BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax.+44 (0)29 2037 1921
E-Mail Bindery@Cardiff.ac.uk

White light interferometry in measurements of micro volumes

A thesis
submitted to
Cardiff University
for the degree of

Doctor of Philosophy

by

Carlo Ferri

Manufacturing Engineering Centre
Cardiff University
United Kingdom

2006

UMI Number: U584932

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U584932
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

In memory of my father and Ines

Synopsis

A procedure for measuring micro-volumes of solids with irregular and complex boundaries of both concave and convex forms has been developed. Advantages, limitations and potential applications of the developed procedure are identified and discussed. The precision of this procedure is demonstrated in the case of a convex form. In the light of the key role of metrology in the current trend toward product miniaturisation, the need for dedicating resources and effort to assessing quantitatively the performances of measuring processes is most apparent.

The developed procedure is based on white light interferometric microscopy. In order to enable the deployment of this measuring system in the developed procedure and in agreement with the centrality of metrology mentioned above, an investigation of white light interferometric microscopy has been carried out from a user perspective. In particular, two sources of variability were identified and experimentally quantified and the precision in repeatability conditions was estimated when measuring length along the z-axis in a pre-specified micro-metric range. A critical analysis of the calibration procedure built in the investigated microscope has also been conducted. In order to overcome the pitfalls discovered in such an analysis, a spline-based calibration procedure has been developed and demonstrated. In addition, the control methods needed for the practical usage of the proposed calibration procedure have been developed. Calibration studies are made possible by the provision of traceable reference materials. Therefore, a cost-effective and versatile procedure for the building of traceable reference samples of length in the micrometric range was developed. The proposed method used standard gauge blocks commonly found in metrology laboratories.

Acknowledgements

I gratefully aknowledge the acceptance of my application for pursuing postgraduate education and the financial support of this investigation obtained via the Manufacturing Engineering Centre under the supervision of Prof. Duc Truong Pham and Prof. Stefan Dimov.

I am also grateful to Atanas Ivanov who provided me assistance in the closing stages of this investigation in finalising and presenting the results already achieved and to Christian Griffiths who patiently and swiftly proofread my manuscript.[1]

[1] This thesis was typed using solely free software: the document processor LyX and the document preparation system LaTeX 2_ε.

Contents

List of Figures

List of Tables

Notation

Abbreviations

$\mu - EDM$	Micro electro discharge machining, *first used on page* 8.
AIC	Akaike information criterion for model selection, *first used on page* 93.
ANCOVA	Analysis of covariance, *first used on page* 99.
ANOVA	Analysis of variance, *first used on page* 48.
BIC	Bayes information criterion for model selection, *first used on page* 94.
BSI	British Standard Institution, *first used on page* 17.
CAD	Computer aided design, *first used on page* 26.
CCD	Charge-coupled device, *first used on page* 13.
CUSUM	Cumulative-sum control chart, *first used on page* 132.
GSL	GNU scientific library, *first used on page* 121.
GUM	Guide to the expression of uncertainty in measurement, *first used on page* 18.
ISO	International Organisation for Standardisation, *first used on page* 17.
NIID	Normally independently distributed random variable, *first used on page* 50.
NIST	National Institute of Standards and Technology, *first used on page* 16.

NPL National Physical Laboratory, *first used on page* 16.

OPD Optical path difference, *first used on page* 13.

PD Published Document, *first used on page* 17.

pixel Picture element, *first used on page* 151.

PPM Parts per million, *first used on page* 156.

RM Reference material, *first used on page* 24.

RSS Residuals sum of squares for a generic linear model, *first used on page* 85.

SEM Scanning electron microscope, *first used on page* 11.

STL Stereolithography format, *first used on page* 151.

VIM International vocabulary of general terms in metrology, *first used on page* 17.

VSI Vertical scanning interferometry, *first used on page* 12.

WLI White light interferometry or white light interferometer, *first used on page* 8.

Greek mathematical symbols

α_{test} [*Dimensionless*] First type error in a single test of a family of m tests with overall first type error α. It is the probability of a false allarm in a single test at the t-th instant when there are m RM in the control method to be checked at the same instant t, *first used on page* 138.

