NPL Report CMSC 46/04

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

From the Software Support for Metrology Programme

The applicability of non-parametric methods of statistical analysis to metrology

By
M G Cox, P M Harris
National Physical Laboratory, UK
Annarita Lazzari
Mitutoyo Italia, Milan, Italy

March 2004

The applicability of non-parametric methods of statistical analysis to metrology

M G Cox, P M Harris
National Physical Laboratory, UK
Annarita Lazzari
Mitutoyo Italia, Milan, Italy

March 2004

ABSTRACT

There are classes of problems in metrology for which little or no distributional knowledge is available concerning the values of the input quantities to a model of measurement. Consequently, non-parametric methods have value when applied to such problems. So-called ‘re-sampling’ methods are discussed and applied to two problems, one concerning environmental and the other dimensional measurement. These problems have very different character, and the re-sampling approaches used are selected accordingly. Further, consideration is given to an approach for the provision of a coverage region for the values of multivariate quantities that makes no assumption about the distribution for the values of these quantities. The approach is applied to a problem in electrical metrology.

ISSN 1471–0005

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

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

Contents

1 Introduction

Metrologists frequently work with models of measurement. These models relate input quantities about which information is available to an output quantity (or quantities) about which information is required. A main concern is to provide an estimate of the value of the output quantity, the measurement uncertainty associated with this estimate, and a coverage interval (corresponding to a stipulated coverage probability) for the value of the output quantity. Approaches based on the use of the Guide to the Expression of Uncertainty in Measurement (GUM) [4] are widely used for this purpose. These approaches require that probability distributions are assigned for the values of the input quantities. The GUM concentrates on the use of estimates of the values of the input quantities and uncertainties associated with these estimates, derived from these distributions, together with the use of the law of propagation of uncertainty and the assignment of a Gaussian distribution to the value of the output quantity, to derive the required information. GUM Supplement 1 [3] applies the propagation of distributions [15] directly to the assigned distributions for the values of the input quantities to determine the distribution for the value of the output quantity (thus avoiding the Gaussian assumption), from which the required information is obtained.

There are problems in metrology where the only knowledge of the values of the input quantities is in the form of samples of measurement of those quantities. If the samples are of adequate size, at least 15, say, such problems can be addressed using non-parametric methods of statistical analysis that make intensive ‘re-use’ of these measurements. They are in contrast to the methods in the companion report, ‘Statistical error modelling’ [12], which are based on assigning statistical models to the deviations associated with the measurements. These so-called ‘re-sampling’ methods [17, 18] (section 2) provide an approximation to the distribution for the value of the output quantity, from which the required information can be extracted.

Two applications of re-sampling methods are given. One application is in the area of air quality (section 3), where a trend model is used to help understand and predict pollution levels of ground-level ozone, given measurements over a period of time [13, 14]. The form of re-sampling used is typical of that applied to regression problems.

The other application of re-sampling is in the area of roundness assessment [1] (section 4), where the profile is measured of a section of a component whose deviation from perfect form is sought. The type of re-sampling used takes account of the fact that the measurements contain mutual dependencies in the form of serial correlation. Failure to account for this effect would provide invalid uncertainties associated with the component and the measuring instrument.

A further non-parametric method considered in this report is a distribution-free approach to derive a coverage interval for the value of a quantity or a coverage

region for the values of a set of quantities (section 5). The method uses estimates of the values of the quantities and the associated standard uncertainties and, if appropriate, covariances, to derive the coverage region.

The method is illustrated by an application in electrical metrology [4] (section 6) relating to the determination of the resistance and reactance of a circuit element.

2 Re-sampling methods

This section reviews the class of non-parametric methods known as re-sampling methods, especially the use of 'bootstrap re-sampling' [5, 6, 17, 18], to evaluate uncertainties associated with measurement results.

Bootstrap re-sampling can be considered as a means for approximating the distribution for the value of a quantity such as one derived from regression parameters (section 2.5). In this approach, no assumption is made about the nature of the uncertainties associated with the measurements.

The bootstrap is based on the use of the measurements to construct further measurements. Such measurements do not contain (because they cannot) further information *per se*. However, they can assist in approximating the distribution for the value of a quantity using estimates of the value of the quantity derived from the 'new' measurements. These estimates are considered to mimic those that would have arisen had it been possible to sample from the population from which the given measurements are deemed to have been drawn.

Let the provided measurements of the values of a quantity $\boldsymbol{X} = (X_1, \ldots, X_N)^{\mathrm{T}}$ be denoted by $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$. Let y denote an estimate of the value of Y, the quantity (the mean or median would be a simple example) to be formed from $\boldsymbol{X}$ using a known model of measurement $Y = f(\boldsymbol{X})$. Consider a discrete probability distribution assigned to the value of each X_i (section 2.1) and sample from each of those distributions to provide a 'new' set of N measurements, $\boldsymbol{x_1} = (x_{1,1}, \ldots, x_{N,1})^{\mathrm{T}}$. Denote the corresponding value of the output quantity for these measurements by y_1.[1] Derive such a 'new' set of measurements many times, M, say, giving in all 're-samples' $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, and corresponding estimates $y_1, \ldots, y_M$ of the value of Y. With the bootstrap, this set of values is taken as the basis of an approximation to the distribution for the value of Y.

Some problems, especially regression problems, have input data that is hierarchical in nature. Rather than there being a single input data set, there is a *set of sets* of input data. For instance, a single data set might constitute N triplets of spatial

[1] For instance, if the original measurements were $\boldsymbol{x} = (1.2, 1.3, 1.0, 1.5)^{\mathrm{T}}$ and Y were the arithmetic mean, then $y = 1.25$. A re-sampled set of measurements is $\boldsymbol{x}_1 = (1.3, 1.5, 1.2, 1.2)^{\mathrm{T}}$, for which $y_1 = 1.30$.

co-ordinates, representing points on a surface in three-dimensional space. The set can be represented as a regular array, e.g., as an $N \times 3$ matrix. An hierarchical data set might consist of a number of data sets each of that type. It would have two hierarchical levels. No clear distinction is made or necessary, however, between single and hierarchical data sets. For instance, suppose there are several, L, say, sets of data, each identical to the above single data set, except for the numerical values of the third co-ordinate. The composite data set can be represented as an $N \times 3 \times L$ matrix, and regarded, if it is helpful to do so, as a single data set.

There can be several hierarchical levels. Hierarchical data will in general have different uncertainty structure at each level. For instance, at one level the data might exhibit mutual dependencies, but mutual independence at another. Any method of evaluating the data would need to respect the various structures in order to provide valid uncertainties associated with the values of the quantities of interest.

No attempt is made here to develop a general theory that can be applied to classes of problems involving hierarchical data. Rather, a particular problem, in roundness assessment, is considered, and a suitable approach based on re-sampling applied. The manner in which this problem is treated may help to provide an indication of how other regression problems involving hierarchical data can be handled using bootstrap re-sampling. For the purpose of this report, each x_i above can where appropriate be regarded as a realization of the 'value' of a vector-valued quantity, for instance. This is the case for the roundness application in section 4. Further comments in the context of regression problems are made at the end of section 2.7.

Another class of re-sampling methods is known as the jack-knife [17, 18]. It is a useful class if the concern is with providing an estimate of the value of the output quantity and the standard uncertainty associated with that estimate. It is not suitable for providing a coverage interval for the value of the output quantity. The use of the jack-knife is not considered further in this report.

2.1 Bootstrap implementation

The basis of a generic bootstrap (algorithm) can be presented as follows.

1. Consider an output quantity Y defined functionally in terms of N input quantities $\boldsymbol{X} = (X_1, \ldots, X_N)$:

$$Y = f(\boldsymbol{X}).$$

2. Given is a sample $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$ of the values of $\boldsymbol{X}$.

3. Assume that the X_i, $i = 1, \ldots, N$, are mutually independent, whose values are each assigned the same (approximate) probability density function (PDF)[2]

$$\widehat{g}(\xi) = \begin{cases} 1/N, & \xi = x_1, \\ \vdots & \vdots \\ 1/N, & \xi = x_N, \\ 0, & \text{otherwise.} \end{cases}$$

4. Select a number M of bootstrap samples to be made.[3]

5. For $r = 1, \ldots, M$, generate a bootstrap sample:

   (a) Take N 'draws' $\boldsymbol{x}_r = (x_{1,r}, \ldots, x_{N,r})^{\mathrm{T}}$ from the PDF $\widehat{g}(\xi)$.

   (b) Form the corresponding value of Y:

$$y_r = f(\boldsymbol{x}_r).$$

6. Apply the considerations of section 2.2 to provide an approximate distribution function $\widehat{G}(\eta)$ for the value of Y.

7. Apply the considerations of section 2.3 to provide an estimate y of the value of Y, the measurement uncertainty $u(y)$ associated with this estimate, and a coverage interval (corresponding to a stipulated coverage probability) for the value of Y.

2.2 Approximate distribution function

In general, suppose a set of realizations of the value of a quantity is available. If all the items in the set can be regarded as equally probable realizations, they

[2]This choice follows from the application of the Principle of Maximum Entropy [31]. Let p_i denote the probability to be assigned to the value $X = x_i$. Then the p_i are to be determined such that the entropy

$$-\sum_{i=1}^{m} p_i \log p_i$$

is maximized subject to the condition that their values sum to unity. For this purpose, the turning point of the Lagrangian

$$L(\boldsymbol{p}, \lambda) = -\sum_{i=1}^{N} p_i \log p_i + \lambda \left(\sum_{i=1}^{N} p_i - 1 \right),$$

where $\boldsymbol{p} = (p_1, \ldots, p_N)^{\mathrm{T}}$ and λ is the Lagrangian parameter, is to be determined. The result is $p_i = 1/N$, $i = 1, \ldots, N$. This argument is a simplification, since the x_i are realizations of the value of a *continuous* quantity. A solution to handle this aspect properly requires a more sophisticated argument.

[3]Generally, M should be chosen to be as large as possible.

can be used to form an approximation to the distribution function for the value of the quantity. The set of realizations may be the consequence of an application of an implementation of the propagation of distributions. It may be the result of the use of bootstrap re-sampling (section 2.1). It may stem from some other calculation or consideration. For the first-mentioned case, the Joint Committee for Guides in Metrology has produced a supplement to the Guide to the Expression of Uncertainty in Measurement (GUM) [4] concerned with the propagation of distributions, and its implementation using Monte Carlo simulation, for the purposes of uncertainty evaluation [3]. The manner in which the approximate distribution is determined from a set of realizations of the value of a quantity is reviewed. Also reviewed is the determination from this distribution of an estimate of the value of the quantity, the associated uncertainty, and a coverage interval for the value of the quantity. Some implementation details of the manner in which bootstrap re-sampling can provide a set of realizations were given in section 2.1.

