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

From the Software Support for
Metrology Programme

Statistical Error Modelling

By
M G Cox and P M Harris

March 2004

Statistical Error Modelling

M G Cox and P M Harris
Centre for Mathematics and Scientific Computing

March 2004

ABSTRACT

Statistical error modelling is regarded as a process that involves developing relationships between measurement data, the required measurement result(s), and the measurement deviations (or 'errors') associated with the measurement data. The problem of evaluating the measurement result(s), and the associated uncertainty, in cases where a statistical model is assigned to the values of the measured quantities is considered. A number of cases are addressed, such as where the measurement deviations arise from random effects only or from a combination of random and systematic effects, and where the statistical model for the effects is completely or only partially specified. The focus is on regression and its application to the calibration of an instrument or measurement system. A number of examples are discussed, concerned with characterising the performance of a clock, developing RF coaxial frequency standards, modelling spectral characteristic data, and assessing the roundness of a nominally circular artefact.

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

i

1   Introduction

This report is about *statistical error modelling* and its application to problems in metrology. Here, statistical error modelling is considered as a process that involves developing relationships between measurement data, the required measurement result(s) and the *errors* associated with the measurement data. We note that the term 'error' is often deprecated because it relates to an 'unknowable' quantity, viz., the difference between the (measured) value of a quantity and the 'true' value of the quantity. The term is here interpreted differently, viz., as the difference between the (measured) value of a quantity and a (best) estimate of the value of the quantity. Usually, the (best) estimate is the expected value of the quantity (regarded as a random variable) [24].[1] In order to avoid confusion, as well as controversy, the terms 'measurement deviation' or 'residual deviation' or (simply) 'deviation' will be used henceforth in place of 'error' with the interpretation given above. One advantage of this interpretation is that the measured value and the associated measurement deviation are straightforwardly related: if x is a measurement of the value of the quantity X with expectation μ and variance σ^2, then $e = x - \mu$ is the measurement deviation associated with x and is a sample of a random variable $E = X - \mu$ with expectation zero and variance σ^2. Often, the expectation μ is represented in terms of a model (as in the regression problem described below).

Statistical error modelling can be translated into the framework of the Guide to the Expression of Uncertainty in Measurement (GUM) [2] in which the measurement deviations become subsumed in input quantities and their influences summarised by uncertainties. The methods for uncertainty evaluation described in the GUM may then be applied to obtain the uncertainties associated with the required measurement result(s). An example to illustrate the relationship between statistical error modelling and the 'input-output' modelling of the GUM is available [14].

This report concerns the problem of evaluating the measurement result(s), with the associated uncertainty, in cases where a statistical model is *assigned* to the values of the measured quantities (or, equivalently, to the values of the measurement deviations associated with those quantities). The model can take the form of estimates of the values of the measurement deviations,[2] with the associated uncertainties, or a probability distribution with known parameters (expectation, variance, etc.). The model is assigned on the basis of prior knowledge about the measurement system or instrument giving rise to the data using considerations of maximum information entropy [25].

The approaches considered in this work are in contrast to *non-parametric* methods, such as the bootstrap and jackknife, which are based on re-using the available data to evaluate the measurement result(s), with the associated uncertainty, without *a*

---

[1]Equivalently, the expectation (mean) of the distribution function for the value of the quantity.
[2]Typically, the estimates will be zero.

priori assigning a statistical model. Non-parametric methods are the subject of a companion report [13].

The focus of this report is on *regression* and its application to the calibration of an instrument or measurement system (section 2). We suppose that a linear regression model is known,[3] of the form

$$v_i = \sum_{j=1}^{n} a_j h_j(t_i) + e_i, \quad i = 1, \ldots, m,$$

in which v_i is a measurement of a response v corresponding to a value t_i of a stimulus t, $\{h_j(t)\}_1^n$ are basis functions defining a model for v in terms of t, $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$ is a vector of model parameters, and $\mathbf{e} = (e_1, e_2, \ldots, e_m)^{\mathrm{T}}$ a vector of measurement deviations associated with the measurements of v.[4]

In some situations it is possible to regard the measurement deviations $\mathbf{e}$ as resulting from *random effects* only. Such effects produce variation in repeated measurements of the response (for each value of the stimulus). Furthermore, the uncertainties associated with estimates of the values of the model parameters $\mathbf{a}$ arising from such effects, as well as those associated with the required measurement results (derived from these estimates), can generally be reduced by increasing the number of measurements. We suppose that in this case $\mathbf{e} = \mathbf{e}_r$ is a sample of random variables $\mathbf{E}_r$ for which the estimate of the value of $\mathbf{E}_r$ is the zero vector with associated uncertainty matrix (covariance matrix) V_r. In the absence of any other knowledge about the measurements, and applying the principle of maximum information entropy, a multivariate normal (Gaussian) probability density function $N(\mathbf{0}, V_r)$ is assigned to the value of $\mathbf{E}_r$.

In other circumstances it may be known that the measurement deviations $\mathbf{e}$ arise from a combination of random and *systematic effects*. The latter give rise to deviations that are *common* to repeated measurements of the response (corresponding to the same and possibly different values of the stimulus). Generally, the uncertainties associated with the estimates of the values of the model parameters $\mathbf{a}$ arising from such effects cannot be estimated from repeated measurements or reduced by increasing the number of such measurements. We write

$$\mathbf{e} = \mathbf{b} + \mathbf{e}_r,$$

with $\mathbf{e}_r$ defined as above, and $\mathbf{b}$ a sample of random variables $\mathbf{B}$, with $\mathbf{E}_r$ and $\mathbf{B}$ mutually independent. We suppose that the estimate of the value of $\mathbf{B}$ is the zero vector[5] with associated uncertainty matrix V_b.

---

[3]A formulation in which the model is nonlinear can also be treated [12].

[4]In this formulation (discussed more fully in section 2) it is assumed that the measurements made of the stimulus are known accurately compared to those of the response. A formulation in which the measurements of both response and stimulus are regarded as inexact can also be treated [12].

[5]In some situations the estimate for the value of $\mathbf{B}$ will not be the zero vector. In these situations a *correction* can be applied to the measurements so that, following the correction, the estimate for the value of $\mathbf{B}$ is the zero vector [2, Clause 3.2].

Consideration is given to methods for the determination of estimates of the values of the model parameters $\mathbf{a}$, together with the uncertainty matrix associated with these estimates, in the cases that the following statistical models for the measurement deviations are assumed to apply:

1. There are random effects only and these effects are quantified, i.e., V_r is known (section 3)

2. There are random and systematic effects and these effects are quantified, i.e., V_r and V_b are known (section 4)

3. There are random effects only and these effects are *unquantified*, i.e., V_r is unknown (section 5).

   A number of methods are considered for problems in this category, but for each *some* knowledge about the statistical model is assumed. Usually, this takes the form of an uncertainty matrix V_r associated with the random effects that can be described by parameters $\boldsymbol{\lambda}$, i.e., $V_r = V_r(\boldsymbol{\lambda})$. Some examples include:

   (a) $V_r = \mathrm{diag}\{\sigma^2, \sigma^2, \ldots, \sigma^2\}$, with $\lambda = \sigma$ unknown[6]

   (b) $V_r = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2\}$, with $\boldsymbol{\lambda} = (\sigma_1, \sigma_2, \ldots, \sigma_m)^{\mathrm{T}}$ unknown[7]

   (c) The components of $\mathbf{E}_r$ are (serially) correlated, with the measurement deviations $\mathbf{e} = \mathbf{e}_r$ satisfying

   $$e_i = \sum_{\ell=1}^{q} c_\ell e_{i-\ell} + z_i, \quad i = q+1, \ldots, m,$$

   where z_i, $i = q+1, \ldots, m$, are independent samples from a (normal) distribution with zero expectation and variance σ^2. Here, $\boldsymbol{\lambda} = (c_1, c_2, \ldots, c_q, \sigma)^{\mathrm{T}}$ and q are unknown.[8]

The methods described are illustrated by a number of metrology problems, concerned with characterising the performance of a clock (section 3.3), developing RF coaxial frequency standards (section 4.2), modelling spectral characteristic data (section 5.3.1), and assessing the roundness of a nominally circular artefact (section 5.4.3). Summary remarks are given in section 6.

---

[6]The components of $\mathbf{E}_r$ are independent and identically distributed with unknown variance.

[7]The components of $\mathbf{E}_r$ are independent but with (generally) different variances. This statistical model may be used to describe, for example, measurement processes for which the standard uncertainty $u(v)$ associated with a measurement v depends on (a) 'where' the measurement is taken, i.e., it depends on the value t of the stimulus, or (b) the measurement itself, i.e., it depends on the value v of the response. For example, the latter arises in cases where the *relative* standard uncertainty $u(v)/v$ associated with a measurement v is constant.

[8]This statistical model defines an *autoregressive* (AR) process. A simple example of an AR process is *random walk* for which $q = 1$ and $c_1 = 1$.

2 Problem formulation

Many metrology experiments involve determining the behaviour of a response v as a function of a stimulus t. *Model building* involves establishing a functional relationship between these quantities, which we write generically as

$$v = h(t, \mathbf{a})$$

in terms of parameters $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$. In the particular case that h is a *linear* function of the parameters $\mathbf{a}$, the functional model takes the form

$$v = \sum_{j=1}^{n} a_j h_j(t), \tag{1}$$

in which $\{h_j(t)\}_1^n$ is a set of linearly independent functions of the stimulus t. The focus of this report is on linear functional models of the form (1).

The metrology experiment yields measurement data (t_i, v_i), $i = 1, \ldots, m$, which we assume here is gathered in such a way that the measurements t_i of t are accurate relative to the measurements v_i of v. *Discrete modelling* involves establishing relationships between the measurement data and the parameters $\mathbf{a}$ that define the behaviour of the instrument or measurement system, supplemented by (perhaps incomplete) information about the statistical nature of the measurements. In the situation described, these relationships take the form

$$v_i = \sum_{j=1}^{n} a_j h_j(t_i) + e_i, \quad i = 1, \ldots, m, \tag{2}$$

where e_i denotes the (unknown) measurement deviation associated with v_i. Equivalently, using matrix-vector notation,

$$\mathbf{v} = H\mathbf{a} + \mathbf{e}, \tag{3}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_m)^{\mathrm{T}}$, $\mathbf{e} = (e_1, e_2, \ldots, e_m)^{\mathrm{T}}$ and the ijth element H_{ij} of H is given by

$$H_{ij} = h_j(t_i), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

Finally, *model solving* is concerned with using knowledge about the statistical nature of the measurements (or, equivalently, about the associated deviations $\mathbf{e}$) to 'solve' the equations (2) or (3) to obtain estimates $\widehat{\mathbf{a}}$ of the values of the parameters $\mathbf{a}$, together with the uncertainty matrix $V_{\widehat{a}}$ associated with these estimates.

A further requirement is often to evaluate the uncertainty associated with a function of the model parameters, such as the value of the functional model corresponding to a specified stimulus $t = t_0$:

$$v_0 = \sum_{j=1}^{n} a_j h_j(t_0).$$

3 Quantified random effects

Suppose the measurement deviations $\mathbf{e}$ in (3) result from random effects only, with $\mathbf{e} = \mathbf{e}_r$ a sample of random variables $\mathbf{E}_r$ with zero expectation and uncertainty matrix V_r. Then, estimates $\widehat{\mathbf{a}}$ of the values of the model parameters $\mathbf{a}$ are obtained by solving the *least-squares* problem:

$$\min_{\mathbf{a}}(\mathbf{v} - H\mathbf{a})^{\mathrm{T}}V_r^{-1}(\mathbf{v} - H\mathbf{a}). \tag{4}$$

The use of least-squares analysis is supported by the Gauss-Markov theorem [21] that states that of all linear unbiased estimates of the values of the model parameters, the least-squares estimate is that with minimum variance.