α_{th} [K^{-1}] Coefficient of linear thermal expansion, *first used on page* 35.

$\alpha_{WLI}(x, y)$ [*Dimensionless*] Phase change on reflection, *first used on page* 14.

β_i [*Response variable over a fuction of the predictor*] Parameter in a regression model associated with the i-th independent variable, for i=1,...,p. For i=0, β_i is the intercept of the model, *first used on page* 85.

$\hat{\beta}$ [*Response variable over a fuction of the predictor*] Estimate of the vector of the parameters associated with the predictors in a linear model, *first used on page* 137.

Δ_i [µm] Deviation of the central length from the nominal length in the $i-th$ gauge block ($i = 1, 2 \ldots$), *first used on page* 31.

δ_{CUSUM} [*Dimensionless*] Ammount of the shift of the mean in units of standard deviations when the mean of process controlled in a CUSUM goes out-of-control, *first used on page* 133.

$\mu_{0,CUSUM}$ [*The same as the statistic controlled*] Expected value of the statistic plotted in a CUSUM chart when the process is in control, *first used on page* 133.

$\mu_{1,CUSUM}$ [*The same as the statistic controlled*] Expected value of the statistic plotted in a CUSUM chart when the process is out-of-control, *first used on page* 133.

μ_h [µm] General mean in the ANOVA for the repeatability analysis, *first used on page* 75.

μ_s [*Dimensionless*] Overall mean in the ANOVA model for the scan length. It is a parameter, *first used on page* 50.

$\mu_{T_{actual}}$ [°C] Mean of the environmental air temperature, *first used on page* 41.

$\hat{\xi}$ [*Dimensionless*] Estimate of the generic unknown parameter ξ, *first used on page* 51.

$\hat{\sigma}_{volume}$ [*unit of volume*] Estimate of the repeatability standard deviation of the procedure for measuring micro-volumes, *first used on page* 165.

σ [*The same dimension as the response variable*] Standard deviation of the random error in an ANOVA model or , more generally, in a linear model. It is a parameter, *first used on page* 50.

$\sigma_{control}$ [*The same as the response variable in calibration*] Standard deviation of the random variable $\hat{e}_{i,j,t,control}$, *first used on page* 137.

σ_{CUSUM} [*The same as the statistic controlled*] Standard deviation of the statistic plotted in a CUSUM chart, *first used on page* 132.

σ_i [*The same dimension as the response variable*] Standard deviation of the response variable given the i-th cell in a linear model, *first used on page* 81.

σ_r [μm] Repeatability standard deviation. It equals σ in the pertinent random effect ANOVA model, *first used on page* 75.

$\sigma_{T_{actual}}$ [°C] Standard deviation of the environmental air temperature, *first used on page* 41.

σ_{t_h} [μm] Standard deviation of the random effect in the ANOVA model for the repeatability analysis. It is a parameter, *first used on page* 75.

σ_{t_s} [*Dimensionless*] Standard deviation of the random effect in the ANOVA model for the scan length. It is a parameter, *first used on page* 50.

θ [*Dimensionless*] Volumetric wear ratio, *first used on page* 9.

$\chi^2(a \cdot (n-1))$ [*Dimensionless*] Chi-square probability density function with degrees of freedom equal to $(a \cdot (n-1))$, *first used on page* 76.

$\chi^2_\alpha(n)$ [*Dimensionless*] Percentage point of the chi-square probability density function with n degree of freedom that has α% of the probability to the right of $\chi^2_\alpha(n)$, *first used on page* 77.

ω_0 [m^{-1}] Spatial frequency of the intensity $g(x, y, z)$ in the z direction, *first used on page* 14.

Roman mathematical symbols

$abs()$ [*The same as the argument*] Absolute value opearator, *first used on page* 153.

$ARL(\delta = \bar{\delta})$ [*Dimensionless*] Average run length for detecting a shift in the process equal to $\bar{\delta}$, *first used on page* 133.

$c(z - 2h(x, y))$ [*Dimensionless*] Interferogram envelop function, *first used on page* 14.

CL [*The same as the statistic plotted*] Centre line, *first used on page* 138.