Suppose M realizations y_r, $r = 1, \ldots, M$, of the value of a quantity Y are available. An approximation $\widehat{G}(\eta)$ to the distribution function $G(\eta)$ for the value of Y is obtained as follows. Sort the values y_r, $r = 1, \ldots, M$, into non-decreasing order. Denote the sorted values by $y_{(r)}$, $r = 1, \ldots, M$. Assign uniformly spaced cumulative probabilities $p_r = (r - 1/2)/M$, $r = 1, \ldots, M$, to the ordered values [11].[4]

Form $\widehat{G}(\eta)$ as the piecewise-linear function joining the M points $(y_{(r)}, p_r)$, $r = 1, \ldots, M$:

$$\widehat{G}(\eta) = \frac{r - 1/2}{M} + \frac{\eta - y_{(r)}}{M(y_{(r+1)} - y_{(r)})}, \quad y_{(r)} \leq \eta \leq y_{(r+1)}, \quad r = 1, \ldots, M - 1. \tag{1}$$

The values of $y_{(r)}$ (or y_r), when assembled into a histogram (with suitable cell widths) form a frequency distribution that, when normalized to have unit area, provides an approximation $\widehat{g}(\eta)$ to the PDF $g(\eta)$ for the value of Y. Calculations are not generally carried out in terms of this histogram, the resolution of which depends on the choice of cell size, but in terms of the approximation $\widehat{G}(\eta)$ to the distribution function $G(\eta)$. The histogram can, however, be useful as a visual aid to understanding the nature of the PDF, e.g., the extent of its asymmetry.

2.3 Estimates, associated standard uncertainties and coverage intervals

The expectation $\widehat{y}$ of the function $\widehat{G}(\eta)$ (expression (1)) approximates the expectation y of the PDF $g(\eta)$ for the value of Y. The standard deviation $u(\widehat{y})$ of $\widehat{G}(\eta)$ approximates the standard deviation of $g(\eta)$ and is taken as the standard

[4] The values p_r, $r = 1, \ldots, M$, are the midpoints of M contiguous probability intervals of width $1/M$ between zero and one.

uncertainty $u(y)$ associated with y. To a good approximation,[5] for an adequately large value of M, $\widehat{y}$ can be taken as the arithmetic mean

$$\widehat{y} = \frac{1}{M}\sum_{r=1}^{M} y_r,$$

formed from the M values $y_1, \ldots, y_M$, and the standard uncertainty $u(\widehat{y})$ determined from

$$u^2(\widehat{y}) = \frac{1}{M-1}\sum_{r=1}^{M}(y_r - \widehat{y})^2.$$

Let α denote any value between zero and $1-p$, where p is the required coverage probability. The endpoints of a $100p$ % coverage interval $I_p(Y)$ for the value of Y are $G^{-1}(\alpha)$ and $G^{-1}(p+\alpha)$, i.e., the α– and $(p+\alpha)$–quantiles of $G(\eta)$. The β–quantile is the value of η such that $G(\eta) = \beta$. The choice $\alpha = 0.025$ gives the coverage interval defined by the 0.025– and 0.975–quantiles, providing an $I_{0.95}(Y)$ that is probabilistically symmetric. The probability is 2.5 % that the value of Y is smaller than the left-hand endpoint of the interval and 2.5 % that it is larger than the right-hand endpoint. If $g(\eta)$ is symmetric about its expectation, $I_p(Y)$ is symmetric about the estimate of the output quantity value Y, and the left-hand and right-hand endpoints of $I_p(Y)$ are equidistant from y.[6]

A value of α different from 0.025 would generally be appropriate were the PDF asymmetric. Usually the shortest $I_p(Y)$ is required, because it corresponds to the best possible location of the value of Y within a specified probability. It is given by the value of α satisfying $g(G^{-1}(\alpha)) = g(G^{-1}(p+\alpha))$, if $g(\eta)$ is single-peaked, and in general by the value of α such that $G^{-1}(p+\alpha) - G^{-1}(\alpha)$ is a minimum. If $g(\eta)$ is symmetric, the shortest $I_p(Y)$ is given by taking $\alpha = (1-p)/2$.

The shortest $I_p(Y)$ can generally be obtained computationally from $\widehat{G}(\eta)$ by determining α such that $\widehat{G}^{-1}(p+\alpha) - \widehat{G}^{-1}(\alpha)$ is a minimum. The law of propagation of uncertainty, as in the GUM [4], provides the shortest such interval for the Gaussian distribution assigned to the value of Y.

Figure 1 shows an example of a distribution function $G(\eta)$ and the equiprobabilistic and shortest 95 % coverage intervals.

2.4 Regression models

Emphasis is placed on the use of the bootstrap within *regression problems*, which constitute a large fraction of mathematically formulated problems in metrology, in-

[5] Exact expressions for $\widehat{y}$ and $u(\widehat{y})$ are available [11].

[6] For a symmetric PDF for the value of Y, the procedure would, for a sufficiently large value of M, give the same coverage interval for practical purposes as taking the product of the standard uncertainty and the coverage factor that is appropriate for that PDF. This PDF is generally not known analytically.

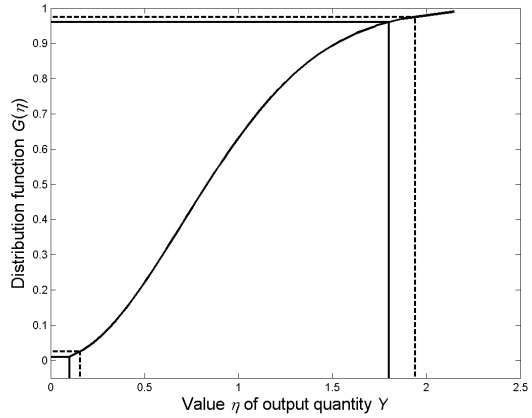

Figure 1: A distribution function $G(\eta)$ corresponding to an asymmetric PDF. Broken lines mark the endpoints of the probabilistically symmetric 95 % coverage interval and the corresponding probability points, viz., 0.025 and 0.975. Solid lines mark the endpoints of the shortest 95 % coverage interval and the corresponding probability points, which are 0.006 and 0.956 in this case. The lengths of the 95 % coverage intervals in the two cases are 1.76 units and 1.69 units, respectively.

cluding calibration, geometric assessment (more on both of which is given below), data fusion, experimental data analysis, conformity assessment and key comparisons [2, 7, 9, 10, 11, 20, 21].

Within regression there are areas where 'immediate' application of the bootstrap brings advantages. There are other areas where straightforward application would be invalid, because, e.g., there are mutual dependencies in the measurements that exist, but which have not been quantified *a priori*. A main purpose of this section is to review the conventional use of the bootstrap for regression problems. Section 4 considers, through a comprehensive example in dimensional metrology concerning roundness assessment, how a more complicated case involving correlated data can be addressed.

Regression models that depend linearly on their parameters $\boldsymbol{A} = (A_1, \ldots, A_n)^{\mathrm{T}}$ of the form

$$F(\boldsymbol{A}; T) = \sum_{j=1}^{n} A_j \phi_j(T),$$

where the $\phi_j(T)$ denote an appropriate set of linearly independent functions, are often used in metrology. They are typically used to model sets of data $(t_i,\ x_i)$, $i = 1, \ldots, m$, where the t_i represent measurements of the values of a stimulus variable and the x_i measurements of the corresponding values of a response variable. Often the uncertainties associated with the t_i are regarded as negligible compared with those associated with the x_i. That is the case considered here, although the approach is capable of extension.

The parameters in the regression model are often determined by minimizing with respect to $\boldsymbol{A}$ the differences between the x_i and the corresponding model values $F(\boldsymbol{A}; t_i)$ in a weighted least-squares sense, i.e., by minimizing the sum of the squares of the weighted residual deviations,

$$\sum_{i=1}^{m} w_i^2 e_i^2(\boldsymbol{A}), \tag{2}$$

where

$$e_i(\boldsymbol{A}) = x_i - F(\boldsymbol{A}; t_i)$$

is the unweighted residual deviation and the weight w_i is ideally taken as

$$w_i = 1/u(x_i),$$

the reciprocal of the standard uncertainty associated with x_i.

Sometimes other measures are used in place of expression (2), for instance, the sum of absolute weighted residual deviations,

$$\sum_{i=1}^{m} w_i |e_i(\boldsymbol{A})|.$$

Denote by $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$ the estimate of $\boldsymbol{A}$ obtained by minimizing one of the measures. Only the least-squares measure is considered subsequently in this report.

2.5 Quantities derived from regression models

In many problems, e.g., calibration and geometric form assessment, it is necessary to consider a quantity (or quantities) that is a function $h(\boldsymbol{A})$, say, of the regression parameters $\boldsymbol{A}$, to form $h(\boldsymbol{a})$ and to evaluate the uncertainty associated with $h(\boldsymbol{a})$. For instance, in calibration, the function $X = F(\boldsymbol{A}; T)$ would represent a calibration curve relating stimulus T and response X. The 'inverse function' $T = \psi(\boldsymbol{A}; X)$, say, would be used to form the value of the stimulus T corresponding to a value of the response X. In geometric assessment, $X = F(\boldsymbol{A}; T)$ would represent an actual geometric feature on a manufactured object. The nominal form of the feature according to the design specification would be a geometric element such as a circle or an ellipse. The departure of the assessed feature from, say, a circle, is a measure of the imperfection of the feature. Thus, if T denotes angle and X radius, and $F(\boldsymbol{A}; T)$ is a Fourier representation of a 'near-circular' feature, the feature imperfection, or 'out-of-roundness' (often simply termed 'roundness') is

$$h(\boldsymbol{A}) = \max_T F(\boldsymbol{A}; T) - \min_T F(\boldsymbol{A}; T). \tag{3}$$

2.6 Regression model uncertainties

By making appropriate assumptions, the uncertainties associated with the estimate $\boldsymbol{a}$ of the values of $\boldsymbol{A}$ obtained as in section 2.4, can be evaluated, and thence the uncertainty associated with a 'model value', i.e., the estimate

$$\sum_{j=1}^{n} a_j \phi_j(t_0) \tag{4}$$

for a prescribed estimate t_0 of the value of the stimulus, can be evaluated. Moreover, a coverage interval (corresponding to a prescribed coverage probability) for the estimate (4) of the value of the model, given t_0, can be formed.