Formally,[9] the solution to (4) is given by [21]

$$\widehat{\mathbf{a}} = \left(H^{\mathrm{T}}V_r^{-1}H\right)^{-1}H^{\mathrm{T}}V_r^{-1}\mathbf{v} \tag{5}$$

with associated uncertainty matrix

$$V_{\widehat{a}} = \left(H^{\mathrm{T}}V_r^{-1}H\right)^{-1}. \tag{6}$$

Furthermore, estimates of the values of the measurement deviations are given by

$$\widehat{\mathbf{e}} = \mathbf{v} - H\widehat{\mathbf{a}}, \tag{7}$$

with associated uncertainty matrix

$$V_{\widehat{e}} = V_r - H\left(H^{\mathrm{T}}V_r^{-1}H\right)^{-1}H^{\mathrm{T}}. \tag{8}$$

In the framework of the GUM, we would consider the models

$$\widehat{\mathbf{A}} = \left(H^{\mathrm{T}}V_r^{-1}H\right)^{-1}H^{\mathrm{T}}V_r^{-1}\mathbf{V} = \left(H^{\mathrm{T}}V_r^{-1}H\right)^{-1}H^{\mathrm{T}}V_r^{-1}\left(\mathbf{v} + \mathbf{E}\right), \tag{9}$$

and

$$\widehat{\mathbf{E}} = \mathbf{V} - H\widehat{\mathbf{A}} = (\mathbf{v} + \mathbf{E}) - H\widehat{\mathbf{A}} \tag{10}$$

in terms of vectors $\mathbf{V}$, $\mathbf{E}$, $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{E}}$ of random variables. $\mathbf{V}$ (or $\mathbf{E}$) constitute the *input quantities* (the measurements or their deviations) and $\widehat{\mathbf{A}}$ (or $\widehat{\mathbf{E}}$) the *output quantities* (the least-squares estimates of the model parameters or the residual deviations associated with least-squares fitted model). Then, (5) is obtained by evaluating the model (9) for the estimates $\mathbf{v}$ of $\mathbf{V}$ (or $\mathbf{0}$ of $\mathbf{E}$), and (6) is obtained by applying the law of propagation of uncertainty [2] to this model. Similarly, (7) and (8) are obtained from (10).

---

[9]For reasons of numerical stability it is generally recommended that these formal expressions are not used to *compute* the least-squares solution and its associated uncertainty matrix, but matrix factorisation methods are used instead [1, 12].

3.1 Example: Mean and mutually independent measurements

Suppose the functional model takes the form

$$h(t, \mathbf{a}) = a, \tag{11}$$

and the statistical model is

$$V_r = \sigma^2 I,$$

where I is the identity matrix or order m. The model (11) is a particularly simple example of the generic model (1), with $n = 1$, $a_1 = a$ and $h_1(t) \equiv 1$. The data v_i, $i = 1, \ldots, m$, represent measurements of a (unknown) constant response a of the instrument or measurement system, which are subject to random deviations that are independently and identically distributed. Using matrix-vector notation,

$$\mathbf{v} = \mathbf{1}a + \mathbf{e}_r,$$

where $\mathbf{1} = (1, 1, \ldots, 1)^{\mathrm{T}}$.

Applying (5) and (6), we obtain

$$\widehat{a} = \left(\mathbf{1}^{\mathrm{T}} \sigma^{-2} I \mathbf{1} \right)^{-1} \mathbf{1}^{\mathrm{T}} \sigma^{-2} I \mathbf{v} = \frac{1}{m} \sum_{i=1}^{m} v_i,$$

and

$$V_{\widehat{a}} = \frac{1}{m} \sigma^2.$$

This solution corresponds to the (familiar) result that, the (least-squares) estimate of the value of a for independent and identically distributed measurements is the sample mean of the data. We see that the variance associated with the estimate is inversely proportional to the number m of measurements, again a familiar result, and, therefore, can be reduced by increasing m.[10]

Figure 1 shows an example of data simulated on the basis of the functional and statistical models described above.[11] The data is presented in the form of a time series of $m = 101$ points (t_i, v_i) where

$$t_i = i - 1, \quad v_i = a + e_{r,i}, \quad i = 1, \ldots, m,$$

with $a = 0$ and $e_{r,i}$, $i = 1, \ldots, m$, independent samples from $N(0, 1)$, the normal distribution with zero expectation and unit variance. Also shown are the values of $\widehat{a}$ (solid line) and $\widehat{a} \pm u(\widehat{a})$ (broken lines), where $u(\widehat{a}) = \sqrt{V_{\widehat{a}}}$ is the standard uncertainty associated with $\widehat{a}$.[12]

---

[10]An important, and related, problem arises when V_r is diagonal but not with equal diagonal elements, i.e., the data represents measurements subject to random deviations that are independently but *not* identically distributed. In this case the least-squares estimate corresponds to a *weighted*
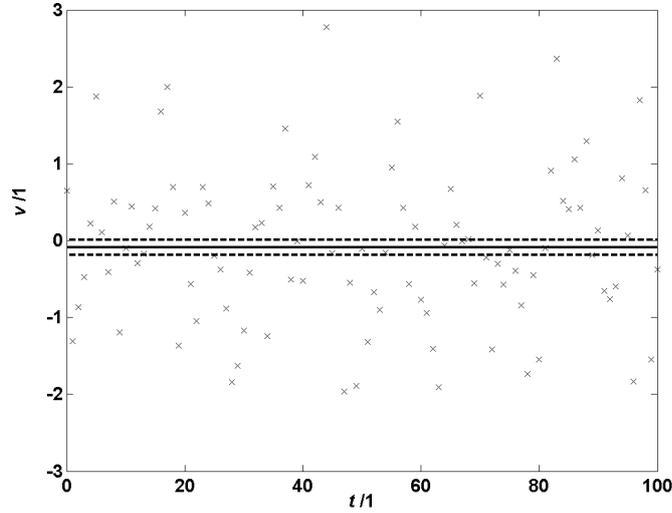
Figure 1: Data simulated on the basis of a constant functional model and mutually independent random measurement deviations. Also shown is the estimate $\widehat{a}$ (solid line) of the value of a and the standard uncertainty 'envelope' defined by $\widehat{a} \pm u(\widehat{a})$ (broken lines) associated with the estimate.

3.2 Example: Mean and mutually dependent measurements

Suppose the functional model is (11), but the measurements satisfy the 'random walk' statistical model

$$v_i = a + e_{r,i}, \quad i = 1, 2, \ldots, m,$$

with

$$e_{r,1} = z_1, \quad e_{r,i} = e_{r,i-1} + z_i = \sum_{k=1}^{i} z_k, \quad i = 2, 3, \ldots, m,$$

and $\mathbf{z} = (z_1, z_2, \ldots, z_m)^{\mathrm{T}}$ is a sample of a random variable $\mathbf{Z}$ with zero expectation and uncertainty matrix

$$V_z = \sigma^2 I.$$

_____

mean of the data. The result is important in the context of interlaboratory comparisons including key comparisons [9].

[11]In this example, as well as those described in sections 3.2 and 3.3, measurements of the stimulus and response variables are assumed to be dimensionless.

[12]It is emphasized that, in this example, the 'dispersion' of possible measurements v_k at $t = k$ is σ, which is (an order of magnitude) greater than $u(\widehat{a}) \approx \sigma/10$, the standard uncertainty associated with the estimate $\widehat{a}$ of a.

By writing

$$\mathbf{E}_r = T\mathbf{Z}, \quad T = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 1 & 0 \\ 1 & \cdots & \cdots & 1 & 1 & 1 \end{pmatrix},$$

it follows that $\mathbf{e}_r = (e_{r,1}, e_{r,2}, \ldots, e_{r,m})^{\mathrm{T}}$ is a sample of a random variable $\mathbf{E}_r$ with zero expectation and, using the law of propagation of uncertainty [14, chapter 6], uncertainty matrix

$$V_r = \sigma^2 T T^{\mathrm{T}}.$$

Applying (5) and (6), we obtain

$$\begin{aligned} \widehat{a} &= \left(\mathbf{1}^{\mathrm{T}}\sigma^{-2}T^{-\mathrm{T}}T^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^{\mathrm{T}}\sigma^{-2}T^{-\mathrm{T}}T^{-1}\mathbf{v} \\ &= \left((T^{-1}\mathbf{1})^{\mathrm{T}}(T^{-1}\mathbf{1})\right)^{-1}(T^{-1}\mathbf{1})^{\mathrm{T}}(T^{-1}\mathbf{v}), \end{aligned}$$

and

$$V_{\widehat{a}} = \sigma^2 \left((T^{-1}\mathbf{1})^{\mathrm{T}}(T^{-1}\mathbf{1})\right)^{-1}.$$

But

$$T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{pmatrix},$$

so that

$$T^{-1}\mathbf{1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and} \quad T^{-1}\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 - v_1 \\ \vdots \\ v_m - v_{m-1} \end{pmatrix},$$

from which we obtain

$$\widehat{a} = v_1,$$

and

$$V_{\widehat{a}} = \sigma^2.$$

In this case, the (least-squares) estimate of the value of a is the first measurement, and the variance associated with the estimate is independent of the number m

of measurements and cannot be reduced by increasing m. The statistical model described here is sometimes referred to as a *random walk* process.[13]

Figure 2 shows an example of data simulated on the basis of the functional and statistical models described above. As in section 3.1, the data is presented in the form of a time series with $a = 0$ and z_i, $i = 1, \ldots, m$, independent samples from $N(0, 1)$. Also shown are the values of $\widehat{a}$ (solid line) and $\widehat{a} \pm u(\widehat{a})$ (broken lines).[14]
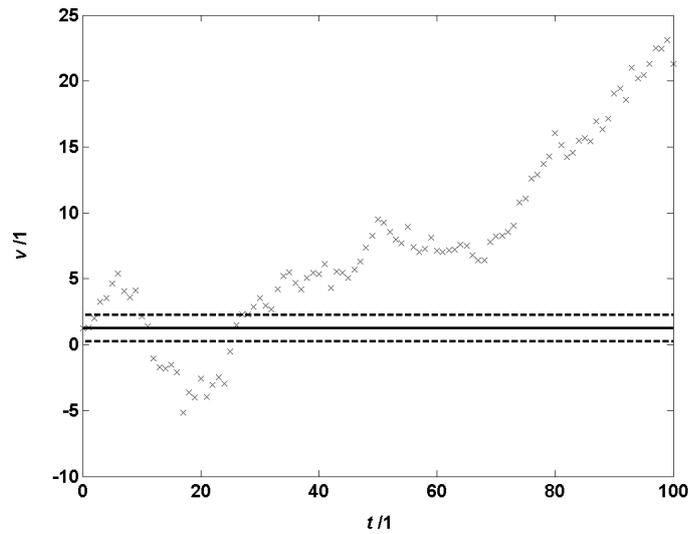

Figure 2: Data simulated on the basis of a constant functional model and mutually dependent random measurement deviations. Also shown is the estimate $\widehat{a}$ (solid line) of the value of a and the standard uncertainty envelope defined by $\widehat{a} \pm u(\widehat{a})$ (broken lines) associated with the estimate.

3.3 Example: Characterising a clock

Let $v(t)$ denote the phase difference, as a function of time t, between two clocks[15] and $\{v_i : i = 1, \ldots, m\}$ a time series of measurements of v made at uniformly spaced times $\{t_i : i = 1, \ldots, m\}$ [19]. We suppose that $v(t)$ can be modelled

---

[13]Although the estimate of the value of a is determined by the first measurement, the uncertainty associated with the estimate depends on 'where' (or 'when') the first measurement is in relation to 'where' (or 'when') the process is regarded as 'starting'. If the only available measurements are those starting at $t = k$ $(1 \le k \le m)$, the measurement at $t = k$ provides the (least-squares) estimate of the value of a. However, the uncertainty associated with the estimate is σ if the process is assumed to start at $t = k$, or $\sigma\sqrt{k}$ if it is assumed to start at $t = 1$.

[14]It is emphasized that, in this example, the standard uncertainty associated with the measurement v_k is $\sigma\sqrt{k}$ and, consequently, the 'dispersion' of possible measurements v_k at $t = k$, $k > 1$, always exceeds $u(\widehat{a}) = \sigma$, the standard uncertainty associated with the estimate $\widehat{a}$ of a.