$d_2(n_i)$ [*Dimensionless*] Expected value of the standardised range distribution, i.e. of $W = R/\sigma$, *first used on page* 139.

d_{CUSUM} [*Dimensionless*] Design parameter for a CUSUM. It represents the relationship between the decision interval H_{CUSUM} and the reference value K_{CUSUM}, *first used on page* 133.

$\hat{\mathbf{e}}$ [*The same as the response variable*] Vector of the residuals in a linear model, *first used on page* 54.

$\hat{e}_{i,j,t,control}$ [*The same as the response variable in calibration*] Difference of the j-th measurement result of the i-th control RM at the t-th control instant from the predicted value of the accepted RM length, *first used on page* 137.

$\mathbf{e}$ [*The same as the response variable*] Vector of the errors in a linear model, *first used on page* 54.

$E(...)$ [*Dimensionless*] Expected value of a random variable, *first used on page* 50.

e_{ij} [*Dimensionless*] The j-th random error at the i-th level of the factor in a random effect ANOVA model, *first used on page* 50.

$g(x, y, z)$ [W/m^2] Intensity of the light at the point $P(x, y)$ of the interferogram recorded when the WLI scanner is located at the point $Q(z)$, *first used on page* 13.

$\mathbf{H}$ [*Dimensionless*] Hat matrix in a linear model, *first used on page* 54.

$\mathbf{h}_k^T$ [*Dimensionless*] Vector representing the k-th row of the hat matrix, *first used on page* 55.

$h(x, y)$ [µm] Profile of a surface, *first used on page* 14.

H_{CUSUM} [*The same as the statistic controlled*] Decision interval in a CUSUM, *first used on page* 132.

h_{CUSUM} [*Dimensionless*] Decision interval in designing a CUSUM chart expressed as multiples of standard deviations of the plotted statistic, *first used on page* 132.

$h_{i,j,t}$ [*The same as the response variable in calibration*] The j-th measurement result of the i-th RM at the generic instant t while operating the measuring system with the proposed control method, *first used on page* 137.

h_{ij} [µm] Response variable in the ANOVA for the repeatability analysis, *first used on page* 75.

HC [*Dimensionless*] Height correction parameter, *first used on page* 105.

$\mathbf{I}$ [*Dimensionless*] Identity matrix, *first used on page* 54.

k [µm] Coverage factor of the expanded uncertainty, *first used on page* 42.

K_{CUSUM} [*The same as the statistic controlled*] Reference value in a CUSUM, *first used on page* 132.

k_{CUSUM} [*Dimensionless*] Reference value in designing a CUSUM expressed as multiples of standard deviations of the plotted statistic, *first used on page* 132.

L [mm] Actual length of a gauge block in a generic point, *first used on page* 35.

$l_{\alpha_{th}}$ [mm] Length of a gauge block after a variation in α_{th}, *first used on page* 42.

$l_{c,i}$ [mm] Central length of the i_{th} gauge block ($i = 1, 2 \ldots$), *first used on page* 31.

L_i [mm] Actual length of the i-th gauge block in a generic point, *first used on page* 32.

$l_{n,i}$ [mm] Nominal length of the i_{th} gauge block ($i = 1, 2 \ldots$), *first used on page* 31.

$l_{n,s}$ [µm] Nominal length of the reference samples, *first used on page* 32.

L_s [mm] Actual length of the reference samples relative to two generic points, *first used on page* 32.

l_{th} [mm] Length of a gauge block after thermal expansion, *first used on page* 41.

LCL [*The same as the statistic plotted*] Lower control limit, *first used on page* 138.

LSL_Q [*The same as Q*] Lower specification limit for the generic quantity Q, *first used on page* 34.

O_{ij} [*Dimensionless*] Position in the run order of the j-th measurement result at the i-th factor level, *first used on page* 85.

p [*Dimensionless*] Number of unknown parameters in a linear model beside σ, *first used on page* 94.

R [*The same as the sample variable*] Range of a sample of random variables, *first used on page* 139.