Uncertainties associated with functions $h(\mathbf{a})$, such as those in section 2.5, can also be formed, by applying the law of propagation of uncertainty or some other appropriate approach.[7]

A common assumption is that the weighted measurement deviations $w_i e_i$ (section 2.4) are realizations of the values of independent normally distributed random variables $N(0,\ 1)$. Then, for the least-squares measure, the above calculations are readily carried out using the well-established theory for such variables.[8]

In particular, a 95 % coverage interval for the value of a quantity Θ, viz., a model parameter A_j or a quantity derived from the model parameters $\boldsymbol{A}$, is given by $[\theta - ku(\theta),\ \theta + ku(\theta)]$, where θ is an estimate of the value of Θ, $u(\theta)$ the associated standard uncertainty, and k a coverage factor [4]. k depends on the distribution for the value of Θ. If the distribution is normal, $k = 1.96$ and the 95 % coverage interval would be $[\theta - 1.96u(\theta),\ \theta + 1.96u(\theta)]$.[9]

In other cases, e.g., where the distribution for the value of Θ is unknown, but distributions are assigned to the values of the stimulus variable, the use of the propagation of distributions can be considered [3].[10]

2.7 Bootstrap regression

When applied to regression problems, the residual deviations from the regression are bootstrapped and used to construct 'new' data and hence 'new' regressions.

[7] The law of propagation of uncertainty cannot readily be applied to the model (3) because it is not differentiable as a consequence of the maximum and minimum operators.

[8] Theory is also available for the case where the weighted measurement deviations are realizations of the values of a joint normally distributed random variable.

[9] The notation $[z - U,\ z + U]$ is used to denote a coverage interval for the value of a quantity Z, for which z is an estimate and U an expanded uncertainty. This form is preferable to the recommendation $z \pm U$ [4], because the latter cannot be generalized to asymmetric coverage intervals.

[10] The propagation of distributions would be a suitable approach for the model (3).

The generic bootstrap algorithm of section 2.1, together with consideration of a quantity derived from the model parameters, becomes:

1. The linear regression model is

   $$\boldsymbol{X} = \boldsymbol{C}\boldsymbol{A},$$

   where $\boldsymbol{C}$ is a fixed $m \times n$ (design or observation) matrix with $m \geq n$ and having rank n, $\boldsymbol{X}$ an m-element observation vector (the input quantities), and $\boldsymbol{A}$ an n-element vector of model parameters (a multivariate counterpart of the output quantity Y in section 2.1).[11]

2. Suppose $\boldsymbol{x}$ is a given vector of measurements of the values of $\boldsymbol{X}$.

3. Determine the least-squares solution $\boldsymbol{a}$ (formally $\boldsymbol{a}$ is given by $\boldsymbol{a} = \boldsymbol{f}(\boldsymbol{x}) \equiv (\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{x}$).[12]

4. Form the model values

   $$\widetilde{\boldsymbol{x}} = \boldsymbol{C}\boldsymbol{a}.$$

5. Form the corresponding residual deviations

   $$\widetilde{\boldsymbol{e}} = \boldsymbol{x} - \widetilde{\boldsymbol{x}}.$$

6. Form the approximate PDF

   $$\widehat{g}(\widetilde{\epsilon}) = \begin{cases} 1/m, & \widetilde{\epsilon} = \widetilde{e}_1, \\ \vdots & \vdots \\ 1/m, & \widetilde{\epsilon} = \widetilde{e}_m, \\ 0, & \text{otherwise.} \end{cases}$$

7. For $r = 1, \ldots, M$,

   (a) Form 'new' residual deviations $\widetilde{\boldsymbol{e}}_r = (\widetilde{e}_{1,r} \ldots, \widetilde{e}_{m,r})^{\mathrm{T}}$, given by m draws from the approximate PDF $\widehat{g}(\widetilde{\epsilon})$.

   (b) Form new observations

   $$\boldsymbol{x}_r = \widetilde{\boldsymbol{x}} + \widetilde{\boldsymbol{e}}_r.$$

   (c) Form $\boldsymbol{a}_r = \boldsymbol{f}(\boldsymbol{x}_r)$.

   (d) Form $q_r = h(\boldsymbol{a}_r)$, where $Q = h(\boldsymbol{A})$ is a quantity derived through a function h from the parameters $\boldsymbol{A}$.

[11] The validity of the model in terms of its consistency with the measurement data should be tested [2].

[12] This step and the subsequent steps can be generalized to *weighted* least-squares regression models.

8. Apply the considerations of section 2.2 to q_r, $r = 1, \ldots, M$, to derive an approximate distribution function for the value of Q.

9. Apply the considerations of section 2.3 to provide an estimate q of the value of Q, the measurement uncertainty $u(q)$ associated with this estimate, and a coverage interval (corresponding to a stipulated coverage probability) for the value of Q.

In terms of the discussion near the start of this section concerning hierarchical data, the key step is the identification of a population (at some level in the hierarchy) for which independent measurements are available. For instance, anticipating the treatment of the roundness assessment application in section 4, the complete set of data constitutes a number of measurement traces, the measurements within each trace being correlated, but the traces themselves uncorrelated. Thus, for this application, each X_i denotes the measurements of the ith trace and f the model that processes the X_i to give the required output quantities ('out-of-roundness' in this case). A further key requirement in the use of the bootstrap is that the model (e.g., least-squares) is appropriate irrespective of the statistical model for the values of the input quantities (which is generally true for least-squares) [12].

3 Trends in ozone concentration

The growth in the extent and scope of air-quality monitoring has resulted in the accumulation of substantial archives of data measured at automatic stations. One of the most important questions for which this data is intended to be used is whether ambient air quality is improving or deteriorating from year to year. Long-term trends are difficult to quantify accurately because of the presence of large seasonal and other effects in the data [19, 25].

The concepts of section 2 are applied to the evaluation of the long-term trend in the concentration of ground-level ozone at a particular site in the UK. The method is based on a regression model incorporating 'seasonal' factors that is fitted to ozone concentration data formed into weekly averages. The uncertainties associated with the results derived from the model are evaluated by bootstrap re-sampling from the residual deviations of the model. No assumption is made about the distribution of the residual deviations that are unexplained by the model.

A report is available [14] that gives a more extensive treatment, also covering a set of sites, aggregation over sites, other time periods, validation of results, and the handling of missing data.

The question addressed is 'What is the long-term trend in ozone concentration, as indicated by the measurements, over a particular seven-year period, and what

is the associated uncertainty?'. The arithmetic mean of hourly data, evaluated over weekly time periods, is used to measure ozone concentration. An underlying straight-line trend for this averaged data, expressed as a gradient with respect to time, is used to indicate the long-term trend.

The approach involves the use of a class of regression models incorporating seasonal factors in addition to an underlying long-term trend. The seasonal factors are used to describe periodic effects. Bootstrap re-sampling is used to approximate the distribution for the value of the gradient, and hence provide a coverage interval for the value of the gradient.

3.1 Regression model with seasonal factors

An approach [14] to estimating the long-term trend in a data series is to calculate seasonal averages of the data and then to fit a suitable model to these averages. If $Y_{i,j}$, $j = 1, \ldots, n$, denote ozone concentration (in parts per billion—ppb) obtained by averaging over n equal time periods in year i, the model takes the form

$$Y_{i,j} = AT_{i,j} + B_j, \qquad T_{i,j} = i - 1 + j/n, \qquad j = 1, \ldots, n.$$

Here A is the gradient of the underlying straight-line trend and $\boldsymbol{B} = (B_1, \ldots, B_n)^{\mathrm{T}}$ are seasonal factors. Estimates a and $\boldsymbol{b}$ of the values of A and $\boldsymbol{B}$ are determined by minimizing the difference between the data and model values in a least-squares sense.

An example of some raw data is illustrated in figure 2, which shows (hourly-averaged) measurements for a seven-year period at a particular site in the UK.

Figure 3 provides the result of fitting the model to data formed as weekly-averaged ozone-concentration values (a model with 53 parameters, given by taking $n = 52$). The upper part of the figure shows the data as small circles and the fitted model as a sequence of straight-line segments joining model values. The nearly horizontal line represents the 'trend' line. Its gradient is the estimate a. (The intercept of this line with any vertical axis is arbitrary.) The lower plot shows the model residuals joined by straight-line segments.

3.2 Evaluation of uncertainties

Data averaging constitutes the initial part of the process of analyzing the data. According to the Central Limit theorem [22, Vol. 1, p193], the distribution for the mean of a sample of values drawn independently from the same distribution tends to normality as the size of the sample tends to infinity. Hence, it might be anticipated that Gaussian statistics would be applicable to the models considered

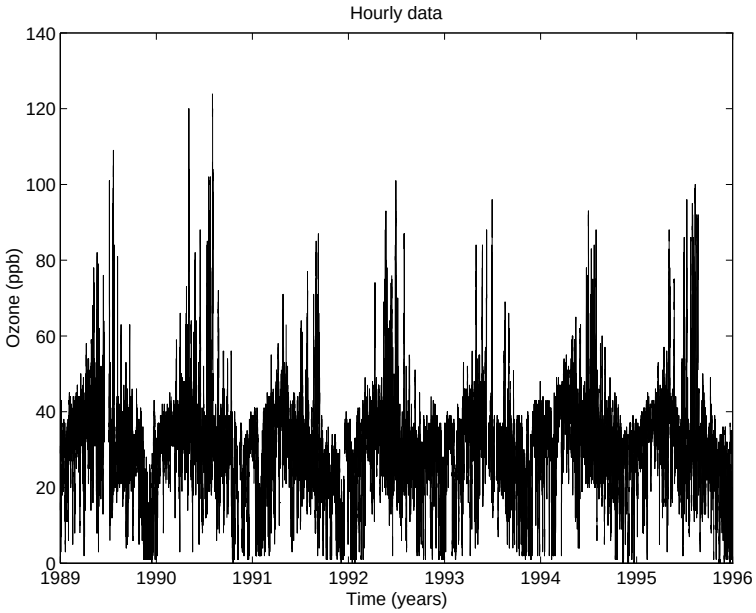

Figure 2: Time series of hourly averaged ozone-concentration measurements for a seven-year period at a particular site in the UK.

