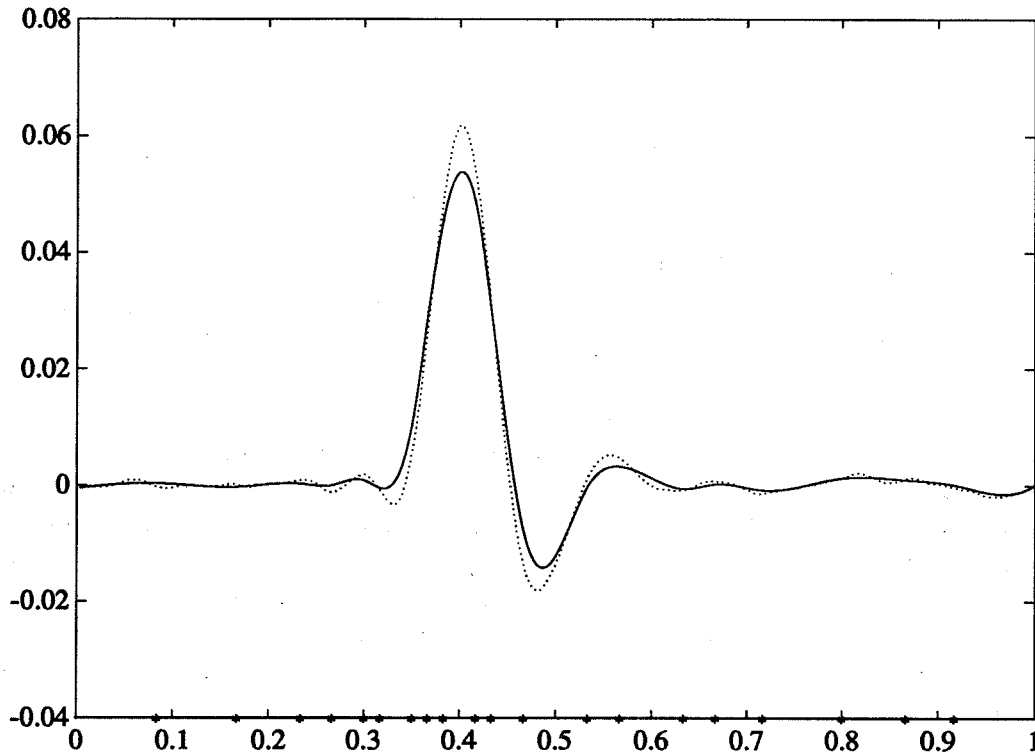


Figure 10 Initial weighted least-squares spline fit (Step 1) to the data shown in Figure 5.

6.2 EXAMPLE 2

For our second example we use the data shown in Figure 5 and constructed with $\omega = 5.0 \times 10^{-3}$ and $\tau = 1.6 \times 10^{-2}$. We use the same model for the true response as used for Example 1, and again set tol to be 1.0×10^{-4} . The initial and final fits are shown in Figures 10 and 11, respectively, and the results produced by the algorithm for each iteration are presented in Table 2. We observe that the final fit provides an acceptable approximation to the underlying function, and a substantial improvement over the initial fit.

7 CONCLUSION

In this paper we have been concerned with measurement processes for which the recorded measured data is contaminated by noise (random error in the response value) and jitter (random