$s_{i,CUSUM}$ [*Dimensionless*] Statistic on purpose defined for monitoring the variability of a normally distributed process with mean zero and variance one via a CUSUM chart, *first used on page* 134.

s_{ij} [*Dimensionless*] Response variable in the ANOVA model for the scan length, *first used on page* 50.

S_{trg} [*unit of area*] Area of the projection on Z=0 of a triangle associated with three points in the cloud of points of a WLI measurement result, *first used on page* 153.

t_e [µm] Limit deviation of length at any point from the nominal length of a gauge block, *first used on page* 31.

T_{actual} [°C] Temperature of a gauge block, under stated hypotheses, *first used on page* 41.

$t_{h,i}$ [µm] The i-th random effect in the ANOVA model for the repeatability, *first used on page* 75.

$t_{s,i}$ [*Dimensionless*] The i-th random effect in the ANOVA model for the scan length, *first used on page* 50.

$U(\ldots)$ [µm] Expanded uncertainty of the generic estimate in parentheses, *first used on page* 42.

$u(T_{actual})$ [°C] Standard uncertainty associated with the input estimate T_{actual}, *first used on page* 41.

$u_c(L_i)$ [µm] Combined standard uncertainty of L_i, *first used on page* 42.

$u_{subscript}(\ldots)$ [µm] Component of the combined standard uncertainty of the output estimate in parentheses due to the input estimate subscript, *first used on page* 42.

UCL [*The same as the statistic plotted*] Upper control limit, *first used on page* 138.

USL_Q [*The same as Q*] Upper specification limit for the generic quantity Q, *first used on page* 34.

V [*unit of volume*] Volume of the solid element associated with a triangle from the cloud of points of a WLI measurement result, *first used on page* 154.

v [µm] Variation in length of a gauge block, *first used on page* 31.

$V(...)$ [*Dimensionless*] Variance of a random variable, *first used on page* 51.

V_{prism} [*unit of volume*] Volume of the right prism associated with the minimum z-coordinate in a triangle from the cloud of points of a WLI measurement result, *first used on page* 153.

$V_{tetrahedron,i}$ [*unit of volume*] Volume of the i-th tetrahedron ($i = 1, 2$) when calulationg the volume associated with a triangle from the cloud of points of a WLI measurement result, *first used on page* 154.

$var(...)$ [*Dimensionless*] Covariance matrix of the random vector in the parentheses, *first used on page* 55.

W_λ [*Dimensionless*] Quantile of the standardised range distribution that leaves λ probability on its left, *first used on page* 139.

$\mathbf{x}_0^T$ [*Dimension of the predictors*] Transpose of a particular instance of the vector of the predictors in the regression model built during the calibration. It is associated with the accepted value of the generic RM used in the control method, *first used on page* 137.

Z_ζ [*Context dependant*] Quantile of the normal standard distribution that leaves ζ probability on its right, *first used on page* 138.

z_{max} (z_{mid}) [*unit of length*] Maximum (intermediate) z-coordinate in a triangle from the cloud of points of a WLI measurement result, *first used on page* 153.

z_{min} [*unit of length*] Minimum z-coordinate in a triangle from the cloud of points of a WLI measurement result, *first used on page* 153.

Chapter 1

Introduction

1.1 Motivation

There is a significant consensus in the scientific community for the existence of a trend towards miniaturisation, both in the consumer and in the industrial product market (cf. Leach et al., 2006, Dowling et al., 2004 and Masuzawa, 2000). On the one hand, smaller sizes allow the consumers to have products with lighter weight, more functionality in the same space and with less energy consumption. On the other hand, the properties of the matter on a small scale can start changing due to larger surface area per unit of mass and due to the fact that quantum effects can begin to dominate the behaviour of the matter (cf. Dowling et al., 2004). This allows the possibility of developing completely new applications.

Metrology has been recognised by the Royal Society as one of the key enabling technologies to support successfully these current efforts in micro- and nano-manufacturing (cf. Dowling et al., 2004). In fact, the science of measurement underpins the characterisation of materials in terms of their physical properties (mechanical, electrical, magnetic and optical). This is relevant when considering the trend towards miniaturisation.

For example, from a dimensional point of view, when reducing the size of a product it is desirable that the size of the dimensional error is also proportionally reduced (cf.