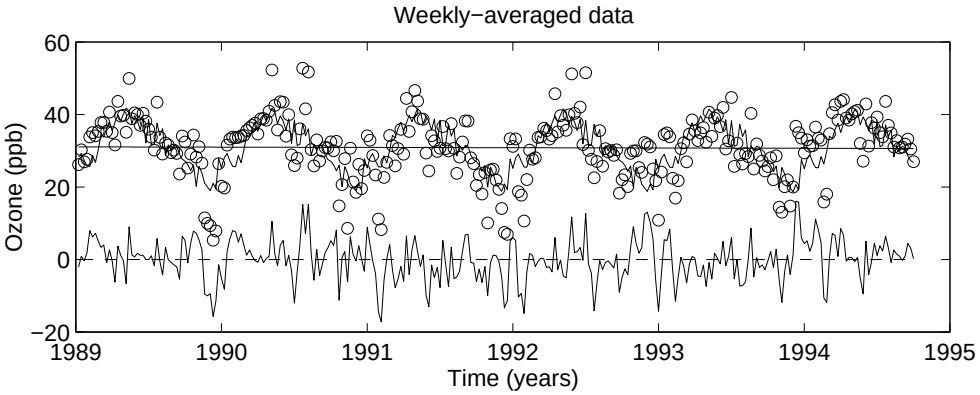

Figure 3: Weekly-averaged ozone-concentration measurements (circles) derived from the data of figure 2, the corresponding 53-parameter regression-model values (joined by straight lines) and the underlying trend line, and (below) the corresponding model residual deviations (also joined by straight lines).

here. However, the hourly figures on which the averages are based are serially correlated, the extent of which is clearly evidenced by forming the set of serial correlations of lag 1, 2, ..., [22, Vol. 1, p362]. Since the assumption of independence does not hold, it is imprudent to assume that the conditions of the Central Limit theorem apply. Hence, no assumption is made concerning the nature of the measurement deviations and accordingly bootstrap re-sampling [17, 18] is used to form a coverage interval.

Bootstrap re-sampling enables non-normally distributed data to be treated robustly, as considered in section 2.7. The quantity Q there is taken as the gradient A quantifying the trend.

3.3 Results

For comparative purposes a 95 % coverage interval for the value of the gradient under an (unsupported) Gaussian assumption is also computed.

The use of the law of propagation of uncertainty of the GUM gave 0.118 ppb yr^{-1} as the estimate of the value of the gradient of the underlying straight-line trend, an associated standard uncertainty of 0.173 ppb yr^{-1}, and a 95 % equiprobabilistic coverage interval of [-0.228, 0.464] ppb yr^{-1}.

Figure 4 shows the approximation to the distribution function for the value of the gradient obtained using bootstrap re-sampling. For this function, the expectation was 0.117 ppb yr^{-1}, the standard deviation 0.155 ppb yr^{-1}, and the 95 % equiprobabilistic coverage interval$[-0.187,\ 0.420]$ ppb yr^{-1}, for comparison with the above results.

Figure 5 shows an approximation to the PDF for the value of the gradient of the long-term trend in ozone concentration. It was obtained by scaling a histogram of the values provided by bootstrap re-sampling. The endpoints of the 95 % equiprobalistic coverage interval are shown as vertical lines.

3.4 Remarks

Although there is a degree of agreement between the results obtained under the Gaussian assumption and those from bootstrap re-sampling, it is expected that the latter would be the more reliable, because of its freedom from distributional assumptions. It is clear from the results, and from their portrayal in figure 5, that they do not provide evidence of a positive or a negative trend in ozone concentration at the site considered. However, when the approach is applied to all the sites in the UK Rural Ozone Monitoring Network [26], a significant positive trend is indicated at some sites [14], suggesting a geographical effect. Moreover, when averaged

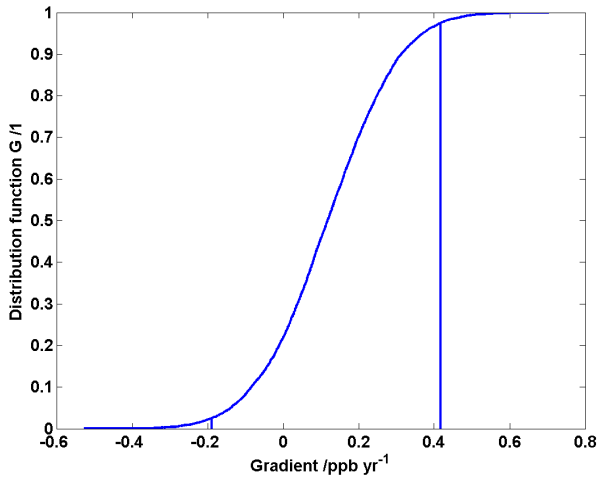

Figure 4: Approximation to the distribution function for the value of the gradient of the long-term trend in ozone concentration provided by bootstrap re-sampling. The endpoints of the equiprobabilistic coverage interval corresponding to a 95 % coverage probability are marked by vertical lines.

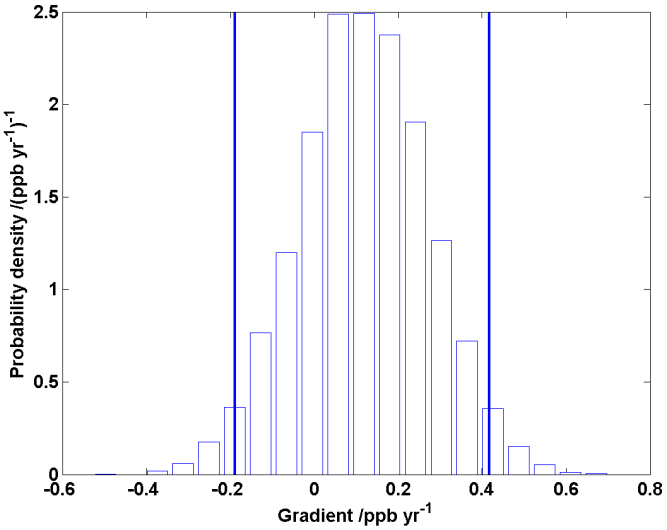

Figure 5: Approximation to the PDF for the value of the gradient of the long-term trend in ozone concentration. The endpoints of the 95 % coverage interval obtained from approximate PDFs are shown as vertical lines.

over all sites to form a ‘national average’, such a trend is also indicated. Had such results been obtained under a Gaussian assumption, there would have been doubts relating to whether this distributional assumption was invalidating the conclusion.

There are alternatives to the above way of modelling ozone-concentration measurements and to the manner in which the model is analyzed. Moreover, the assumptions currently made or implied require further investigation. Some of these issues are indicated below. A treatment of other issues is available [14].

1. Use Fourier series to provide a continuous rather than a discrete model of the measurements.
2. Although a least-squares analysis does not require the measurements to be normally distributed, the assumption is made that the residual deviations have
   (a) zero mean and constant variance,[13]
   (b) are mutually independent,
   (c) are uncorrelated with the measurements themselves.

   These assumptions imply that the trend is deterministic, the seasonal component is deterministic and the seasonal and non-seasonal effects are independent [30].

4 Roundness assessment

A roundness-measuring instrument with a precision rotary stage [23, 27] is used to provide the profile of a nominally circular section of a component and then to quantify the departure of that profile from a perfect circle (the component ‘out-of-roundness’). 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 typically taken. This set of traces is analyzed to determine the component form deviation (CFD), the angle-dependent departure from circularity, and the instrument spindle deviation (ISD), correspondingly, 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 $Y(\theta)$,

[13] Weights were incorporated so that this property applied for the weighted residual deviations.

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 $\phi_1, \ldots, \phi_q$, traces can be obtained having the form

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

These expressions constitute an over-determined 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'.

There are appreciable positive correlations present in the measurements, which, 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 NPL-Taylor Hobson Talyrond 73HPR roundness-measuring instrument.

Figure 6 shows a typical trace, as a polar plot, 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 residual deviations, viz., the radial departures of the measured trace from the model trace, exhibit systematic behaviour. This behaviour is seen more clearly in Figure 7, which shows the residual deviations as a function of angular position as a Cartesian plot. 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.

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

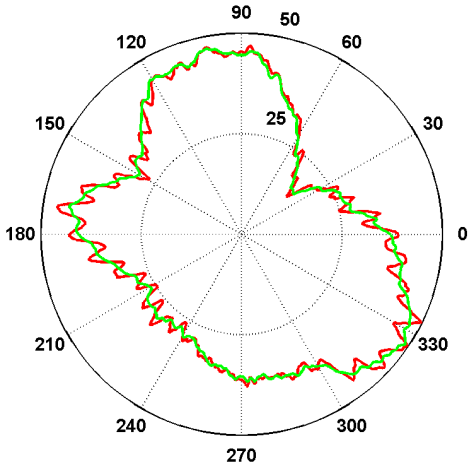

Figure 6: A typical measurement trace, trace 15 of the roundness data (red), and corresponding reconstructed 'model' trace (green).

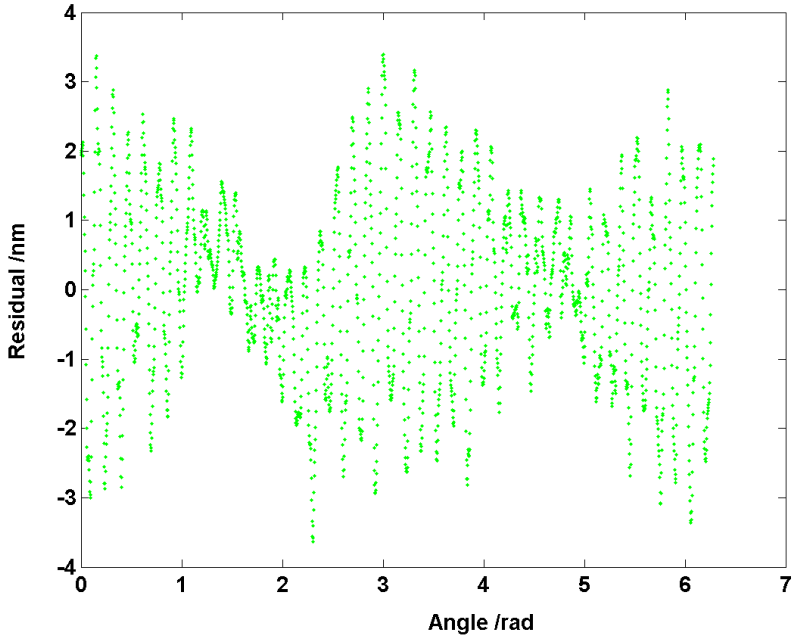

Figure 7: Residual deviations derived from the measurement trace and reconstructed trace shown in Figure 6.

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.

This problem has also been addressed with a very different approach, the use of autoregressive modelling [12]. The treatment there may be compared with that here.

First, a treatment is given of measurements provided by instruments of the type under consideration here, in cases where the measurement deviations are dominated by random effects. Then, a treatment that accounts for the serial correlation effects observed for the Talyrond 73HPR measurements is considered. They are contrasted.

4.1 Models of measurement

The uncertainty associated with the out-of-roundness of the component and of the spindle can be evaluated from a knowledge of the uncertainties associated with the raw measurements (the suppressed radius values in the measurement traces).