[15]One clock may be regarded as a reference clock or timescale.

in terms of parameters p_0 and f_0 that describe, respectively, the phase offset and normalised frequency offset between the clocks as follows:[16]

$$v(t) = p_0 + f_0 t. \tag{12}$$

Furthermore, we suppose the measurements can be modelled as

$$v_i = p_0 + f_0 t_i + e_{r,i}, \quad i = 1, \ldots, m,$$

or, using matrix-vector notation,

$$\mathbf{v} = H\mathbf{a} + \mathbf{e}_r,$$

where

$$H = \begin{pmatrix} \mathbf{1} & \mathbf{t} \end{pmatrix}, \quad \mathbf{t} = (t_1, t_2, \ldots, t_m)^{\mathrm{T}}, \quad \mathbf{a} = (p_0, f_0)^{\mathrm{T}}.$$

It is usual in time metrology to regard $\mathbf{e}_r$ as the sum of independent samples of known 'noise types', such as *white phase modulation* (wpm), *white frequency modulation* (wfm)[17] and *random walk frequency modulation* (rwfm).[18] Here, white phase noise may usually be attributed to the phase comparator used to measure the phase difference, and white frequency and random walk frequency noise processes to the stochastic nature of the clocks themselves.

Figure 3 shows an example of a time series representing measurements of the phase difference between a commercial caesium clock and the timescale Universal Coordinated Time (UTC). Figure 4 shows the same time series but with a linear trend removed, in order to expose the structure of the noise associated with the measurements.

We consider a model for the measurement deviations $\mathbf{e}_r = (e_{r,1}, e_{r,2}, \ldots, e_{r,m})^{\mathrm{T}}$ that takes the form

$$\mathbf{e}_r = \mathbf{e}_{\mathrm{wpm}} + \mathbf{e}_{\mathrm{wfm}} + \mathbf{e}_{\mathrm{rwfm}},$$

where

- $\mathbf{e}_{\mathrm{wpm}}$ is a sample of random variables $\mathbf{E}_{\mathrm{wpm}}$ with zero expectation and uncertainty matrix

$$V_{\mathrm{wpm}} = \sigma^2_{\mathrm{wpm}} I$$

(as in section 3.1)

---

[16]A generalisation is possible in which an additional term, $(f_d/2)t^2$, is included, where f_d describes the normalised frequency drift between the clocks.

[17]Also described as random walk phase modulation.

[18]Other noise types are sometimes also considered, including *flicker phase modulation* (fpm), *flicker frequency modulation* (ffm) and *random run frequency modulation* (rrfm). The noise types encountered here can be distinguished (and characterised) by their power spectral density functions.
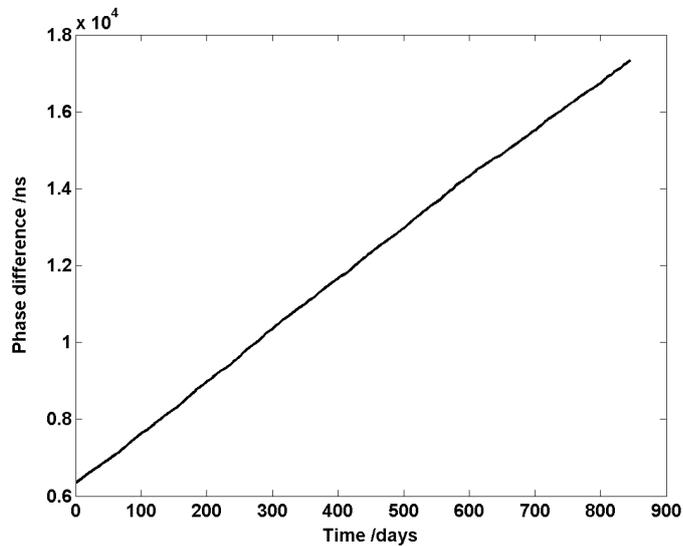
Figure 3: Time series (spanning almost two and a half years) representing measurements of the phase difference between a commercial caesium clock and the timescale UTC. The functional model describing phase difference as a function of time takes the form of a straight-line model.
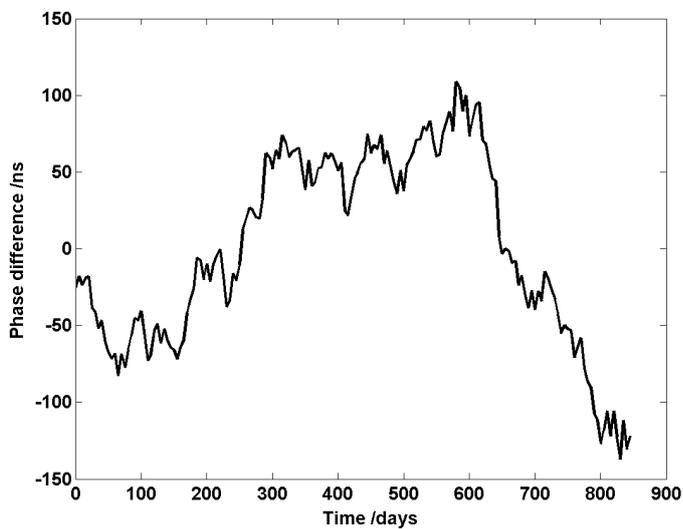

Figure 4: As Figure 3, but with a linear trend removed. Note the two orders of magnitude difference between the vertical scales in this figure and Figure 3. The statistical model describing the measurements of phase difference takes the form of a combination of white phase, white frequency and random walk frequency noise processes.

- $\mathbf{e}_{\mathrm{wfm}}$ is a sample of random variables $\mathbf{E}_{\mathrm{wfm}}$ with zero expectation and uncertainty matrix

$$V_{\mathrm{wfm}} = \sigma_{\mathrm{wfm}}^2 T T^{\mathrm{T}}$$

(as in section 3.2)

- $\mathbf{e}_{\mathrm{rwfm}}$ is a sample of random variables $\mathbf{E}_{\mathrm{rwfm}}$ with zero expectation and uncertainty matrix

$$V_{\mathrm{rwfm}} = \sigma_{\mathrm{rwfm}}^2 (T^2)(T^2)^{\mathrm{T}}$$

(the natural extension of white phase and white frequency noise processes)

- $\mathbf{E}_{\mathrm{wpm}}$, $\mathbf{E}_{\mathrm{wfm}}$ and $\mathbf{E}_{\mathrm{rwfm}}$ are mutually independent.

It follows that $\mathbf{e}_r$ is a sample of random variables $\mathbf{E}_r$ with zero expectation and uncertainty matrix

$$V_r = \sigma_{\mathrm{wpm}}^2 I + \sigma_{\mathrm{wfm}}^2 T T^{\mathrm{T}} + \sigma_{\mathrm{rwfm}}^2 (T^2)(T^2)^{\mathrm{T}}.$$

Provided σ_{wpm}, σ_{wfm} and σ_{rwfm} are *known*, (5) and (6) may be applied to obtain an estimate $\widehat{\mathbf{a}} = (\widehat{p}_0, \widehat{f}_0)^{\mathrm{T}}$ of the value of $\mathbf{a} = (p_0, f_0)^{\mathrm{T}}$ with associated uncertainty matrix $V_{\widehat{a}}$.

The estimates, together with the model (12), are then used to provide an estimate $\widehat{v}$ of the value of the phase difference v between the two clocks at any given time t:

$$\widehat{v} = \widehat{p}_0 + \widehat{f}_0 t.$$

Furthermore, by applying the law of propagation of uncertainty [2], the standard uncertainty $u(\widehat{v})$ associated with the estimate $\widehat{v}$ is obtained from

$$u^2(\widehat{v}) = J V_{\widehat{a}} J^{\mathrm{T}},$$

where

$$J = \begin{pmatrix} 1 & t \end{pmatrix}$$

contains the sensitivity coefficients for the phase difference v with respect to the model parameters p_0 and f_0.

Figure 5 shows an example of data simulated on the basis of the functional and statistical models described above, with

$$p_0 = 0, \quad f_0 = 0.01, \quad \sigma_{\mathrm{wpm}} = \sigma_{\mathrm{wfm}} = 1, \quad \sigma_{\mathrm{rwfm}} = 0.$$

Also shown are the estimates $\widehat{v}$ of the values of the phase difference corresponding to each t_i (solid line) together with the standard uncertainty envelope $\widehat{v} \pm u(\widehat{v})$ (broken lines) associated with these estimates.
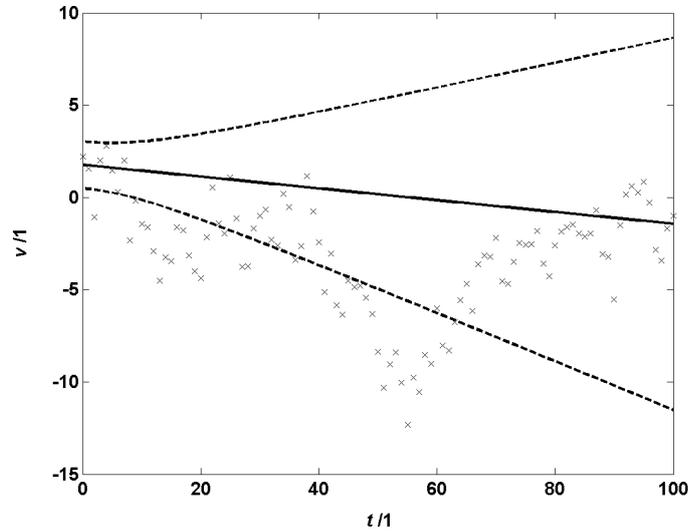
Figure 5: Data simulated on the basis of a straight-line functional model and a statistical model that is a combination of white phase and white frequency noise processes. Also shown are the estimates $\widehat{v}$ of the value of phase difference corresponding to times t_i (solid line) and the standard uncertainty envelope $\widehat{v} \pm u(\widehat{v})$ (broken lines) associated with these estimates.

4 Quantified random and systematic effects

Suppose the measurement deviations $\mathbf{e}$ result from a combination of random and systematic effects, so that (3) takes the form

$$\mathbf{v} = H\mathbf{a} + \mathbf{e} = H\mathbf{a} + \mathbf{b} + \mathbf{e}_r.$$

If the systematic effects are genuinely repeatable (but unkown), it is generally not possible, without further information, to 'separate' them from the functional part of the model. This is because the effects may be *confounded* with the functional model, i.e.,

$$H\mathbf{a} + \mathbf{b} = H\tilde{\mathbf{a}}$$

for different values $\tilde{\mathbf{a}}$ of the model parameters,[19] or there may be insufficient information (data) to make the separation. In other words, it is not possible to obtain estimates for the values of the systematic measurement deviations from an analysis of the measurement data, as can be done for random measurement

---

[19]For the confounding to be 'perfect' $\mathbf{b}$ must be in the range space of H. In other cases the systematic effects and the functional model may be 'nearly' confounded, i.e., $\mathbf{b}$ is 'close to' this range space, and sufficiently close to make it difficult (computationally) to separate the systematic effects from the functional model. Often, in practice, a functional model is chosen to represent the *combination* of instrument response and systematic effects.

deviations in the case that the measurements are subject only to random effects (section 3).[20]

Suppose the only information known about the systematic effects are estimates of their values[21] and the standard uncertainties associated with the estimates. Consider the problem of evaluating an estimate $\hat{v}$ of the value v of the functional model, with the associated uncertainty, for a specified value t of the stimulus.[22] An approach to addressing this problem is as follows:

- Setting $\mathbf{b} = \mathbf{0}$, obtain estimates $\widehat{\mathbf{a}}$ of the values of the model parameters $\mathbf{a}$, with the associated uncertainty matrix $V_{\widehat{a}}$, by solving the least-squares problem (4).

- An estimate $\widehat{v}$ of the value of the functional model is then given by

$$\widehat{v} = h(t, \widehat{\mathbf{a}}) + b,$$

with associated standard uncertainty $u(\widehat{v})$ obtained from

$$u^2(\widehat{v}) = u_r^2 + u^2(b).$$

Here, u_r is the standard uncertainty associated with the term '$h(t, \widehat{\mathbf{a}})$' evaluated in terms of the results $\widehat{\mathbf{a}}$ and $V_{\widehat{a}}$ of the least-squares analysis. This term accounts for the random effects to which the measurements are subject. Furthermore, b represents the systematic effect to which a measurement corresponding to the value t of the stimulus is subject, and is regarded as a sample of a random variable B with zero expectation and variance $u^2(b)$.[23]

The approach is illustrated using the examples described in sections 4.1 and 4.2.

In some cases more will be known about the systematic effects. An example is where the statistical model for the measurements can be expressed in terms of random effects and *common* systematic effects, all of which are (statistically) independent and quantified individually. In this case it is possible to determine *completely* the uncertainty matrix associated with the measurements (or, equivalently, associated with their deviations) and to 'propagate' together the standard uncertainties associated with the random and systematic effects. This approach is illustrated using the example described in section 4.1, and compared with the above approach.