Masuzawa, 2000 and Table 1 in BS EN 20286-1 for a relationship between dimensions and standard tolerance grades in the dimensional specifications of products). The specifications in terms of precision and trueness (cf. BS ISO 5725-1) associated with the manufacturing equipment have therefore of necessity to be at the centre of the research effort. Masuzawa (2000) stresses this concept by describing the equipment precision as a 'necessary condition for micromachining'.

Consequently, measuring and inspection methods with an even higher level of precision should be available to characterise both products and manufacturing processes in the micrometric and the nanometric ranges (cf. Montgomery and Runger, 1993). From this perspective, it is believed that effort and resources spent on quantifying precision and trueness of commercially available measuring equipment, on critically analysing the existing procedures for instrument calibration and on proposing innovative approaches to calibration and precision assessment are unavoidable and essential if the trend towards miniaturisation is to be sustained. In fact, on the one hand, precision studies aim to evaluate the variability of a measuring process, and on the other hand, calibration studies aim to assess the central tendency or location of a measuring process by comparing it with a trusted reference system.

In the light of this framework of ideas, this investigation could not therefore avoid analysing both the precision and the calibration issues of the deployed measurement system, while proposing a novel procedure for measuring volumes on the micrometric scale.

In particular, while investigating the micro-electric discharge machining process, strong evidence supporting the fundamental role of the electrode tool wear was found. Even so, little investigation has concerned the assessment of the presently available methods of measuring such a volumetric wear which, ultimately, hinges on measurements of micro-volumes. Even less interest has been shown in proposing new approaches aimed at overcoming the pitfalls of the currently available procedures for measuring small volumes of solid parts. In particular, the difficulties connected with the fact that the measuring systems are very often separated from the micro manufacturing equipment, have to be addressed when proposing a measuring

procedure.

Some form of quantitative assessment of a proposed measuring procedure in terms of variability has also to be provided. This is essential in order to identify the capability of the measuring process and, ultimately, which measurement tasks can be assigned to the developed measuring process (about the concept of process capability cf. chapter 9 in Montgomery, 1996 and Montgomery and Runger, 1993).

In addition, the study of micro-fabrication processes apart from electric discharge machining would also benefit from the availability of a novel procedure for measuring micro-volumes. For example, the investigation of wear phenomena in the micro milling process and, more generally, in micro-machining could progress further by the availability of a procedure for measuring micro-volumes. More insight into wear phenomena means better evaluation of the expected tool life, with two main consequences. The first is the reduced risk of breaking the tool while machining, with the consequent irreversible damage of the part. The second is an increased efficiency in the exploitation of the tools. These can then be deployed with lower safety factors, commonly introduced for avoiding breakage during machining.

Such a method can also be beneficial to the study of manufacturing processes on a larger dimensional scale, where influential elements of the process cause localised phenomena on a micro scale. For example, a measuring procedure that has similarities with that presented in this study is proposed by Dawson and Kurfess (2005) for the investigation of the flank and crater wear of two types of polycrystalline cubic boron nitrate tools (not coated and TiAlN-coated) in machining operations (turning) of a hardened steel (AISI 52100). Incidentally, it is inferred from the definitions and examples provided by Dowling et al. (2004) that the investigation of Dawson and Kurfess (2005) can rightly be named micro- or nano-technological, although the authors do not do so.

More generally, all the micro fabrication processes that are affected by wear phenomena can benefit from the availability of a procedure for measuring small volumes. For instance, in replication techniques such as micro injection moulding and hot

embossing, this procedure for measuring small volumes could be beneficial in evaluating the amount of volumetric wear in functional localised areas of the moulds (corners and edges, for instance) as a function of the number of parts produced.

In the study of replication techniques on a micro scale, such as micro injection moulding and hot embossing, the proposed procedure for measuring small volumes with complex and/or irregular shapes could also have other usages. For instance, such a method could be deployed by the process designers to compute the exact volumetric thermal shrinkage of the parts, or elements on the parts. In this way they could build an experimental base of knowledge that they could use to improve the designing process of future moulds or to validate theoretical models of the process.

