UNCERTAINTY ANALYSIS IN EXPERIMENTAL MECHANICS

Alcir de Faro Orlando
Mechanical Engineering and Metrology Departments, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil

Keywords: Uncertainty Analysis, Data Analysis, Laboratory Testing, Measuring Instrument Calibration, Metrological Reliability, Testing Data Interpretation, Test Planning.

Contents

Summary

In this text the nature of experimental errors is discussed. Initially, the construction of a

histogram is described and detailed for analyzing the distribution of errors. Their modeling is analyzed with available statistical distributions. Besides discussing conceptually how they fit the measurement errors, a test of goodness of fit is described to test the adherence of the experimental distribution to each one. Then, the uncertainty analysis procedure is described to analyze the experimental data and to report the result of the measurement. As a first requirement for a reliable analysis, a procedure is described to eliminate outliers from the data bank. Single and multiple input quantity measurement procedures are analyzed aiming a rational data taking and reliable result of measurement. On the basis of the described procedures, a methodology is detailed for calibrating measuring systems and interpolation of results. Finally, the measurement of material properties is analyzed as an application of the uncertainty analysis procedure. Several numerical examples are included in the text, with analysis of the results.

1. Introduction

1.1. The Measurement Procedure

Quantity can be defined as a property of a phenomenon, body, or substance, to which a magnitude can be assigned. The realization of the definition of a given quantity, with a stated value and measurement uncertainty, and used as reference, is called measurement standard. The measurand is the quantity intended to be measured.
The process of measurement is a result of the need of quantifying the physical phenomena. It can be basically defined as the transfer of information from the source system through a more convenient measuring system, thus resulting in an observed quantity not readily available by the former one.

Unfortunately, however, the transfer of information always changes the behavior of the source system. Therefore, the measurement science tries to minimize this interaction, or, at least, to estimate the difference between the measured and the source values, by quantifying the contribution of every parameter that takes part in it. The undisturbed value of the source system can thus be estimated by the measuring system, and is called the true value the quantity, which is consistent with its definition

The measurement procedure consists of five sequential steps, so that the true value can be accurately estimated.

- *Identification of the physical quantity to be measured*. It is important to understand the physical phenomenon on which the measurement is made, so that the interaction between the source and measuring systems be under control.
- *The transducer*. It is considered the interface between the source and measuring systems. Its function is to transform the physical quantity to be measured, as existing in one form of energy, into some other physical quantity more easily measured. The new physical quantity may be in the same form of energy as the original phenomenon, or it may be in a different form. The better the knowledge of the properties of the materials used as transducers, the more accurate will be the transfer of information. Because of the fact that they are not completely known, it is only possible to specify a range around the measured value to include the true value within a desired confidence level, thus resulting in an uncertainty band of measurement.

- *Information, Acquisition and Transmission*. Information contained in the energy transfer procedure may be carried on any of the properties of the energy function, such us amplitude, frequency and phase. The acquisition of information must be made in such a manner as to alter the phenomenon to be observed in the least possible manner. The transmission of information must be made in such a manner as to alter the information to be transmitted in the least possible manner.
- *Data analysis*. Data are no good to anyone unless they are analyzed and evaluated in the light of the engineering problem to be solved. The use of statistics for data analysis results in determining the reliability of the measured value, or, in other words, an interval around the measured value to include the true value within a desired confidence level. Interpolation of results can be done by a curve fitting procedure.
- *Interpretation, Control and Feedback*. The interpretation of the properly acquired, transmitted and analyzed data can be used for redesigning a system, or even to check if the performance of a system in operation has been drifted, as compared to the designed value. Also, because of the fact that measurement is a time consuming procedure, the validation of a theory by a measurement procedure is an important tool for an efficient design purposes.

1.2. Metrological Characteristics of a Measuring System

The International Vocabulary of Basic and General Terms in Metrology, called ISO VIM (2004), is used through out this text and has the potential to become a document to which reference may be made in national regulations. The Guide to the Expression of Uncertainty in Measurement, called ISO GUM (1995), is also used through out this text, and was prepared by several organizations to provide rules on the expression of measurement uncertainty for use within standardization, calibration, laboratory accreditation and metrology services.