---

[20]The exception is in cases for which the measurement data is collected in a way *designed* to make these effects estimable. If the presence, and likely cause (e.g., operator, machine, time of day, etc.), of systematic effects is suspected before any measurements are made, it may be possible to collect the data in such a way that confounding of the functional model with the systematic effects is avoided, and estimates of the values of the parameters of the functional model, the systematic and the random effects are obtained.

[21]Assumed here to be zero [2, Clause 3.2].

[22]Other measurement results derived from the model parameters may also be considered.

[23]The example described in section 4.2 illustrates the application of a 'pragmatic' procedure to derive $u(b)$ from information about the measurements.

4.1 Example: Mean and quantified random and systematic effects

Suppose the functional model is (11), but the measurements satisfy

$$v_i = a + e_i = a + b + e_{r,i},$$

where b is a sample of random variable B with zero expectation and variance σ_b^2, and $\mathbf{e}_r = (e_{r,1}, e_{r,2}, \ldots, e_{r,m})^{\mathrm{T}}$ is a sample of random variable $\mathbf{E}_r$ with zero expectation and associated uncertainty matrix

$$V_r = \sigma^2 I.$$

In addition, we assume that B and $\mathbf{E}_r$ are mutually independent.

Defining

$$\alpha = a + b,$$

we write the model equations in the form

$$v_i = \alpha + e_{r,i}.$$

Then, as for the case of mutually independent measurements, the (least-squares) estimate $\widehat{\alpha}$ of the value of α is the sample mean of the data with associated variance

$$V_{\widehat{\alpha}} = \frac{1}{m} \sigma^2.$$

It follows that the estimate $\widehat{a}$ of the value of a, given by

$$\widehat{a} = \widehat{\alpha} - b,$$

is also the sample mean of the data (because the estimate of b is zero),

$$\widehat{a} = \frac{1}{m} \sum_{i=1}^{m} v_i,$$

with associated variance

$$V_{\widehat{a}} = \frac{1}{m} \sigma^2 + \sigma_b^2, \tag{13}$$

obtained by combining the variances associated with $\widehat{\alpha}$ and b in the usual manner. The variance associated with the estimate is the sum of two terms. The first corresponds to the variance associated with the sample mean and depends on the random effects and the number of measurements. The second is the variance associated with the systematic effect common to all measurements. It is clear that the contribution of the first term can be reduced by increasing m, whereas the second term makes a contribution that is independent of m.

An alternative derivation of the above results is as follows. Writing

$$\mathbf{E} = \left(\begin{array}{cc} I & \mathbf{1} \end{array} \right) \left(\begin{array}{c} \mathbf{E}_r \\ B \end{array} \right),$$

it follows that $\mathbf{e}$ is a sample of random variable $\mathbf{E}$ with zero expectation and uncertainty matrix

$$V_e = \sigma^2 I + \sigma_b^2 \mathbf{1}\mathbf{1}^{\mathrm{T}},$$

with, after some linear algebra, inverse

$$V_e^{-1} = \frac{1}{\sigma^2}I - \frac{\sigma_b^2}{\sigma^2(\sigma^2 + m\sigma_b^2)}\mathbf{1}\mathbf{1}^{\mathrm{T}}.$$

Now,

$$\mathbf{1}^{\mathrm{T}}\left(\frac{1}{\sigma^2}I - \frac{\sigma_b^2}{\sigma^2(\sigma^2 + m\sigma_b^2)}\mathbf{1}\mathbf{1}^{\mathrm{T}} \right)\mathbf{1} = \frac{m}{\sigma^2} - \frac{m^2\sigma_b^2}{\sigma^2(\sigma^2 + m\sigma_b^2)} = \frac{m}{\sigma^2 + m\sigma_b^2},$$

and

$$\begin{aligned} \mathbf{1}^{\mathrm{T}}\left(\frac{1}{\sigma^2}I - \frac{\sigma_b^2}{\sigma^2(\sigma^2 + m\sigma_b^2)}\mathbf{1}\mathbf{1}^{\mathrm{T}} \right)\mathbf{v} &= \left(\frac{1}{\sigma^2} - \frac{m\sigma_b^2}{\sigma^2(\sigma^2 + m\sigma_b^2)} \right)\mathbf{1}^{\mathrm{T}}\mathbf{v} \\ &= \frac{1}{\sigma^2 + m\sigma_b^2}\sum_{i=1}^{m} v_i. \end{aligned}$$

It follows, applying (5) and (6), that

$$\widehat{a} = \frac{\sigma^2 + m\sigma_b^2}{m}\frac{1}{\sigma^2 + m\sigma_b^2}\sum_{i=1}^{m} v_i = \frac{1}{m}\sum_{i=1}^{m} v_i,$$

the sample mean of the data, and

$$V_{\widehat{a}} = \frac{\sigma^2 + m\sigma_b^2}{m} = \frac{1}{m}\sigma^2 + \sigma_b^2,$$

as in expression (13).

4.2 Example: Development of RF coaxial frequency standards

In the context of the development of RF coaxial frequency standards, measurements Γ_i, $i = 1, \ldots, m$, are made at frequencies f_i, $i = 1, ..., m$, of the (complex-valued) voltage reflection coefficient Γ for an impedance standard, such as a short-circuit, open-circuit or matched load. Typically, measurements are made at frequencies spanning a range from 100 MHz to 1000 MHz (in steps of 50 MHz or 100 MHz). In addition, a measurement may be available at zero frequency (corresponding to DC).

The measurement Γ_i is subject to both random and systematic effects. The random effect is quantified by a (complex-valued) quantity u_i whose real and imaginary parts are the standard uncertainties associated with the real and imaginary parts of Γ_i obtained from a Type A evaluation of uncertainty.[24] The systematic effect is similarly quantified by a (complex-valued) quantity v_i obtained from a Type B evaluation of uncertainty.[25] The nature of the measurements is such that for each i, the real and imaginary parts of u_i are equal (to $\sigma_{r,i}$, say) and, similarly, those of v_i are equal (to $\sigma_{b,i}$, say). It is further assumed that the random effects associated with the real and imaginary parts of the measured voltage reflection coefficients are mutually independent.

Figures 6 and 7 illustrate an example of measurements made of an open-circuit impedance standard. Figure 6 shows the real part of the measured data Γ_i, $i = 1, \ldots, m$, together with the standard uncertainties $\sigma_{r,i}$, $i = 1, \ldots, m$, associated with the random effects. The information is presented as standard uncertainty bars, i.e., (vertical bars) centred symmetrically on the measured value of (total) length $2\sigma_{r,i}$. Figure 7 shows together the standard uncertainties $\sigma_{r,i}$ and $\sigma_{b,i}$, $i = 1, \ldots, m$, associated with the random and systematic effects. The former are presented as standard uncertainty bars centred on the zero line, and the latter by 'dots' at $\pm\sigma_{b,i}$. In this example, some information is available about the systematic effects at frequencies other than those at which measurements are made (in particular, at frequencies less than 100 MHz), and is shown.

The aim is to obtain estimates of the value of Γ, with the associated standard uncertainties, at frequencies where Γ cannot readily be measured (especially between DC and 100 MHz), and to reproduce values elsewhere that are consistent with measurement values in terms of their associated uncertainties. An approach is proposed [10] that involves

- modelling the data using a family of empirical functions incorporating physical information, where it is available, to help ensure the model is physically feasible[26]

- selecting a valid model from the family, i.e., one that is consistent with the available measurement data

- evaluating the selected model at any required frequency to obtain at estimate of the value of Γ at that frequency, together with the associated standard uncertainty.

---

[24] A Type A evaluation of uncertainty involves the statistical analysis of a series of (repeated) observations [2].

[25] A Type B evaluation of uncertainty means an evaluation *other* than statistical [2].

[26] Such physical information takes the form of prescribed values at DC for the model and a number of its derivatives with respect to frequency.
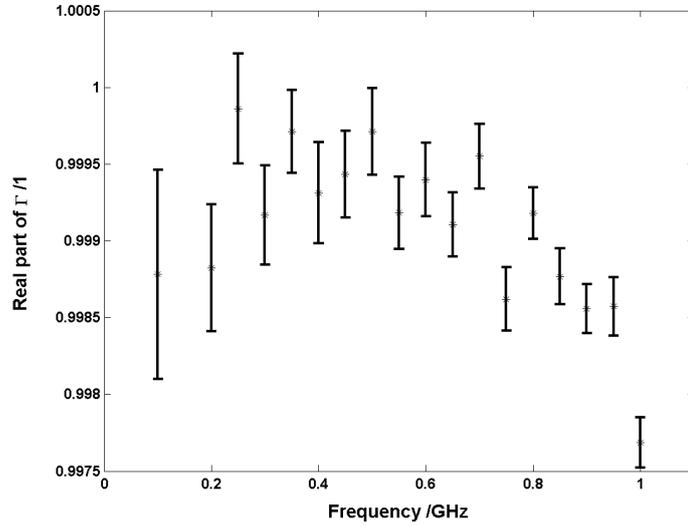
Figure 6: The real part of measured voltage reflection coefficient data for an open-circuit impedance standard, together with the standard uncertainties $\sigma_{r,i}$ associated with the random effects, the latter presented as standard uncertainty bars.
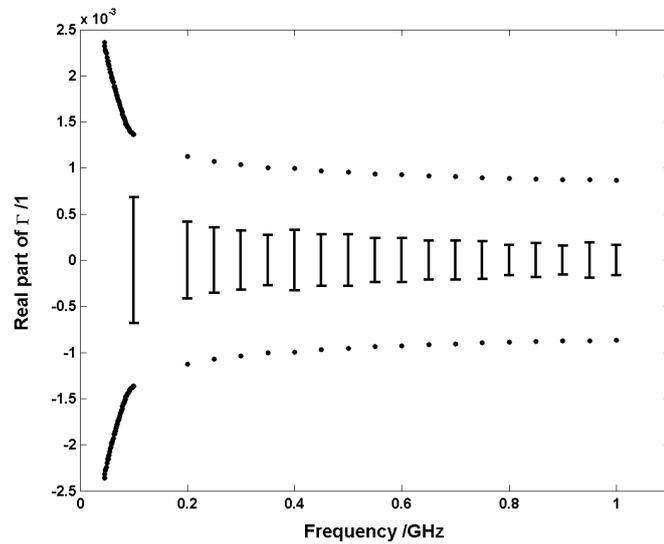

Figure 7: Standard uncertainties $\sigma_{r,i}$ and $\sigma_{b,i}$, $i = 1, \ldots, m$, associated with the random and systematic effects for the data shown in Figure 6. The former are presented as standard uncertainty bars centred on the zero line, and the latter by 'dots' at $\pm\sigma_{b,i}$.

If the systematic effects are genuinely repeatable, it is not possible, without further information, to model the required voltage reflection coefficient on its own. It is only possible to model the combination of the voltage reflection coefficient and the accompanying systematic measurement deviations. Consequently, the proposed approach [10] adopts a pragmatic procedure, motivated also by the alternative derivation presented in section 4.1, as follows.

A functional model is chosen for the data that takes the form

$$h(f, \mathbf{a}) = \mu(f) + z(f)p(f, \mathbf{a}),$$

where $\mu(f)$ is a (complex-valued) polynomial that satisfies the specified conditions at DC,[27] $z(f)$ is a (real-valued) polynomial that satisfies the same constraints but with value zero,[28] and $p(f, \mathbf{a})$ is a (complex-valued) polynomial with (complex-valued) adjustable parameters $\mathbf{a}$. Then,

$$\Gamma_i = h(f_i, \mathbf{a}) + e_{r,i}, \quad i = 1, \ldots, m,$$

where $\mathbf{e}_r = (e_{r,1}, e_{r,2}, \ldots, e_{r,m})^{\mathrm{T}}$ is a sample of random variable $\mathbf{E}_r$ with zero expectation and associated uncertainty matrix

$$V_r = \mathrm{diag}\{\sigma_{r,1}^2, \sigma_{r,2}^2, \ldots, \sigma_{r,m}^2\}.$$

Estimates $\widehat{\mathbf{a}}$ of the values of the parameters $\mathbf{a}$ are obtained by solving the least-squares problem (4), which takes the form

$$\min_{\mathbf{a}} \bar{\mathbf{r}} V_r^{-1} \mathbf{r},$$

where the ith elements of $\mathbf{r}$ and $\bar{\mathbf{r}}$ are

$$r_i = \Gamma_i - h(f_i, \mathbf{a})$$

and the complex conjugate of r_i, respectively. Formally, the solution $\widehat{\mathbf{a}}$ and its associated uncertainty matrix $V_{\widehat{a}}$ are given by (5) and (6), with the matrix H comprising the values at the frequencies f_i, $i = 1, \ldots, m$, of the polynomial basis functions used to represent $p(f, \mathbf{a})$.[29]

Corresponding to each point on the fitted model an associated standard uncertainty can be evaluated from the results of the fitting process in terms of the uncertainty matrix $V_{\widehat{a}}$. This standard uncertainty corresponds to a Type A evaluation of uncertainty because it derives from the propagation of the random effects associated with the measurement data. However, the data also contains systematic effects

---

[27]For example, for an open-circuit impedance standard, $\Gamma = 1$ at DC, so that $\mu(f)$ is chosen to satisfy the constraint $\mu(0) = 1$.