A sequence of models of measurement is involved, viz.,

1. The coefficients in the Fourier representation of each trace as a function of the suppressed radius measurements constituting the trace.

2. The coefficients in the Fourier representations of the CFD and the ISD as functions of the coefficients in the Fourier representation of the traces.

3. The values of the Fourier representations of the CFD and the ISD as functions of the coefficients in those representations.

4. The out-of-roundness of the component and of the spindle as functions of the values of the Fourier representations of the CFD and the ISD.

The uncertainty evaluation can accordingly, in principle, be carried out in four main stages by applying the law of propagation of uncertainty to these models. Doing so would provide successively the uncertainties associated with (1) the Fourier representations of the traces, (2) the coefficients in the Fourier representations of the CFD and the ISD, (3) the values of these Fourier representations, and finally (4) the out-of-roundness of the component and of the spindle.

4.2 Mutually independent measurements

Consider roundness-measuring instruments for which the measurement deviations can be regarded as mutually independent and identically distributed, with (unspecified) associated standard uncertainty $u(y)$. Details of the analysis are available in this circumstance [11], the results for which corresponding to the above four stages are summarized below.

1. The standard uncertainty associated with each coefficient in the Fourier representations of the traces is equal to $(2/m)^{1/2}u(y)$, where m is the number of measurements in each trace.[14] Furthermore, the covariances associated with the Fourier coefficients are all equal to zero, and hence these coefficients are mutually independent.

2. The standard uncertainty associated with the kth coefficient in the Fourier representations of the CFD and the ISD is equal to the product of the above standard uncertainty, $(2/m)^{1/2}u(y)$, and the factor

$$\rho_k = 1/(1 - \mu_k^2/q^2)^{1/2},$$

where q is the number measurement traces and

$$\mu_k = (c_k^2 + s_k^2)^{1/2}$$

with

$$c_k = \sum_{\ell=1}^{q} \cos k\phi_\ell, \qquad s_k = \sum_{\ell=1}^{q} \sin k\phi_\ell,$$

the ϕ_ℓ denoting the indexing angles used.

3. The standard uncertainties $u(C(\theta))$ and $u(S(\theta))$ associated with (values of) the Fourier representations $C(\theta)$ and $S(\theta)$ of the CFD and the ISD, respectively, are equal to the product of the above standard uncertainty, $(2/m)^{1/2}u(y)$, and the factor

$$\tau = \left(\sum_{k=1}^{n} \rho_k^2 \right)^{1/2},$$

where n is the number of harmonics in the Fourier representations of $C(\theta)$ and $S(\theta)$, i.e.,

$$u(C(\theta)) = u(S(\theta)) = (2/m)^{1/2}\tau u(y) = u(C) = u(S), \tag{5}$$

say, a result that is independent of θ.

[14] There is an exception: the standard uncertainty associated with the constant terms in the Fourier representations is equal to $(1/m)^{1/2}u(y)$ [11].

4. The component out-of-roundness

$$\delta C = C(\theta_{\max}) - C(\theta_{\min}),$$

where $\theta_{\max}$ and $\theta_{\min}$ are the angles at which $C(\theta)$ takes its maximum and minimum values, respectively, with an analogous result for the spindle out-of-roundness. Letting $\rho(a, b)$ denote the correlation coefficient associated with estimates a and b, the standard uncertainty $u(\delta C)$ associated with δC is therefore given by

$$\begin{aligned} u^2(\delta C) &= u^2(C(\theta_{\max})) + u^2(C(\theta_{\min})) \\ &- 2\rho(C(\theta_{\max}),\ C(\theta_{\min}))u(C(\theta_{\max}))u(C(\theta_{\min})), \end{aligned}$$

which, using expression (5), gives

$$u^2(\delta C) = (2 - 2\rho(C(\theta_{\max}),\ C(\theta_{\min})))u^2(C).$$

Hence, since

$$|\rho(C(\theta_{\max}),\ C(\theta_{\min}))| \leq 1,$$

it follows that

$$u(\delta C) \leq 2u(C).$$

Without attempting to quantify the above correlation coefficient, the standard uncertainty associated with each of the out-of-roundness estimates is no greater than the product of $2(2/m)^{1/2}\tau$ and the standard uncertainty $u(y)$ associated with the raw measurements. Typical values are $m = 2\ 000$, $q = 20$ and $n = 150$. For sensibly selected indexing angles, τ proves to be close to three: $\tau = 2.8$ for the instance reported here. This multiplicative factor is then, taking the value $\tau = 2.8$,

$$2 \times (2/m)^{1/2}\tau \approx 2 \times (2/2000)^{1/2} \times 2.8 \approx 0.18.$$

The root-mean-square residuals for the Fourier representations of the measurement traces all lay between 0.30 nm and 0.35 nm, with a mean value of 0.33 nm, which estimates the standard uncertainty $u(y)$. Thus, the standard uncertainty associated with each of the out-of-roundness estimates is equal to $0.18 \times 0.33 \approx 0.06$ nm. This value is exceedingly small,[15] even for ultra-high accuracy roundness measurement, and hence the basic assumption behind the uncertainty evaluation is highly questionable.

4.3 Mutually dependent measurements

The uncertainty evaluation in Section 4.2 would be valid for instruments that provide traces consisting of mutually independent measurements. However, for the

[15] It would be even smaller were the above correlation coefficient appreciably less than unity.

Talyrond 73HPR roundness-measuring instrument, this is not the case. It would be possible to use such an approach if the covariance effects could be quantified, but such information is not available, at least not directly, from knowledge of the instrument.

Strong serial correlation is observed in the residual deviations when the traces are reconstructed from the Fourier model. This effect indicates the presence of mutual dependence (covariance) in the measurements.

An approach is proposed in terms of the hierarchical considerations in section 2 (also see the end of section 2.7). A two-level data hierarchy is considered, the lower level corresponding to the measurements within a trace, and the higher to the traces themselves. Although serial correlation is present within each trace, there is no evidence for dependencies between traces.

Were bootstrap re-sampling to be used to generate new data sets (traces) as the sum of the model (re-constructed) traces and re-sampled residuals (as in section 2.7), the serial correlation would be destroyed, and the resulting uncertainty evaluation would be invalid. Instead, re-sampling is carried out at the trace level rather than at the level of the individual points in each trace. In this way the uncertainty structure in the different traces is retained, but, by randomly selecting traces, with replacement from the complete set of measurement traces, 'new' complete roundness data sets having characteristics consistent with those of the measured traces can be generated.

$M = 10\ 000$ trials were made. Each trial consisted of a complete re-sampled roundness data set, always re-sampling from the original data with replacement. For the number of traces, q, equal to 22, in the case here, the original set of traces is indexed 1, 2, 3, ..., 21, 22. An instance of the traces used for one of the 10 000 trials (in fact the last of these) is 17, 9, 22, 3, 13, 3, 19, 9, 17, 6, 14, 17, 16, 7, 12, 15, 1, 13, 16, 7, 16, 8. When sorted, these indices are 1, 3, 3, 6, 7, 7, 8, 9, 9, 12, 13, 13, 14, 15, 16, 16, 16, 17, 17, 17, 19, 22, thus indicating that original trace 1 was used, as was trace 3 (twice), trace 6, etc. Original traces 16 and 17 were in fact each used three times. Eight of the original traces (2, 4, 5, 10, 11, 18, 20, 21) were not used at all.

For each of the above M bootstrapped data sets, the complete problem was solved, first assigning of course the correct index angle to each re-sampled trace.

The result of this computation was M values of δC and M values of δS, the out-of-roundness quantities. These values were used to provide approximations to the distribution functions for the values of the CFD and ISD, using the approach described in section 2.2. These approximate distribution functions are depicted as figures 8 and 9. The tail behaviour of these distribution functions is different.

As also indicated in section 2.2, approximate PDFs for the values of the CFD

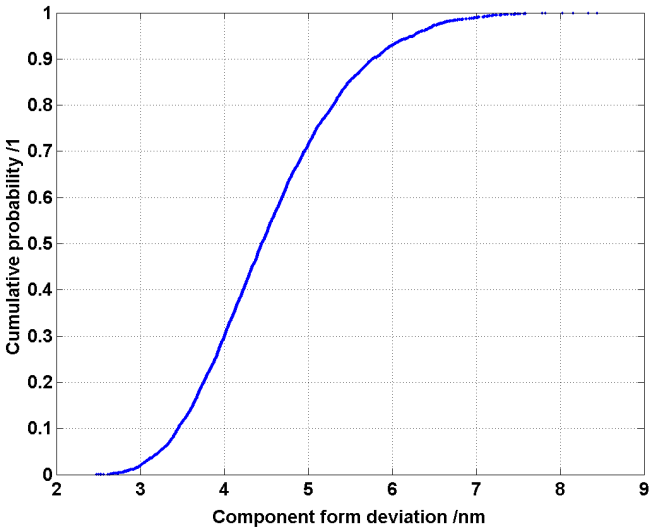

Figure 8: Approximation to the distribution function for the value of the component form deviation.

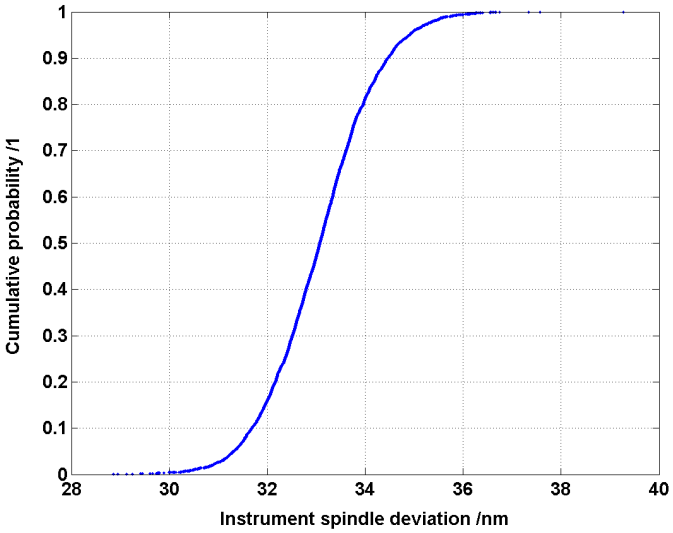

Figure 9: Approximation to the distribution function for the value of the instrument spindle deviation.

and ISD can be formed. Figures 10 and 11 show the approximate PDFs so obtained. They were determined by first forming histograms (with 20 bins) and then converting them to bar charts by applying a scaling factor such that the areas of the resulting charts were unity.