According to them, the difference between the measured value of the measurand and its true value is called *error of measurement*, and is most of the time different from zero. A well established rule in the scientific investigation says that whenever an experiment is made for the first time the results are much different from what they should be. By gradually refining the experimental technique and measurement method, they asymptotically approach what it can be considered a reliable description of the phenomenon. However, because of the fact that the true value is seldom known, the *error of measurement* can be calculated using the so called conventional value of a quantity, which is attributed by formal agreement to a quantity for a given purpose. It can also be considered as the value indicated by a measurement standard, whenever it is much more accurate than the measuring system.

There are basically three types of errors. The first one can be easily recognized and eliminated. It comes mostly from wrong calculations and measurements. A statistical analysis of the results can show that the data points do not belong to the same population, and, therefore, they should be discarded off. Chauvenet and Grubbs criterion are examples of statistical analysis procedures.

The second one is called systematic error, and cannot be easily detected. The statistical analysis is not normally useful. It comes from the fact that the measurement and calibration conditions are different, causing what is called installation effects. *Calibration* is defined as the operation establishing the relation between quantity values provided by measurement standards and the corresponding indications of a measuring system, carried out under specified conditions and including evaluation of measurement uncertainty. As a result of this comparison, a *Correction* must be applied to the indication of the measuring system, to compensate for systematic effects. It can take two different forms, such as an addend or a multiplicative factor. The misinterpretation of the physical phenomenon can be a cause of this error, which can be estimated on the basis of the governing theory. However, it can be minimized by calibrating the measuring system under the same measurement conditions. Usually, only a limited number of data points is used when calibrating a measuring system. However, most possibly the measured quantity will not be the same as during calibration. An interpolation procedure has to be used to relate the indication of the measuring system with the quantity value. Curves, like polynomials, are usually fitted to the experimental data by the least square method, resulting in the average performance of the measuring system in the range of operation. The choice of the type of curve must be based on the how close the performance of the measuring system is reproduced in the range of operation.
The third one is related to the lack of information about the performance of the measuring system, or even about the controlling conditions of the experiment. It is very difficult to be identified, and its modeling is usually done using probability distribution models. That is why they are called random errors. In the absence of further information, the Gaussian or normal distribution has been successfully utilized to model real error distributions, although lacking physical meaning in its upper and lower ends, where the measurand is supposed to go, unrealistic and respectively, to $\pm\ \infty$. Rectangular and triangular probability distributions have been also used, to account for the fact that measurement errors are smaller than a certain value, called *maximum measurement errors*, and their distribution is not completely known. Any other error distribution can be used provide that information is available. The importance of the misfit and the overall adherence to the real distribution can be usually quantified by the chi-square test of goodness of fit.

Uncertainty of measurement is a parameter that characterizes the dispersion of the quantity values that are being attributed to a measurand, based on the available information. The dispersion is due to a definitional uncertainty of the measurand, and random and systematic effects in the measurement. Because of the fact that the error of measurement is seldom known, uncertainty is a more realistic approach to quantifying the physical phenomenon and defines an interval around the measured value to include the true value within a desired confidence level. The evaluation method of the uncertainty of measurement is called Type A procedure when statistical methods are used for estimating the interval. Otherwise, it is called Type B procedure, which is based on *a priori* assumed distribution.

Standard uncertainty is often defined as the uncertainty of the result of a measurement expressed as a standard deviation.

Expanded uncertainty is a quantity defining an interval about the result of a measurement that be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

Coverage factor is a numerical factor used as a multiplier of the standard uncertainty in order to obtain an expanded uncertainty.