[28]For example, corresponding to the constraint $\mu(0) = 1$, $z(f)$ is chosen to satisfy $z(0) = 0$.

[29]It is recommended, for reasons of numerical stability, that a representation in terms of Chebyshev polynomials is chosen [1].

characterised by Type B evaluations of uncertainty. At each frequency corresponding to a specified measurement point the combination in quadrature of the *given* uncertainty associated with systematic effects and the *computed* uncertainty associated with random effects can be expected to yield an improved combined standard uncertainty at that point (*cf.* formula (13)).

The Type B evaluation of uncertainty associated with an interpolated (fitted) value for a frequency *different* from those for the measurement points necessarily involves making some assumptions about the nature of such uncertainties. As described above, a Type B evaluation of uncertainty is provided for each measurement point. In the absence of other information it seems reasonable to assume that the standard uncertainty from this source would vary smoothly with frequency. In that sense, the uncertainty associated with systematic effects at a measurement point could be predicted reasonably from the uncertainties associated with systematic effects at neighbouring measurement points. As a consequence, it is expected that the standard uncertainty from this source at a non-measurement point can also be predicted from the uncertainties associated with systematic effects at neighbouring measurement points. Various schemes could be used for the prediction, and indeed an empirical model could be built for the behaviour of the standard uncertainty from this source as a function of frequency, and used to provide intermediate values. Instead, arguably one of the simplest possible schemes is employed, viz., linear interpolation.[30] The standard uncertainty from a Type B evaluation of uncertainty at any frequency value is given by interpolating between the corresponding values at the measurement points that bracket this frequency. The standard uncertainty associated with *any* interpolated value is given by the combination in quadrature of the uncertainty associated with systematic effects computed by linear interpolation and the uncertainty associated with random effects computed from the fitting process.

Figures 8 and 9 illustrate the results of applying this approach to the measurement data shown in Figures 6 and 7. Figure 8 shows the real part of the least-squares best-fit model for the data shown in Figure 6 together with the standard uncertainty envelope corresponding to a Type A evaluation of uncertainty for the fitted values. The model takes the form of a cubic polynomial with a constraint on its value at DC. Figure 9 shows the real part of the least-squares best-fit model for frequencies in the range 0 – 200 MHz together with the standard uncertainty envelope corresponding to a *combined* (Type A and Type B) evaluation of uncertainty for the fitted values.

---

[30]A practical advantage of linear interpolation is that the interpolated values are convex combinations of the values for the measurement points. Curvilinear interpolation functions do not have this guarantee and might on occasions provide spurious results.
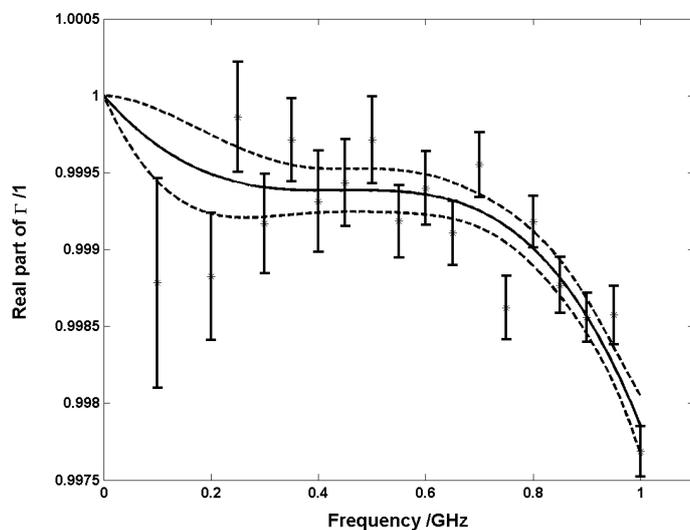
Figure 8: The real part of the least-squares best-fit model for the data shown in Figure 6 (solid line). The standard uncertainty envelope corresponding to a Type A evaluation of uncertainty for the fitted values is also shown (broken lines).
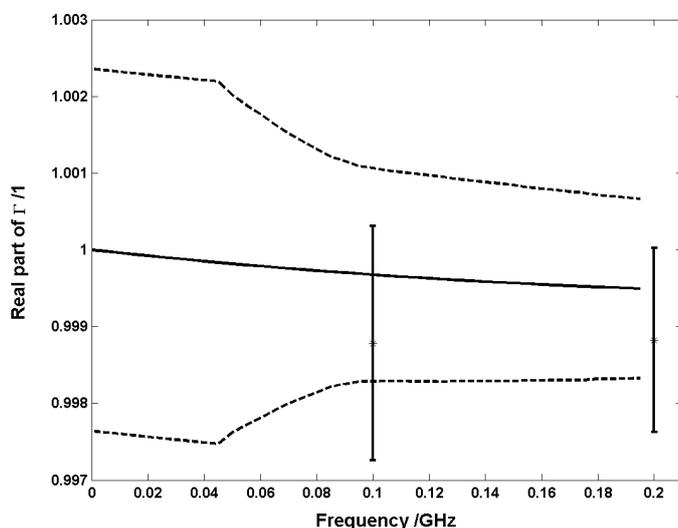

Figure 9: The real part of the least-squares best-fit model for the data shown in Figure 6 for the frequency range 0 – 200 MHz (solid line). The standard uncertainty envelope corresponding to a *combined* (Type A and B) evaluation of uncertainty for the fitted values is shown (broken lines). The measurement data within the frequency range, with the associated (combined) uncertainty bars are also indicated.

5 Unquantified random effects

This section is concerned with problems for which the measurement deviations $\mathbf{e}$ in (3) result from random effects *only* and the statistical model for the measurements is *incomplete*. Typically, information is known about the mutual dependencies between the measurements, so that the 'structure' of the uncertainty matrix V_r associated with the measurements is known, but V_r depends on parameters $\boldsymbol{\lambda}$ whose values are unknown. This is in contrast to the considerations of Sections 3 and 4 for which the uncertainty matrices associated with the random and systematic effects are (completely) quantified.

A number of 'uncertainty structures' and approaches are indicated. In some of the approaches (sections 5.1 and 5.2), estimates of the values of the parameters $\boldsymbol{\lambda}$ are obtained as part of the *model solving* process (section 2). In others (sections 5.3 and 5.4), some pre-processing of the measurements is undertaken in order to obtain estimates of $\boldsymbol{\lambda}$ prior to applying the methods of section 3.

5.1 Ensuring conformity

Suppose the measurement deviations $\mathbf{e} = \mathbf{e}_r$ in (3) are a sample of random variable $\mathbf{E}_r$ with zero expectation and uncertainty matrix $V_r = \sigma^2 I$, i.e., the deviations are independent and identically distributed. Define the random variable S^2 by

$$S^2 = \frac{1}{\sigma^2}\widehat{\mathbf{E}}^{\mathrm{T}}\widehat{\mathbf{E}},$$

in terms of the random vector of residual deviations $\widehat{\mathbf{E}}$ given by (10). Then, provided the functional model and the measurement data *conform* and the measurement deviations are a sample from a multivariate normal distribution with the above parameters, it follows that S^2 has a χ^2–distribution with $\nu = m - n$ degrees of freedom, expectation ν and variance 2ν [21]. Consequently, for a functional model and measurement data that conform, we expect

$$s^2 = \frac{1}{\sigma^2}\widehat{\mathbf{e}}^{\mathrm{T}}\widehat{\mathbf{e}} \approx \nu.$$

The result may be used to (a) test the conformity of the functional model with the measurement data,[31] or (b) obtain an estimate of the value of σ in the case that the functional model and the measurement data are assumed to conform and the above statistical model applies.

---

[31]An indication that the functional model and measurement data do *not* conform is that the value of s^2 is far from ν relative to $\sqrt{2\nu}$.

To undertake (b), note that in the case $V_r = \sigma^2 I$, the least-squares solution given by (5) is

$$\widehat{\mathbf{a}} = \left(H^{\mathrm{T}}H\right)^{-1} H^{\mathrm{T}}\mathbf{v},$$

and is independent of σ. It follows that an estimate $\widehat{\sigma}$ of the value of σ is obtained by requiring that

$$s^2 = \frac{1}{\widehat{\sigma}^2}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right)^{\mathrm{T}}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right) = m - n,$$

i.e.,

$$\widehat{\sigma}^2 = \frac{1}{m - n}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right)^{\mathrm{T}}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right). \tag{14}$$

5.2 Maximum likelihood estimation

The use of least-squares analysis is supported by the Gauss-Markov theorem (section 3). In addition, in the case that a multivariate normal probability density function is assigned to the value of $\mathbf{E}_r$, the least-squares problem is equivalent to maximising with respect to the parameters $\mathbf{a}$ the *likelihood function*[32]

$$\ell(\mathbf{v}|\mathbf{a}) = \frac{1}{\sqrt{|2\pi V_r|}}\exp\left\{-\frac{1}{2}(\mathbf{v} - H\mathbf{a})^{\mathrm{T}}V_r^{-1}(\mathbf{v} - H\mathbf{a})\right\},$$

or its logarithm, the *log-likelihood function*

$$L(\mathbf{v}|\mathbf{a}) = -\frac{1}{2}\log\left(|2\pi V_r|\right) - \frac{1}{2}\left(\mathbf{v} - H\mathbf{a}\right)^{\mathrm{T}}V_r^{-1}\left(\mathbf{v} - H\mathbf{a}\right).$$

Consequently, in the case that V_r is fully specified, the least-squares estimate corresponds to a maximum-likelihood estimate.

Now suppose V_r is not fully quantified, but is described in terms of parameters $\boldsymbol{\lambda}$, i.e., $V_r = V_r(\boldsymbol{\lambda})$. Then, maximum-likelihood estimates of the values of the parameters $\mathbf{a}$ and $\boldsymbol{\lambda}$ are obtained by maximising with respect to these parameters the log-likelihood function given above. The use of maximum-likelihood estimation has been proposed, for example, in the contexts of undertaking least-squares adjustment in the presence of discrepant data [11] and data fusion [3].

The following is a simple illustration. Suppose $V_r = \sigma^2 I$ (as in section 5.1), so that $\lambda = \sigma$. Then, the log-likelihood function is given by

$$L(\mathbf{v}|\mathbf{a}) = -\frac{m}{2}\log\left(2\pi\right) - m\log\sigma - \frac{1}{2\sigma^2}\left(\mathbf{v} - H\mathbf{a}\right)^{\mathrm{T}}\left(\mathbf{v} - H\mathbf{a}\right).$$

It follows that $\widehat{\mathbf{a}}$ and $\widehat{\sigma}$ maximise the log-likelihood function if $\widehat{\mathbf{a}}$ minimises

$$\left(\mathbf{v} - H\mathbf{a}\right)^{\mathrm{T}}\left(\mathbf{v} - H\mathbf{a}\right)$$

---

[32]The notation $|V|$ is used to denote the determinant of the (square) matrix V.

with respect to $\mathbf{a}$, and

$$\widehat{\sigma}^2 = \frac{1}{m}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right)^{\mathrm{T}}\left(\mathbf{v} - H\widehat{\mathbf{a}}\right). \tag{15}$$

The maximum-likelihood estimate of $\mathbf{a}$ corresponds to the least-squares estimate (as before). However, the maximum-likelihood estimate (15) of σ^2 differs (in the denominator) from that (14) given by considerations of conformity. For $m \gg n$ the difference is not likely to be significant.