An analysis of the replicated parts for defects could also be carried out based on measurements of micro-volumes. This could include a quantitative assessment of the mechanism of formation of defects due to poor filling of micro-cavities. In fact, it would be possible to measure the micro-volume of the cavity on the mould and compare it to the corresponding volume on the replicated part.

A procedure for measuring small volumes can also be of help in the assessment of the efficiency of competing micro-manufacturing processes. Among them, it could, for example, be interesting to compare and contrast focused ion beam machining, laser machining, electro-chemical machining and electron beam machining. By directly measuring the micro-volumes removed by competing processes in the same unit of time, the material removal rate of these processes can, in fact, be seamlessly and directly evaluated.

Finally, it is envisaged that not only manufacturing process designers but also product designers, such as designers of microfluidic devices (micro-pumps, micro-valves, micro-flowmeters and micro-mixers) could benefit from a procedure for measuring small volumes. In fact, they could inspect the parts delivered to them by manufacturers in order to assess the compliance to the specifications of tolerances that they have prescribed, where these specifications could be in volumetric terms rather then dimensional. In this way, the process of tolerance specifications might be facilitated

by being expressed in volumetric units, closer to the functional requirements of the part.

In order to keep the investigation focussed, the benefits of a procedure for measuring small volumes have been extensively supported by a literature survey in the sole case of the micro-electric discharge machining process.

1.2 Aims and Objectives

The remit of this investigation was to:

- provide a procedure for measuring volumes of solid material with irregular and/or complex boundary form, both of convex and concave shape, without the need of any reference system external to the measurand part itself;
- assess quantitatively the precision of the developed procedure for measuring volumes and to highlight its limitations;
- identify the sources of variability of the measuring system upon which the procedure for measuring volumes is based and to quantify their effects on the measurement results. Namely, to identify the sources of variability of measurement results of lengths taken by a white light interferometric microscope along the z axis;
- investigate the calibration procedure deployed in the measuring instrument upon which the procedure for measuring volumes is based. Namely, to identify the weaknesses of the current calibration procedure of a white light interferometric microscope and to propose improvements;
- develop a cost-effective and versatile procedure for making available traceable reference materials for calibration purposes in the micrometric range.

1.3 Outline of this thesis

This study has seven chapters, of which Chapters 3 to 6 encompass the main investigations, whereas Chapters 2 and 7 are a literature review and a summary of the contributions of this work, respectively.

In Chapter 2, the context of this investigation is set by making provision of background knowledge for Chapters 3 to 6. This chapter consists of three interconnected sections. In the first section, evidence of the critical nature of wear in micro-electric discharge machining are provided and the need for a new procedure for measurements of small volumes is highlighted. An approach based on the white light interferometric microscope is also identified as a viable approach to wear evaluation. In the second section, therefore, the main characteristics and fundamental principles of white light interferometric microscopy are presented and critically analysed. In the third section, the metrological concepts used in describing this measuring system and also employed throughout this study are presented and comparatively scrutinised.

In Chapter 3, the tools for a correct use and for a performance assessment of the white light interferometric microscope are introduced. In particular, a cost-effective and versatile procedure for building traceable reference samples of length in the micrometric range is presented and analysed in terms of its uncertainty. A study of the effects of discretionary setup parameters on the measurement results is then described.

In Chapter 4, by deploying and building upon the findings of Chapter 3, the precision of the white light interferometric microscope in repeatability conditions is assessed in a pre-specified micrometric range. In this investigation, a dependence of the measurement results on the sequence of the tests was revealed and gave rise to a subsequent quantitative study also presented in this chapter.

In Chapter 5 the calibration procedure of the white light interferometric microscope is addressed. This chapter is divided into four main sections. In the first section,

the calibration procedure built in the measuring system is examined and interpreted in the light of the International Standards. In the second section, the bias induced by the built-in calibration procedure is experimentally highlighted. In the third section, in order to overcome the problems generated by the built-in procedure, a novel alternative calibration procedure is described and implemented. In the fourth section, two control methods necessary for the practical use of every calibration procedure are introduced.