When reading the indication of a displaying device of a measuring system, the interpolation uncertainty is a source of errors. It arises from the inability of human observers to assess without uncertainty the numerical value to be assigned to the position of a pointer located between two successive marks on a scale of the displaying device, which consists of an ordered set of marks, together with any associated numbers or quantity values. In principle, the smallest scale subdivision for interpolation purposes is the *resolution* of the measuring system, which is defined as the smallest change in the value of a quantity being measured by a measuring system that causes a perceptible change in the corresponding indication. The random fluctuations of the reading, however, can be larger than the resolution. That is why, in practice, one prefers to have larger scale subdivisions. If the width of the interval between two successive marks on the scale is chosen as the smallest difference between indications of the displaying device, all positions of the pointer in the interval with the same width, having each mark as the center, are read to be the same and equal to the quantity indication of the mark. A better resolution can frequently be achieved if half the width of the interval between two successive marks on the scale is chosen to be the smallest difference between indications of the displaying device. For a digital displaying device, it is equal to the variation of its indication when the least significant digit varies one unit. Anyway, this is the smallest uncertainty which will ever be associated with a particular measuring system, and is called uncertainty of the displaying device.

Because of the nature of the measuring system and the knowledge of its performance, whenever replicated measurements of the same quantity under repeatability conditions are taken, different values will be indicated by the measuring system. Their dispersion can be a measure of the property of the measuring system called *repeatability*. Its numerical value is closely related to the uncertainty of the displaying device. Thus, if the smallest difference between indications of the measuring system is chosen to be much higher than its error, the dispersion of replicated measurements can be numerically equal to zero. On the other hand, if it is chosen to be of the same order of magnitude of the resolution of the measuring system, the dispersion of the indications can be numerically much more representative of the repeatability of the results of measurement. The right procedure is to combine the uncertainty of the displaying device with the dispersion of replicated measurements, so that a better estimate of the repeatability could be done. Experience has shown that if half the width of the interval between two successive marks on the scale is chosen to be the smallest difference between the indications of the displaying device, a good compromise between good estimate of repeatability and time required for measurements is obtained.

Repeatability is closely related to uncertainty of measurement if systematic errors are minimized by calibration of the measuring system. In other words, it approximately represents an interval around a measured value in which it is believed the true value lays. The theory shows that when the number of replicated measurements increases, there is a reduction of the difference between the true value and its estimate, as defined by their arithmetic mean, down to a lower limit defined by the uncertainty of the displaying device. Below this point, it is useless to increase the number of replicated measurements. In principle, the same uncertainty of measurement can be achieved by two different measuring systems, if the number of replicated measurements can be varied. Unfortunately, in practice, only one measurement can be made to quantify a physical phenomenon. Thus, if small uncertainties of measurements are required, repeatable measuring systems must be chosen.

The measurement of a quantity can be made by different measurement procedures, measuring systems, operators and at different locations. The results can be different because the errors of measurement are not the same. The agreement between different results of measurement can be indicated by a property of the measuring system called *Reproducibility*. Usually, the result of measurement can be summarized by a single quantity value and its measurement uncertainty. A statistical procedure called hypothesis test can be used to show if the results are the same to within a given significance level. In order to check if the errors of measurement are under control, experimentalists use different measurement procedures and compare the results of measurement.

*Sensitivit*y is a property of a measuring system that relates the change of its indication to the corresponding change in the value of the quantity being measured. It can be calculated if a relationship between the quantity being measured and the parameters of the measuring system is available. Or, it can be measured by varying the parameters of the measuring system during an experiment. Sensitivity is an important property of the measuring system used to estimate the uncertainty of measurement of the quantity, and is also used for experiment planning purpose.

Sometimes, mainly in mechanical systems, the response of a measuring system is different if the measurand is being continuously increased or reduced, because the transducer materials react differently to the type of efforts they are submitted to. If the direction of the continuous variation of the effort can be known a priori, two curves can be used to model its behavior and reduce the uncertainty of measurement. Otherwise, a single curve is used. This phenomenon is called *hysteresis*.

Finally, the knowledge of these properties is useful to plan an experiment, or to control the uncertainty of measurement of a measuring system.

1.3. Objectives

The objectives of this paper are to present different methodologies to analyze the experimental data, to estimate its uncertainty of measurement and to supply information so that the data taking procedure is under control.