5.3 Local (empirical) analysis

Suppose the measurement deviations $\mathbf{e} = \mathbf{e}_r$ in (3) are a sample of random variable $\mathbf{E}_r$ with zero expectation and uncertainty matrix $V_r = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2\}$, i.e., the deviations are independent but have (generally) different variances. This model is (possibly) the next in complexity following that considered in section 5.1.

A procedure is described [15] for obtaining estimates $\widehat{\sigma}_i$ of the values of σ_i, $i = 1, \ldots, m$, in cases where such information is not available. The motivation for the approach is that, whereas assigning an uncertainty matrix of the form $V_r = \sigma^2 I$ to the *complete* set of measurements may not be valid, assigning an uncertainty matrix of this form to (local) subsets of the data may be acceptable, particularly if the values of σ_i change smoothly (and slowly) across the measurements. The approach also relies on the observation that over a small interval any sufficiently smooth function can be modelled by a polynomial of low order.

The approach is summarised as follows. Consider the measurement data arranged so that the corresponding values t_i of the stimulus are in increasing order. For each data point (excepting the first few and last few), consider the subset of the data given by the data point itself and a small number (for example, two) to the left and the same number to the right. Model the subset by a low order polynomial (for example, a quadratic polynomial) under the assumption that the measurement deviations associated with the subset are independent and identically distributed. Finally, apply the considerations of section 5.1 regarding conformity to obtain an estimate of the standard uncertainty associated with the measured data at and in the neighbourhood of the central point.

5.3.1 Example: Spectral characteristic modelling

An illustration is given [16] of the procedure applied to data arising in the modelling of spectral irradiance data. The data consists of measurements E_i of spectral irradiance E (in $\mathrm{Wm}^{-2}\mathrm{nm}^{-1}$) for a lamp corresponding to values λ_i of wavelength (in nm). A model for the measurements takes the form of the product of the Planck

function $L(\lambda)$ and a polynomial $p_n(\lambda, \mathbf{a})$ of order n (degree $n - 1$) with adjustable parameters $\mathbf{a}$, viz.,

$$h(\lambda, \mathbf{a}) = L(\lambda) p_n(\lambda, \mathbf{a}).$$

Figure 10 shows the results for a particular lamp. The crosses denote the estimates $\widehat{\sigma}_i$ obtained using the local analysis described above. The broken lines represent a segmented standard uncertainty profile based on these estimates, and the solid lines show for comparison the specified profile for the measurements.

The reason for replacing the individual estimates $\widehat{\sigma}_i$ by 'smoothed' values is that each estimate, being derived from a small number of measurements, might not be expected to be very reliable. In this example, the known piecewise nature of the specified standard uncertainty profile has been used to inform the 'smoothing', i.e., the same breakpoints between straight-line segments have been used. It is evident that the specified uncertainties at the lower end of the spectral region would appear to be too low, whereas those for the bulk of the spectral region are too large. A model based on the estimated data uncertainties differs little functionally in its behaviour from the model obtained using the specified data uncertainties, although the uncertainties associated with the model values are different [16].
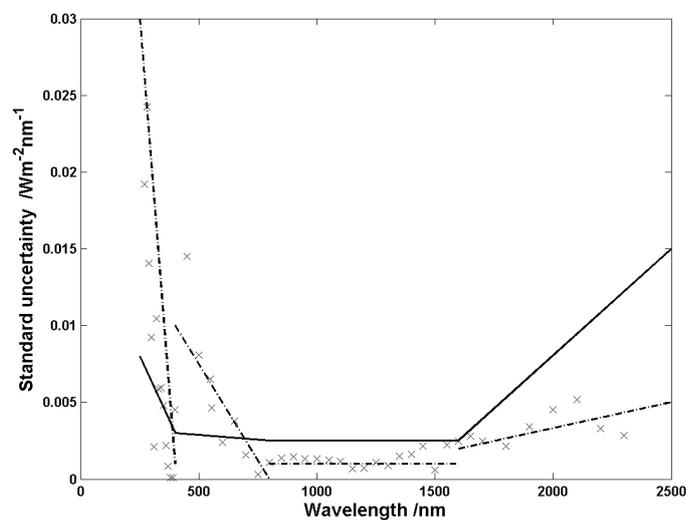

Figure 10: The crosses denote estimates of the standard uncertainties associated with spectral irradiance data obtained using local polynomial models. The broken lines represent a segmented standard uncertainty profile based on these estimates, and the solid lines the specified standard uncertainty profile for the measurements.

5.4 Autoregressive modelling

This section is concerned with the use of autoregressive (AR) modelling to treat measurement data sets containing *serial correlation*.

The *ordinary* least-squares problem corresponding to the model (2) or (3) is

$$\min_{\mathbf{a}}(\mathbf{v} - H\mathbf{a})^{\mathrm{T}}(\mathbf{v} - H\mathbf{a}), \tag{16}$$

which is the least-squares problem (4) with $V_r = I$, the identity matrix of order m. Estimates of the model parameters $\mathbf{a}$ obtained as the solution to (16) (and predictions based on them) are unbiased and consistent [21] *even* when the above uncertainty matrix V_r does not describe the statistical model for the measurement deviations as is the case, e.g., when those deviations are serially correlated. However, such estimates are inefficient, i.e., they do not correspond to the linear unbiased estimates of minimum variance *unless* the statistical model $V_r = I$ applies. There are important classes of problems in metrology for which the measurement data is serially correlated. In cases where the serial correlation is positive the use of ordinary least-squares will tend to deliver uncertainties associated with estimates of the model parameter values that are underestimates [17].[33] Furthermore, the 'stronger' is the correlation, the greater the underestimate will be. AR modelling gives the possibility to provide realistic uncertainty evaluations.

Consider the model (2) in the case where the t_i are uniformly spaced, with $\Delta t = t_{i+1} - t_i$, and the e_i satisfy the relationship

$$e_i = \sum_{\ell=1}^{q} c_\ell e_{i-\ell} + z_i, \quad i = q+1, \ldots, m, \tag{17}$$

where z_i, $i = q+1, \ldots, m$, are independent samples from a (normal) distribution with zero expectation and variance σ^2.

5.4.1 Parameter estimation

The values of the parameters $\mathbf{a}$ in model (2) and the AR model parameters $\mathbf{c} = (c_1, \ldots, c_q)^{\mathrm{T}}$ in (17) can be estimated as follows:

1. Form estimates $\hat{\mathbf{a}}$ of the parameters $\mathbf{a}$ in model (2) by solving the ordinary least-squares problem (16).

---

[33]A study of parameter estimation, and the associated uncertainties, when measurements are serially correlated and the measurement deviations are modelled as a second order autoregressive process is available [17]. An approach to accounting for the effect on the uncertainty due to the correlation is presented, whereby a compensating 'scale factor' is applied.

2. Form estimates $\widehat{e}_i$ of the measurement deviations e_i:

$$\widehat{e}_i = v_i - \sum_{j=1}^{n} \widehat{a}_j h_j(t_i), \quad i = 1, \ldots, m.$$

3. Provide estimates $\widehat{c}_\ell$ of the parameters c_ℓ in the model

$$\widehat{e}_i = \sum_{\ell=1}^{q} c_\ell \widehat{e}_{i-\ell} + z_i, \quad i = q+1, \ldots, m,$$

the counterpart of expression (17) with $\widehat{e}_i$ in place of e_i.

In step 3, the $\widehat{c}_\ell$ can be determined by regressing the vector $(\widehat{e}_{q+1}, \ldots, \widehat{e}_m)^{\mathrm{T}}$ on the q lagged vectors $(\widehat{e}_q, \ldots, \widehat{e}_{m-1})^{\mathrm{T}}, \ldots, (\widehat{e}_1, \ldots, \widehat{e}_{m-q})^{\mathrm{T}}$. A conventional least-squares algorithm can be used for this purpose. The provision of a suitable value for q is considered below.

5.4.2 Order of process

The order q of the AR process can be determined in the following way. For each candidate order $q = 1, 2, \ldots, q_{\max}$, where $q_{\max}$ is a prescribed 'maximum' order that is unlikely to be exceeded (e.g., $q_{\max} = 50$ if there is adequate data), fit the AR model of order q to the estimated measurement deviations $\widehat{e}_i$.[34] Form the root-mean-square (RMS) residual for each order (for example, as in section 5.1) and examine its behaviour as a function of order. Like their counterparts in data approximation [8], the RMS residuals have a tendency to decrease, often rapidly, through the low orders, then eventually saturating to an essentially constant level. Figure 13 in section 5.4.3 shows a graph of such behaviour for the roundness measurement application subsequently discussed. The lowest order in the 'saturation zone' is usually selected.

This procedure is different from that proposed in many texts on AR. Typically, a *statistical* test is used to decide whether the (estimated) measurement deviations can adequately be explained by an AR process of some order. Originally the Durbin-Watson test [18] was used, but more recently the Lagrange multiplier test [6] and especially the Breusch-Godfrey [4] and Breusch-Pagan tests [5] have found favour because of the inconclusiveness of Durbin-Watson in some situations. A suggested approach for metrological data, which typically has very different characteristics from economic and other data to which such tests are usually applied, is as follows. Apply the indicated procedure, but support it with tests of the above nature in cases where the choice of order of the AR model is problematic.

---

[34]The complete set of models from order one to order 50 may be fitted 'simultaneously', using a procedure successfully applied in data approximation [8].

5.4.3 Example: Roundness assessment

A roundness-measuring instrument with a precision rotary stage [20, 23] is used to establish the profile of a nominally circular section of a component and then to quantify the departure of that profile (the component 'out-of-roundness') from a perfect circle. The instrument enables the component to be rotated through any specified angle before measurement. Measurement traces, approximately 20 in number, each consisting of, say, 2 000 suppressed-radius values for a particular position of the stage, are taken. These traces are combined to determine the component form deviation (CFD), the angle-dependent departure from circularity, and the instrument spindle deviation (ISD) in terms of their Fourier harmonics. The measurement uncertainties associated with these deviations are to be evaluated.

The ISD has a systematic angle-dependence, because the instrument does not rotate perfectly, although highly repeatably. The measurement traces are hence not those of the component alone, but a superimposition of CFD and ISD. The trace $Y(\theta)$, say, is thus of $C(\theta) + S(\theta)$, $C(\theta)$ denoting the CFD, $S(\theta)$ the ISD and θ the angular position of the gauge head.

By rotating the component through angles $\phi_1, \ldots, \phi_M$, traces can be obtained having the form

$$
\begin{aligned}
Y_1(\theta) &= C(\theta - \phi_1) + S(\theta), \\
&\vdots \\
Y_M(\theta) &= C(\theta - \phi_M) + S(\theta).
\end{aligned}
$$

These expressions constitute an overdetermined set of defining equations for $C(\theta)$ and $S(\theta)$.

An approach for determining $C(\theta)$ and $S(\theta)$ can be summarized as follows:

1. Form the Fourier harmonics for each trace (using the FFT).

2. Express $C(\theta)$ and $S(\theta)$ algebraically as Fourier series with adjustable coefficients (the model parameters in the context of this work).

3. Construct a model relating these Fourier representations and solve it for the Fourier harmonics (model parameters) of $C(\theta)$ and $S(\theta)$.

4. Propagate the measurement uncertainties associated with the traces through the model to evaluate the uncertainties associated with the harmonics and thence the CFD and ISD 'out-of-roundness'.

The appreciable positive correlations present in the measurements, if ignored, would lead to a gross underestimate of the uncertainties in step 4.

That the measurement deviations are uncorrelated is an extremely poor assumption for the data of concern here, taken from the ultra-high precision Talyrond 73HPR roundness-measuring instrument. Figure 11 shows a typical trace and the corresponding 'model trace', i.e., the trace reconstructed from the Fourier coefficients determined by the above solution process. Although in general terms the reconstructed trace is close to the measured trace in a practical sense, the measurement deviations, viz., the radial departures of the model trace from the measured trace, exhibit systematic behaviour. This behaviour is seen more clearly in Figure 12, which shows the measurement deviations as a function of angular position. These deviations have a behaviour very different from that of Gaussian noise, and can clearly be seen to be strongly serially correlated. Behaviour that is similar in character is observed for the other traces.
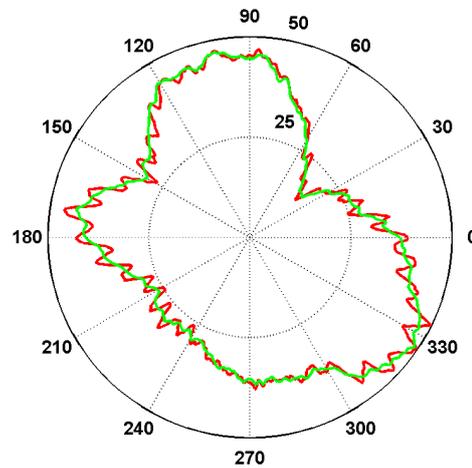

Figure 11: A typical measurement trace (red), Trace 15 of the roundness data, and corresponding reconstructed 'model' trace (green).