In Chapter 6 a procedure for measuring volumes with complex and irregular boundaries on a micro scale is developed. First, the overall procedure, with the pertinent set of implemented computer programs is described. The algorithm developed and implemented for measuring volumes is then illustrated and its performance is assessed. The entire procedure is then demonstrated in the case of micro-volumes with both a convex and a concave boundary. An assessment of the precision of the procedure is then carried out in the case of a convex micro form (a shaped micro-electrode tool of the micro-electric discharge machining process). Finally, the limitations of the procedure in term of biasing factors and the potential further developments are presented.

In Chapter 7 the main contributions and conclusions of this study are summarised. Some possible directions for further investigations in this area are also suggested.

Chapter 2

Literature review

The first section of this chapter is dedicated to the centrality of the concept of volumetric wear ratio in micro-electric discharge machining ($\mu - EDM$) operations. A white light interferometer (WLI) approach to measuring micro-volumes is also proposed as a viable possibility to improve the accuracy of the volumetric wear ratio measurement.

In the second section, fundamental principles in the field of microscopic interferometry have been surveyed. They are, in fact, most relevant to the use of this measuring instrument.

The proposal of a volume measuring method and a critical analysis of a measurement instrument, the white light interferometer microscope, rely on some general metrology concepts such as accuracy, precision, trueness, bias, uncertainty and calibration. These concepts are considered in the third section.

This volume measuring method is also based on fitting surface elements to cloud of points whose coordinates are experimentally measured. A number of fundamental concepts in the domain of geometrical modelling and reverse engineering are therefore briefly mentioned in the fourth section.

2.1 The necessity of measuring micro-volumes

In electric discharge machining the reduction of the discharge energy during each breakdown of the dielectric liquid is a necessary condition for such a process to be considered a micromachining technology (cf. Masuzawa, 2000 and Tsai and Masuzawa, 2004).

In working conditions providing low discharge energy, the ratio of the volume removed from the tool electrode to the volume removed from the work piece electrode, also known as volumetric wear ratio (θ), increases significantly. This is shown in Dibitonto et al. (1989) and in Patel et al. (1989) via experimental data for a copper-made anode tool and a steel-made cathode work piece. The range of data for low discharge energies, although still large for $\mu - EDM$ operations, has been rearranged in Table 2.1, where energy and wear ratio values have also been derived.

In manufacturing micro through-holes with cylindrical electrodes and with constant low discharge energy settings, there is experimental evidence that the smaller the electrode diameter, the larger the volumetric wear ratio becomes (cf. table 1 and figure 1 in Tsai and Masuzawa, 2004).

The main consequence of a high volumetric wear ratio is a quick change in the shape of the tool electrode (cf. Yu et al., 1998).

Several authors have proposed different techniques aimed at overcoming the accuracy problems caused by the rapidly changing shape of the tool electrode during the die-sinking micro-electric discharge machining of complex 3D cavities (for instance, cf. Yu et al., 1998, Yu et al., 2003 and Pham et al., 2004).

These techniques, however, apart from the method presented in Pham et al. (2004), have relied on an a priori perfect knowledge of the volumetric wear ratio, which has been considered a deterministic constant quantity. In order for those approaches to represent viable methods, a measurement procedure of the micro-volumes removed from the tool and from the workpiece is therefore needed. Such a procedure must

Current I [A]	Pause time P [μs]	Pulse time T [μs]	Discharge Voltage, V [V]	Pulse Energy E [μJ]	Cathode Material removal rate M.R.R. [$mm^3/\ min$]	Anode	Wear ratio θ [%]
3.67	2.4	13	25	1192.75	3.1	0.026	0.84
2.85	1.3	7.5	25	534.38	1.6	0.024	1.5
2.34	1.0	5.6	25	327.60	0.3	0.009	3

Table 2.1: Rearrangement of the Experimental data from Dibitonto et al. (1989) and Patel et al. (1989). θ has been derived from M.R.R., E from I, V and T.

also have trueness and precision adequate to the usage in operations on a micro scale (cf. BS ISO 5725-1, for the definitions of trueness and precision).