2. Modeling the Measurement

2.1. Histogram

Modeling the measurement can be made by ordering the data of n replicated measurements of a given measurand and counting the number of events k_i that belong to the same ith interval of width Δ, or a bin size, calculated as the ratio between the data range T and the number of intervals N.

$$T = x_{\max} - x_{\min} \qquad (1)$$

$$\Delta = \frac{T}{N} \qquad (2)$$

A bar-graph histogram can thus be constructed to represent the shape of the distribution. In statistics, a histogram is a graphical display of tabulated frequencies. The bin size is a compromise between sample error and the resolution. If a too small bin size is chosen, the bar height of each bin suffers from a significantly large fluctuation due to the paucity of data samples in each bin. If a too large bin size is chosen, the histogram cannot represent the shape of the distribution because the resolution is not good enough.

As an example, let us suppose that the length of a rod was measured $n = 436$ times by a micrometer. Its length was also measured by an optical comparator, resulting in the so called conventional true value. The error of the measurement was calculated, laying in the ($x_{\min}$ = -0.65 mm to $x_{\max} = 0.46$ mm) interval. The error interval was divided into $N = 20$ parts, and the error frequency (k_i) in each bin was counted. Figure 1 presents a bar-graph histogram for the shape of the distribution, in which the middle point (y_i) of the ith error interval was selected for representing the *x*-coordinate.

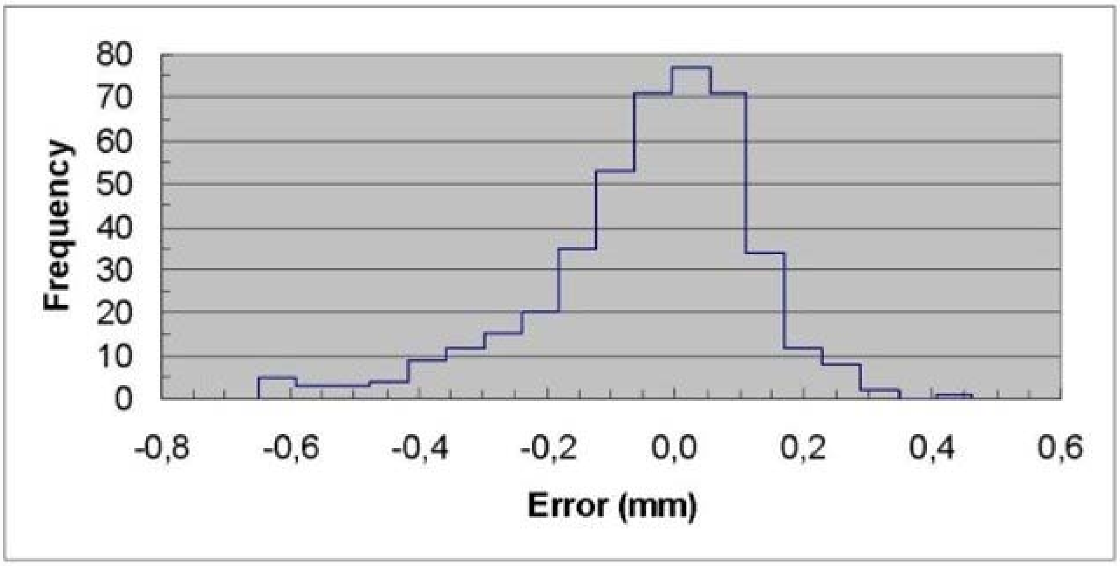

Figure 1. Frequency Distribution

The average value of the error $(\bar{x})$ and the standard deviation (s) of the distribution can be calculated by the following relationships.

$$\bar{x} = \frac{1}{n}.\sum_{i=1}^{n} x_i \tag{3}$$

$$s = .\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{4}$$

resulting in $\bar{x} = -0.10$ mm and $s = 0.17$ mm.

Several formulations are available in the literature for determining the optimum number of bins. Scott's formula has been used for calculating the bin width Δ. The number of bins is then calculated using Eq. (2) and rounding it off to next integer number.

$$\Delta = \frac{3.5s}{n^{1/3}} \tag{5}$$

Thus, $\Delta = 0.078$ mm, and N = 15.

Another method for selecting the bin size of a histogram is available in the literature, called the Shimazaki and Shinomoto's procedure (2007), and follows the steps.