Although the measured traces can each *individually* be closely represented by a Fourier series, the same is not true for the traces reconstructed from the Fourier series determined for $C(\theta)$ and $S(\theta)$. The reason is two-fold. First, these two Fourier series have a number of Fourier coefficients that is much smaller than the total number of measurements in the original traces. Second, and more importantly, the nature of the measurement deviations in the traces is such that they can be modelled individually but not collectively by Fourier series.

Figure 13 depicts the root-mean-square (RMS) residuals for AR processes of all orders up to 50, when applied to one of the traces, Trace 15 (shown in Figures 11 and 12), of the 21 traces in a particular roundness measurement. Order '0' corresponds to there being no modelling: the RMS residual denotes the RMS value
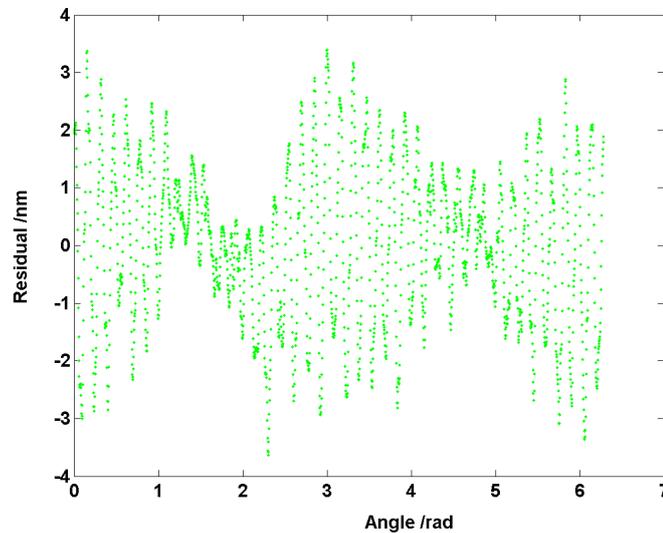
Figure 12: Measurement deviations derived from the measurement trace and reconstructed trace shown in Figure 11.

for the original deviations. The RMS residual reaches a saturation level at order (approximately) eight. The residuals of the AR model of order eight for Trace 15 are shown in figure 14. They would appear to have 'white noise' behaviour.

A histogram of the measurement deviations for Trace 15 and of the residuals for that trace following the use of the AR model of order eight are shown in figures 15 and 16. Note the two scales for the measurement deviations, differing by two orders of magnitude. The values of the modelled deviations for the AR model of order eight (red) for roundness trace 15, together with the measurement deviations (green) of figure 12, are shown superimposed in figure 17. It is difficult visually to distinguish between them.

Figure 18 depicts the RMS residuals for AR processes of all orders up to 15 for all 21 traces. A similar pattern is observed across the traces. Saturation occurs at essentially the same level (between 0.007 3 nm and 0.007 8 nm) for all traces, although the initial values of the RMS deviations ranged from 0.6 nm to 1.5 nm. An order eight AR model appears to be suitable for all 21 traces.

Table 1 lists the first eight AR coefficients for each of the 21 traces, with the respective mean values, and Figure 19 depicts the deviations of the values of the coefficients from the mean values. There is no apparent dependence of these deviations on the trace number. The dispersions of the individual values about the means are such that it is reasonable to model all traces by the same AR process (with the same numerical values for the parameters).
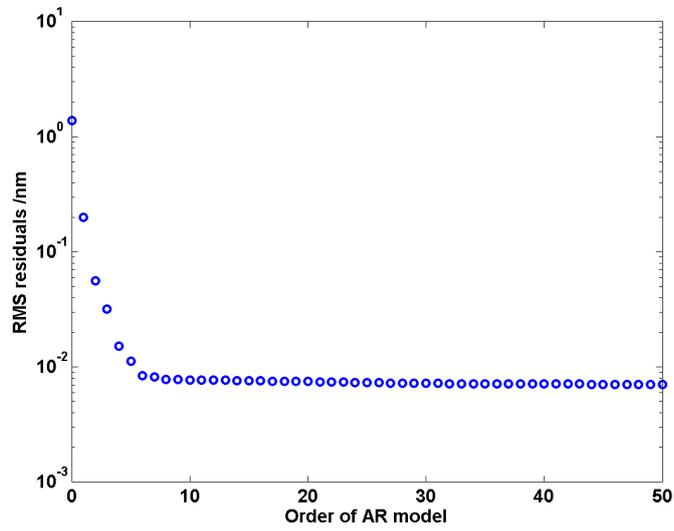
Figure 13: The root-mean-square (RMS) residuals for AR processes of all orders up to 50, when applied to Trace 15 of the roundness data.
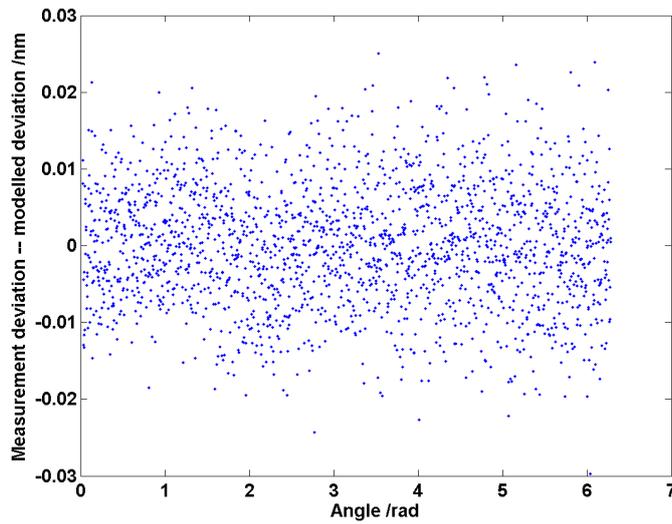

Figure 14: The residuals of the AR model of order eight for Trace 15 of the roundness data.
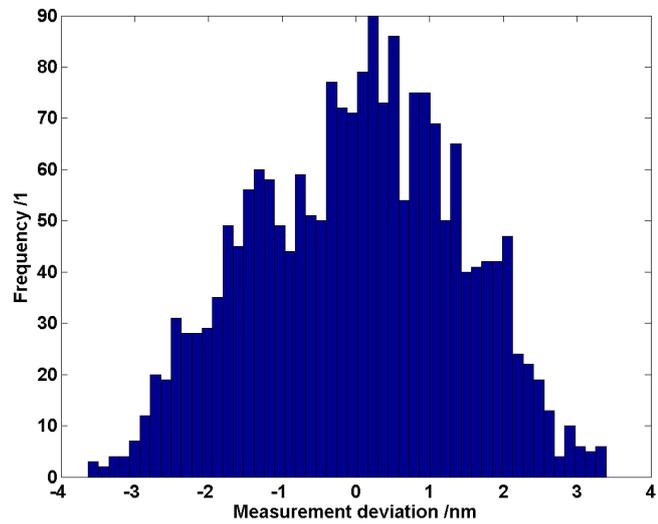
Figure 15: A histogram of the measurement deviations for roundness trace 15.
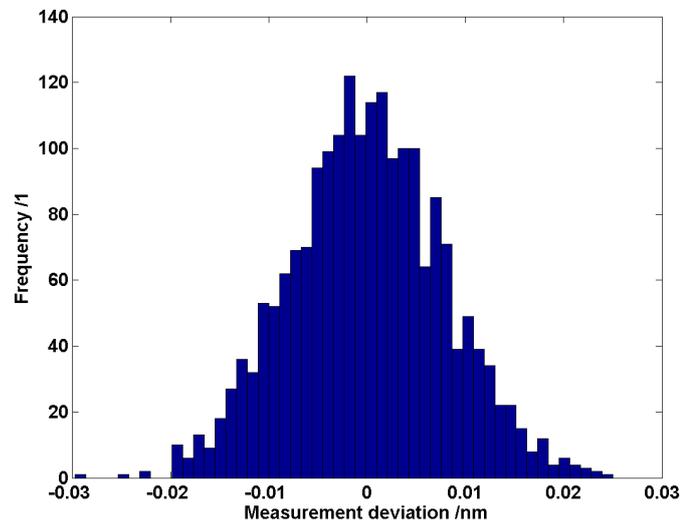

Figure 16: A histogram of the residuals for roundness trace 15 following the use of an AR model of order eight.
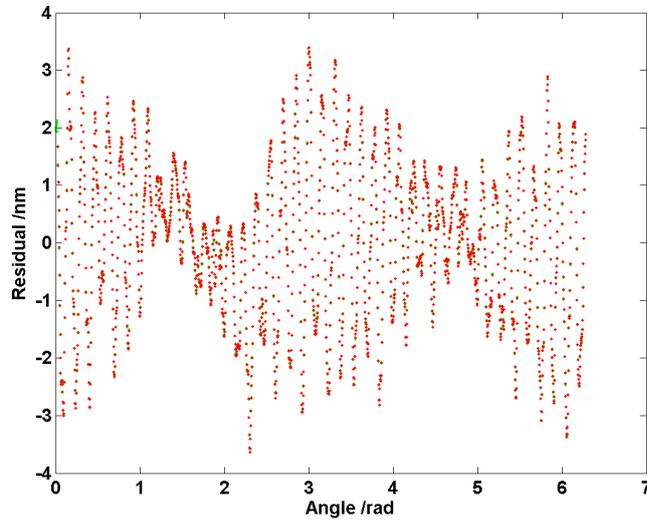
Figure 17: The values of the model deviations for the AR model of order eight (red) for roundness trace 15, together with the measurement deviations (green). To graphical accuracy the values of the model deviations and those for the measurement deviations coincide and, consequently, the latter are hidden by the former.
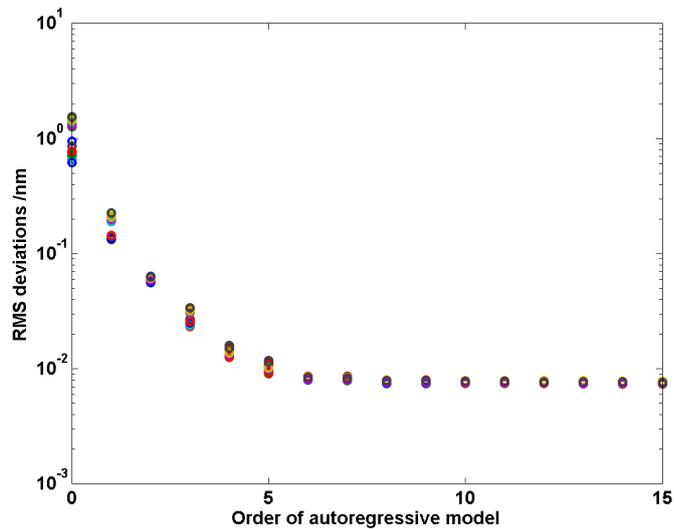

Figure 18: The root-mean-square residuals for AR processes of all orders up to 15 for all the 21 traces.