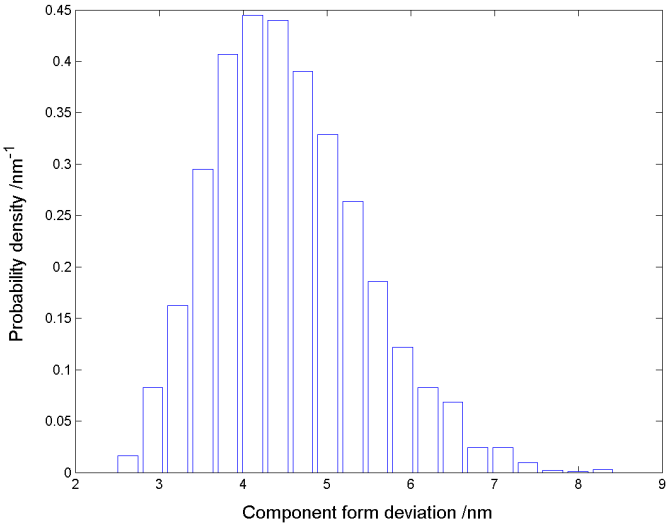

Figure 10: Approximate PDF for the value of the component form deviation.

It is observed that the approximate PDF for the value of the CFD is asymmetric, whereas that for the ISD would appear to be symmetric. Since the value of the CFD is much closer to zero than that of the ISD (relative to the associated standard uncertainties), such behaviour might be anticipated. A symmetric PDF that can take all possible real values, such as a Gaussian, would, with some non-zero probability, take negative values. A PDF for the value of a positive quantity, such as the out-of-roundness here, that is close to the origin is necessarily asymmetric.

The results obtained from the approximations to the distribution functions for the values of the CFD and ISD were

$$\delta C = 4.54 \text{ nm}, \qquad u(\delta C) = 0.91 \text{ nm},$$
$$I_{\text{equi}}(\delta C) = [3.03,\ 6.55] \text{ nm}, \qquad I_{\text{short}}(\delta C) = [2.98,\ 6.43] \text{ nm},$$

and

$$\delta S = 33.07 \text{ nm}, \qquad u(\delta S) = 1.09 \text{ nm},$$
$$I_{\text{equi}}(\delta S) = [30.96,\ 35.29] \text{ nm}, \qquad I_{\text{short}}(\delta S) = [30.89,\ 35.18] \text{ nm},$$

where $I_{\text{equi}}(\delta C)$ and $I_{\text{short}}(\delta C)$ denote the equiprobabilistic and shortest coverage intervals (section 2.2) for the value of δC, and similarly for δS. An additional digit

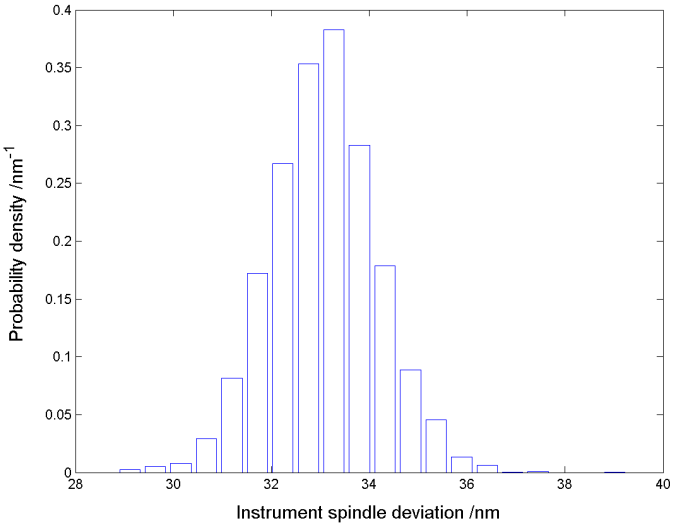

Figure 11: Approximate PDF for the value of the instrument spindle deviation.

beyond the number that would be used in practice is quoted in these results for comparison purposes.

These results can be compared with those of section 4.2, which applied under the (invalid) assumption that the measurements were mutually independent. There, $u(\delta C) = u(\delta S) = 0.06$ nm was obtained, in contrast to the above standard uncertainties of 0.9 nm and 1.1 nm. The later values, obtained using bootstrap re-sampling, were relatively very close to those anticipated by those responsible for the design of the roundness instrument.

To help appreciate the degree of *numerical* approximation to which bootstrap re-sampling had produced results, the exercise was repeated three times, each with $M = 1\,000$ trials. The following values were produced:

$$\begin{aligned}
&\delta C = 4.59,\ 4.52,\ 4.58 \text{ nm}, \qquad u(\delta C) = 0.90,\ 0.89,\ 0.90 \text{ nm},\\
&I_{\text{equi}}(\delta C) = [3.06,\ 6.58],\ [3.11,\ 6.40],\ [3.09,\ 6.55] \text{ nm},\\
&I_{\text{short}}(\delta C) = [2.94,\ 6.35],\ [2.97,\ 6.16],\ [2.99,\ 6.37] \text{ nm},
\end{aligned}$$

and

$$\begin{aligned}
&\delta S = 33.06,\ 33.09,\ 33.08 \text{ nm}, \qquad u(\delta S) = 1.07,\ 1.09,\ 1.13 \text{ nm},\\
&I_{\text{equi}}(\delta S) = [30.99,\ 35.21],\ [30.96,\ 35.34],\ [31.01,\ 35.30] \text{ nm},\\
&I_{\text{short}}(\delta S) = [30.96,\ 35.14],\ [30.82,\ 35.13],\ [31.10,\ 35.36] \text{ nm}.
\end{aligned}$$

For δS the equiprobabilistic and shortest coverage intervals appear to be substantially the same. For δC the shortest coverage interval appears to be shifted left

slightly relative to the equiprobabilistic coverage interval. These observations are consistent with the apparent symmetry and asymmetry of the respective distributions.

The extent of the agreement of the above results across the three sets of trials provides an indication of the stability of the results of bootstrap re-sampling for this problem. Similar considerations apply in the context of the propagation of distributions in terms of an adaptive procedure for deciding a suitable number of trials [3].

5 Coverage regions

The determination of a coverage interval in an uncertainty evaluation is a common requirement in all metrology disciplines. There would be value in having a counterpart of these results for multivariate quantities. Such quantities occur in areas such as optical, electrical, acoustical and length metrology. Distribution-free coverage regions are given for the univariate case and the (correlated) multivariate case. The results constitute generalizations of Chebyshev's inequality, which applies to the case of a single output quantity (appendix A.1). They are compared with the corresponding coverage regions under a Gaussian assumption.

In the case of a single quantity, the GUM approach is to assign a Gaussian distribution to the value of the output quantity, having an expectation equal to the estimate of the value of the quantity and standard deviation equal to the standard uncertainty associated with the estimate, and using this distribution to provide a coverage interval for the value of the quantity. Appeal to the Central Limit theorem is made in the GUM to motivate this approach.

For a multivariate quantity, the natural extension of the GUM approach would be to assign a multivariate Gaussian distribution to the values of the quantity, with expectation equal to the estimates of those quantities and covariance equal to the uncertainty matrix associated with those estimates. Coverage regions would then be given by ellipsoids centred on the estimates whose semi-axes and orientations are derived from the uncertainty matrix. The results provided here are a non-parametric counterpart of this approach.

The estimate(s) and the associated standard uncertainty (uncertainty matrix) would have been provided by an application of the law of propagation of uncertainty or some other approach. Here, this same information is used to provide a distribution-free coverage interval (region).

For a single output quantity Y, an estimate y of the value of Y and the standard uncertainty $u(y)$ associated with y are assumed to be available.

For a multivariate output quantity[16] $\boldsymbol{Y} = (Y_1, \ldots, Y_M)^{\mathrm{T}}$, estimates y_i of the values of the Y_i and the uncertainty matrix $\boldsymbol{V}$ associated with the estimates are assumed to be available. $\boldsymbol{V}$ has diagonal elements equal to the variances (squared standard uncertainties) $u^2(y_i)$ and off-diagonal elements equal to the covariances $u(y_i,\ y_j)$ associated with the y_i. If the off-diagonal elements are all zero, $\boldsymbol{Y}$ is an uncorrelated multivariate output quantity; otherwise it is a correlated multivariate output quantity.

Let $100p\ \%$ be the required coverage probability (so $p = 0.95$ for 95 % coverage, e.g.).

The results, with derivations for the multivariate case in appendix A, are as follows. In each case the coverage interval or coverage region is stated in terms of a coverage factor k. The manner in which k is obtained is then given.

Coverage interval for single output quantity. The interval $0 \pm k$, or $[-k,\ k]$, contains $100p\ \%$ (under a Gaussian assumption) or at least $100p\ \%$ (in the distribution-free case) of the distribution for the value of (the standardized quantity) $(Y - y)/u(y)$.

Equivalently, in terms of the original quantity Y, the interval $y \pm ku(y)$, or $[y - ku(y),\ y + ku(y)]$, contains $100p\ \%$ (under a Gaussian assumption) or at least $100p\ \%$ (in the distribution-free case) of the distribution for the value of Y.

Coverage factor under Gaussian assumption. The value of k is obtained from the normal probability integral. Specifically, the equation

$$\frac{1}{(2\pi)^{1/2}} \int_{-k}^{k} e^{-Z^2/2} dZ = p, \tag{6}$$

the left-hand side of which is the normal probability integral, is solved for k. Tables and software are widely available for obtaining k.

Distribution-free coverage factor. The value of k is obtained from Chebyshev's inequality. Specifically, the equation

$$1 - \frac{1}{k^2} = p.$$

is solved for k. It has the solution

$$k = \left(\frac{1}{1-p}\right)^{1/2}. \tag{7}$$

[16] Locally to this section and appendix A, M is used to denote the number of output quantities for consistency with material elsewhere on the multivariate case. In previous sections it denoted the number of bootstrap samples or Monte Carlo trials.

Example. Calibration of gauge block. Example H.1 in the GUM [4, p67] gives a model for the length L at 20 °C of a gauge block (there termed end gauge). The result of an uncertainty evaluation is the estimate $\ell = 50.000\,838$ mm of the value of L and an associated standard uncertainty $u(\ell) = 0.000\,032$ mm.

For a coverage probability of $p = 0.95$, under a Gaussian assumption, the coverage factor from expression (6) is $k = 1.96$. Then, since $ku(y) = 1.96 \times 0.000\,032 = 0.000\,063$ mm, the resulting coverage interval is $50.000\,838 \pm 0.000\,063$ mm or, equivalently, $[50.000\,775,\ 50.000\,901]$ mm.