- Given a number N of bins, count the number of events k_i that belong to the same i th interval of width Δ.
- Calculate the average $\bar{k}$ and the variance υ of the number of events k_i in each bin.

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i \tag{6}$$

$$\upsilon = \frac{1}{N}\sum_{i=1}^{N}\left(k_i - \bar{k}\right)^2 \tag{7}$$

- Calculate the function C_N

$$C_N = \frac{2.\bar{k} - \upsilon}{\left(n.\Delta\right)^2} \tag{8}$$

- Calculate C_N for each number of bins N, choosing the value that minimizes the function C_N

Figure 2 shows a plot of the function C_N with the number of bins (N). It can be seen that a number of bins of approximately equal to $N = 20$ minimizes the function C_N and thus can be chosen for constructing the histogram that represents the shape of the distribution.

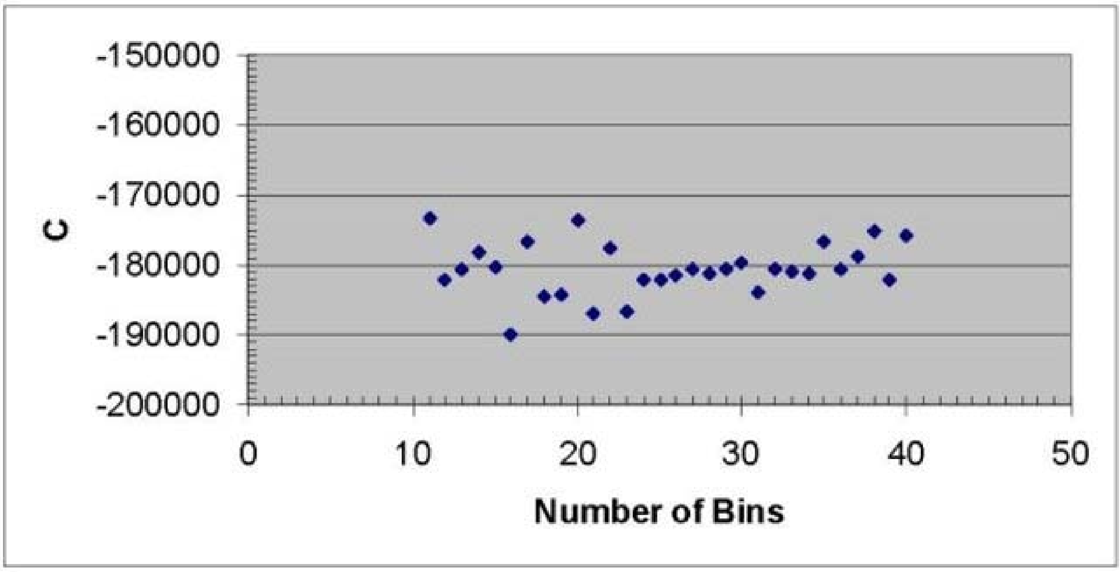

Figure 2. Optimization of the number of bins

If the number of events in each bin k_i is divided by the total number of replicated measurements n, the normalized histogram represents the so called probability density function (p), which determines the probability of getting any particular measurement in the range defined by each bin width. When the number of replicated measurements gets larger, and the bin width gets smaller, the probability density function approaches a continuous curve.

When analyzing the experimental data it is sometimes desired to know the number of measurement errors in a two standard deviation width range, having the average value $\bar{x}$ at its mid point, or in the $\bar{x} \pm s$ range. It can be done by adding the number of events of all bins that belong to the interval defined by the desired range, including the ones that contain the upper and lower limits of the range. For $N = 20$ bins, 7 bins are selected and 361 measurement errors were counted in that range, out of $n = 436$ replicated measurements, or 83.0 %. According to the statistical model, there is a 83.0 % probability that the measurement errors be in the $\bar{x} \pm s$ range. If the $\bar{x} \pm 2s$ is considered, 13 bins are selected, 419 measurement errors are counted and the probability is 96,1 %. It is interesting to observe that if the total number of bins increases, the probabilities are better estimated. For $N = 30$, the probabilities are, respectively, 78.4 % and, 96.1 %.