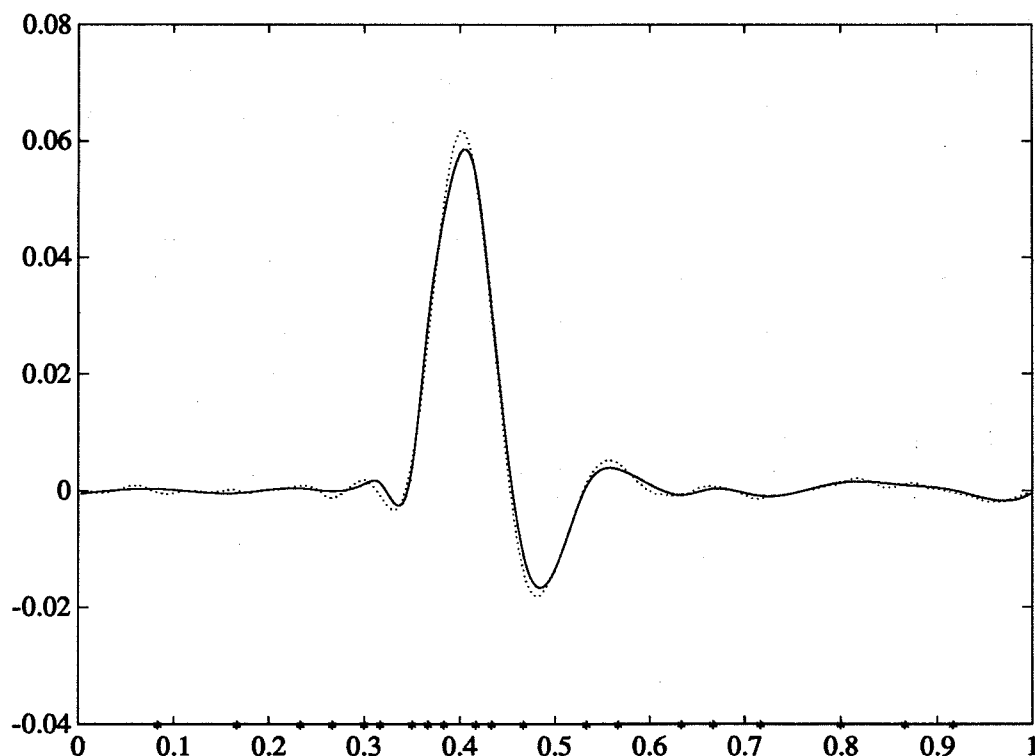


Figure 11 Final fit (after 6 iterations) to the data shown in Figure 5; this takes into account the bias and correct weighting of the measurements using estimates of ω and τ .

error in the control value at which a measurement is taken). An analysis of such measurement processes has shown the existence of a systematic bias in the measured data, and has related the variance of a measured value to the amount of noise and jitter that is present. Consequently, we have presented an algorithm for removing, or at least reducing, the effects of noise and jitter from a measured signal. To apply the algorithm we require that repeat measurements taken at each of a number of values of the control variable are available. Such replicate observations are readily available in the electrical measurement application considered. The algorithm returns estimates of the amount of noise and jitter present as well as an approximation to the true response underlying the data.

We conclude by listing a number of aspects of the work and related ideas that we believe warrant further consideration.

1. When computing estimates of ω and τ at Step 3 of the algorithm we have taken account of the uncertainty in the measured variance values σ_i^2 but not of the uncertainty in the values $(s^{(1)}(t_i))^2$. If these uncertainties were available, it might be more appropriate to compute a linear fit to the data by *total* least-squares. Estimates of these uncertainties can be obtained using DASL [1].
2. We have indicated (Section 4.3) that a limitation of the proposed algorithm is the ability of the model s to represent adequately the underlying function g . If s is taken from the class of polynomial splines, this limitation can be minimized by suitable choice of the number and distribution of interior knots used to define s . The topic of automatic knot placement in the context of least-squares data fitting by splines is the subject of [2] and [3], and many other useful references are given therein. An automatic knot placement algorithm could be used at Step 1 to generate an initial approximation s together with an appropriate set of interior knots, although we have no experience of doing this. Moreover,

although desirable, it is not obvious how the knots themselves should be updated as part of each iteration of the algorithm.

3. We have no experience of how the results produced by the algorithm depend on the quantity of data available, in particular the dependence on the number R of replicate measurements taken. It would be useful to the experimenter to quantify, if possible, this dependence.
4. Finally, Humphreys ([9]; Part II, Chapter 3) discusses the possibility of signal dependent jitter which arises in his measurement system where the trigger signal for the sampling oscilloscope is derived from the signal being measured. As a consequence, fluctuations in the signal amplitude give rise to variations in the triggering point, and the jitter distribution is no longer a fixed probability distribution. An analysis of this type of measurement process has not been considered, but is of practical interest.

8 ACKNOWLEDGEMENT

We thank Miss A P Cracknell for her contribution to the software implementation of the algorithm described in this report, and to our colleague Prof G T Symm for commenting on a draft of this paper.

9 REFERENCES

- [1] G.T. Anthony and M.G. Cox. The National Physical Laboratory's Data Approximation Subroutine Library. In J.C. Mason and M.G. Cox, editors, *Algorithms for Approximation*, pages 669–687. Clarendon Press, Oxford, 1987. Also *National Physical Laboratory DITC Report No. 71/86*.
- [2] M. G. Cox, P. M. Harris, and H. M. Jones. Strategies for knot placement in least squares data fitting by splines. Technical Report DITC 101/87, National Physical Laboratory, Teddington, UK, 1987.
- [3] M. G. Cox, P. M. Harris, and H. M. Jones. A knot-placement strategy for least-squares spline fitting based on the use of local polynomial approximations. In J.C. Mason and M.G. Cox, editors, *Algorithms for Approximation II*, pages 37–45. Chapman and Hall, London, 1988. Also *National Physical Laboratory DITC Report No. 148/89*.
- [4] M.G. Cox. The numerical evaluation of B-splines. *J. Inst. Math. Appl.*, 10:134–149, 1972.
- [5] M.G. Cox. Practical Spline Approximation. In P.R. Turner, editor, *Lecture Notes in Mathematics 965: Topics in Numerical Analysis*, pages 79–112. Springer-Verlag, Berlin, 1982. Also *National Physical Laboratory DITC Report No. 1/82*.
- [6] C. de Boor. On calculating with B-splines. *J. Approx. Theory*, 6:50–62, 1972.
- [7] C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.
- [8] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1982.
- [9] D. A. Humphreys. *Measurement Techniques, Numerical Methods and Photodiode Rise-time Transfer Standards for Picosecond Optoelectronic Device Metrology at 1 – 1.6 μ m*. PhD thesis, University of London, 1990.