For a distribution-free coverage interval, the coverage factor is given by formula (7), viz., $k = (1/(1 - 0.95))^{1/2} = 4.47$. Since $ku(y) = 4.47 \times 0.000\,032 = 0.000\,143$ mm, the resulting coverage interval is $50.000\,838 \pm 0.000\,0143$ mm or, equivalently, $[50.000\,695,\ 50.000\,981]$ mm.

Coverage region for uncorrelated multivariate output quantities. The sphere having radius k and centred at the origin contains $100p$ % (under a Gaussian assumption) or at least $100p$ % (in the distribution-free case) of the distribution for the values of (the standardized quantities) $Z_1, \ldots, Z_M$, where $Z_i = (Y_i - y_i)/u(y_i)$.

Equivalently, in terms of the original quantity $\boldsymbol{Y}$, the ellipsoid with semi-axes $ku(y_1), \ldots, ku(y_M)$ centred at $(y_1, \ldots, y_M)$ contains $100p$ % (under a Gaussian assumption) or at least $100p$ % (in the distribution-free case) of the distribution for the values of $Y_1, \ldots, Y_M$.

Coverage factor under Gaussian assumption. The value of k satisfies

$$\int_0^k g(\zeta)d\zeta = p, \tag{8}$$

and $g(\zeta)$ is the PDF for the χ^2 distribution with M degrees of freedom, given by

$$g(\zeta) = \frac{1}{2^{M/2}\Gamma(M/2)}\zeta^{M/2-1}e^{-\zeta/2}, \qquad \zeta > 0. \tag{9}$$

Tables and software are widely available for obtaining k from expression (8) given p.

An important case is the bivariate case ($M = 2$). Then, since $\Gamma(M/2) = \Gamma(1) = 1$, the PDF (9) reduces to

$$g(\zeta) = \frac{1}{2}e^{-\zeta/2}.$$

Consequently, expression (8) becomes

$$\int_0^k \frac{1}{2}e^{-\zeta/2}d\zeta = \left[-e^{-\zeta/2}\right]_0^k = 1 - e^{-k/2} = p,$$

from which

$$k = -2\log_e(1 - p). \tag{10}$$

Distribution-free coverage factor. The value of k is given by

$$k = \left(\frac{M}{1-p}\right)^{1/2}. \tag{11}$$

Coverage region for correlated multivariate output quantities. The sphere having radius k and centred at the origin contains at least $100p$ % of the distribution for the values of (the transformed quantities) $\boldsymbol{Z} = (Z_1, \ldots, Z_M)^{\mathrm{T}} = \boldsymbol{V}^{-1/2}\boldsymbol{Y}$. The transformation $\boldsymbol{V}^{-1/2}$ converts the correlated multivariate output quantity $\boldsymbol{Y}$ into uncorrelated standardized multivariate output quantities Z_i, $i = 1, \ldots, M$.[17]

Coverage factor under Gaussian assumption. The material above concerning the χ^2 distribution with M degrees of freedom applies generally, and the result (10) specifically in the two-dimensional case.

Distribution-free coverage factor. The result (11) for uncorrelated multivariate output quantities applies.

An application of these results to circuit element measurements is given in section 6.

6 Circuit element measurement

To demonstrate the application of the results of section 5, consider the measurement problem in Annex H.2 of the GUM [4]. Using a notation local to this section, in this problem the resistance R and reactance X of a circuit element are determined by measuring simultaneously the amplitude V of a sinusoidally-alternating potential difference across its terminals, the amplitude I of the alternating current passing through it, and the phase-shift angle ϕ of the alternating potential difference relative to the alternating current.

Table 1 gives the output quantities $Y_1 \equiv R$ and $Y_2 \equiv X$ for the bivariate model $R = V\cos\phi/I$, $X = V\sin\phi/I$.[18]

The uncertainty matrix associated with y_1 and y_2 is

$$V = \begin{bmatrix} u(y_1) & 0 \\ 0 & u(y_2) \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} u(y_1) & 0 \\ 0 & u(y_2) \end{bmatrix} = \begin{bmatrix} 0.005\,0 & -0.012\,3 \\ -0.012\,3 & 0.087\,0 \end{bmatrix}.$$

[17] Each component Z_i of $\boldsymbol{Z}$ cannot be expressed only in terms of the corresponding component Y_i of $\boldsymbol{Y}$, as was possible for uncorrelated multivariate output quantities.

[18] The GUM also considers a further output quantity $Z = V/I$, but since this is related to the other output quantities R and X by $Z^2 = R^2 + X^2$, consideration of Z is not given here. In fact, it can be addressed by regarding $Z = (R^2 + X^2)^{1/2}$ as a *further* model that can be treated following the evaluation of the uncertainties associated with R and X.

Output quantity	$Y_1 \equiv R\,/\Omega$	$Y_2 \equiv X\,/\Omega$
Estimate	$y_1 = 127.732$	$y_2 = 219.847$
Standard uncertainty	$u(y_1) = 0.071$	$u(y_2) = 0.295$
Correlation coefficient /1	$\rho = r(y_1, y_2) = -0.588$	

Table 1: Estimates, associated standard uncertainties and correlation coefficient for the values of the output quantities for the circuit element application.

Application of the results in section 5 for multivariate output quantities and a coverage probability $p = 0.95$, gives $k = 2.45$ for a Gaussian PDF assigned to the values of R and X and $k = 6.32$ in the distribution-free case. Figure 12 shows the two elliptical coverage regions obtained.[19]

The price to be paid for a distribution-free coverage region is evident. The normality assumption can, however, be a strong assumption to make, especially for highly asymmetric distributions [8].

Account has not been taken in this example of the use of the effective degrees of freedom associated with the standard uncertainties [4, p61]. Indeed, the GUM does not provide relevant guidance in the case, as here, for correlated estimates of the values of the output quantities. Taking such account would be expected to have the effect of inflating the coverage factors both for the Gaussian case [4, Clause 6] and the distribution-free case (section A).

7 Concluding remarks

Non-parametric methods of statistical analysis have value in metrology in situations where information concerning the distributions for the values of the quantities involved are unavailable. So-called re-sampling methods make intensive re-use of the available data in order to infer approximate distributional knowledge from which estimates and associated standard uncertainties for the values of the quantities can be inferred. Coverage intervals or coverage regions for the values of the quantities can also be obtained.

This report has described some approaches, which cover two types of bootstrap

[19] These ellipses were plotted by (a) defining a (large) set of points lying on the unit circle centred at the origin in the $Z_1 - Z_2$ plane, (b) transforming these points to the $Y_1 - Y_2$ plane using the inverse of expression (12), after apply a scaling factor of k, viz.,

$$\boldsymbol{Y} = \boldsymbol{y} + k\boldsymbol{V}^{1/2}\boldsymbol{Z},$$

and (c) plotting the resulting points joined by straight-line segments to represent the coverage ellipse corresponding to coverage factor k.

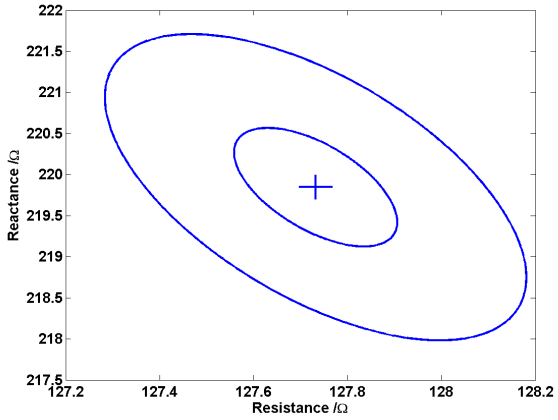

Figure 12: Elliptical coverage regions corresponding to a 95 % coverage probability for the resistance-reactance example. The inner ellipse is that under a Gaussian assumption and the outer corresponds to the distribution-free approach. The cross marks the centre of the ellipse, given by the estimates of the values of the output quantities.

re-sampling, one for 'conventional' problems, and one for hierarchical data that may contain mutual dependencies. It has also considered the determination of distribution-free coverage regions given only estimates of a set of quantities and the associated standard uncertainties and possibly covariances.

Three examples to illustrate the approaches have been given. They cover air quality, roundness assessment and circuit elements.

Acknowledgements

The work described here was supported by the National Measurement System Directorate of the UK Department of Trade and Industry as part of its NMS Software Support for Metrology programme. Part of the work was carried out in collaboration with the NMS Valid Analytical Measurement programme (air quality application) and part with the NMS Length Metrology programme (roundness assessment application). Appendix A greatly benefited from collaboration with Gaetano Iuculano and Gabriella Pelligrini at the University of Florence. Sarantis Kamvissis gave valuable statistical advice. Martyn Byng (NAG Ltd.) provided extremely helpful review comments.

References

[1] G. T. Anthony and M. G. Cox. Reliable algorithms for roundness assessment according to BS3730. In M. G. Cox and G. N. Peggs, editors, *Software for Co-ordinate Measuring Machines*, pages 30–37, Teddington, UK, 1986. National Physical Laboratory.

[2] 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.

[3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement. Supplement 1: Numerical Methods for the Propagation of Probability Distributions. Technical report, Joint Committee for Guides in Metrology, 2004. Draft.

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

[5] M. R. Chernick. *Bootstrap Methods. A Practioner's guide*. Wiley, New York, 1999.

[6] P. Ciarlini. Bootstrap algorithms and applications. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology III*, pages 12–23, Singapore, 1997. World Scientific.

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

[8] M. G. Cox, T. J. Esward, P. M. Harris, N. McCormick, N. M. Ridler, M. J. Salter, and J. P. R. B. Walton. Visualization of uncertainty associated with classes of electrical measurement. Technical Report CMSC 44/04, National Physical Laboratory, Teddington, UK, 2004.

[9] M. G. Cox, A. B. Forbes, J. Flowers, and P. M. Harris. Least squares adjustment in the presence of discrepant data. In Patrizia Ciarlini, M. G. Cox, F. Pavese, D. Richter, and G. B. Rossi, editors, *International Conference on Advanced Mathematical and Computational Tools in Metrology VI, Torino, September, 2003*. In press.

[10] M. G. Cox, A. B. Forbes, P. M. Harris, and G. N. Peggs. Experimental design in determining the parametric errors of CMMs. In V. Chiles and D. Jenkinson, editors, *Laser Metrology and Machine Performance IV*, pages 13–22, Southampton, 1999. WIT Press.

[11] M. G. Cox and P. M. Harris. Best Practice Guide No. 6. Uncertainty evaluation. Technical report, National Physical Laboratory, Teddington, UK, 2004.

[12] M. G. Cox and P. M. Harris. Statistical error modelling. Technical Report CMSC 45/04, National Physical Laboratory, Teddington, UK, 2004.