The probability distribution function (P) is defined by the probability that a random variable be less or equal than a given value. When the number of replicated measurements gets larger, and the bin width gets smaller, the probability distribution function approaches a continuous curve. Thus, the probability that the measurement errors be in a given interval is calculated by subtracting the values of the probability distribution function at the upper and lower limits of the interval.

-
-
-

Bibliography

Abramowitz, M. and Stegun, I.A. (1965). Handbook of Mathematical Functions. 1046 pp. Dover Publications, Inc, New York. [A great variety of functions and high accuracy of tabulation, useful for preliminary survey of problems before programming for machine operation].

Bevington, P.R. (1969). *Data Reduction and Error Analysis for the Physical Sciences.* 336 pp. McGraw-Hill Book Company, New York. [A detailed introduction to the methods used in experimental research for the reduction and error analysis, including techniques for fitting the physical data].

Grubbs, F. (1969). *Procedures for Detecting Outlying Observations in Samples.* Technometrics 11, No. 1, 1-21 pp. [The fundamentals of the methodology for detecting outliers].

Holman, J.P. (1971). *Experimental Methods for Engineers.* 423 pp.McGraw-Hill Book Company, New York. [A survey of a rather broad range of instruments and experimental techniques, with strong emphasis on problem solving and the importance of accuracy, error and uncertainty in experimental measurement].

ISO GUM (1995). *Guide to the Expression of Uncertainty in Measurement*, 101 pp. International Organization for Standardization. Switzerland. [It establishes general rules for evaluating and expressing uncertainty in measurement that are intended to be applicable to a broad spectrum of measurements].

ISO VIM (2004). *International vocabulary of basic and general terms in metrology (VIM).* 55 pp. International Organization for Standardization, Switzeland. [A vocabulary of conceptual, basic and general terms in metrology, including notes and examples for better establishing the differences among them].

ISO 5725-2 (1994). *Accuracy (trueness and precision) of measurement methods and results—Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method.* International Organization for Standardization, Switzeland. [General procedures and recommendations for metrologically characterizing a measurement].

Moffat, R.J. (1978). *Planning Experimental Programs.* 131 pp. Department of Mechanical Engineering Report, Stanford University, USA [General principles for planning and analyzing an experimental work, discussing conceptually and analytically the contribution of each parameter in the result, with several commented examples].

Shimazaki, H. and Shinomoto, S. (2007). *A Method for Selecting the Bin Size of a Time Histogram.* Neural Computation 19, 1-25 pp. [The fundamentals of the methodology for critically determining the goodness of the fit of a time histogram].

Biographical Sketch

Alcir de Faro Orlando was born in Rio de Janeiro, Brazil, December 24, 1943. He graduated in mechanical engineering from the Aeronautics Institute of Technology (ITA) in 1967, S.Paulo, Brazil. He earned a MSc degree in thermosciences from COPPE (Federal University of Rio de Janeiro), Rio de Janeiro, Brazil, in 1969, and PhD degree in thermosciences from Stanford University, USA, in 1974.

As Associated Professor, he worked for Federal University of Rio de Janeiro (Rio de Janeiro, Brazil), University of Campinas (Campinas, S.Paulo, Brazil) and is presently working for Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, in the Department of Mechanical Engineering and in the Department of Metrology since 1979. He basically teaches courses on thermodynamics, heat transfer, energy analysis, uncertainty analysis and measurement of temperature, pressure and flow rate. He has been supervising more than 50 MSc and PhD theses in metrology and experimental methods. He has been very active in projects with electric energy utility companies and with the Brazilian oil company Petrobras, with emphasis on custody transfer measurement of flow rate. He manages a laboratory in temperature and pressure, accredited by the Brazilian Calibration Network for calibrating industry measuring instruments. He has about 100 published papers in international and Brazilian conferences and journals. Before working for PUC-Rio, he worked for an engineering company for 3 years, developing solar energy projects. Presently he is a consultant for the Brazilian Government for commissioning solar energy systems.

Dr. Orlando is a member of the Brazilian Society for Mechanical Sciences (ABCM) and Brazilian Society for Metrology (SBM).