Trace	Model coefficient order							
No.	1	2	3	4	5	6	7	8
1	4.4	-8.3	8.4	-3.5	-1.7	3.2	-1.7	0.4
2	4.4	-8.3	8.4	-3.7	-1.4	2.8	-1.5	0.3
3	4.4	-8.3	8.5	-3.9	-1.1	2.7	-1.5	0.3
4	4.4	-8.3	8.3	-3.4	-1.8	3.2	-1.7	0.4
5	4.4	-8.6	9.0	-4.5	-0.9	2.7	-1.6	0.3
6	4.4	-8.5	8.6	-3.9	-1.4	3.0	-1.7	0.4
7	4.4	-8.4	8.6	-3.7	-1.7	3.2	-1.8	0.4
8	4.4	-8.4	8.7	-4.0	-1.1	2.7	-1.5	0.3
9	4.4	-8.4	8.6	-3.8	-1.4	2.9	-1.6	0.3
10	4.4	-8.2	8.2	-3.4	-1.4	2.7	-1.4	0.3
11	4.4	-8.5	8.5	-3.6	-1.5	2.9	-1.5	0.3
12	4.4	-8.5	8.7	-4.2	-0.7	2.2	-1.2	0.2
13	4.5	-8.7	9.0	-4.3	-0.9	2.6	-1.5	0.3
14	4.5	-8.7	9.2	-4.7	-0.6	2.4	-1.4	0.3
15	4.4	-8.5	8.8	-4.0	-1.3	2.9	-1.6	0.3
16	4.5	-8.7	9.1	-4.5	-0.9	2.6	-1.5	0.3
17	4.5	-8.6	8.9	-4.2	-1.1	2.7	-1.5	0.3
18	4.5	-8.6	8.9	-4.3	-0.7	2.3	-1.3	0.3
19	4.4	-8.5	8.7	-4.0	-1.1	2.6	-1.4	0.3
20	4.5	-8.6	8.7	-3.8	-1.4	2.9	-1.5	0.3
21	4.4	-8.6	8.8	-4.1	-1.2	2.7	-1.5	0.3
Mean	4.4	-8.5	8.7	-4.0	-1.2	2.8	-1.5	0.3

Table 1: Coefficients of AR models of order eight for the 21 roundness traces, and the mean coefficient values over the traces.
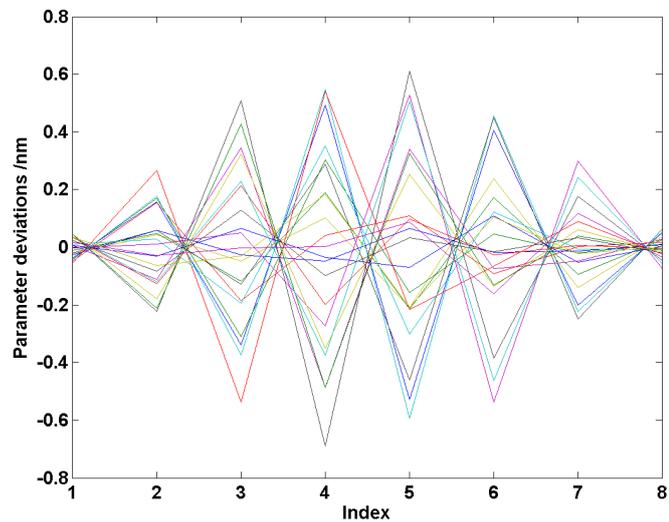

Figure 19: Deviations from their mean values of the coefficients of AR models of order eight for the 21 roundness traces (Table 1). The points indicating the deviations for each trace are joined by straight lines, and show there is no apparent dependence of the deviations on trace number.

It can be concluded that for this roundness application the measurement deviations for all traces have been satisfactorily modelled by the same AR process (of order eight). For this and other applications that can be so modelled, the problem can then be transformed into a replacement problem to which ordinary least-squares can be applied (section 5.4.4). Moreover, for certain classes of problems of this type, which are likely to arise in metrology applications, the transformed problem can be solved by the same ordinary least-squares algorithm that was designed to solve the original problem, but that did not pay regard to the uncertainty structure. A different parametrization of the solution is obtained, however. If a solution in terms of the original parametrization is required, it can be transformed accordingly, using a linear transformation for the models considered here.

For the roundness application, the original model involved a Fourier series, i.e., a linear combination of basis functions consisting of cosine and sine terms in the fundamental, second, third, etc. harmonics, whose coefficients were required. The transformed problem involves basis functions that are linear combinations of these terms. So the problem can be solved as Fourier series in terms of the original basis functions, and then converted to the required form.

5.4.4 Transforming the problem

The three steps in section 5.4.1 in fact constitute the first stages in a well-established procedure [7] for treating serial correlation that has been available for over half a century. It has been applied extensively in areas such as financial data modelling, econometrics and geostatistics, but it would appear that is has been used relatively little in metrology.

In outline, the procedure, including the above steps, is as follows:

1. Form estimates $\widehat{\mathbf{a}}$ of the parameters $\mathbf{a}$ in model (2) by solving the ordinary least-squares problem (16).

2. Form estimates $\widehat{e}_i$ of the measurement deviations e_i:

$$\widehat{e}_i = v_i - \sum_{j=1}^{n} \widehat{a}_j h_j(t_i), \quad i = 1, \ldots, m.$$

3. Provide estimates $\widehat{c}_\ell$ of the parameters c_ℓ in the model

$$\widehat{e}_i = \sum_{\ell=1}^{q} c_\ell \widehat{e}_{i-\ell} + z_i, \quad i = q+1, \ldots, m,$$

the counterpart of expression (17) with $\widehat{e}_i$ in place of e_i.

4. Use these AR parameter estimates to transform the original measurement deviations into a form suitable for ordinary least-squares.

5. Use ordinary least-squares to estimate (modified) model parameters for the transformed data.

6. Transform these modified parameters into the required form.

The transformation required in Step 4 of the procedure above can be carried out as follows. Form transformed measurements

$$\widetilde{v}_i = v_i - \sum_{\ell=1}^{q} c_\ell v_{i-\ell}.$$

The use of expression (2) then gives

$$\widetilde{v}_i = \sum_{j=1}^{n} a_j h_j(t_i) - \sum_{\ell=1}^{q} c_\ell \sum_{j=1}^{n} a_j h_j(t_{i-\ell}) + e_i - \sum_{\ell=1}^{q} c_\ell e_{i-\ell}.$$

Let

$$\widetilde{h}_j(t) = h_j(t) - \sum_{\ell=1}^{q} c_\ell h_j(t - \ell \Delta t).$$

Then, using expression (17),

$$\widetilde{v}_i = \sum_{j=1}^{n} a_j \widetilde{h}_j(t_i) + z_i.$$

Since the z_i are regarded as values drawn from a normal distribution, the transformed measurement deviations can be processed using ordinary least-squares. Once new estimates of the values of the model parameters have been obtained, improved estimates $\widehat{e}_i$ of the measurement deviations are obtained and the process iterated.

A further issue relates to the loss of information associated with the fact that there is less transformed than original data. There are methods for estimating these missing observations, e.g., the Prais-Winsten transformation [22] can be used for AR models of order one. The loss of data is of little consequence in this roundness application, since there is adequate data remaining, with 2 000 points per trace and an AR model order of eight, leaving 1 992 points remaining. In some other areas, with smaller numbers of measurements, the modification might be important.

6   Summary

This report has been about statistical error modelling, here regarded as a process that involves developing relationships between measurement data, the required

measurement result(s), and the measurement deviations (or 'errors') associated with the measurement data. The relationships are then used as the basis of evaluating the measurement result(s) and the associated uncertainty, in a manner that is consistent with the framework for uncertainty evaluation given in the Guide to the Expression of Uncertainty in Measurement (GUM).

The concern has been with problems for which a statistical model, specified completely or perhaps only partially, is assigned to the values of the measured quantities. The focus has been on (least-squares) regression and its application to the calibration of an instrument or measurement system. In particular, methods of analysis have been proposed for problems for which the measurement deviations arise from

- random effects for which the statistical model is completely specified

- random and systematic effects for which the respective statistical models are completely specified

- random effects for which the statistical model is only partially specified.

For the last-mentioned category of problems, methods based on ensuring conformity, maximum likelihood estimation, empirical analysis, and autoregressive modelling have been considered. Although the methods have not been used extensively in metrology (but widely applied in some other disciplines) they have been shown to provide useful approaches to solving metrology problems.

A number of examples have been discussed, concerned with characterising the performance of a clock (in time and frequency metrology), developing RF coaxial frequency standards (in electromagnetic metrology), modelling spectral characteristic data (in optical radiation metrology), and assessing the roundness of a nominally circular artefact (in dimensional metrology).

Acknowledgements

ample. Martyn Byng (Numerical Algorithms Group Ltd.) made useful comments on a draft of this report.

References

[1] R. M. Barker, M. G. Cox, A. B. Forbes, and P. M. Harris. Best Practice Guide No. 4. Discrete modelling and experimental data analysis. Technical report, National Physical Laboratory, Teddington, UK, 2004. Available for download from the SS*f*M website at www.npl.co.uk/ssfm/download/bpg.html.

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

[3] R. Boudjemaa and A. B. Forbes. Parameter estimation methods for data fusion. Technical Report CMSC 38/04, National Physical Laboratory, Teddington, UK, 2004.

[4] T. S. Bresch and L. Godfrey. A review of recent work on testing for autocorrelation in dynamic simultaneous models. In D. Currie, R. Nobay, and D. Peel, editors, *Macroeconomic Analysis: Essays in Macroeconomics and Econometrics*. Croom Helm (London), 1981.

[5] T. S. Bresch and A. R. Pagan. A simple test of heteroskedasticity and random coefficient variation. *Econometrica*, 47:1287–1294, 1979.

[6] T. S. Bresch and A. R. Pagan. The lagrange multiplier test and its application to model specification in econometrics. *Review of Economic Studies*, 47:239–254, 1980.

[7] D. Cochrane and G. H. Orcutt. Application of least squares regressions to relationships containing autocorrelated error terms. *J. Amer. Statist. Assoc.*, 44:32–61, 1949.

[8] M. G. Cox. Constructing and solving mathematical models of measurement. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology II*, pages 7–21, Singapore, 1996. World Scientific.

[9] M. G. Cox. The evaluation of key comparison data. *Metrologia*, 39:589–595, 2002.

[10] M. G. Cox, M. P. Dainton, N. M. Ridler, M. J. Salter, and P. R. Young. An interpolation scheme for precision intermediate frequency reflection coefficient measurement. *IEEE Trans. IM*, 52:27–37, 2003.

[11] M. G. Cox, A. B. Forbes, J. Flowers, and P. M. Harris. Least squares adjustment in the presence of discrepant data. In Advanced Mathematical and Computational Tools in Metrology VI, 2004. To appear.

[12] M. G. Cox, A. B. Forbes, P. M. Harris, and I. M. Smith. The classification and solution of regression problems for calibration. Technical Report CMSC 24/03, National Physical Laboratory, Teddington, UK, 2003.

[13] M. G. Cox and P. M. Harris. The applicability of non-parametric methods of statistical analysis to metrology. Technical Report CMSC 46/04, National Physical Laboratory, Teddington, UK, 2004.

[14] M. G. Cox and P. M. Harris. Best Practice Guide No. 6. Uncertainty evaluation. Technical report, National Physical Laboratory, Teddington, UK, 2004. Available for download from the SS*f*M website at www.npl.co.uk/ssfm/download/bpg.html.

[15] M. G. Cox, P. M. Harris, P. D. Kenward, and I. M. Smith. Exploratory data analysis. Technical Report CMSC 47/04, National Physical Laboratory, Teddington, UK, 2004.

[16] M. G. Cox, P. M. Harris, P. D. Kenward, and E. Woolliams. Spectral characteristic modelling. Technical Report CMSC 27/03, National Physical Laboratory, Teddington, UK, 2003.

[17] G. Durando and G. Mana. Propagation of error analysis in least-squares procedures with second-order autoregressive measurement errors. *Meas. Sci. Technol.*, 13:1505–1511, 2002.

[18] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. *Biometrika*, 37:409–428, 1950.

[19] P. M. Harris, J. A. Davis, M. G. Cox, and S. L. Shemar. Least-squares analysis of time series data and its application to two-way satellite time and frequency transfer measurements. *Metrologia*, 40:S342–S347, 2003.

[20] Cao Linxiang, Wang Hong, Li Xiongua, and Shen Qinghing. Full-harmonic error separation technique. *Meas. Sci. Technol.*, 3:1129–1132, 1995.

[21] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[22] S. J. Prais and S. J. Winsten. Trend estimators and serial correlation. Technical Report Discussion Paper No 383, Cowles Commission, Chicago, 1954.

[23] C. P. Reeve. The calibration of a roundness standard. Technical Report NBSIR 79, National Bureau of Standards, 1979.

[24] A. M. H. van der Veen and M. G. Cox. Error analysis in the evaluation of measurement uncertainty. *Metrologia*, 40:42–50, 2003.

[25] K. Weise and W. Wöger. A Bayesian theory of measurement uncertainty. *Measurement Sci. Technol.*, 3:1–11, 1992.