[13] M. G. Cox, P. M. Harris, M. J. T. Milton, and P. T. Woods. A method for evaluating trends in ozone-concentration data and its application to data from the UK Rural Ozone Monitoring Network. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology III*, pages 171–177, Singapore, 1997. World Scientific.

[14] M. G. Cox, P. M. Harris, M. J. T. Milton, and P. T. Woods. Method for evaluating trends in ozone concentration data and its application to data from the UK rural ozone monitoring network. Technical Report CMSC 15/02, National Physical Laboratory, Teddington, UK, 2002.

[15] M. G. Cox, P. M. Harris, and B. R-.L. Siebert. Evaluation of measurement uncertainty based on the propagation of distributions using Monte Carlo integration. Izmeritelnaya Technika, 2003. In Press.

[16] M. G. Cox, G. Iuculano, Annarita Lazzari, and Gabriella Pellegrini. Distribution-free bound for the level of confidence of a measurement process. In *Proceedings of UNCERT 2003, International Conference on the Uncertainty of Measurement*, Oxford, UK, 2003. National Physical Laboratory.

[17] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1982.

[18] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability 57. Chapman and Hall, New York, 1993.

[19] R. A. Eldred and T. A. Cahill. Trends in elemental concentrations of fine particles at remote sites in the United States of America. *Atmospheric Environment*, 28:1009–1019, 1994.

[20] A. B. Forbes. Efficient estimators in data fusion. In P. Ciarlini, M. G. Cox, E. Filipe, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology V. Series on Advances in Mathematics for Applied Sciences Vol. 57*, pages 163–168, Singapore, 2001. World Scientific.

[21] A. B. Forbes. Efficient algorithms for structured self-calibration problems. In J. Levesley, I. Anderson, and J. C. Mason, editors, *Algorithms for Approximation IV*, pages 146–153, Huddersfield, UK, 2002. University of Huddersfield.

[22] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Charles Griffin, London, 1968.

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

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

[25] S. J. Oltmans. Trends in ozone in the troposphere. *Geophysical Research Letters*, 25:139–142, 1998.

[26] PORG. Ozone in the United Kingdom 1992. Third report of the United Kingdom Photochemical Oxidants Review Group. Technical report, Harwell Laboratory, Oxfordshire, 1993.

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

[28] J. R. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, Ca., USA, second edition, 1995.

[29] J. G. Saw, M. C. K. Tang, and T. C. Mo. Chebychev's inequality with estimated mean and variance. *American Statistician*, 38:130–132, 1984.

[30] W. W. S. Wei. *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley, 1990.

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

A Expressions for distribution-free coverage regions

This appendix provides the basis for the statements made in section 5.

A.1 Single quantities

Suppose it is required to provide a coverage interval for the value of the quantity Y corresponding to a coverage probability of 95 %, and that nothing is known or assumed about the distribution.

> The coverage interval $y \pm ku(y)$, where $k = 4.47$, contains at least 95 % of the distribution for the value of Y.

This result is derived from *Chebyshev's inequality* [28, p125], which states that the probability that the value of Y lies in the interval $y \pm ku(y)$ is at least $1 - k^{-2}$. The value of k for which $1 - k^{-2} = 0.95$ is 4.47. By its nature it cannot be as sharp as an interval derived from knowledge of the PDF for the value of Y [11]. For instance, if the distribution for the value of Y is Gaussian, this interval is $y \pm 1.96u(y)$. The length of the coverage interval derived from Chebyshev's inequality is 2.3 times that for a Gaussian distribution for the value of Y.

These results apply if the effective degrees of freedom associated with $u(y)$ is infinite, or in practice large. Otherwise, the k-factor becomes inflated, as in the case of the t–distribution [29].

If it is reasonable to regard the distribution as *symmetric* and *unimodal* (single-peaked), tighter results based on *Gauss's inequality* are possible. The coverage interval $y \pm ku(y)$, where $k = 2.98$, contains at least 95 % of the distribution for the value of Y. Gauss's inequality states that the probability that the value of Y lies in the interval $y \pm ku(y)$ is at least $1 - \frac{4}{9}k^{-2}$. The value of k for which $1 - \frac{4}{9}k^{-2} = 0.95$ is 2.98. This interval is only approximately 50 % longer than that when the distribution for the value of Y is taken as Gaussian. Again these results apply only if the degrees of freedom is infinite, or in practice large.

A.2 Multivariate quantities

Consider quantities $\boldsymbol{Y} = (Y_1, \ldots, Y_M)^{\mathrm{T}}$, treated as random variables. Suppose $\boldsymbol{y} = (y_1, \ldots, y_M)^{\mathrm{T}}$ are estimates of $\boldsymbol{Y}$ and the covariance matrix associated with $\boldsymbol{Y}$ is $\boldsymbol{V}$. Regard $\boldsymbol{y}$ as the expectation of $\boldsymbol{Y}$ and $\boldsymbol{V}$ as the covariance of $\boldsymbol{Y}$.

Introduce standardized variables (with zero expectation and unit variance):

$$\boldsymbol{Z} = (Z_1, \ldots, Z_M)^{\mathrm{T}},$$

where[20]

$$\boldsymbol{Z} = \boldsymbol{V}^{-1/2}(\boldsymbol{Y} - \boldsymbol{y}). \tag{12}$$

$\boldsymbol{V}$ can be expressed as

$$\boldsymbol{V} = \boldsymbol{U}\boldsymbol{R}\boldsymbol{U},$$

where $\boldsymbol{U}$ is the diagonal matrix with the standard uncertainties $u(y_i)$ on its diagonal and $\boldsymbol{R}$ is the correlation matrix associated with $\boldsymbol{y}$.

The multivariate Gaussian distribution has constant density on ellipses or ellipsoids of the form

$$(\boldsymbol{Y} - \boldsymbol{y})\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{y}) = c^2,$$

where c is a constant [24, p38]

The square of the distance of a *point* $\boldsymbol{Z}$ from the origin, viz.,

$$\|\boldsymbol{Z}\|_2^2 = Z_1^2 + \cdots + Z_M^2,$$

is the sum of squares of M independent Gaussian variates and therefore distributed as chi-squared with M degrees of freedom. The function

$$Z_1^2 + \cdots + Z_M^2 = c^2$$

describes a circle or sphere of radius c. The probability of a value lying in the above ellipse or ellipsoid is identical to its lying in this circle or sphere. Hence the use of a value of chi-squared provides a coverage region in the form of an ellipsoid corresponding to a prescribed coverage probability p.

Now consider the derivation of a distribution-free coverage region (following recent work [16]).

Define the distance function

$$d(\boldsymbol{Z}) = \frac{\|\boldsymbol{Z}\|_2}{M^{1/2}} = \left(\frac{1}{M}\sum_{i=1}^{M} Z_i^2\right)^{1/2} \tag{13}$$

and the region

$$\Omega_k = \{\boldsymbol{Z} : \ d(\boldsymbol{Z}) < k\},$$

[20] $\boldsymbol{V}^{-1/2}$ can be formed numerically as follows. Compute the eigendecomposition

$$\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}},$$

where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements equal to the eigenvalues of $\boldsymbol{V}$ and $\boldsymbol{Q}$ is an orthogonal matrix ($\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$) containing the corresponding eigenvectors. Then, as can immediately be verified,

$$\boldsymbol{V}^{-1/2} = \boldsymbol{Q}\boldsymbol{D}^{-1/2}\boldsymbol{Q}^{\mathrm{T}}.$$

viz., the interior of the sphere in R^M of radius $M^{1/2}k$ centred at the origin.

$d(\boldsymbol{Z})$ is itself a random variable, being a function of the standardized random variables Z_i. The coverage probability for the region Ω_k is

$$P\{d(\boldsymbol{Z}) < k)\} \equiv \int_{\Omega_k} g(\boldsymbol{\zeta})d\boldsymbol{\zeta} = 1 - \int_{\widetilde{\Omega}_k} g(\boldsymbol{\zeta})d\boldsymbol{\zeta},$$

where $g(\boldsymbol{\zeta})$ is the PDF for the values of $\boldsymbol{Z}$ and $\widetilde{\Omega}_k = \mathrm{R}^M \setminus \Omega_k$.

For any non-negative function $\alpha(t)$, denoting expectation by E,

$$\mathrm{E}\{\alpha(d(\boldsymbol{Z}))\} = \int_{\mathrm{R}^M} \alpha(d(\boldsymbol{\zeta}))g(\boldsymbol{\zeta})d\boldsymbol{\zeta}.$$

When α is non-negative and non-decreasing,

$$\mathrm{E}\{\alpha(d(\boldsymbol{Z}))\} \geq \int_{\widetilde{\Omega}_k} \alpha(d(\boldsymbol{\zeta}))g(\boldsymbol{\zeta})d\boldsymbol{\zeta} \geq \alpha(k) \int_{\widetilde{\Omega}_k} g(\boldsymbol{\zeta})d\boldsymbol{\zeta}.$$

Thus,

$$P\{d(\boldsymbol{Z}) < k\} > 1 - \mathrm{E}\{\alpha(d(\boldsymbol{Z}))\}/\alpha(k), \tag{14}$$

a distribution-free lower bound for the level of confidence for a coverage region for a multivariate model. By taking $\alpha(t) = t^2$, inequality (14) becomes

$$P\{d(\boldsymbol{Z}) < k\} > 1 - \mathrm{E}\{d^2(\boldsymbol{Z})\}/k^2. \tag{15}$$

Inequality (15) constitutes a generalization to any number of quantities of the Chebyshev inequality for the univariate case. A distribution-free lower bound for the coverage probability for a region is now determined in two specific cases from this result.

Now, using expression (12) and the distance function (13),

$$\mathrm{E}\{d^2(\boldsymbol{Z})\} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{E}\{(Y_i - y_i)^2\}/u^2(y_i) = 1,$$

and hence inequality (15) gives

$$P\{d(\boldsymbol{Z}) < k\} > 1 - \frac{1}{k^2}, \tag{16}$$

a result that is identical to that for a single quantity.

As in the one-dimensional case, this result can be 'inverted'. Let p be the required coverage probability. Then, setting

$$1 - \frac{1}{k^2} = p,$$

it follows that

$$k = \left(\frac{1}{1-p}\right)^{1/2}.$$

Inequality (16) can be expressed as

$$P\left\{Z_1^2 + \cdots + Z_M^2 \leq \left(\frac{M}{1-p}\right)^{1/2}\right\} > p.$$

Thus, the sphere of radius $(M/(1-p))^{1/2}$ centred at the origin is a coverage region for the values of the standardized quantities corresponding to a coverage probability of at least p.