On the one hand, Pham et al. (2004) pointed out the need for more accurate measurement procedures for the volumetric wear ratio. In particular, these procedures should address the issue of estimating not only the reduction in length of the tool electrode, but also the changes in the shape of the tool electrode that has occurred during machining. On the other hand, by proposing an estimation of the volume removed from the tool electrode that is based only on measurements of length, Pham et al. (2004) did not directly address the issue they themselves raised.

Tsai and Masuzawa (2004) introduced a method for assessing the micro-volumes removed from rotating cylindrical tool electrodes and from workpiece electrodes in machining micro through-holes. On the tool electrode, the removed micro volume was divided in two parts, corresponding to the reduction in the tool length and to the corner rounding. This approach was based on three different instruments. Firstly, an optical microscope with a resolution of $0.1\ \mu m$ was used for measuring the original diameter of the tool electrode. Secondly, on board the EDM machine, the electrical contact system enabled the measurement of the difference in the end-of-tool position before and after machining. It was in this way to estimate the reduction in length of the tool electrode. Thirdly, scanning electron microscope (SEM) pictures were used to estimate the micro-volume removed from the corner of the tool electrode. In particular, this volume was evaluated on the basis of an interpolating polynomial function passing through two points of the rounded corner in the SEM picture of the tool longitudinal profile. The micro-volume removed from the work piece electrode was estimated by considering the micro through-hole as a conical frustum with a height equal to the thickness of the workpiece. No further details are provided on how the radii of the frustum were measured.

In this thesis, a volume measurement procedure that is based on white light interferometry is proposed. WLI has already been proved a valuable means for investigating the $\mu - EDM$ process. In fact, WLI was utilised for characterising the lateral surface of micro-tool electrodes grounded by $\mu - EDM$, in terms of roughness, waviness and

form profiles (Vallance et al., 2004). Nevertheless, the usage of white-light interferometric techniques for measuring micro volumes removed both from the tool and the work piece electrodes in order to assess the volumetric wear ratio, does not appear to have been investigated yet.

WLI techniques make possible the description of a surface by a cloud of three-dimensional points $P(x, y, z)$ with their projection$P_0(x, y)$ evenly spaced on the plane $z = 0$. This characteristic makes WLI, in contrast to other techniques, potentially suitable for the seamless measurement of the overall volume removed from an electrode, accounting for variations both in its length and in its shape. This advantage over other measuring techniques holds as long as the shape of the part to be measured allows the light coming from the instrument to be reflected back into it.

More insight into white light interferometric microscopy is provided in the next section.

2.2 White Light Interferometer microscopy

2.2.1 General principles

White light interferometry, also known as vertical scanning interferometry (VSI), is one of the interferometric techniques suitable for shape measurements of microscopic features (O'Mahony et al., 2003 and Schmit and Olszak, 2002).

A beam of light emitted by a spatially and temporally incoherent[1] source is split into two separated beams: the object beam, which is reflected by the part to be measured, and the reference beam, which is reflected by a reference mirror. The separation into two beams is carried out by an optical component called a beam splitter. After being

[1]On the concept of coherent sources see , for instance,COHER and pp. 860, 861 in Alonso and Finn (1967).

reflected, the beams combine, undergoing constructive and destructive interference and producing bright and dark bands named *fringes* (see O'Mahony et al., 2003 and, for more details on constructive and destructive interference, see for instance pp. 857 - 862 in Alonso and Finn, 1967).

The image containing the fringes, also known as the *interference pattern of the two beams*, is called *interferogram* (O'Mahony et al., 2003). A detector, usually a charge-coupled device (CCD), records changes of the light intensity in a predefined number of interferograms, while the optical path difference (OPD) between the object and the reference beams is modified according to a pre-specified setting (Schmit and Olszak, 2002). The optical path or optical path length is defined as the distance travelled in an optical system by light (Weisstein, 2006e). In this context, the path length travelled by the object beam is the distance between the beam splitter and the point on the surface of the part to be measured where it is reflected. In contrast, the path length travelled by the reference beam is the distance between the beam splitter and the point on the reference surface where the beam is reflected. Schemata of the white light interferometer in Michelson and in Mirau configurations are illustrated in Pavlicek and Soubusta (2003) and in Chim and Kino (1991), respectively.

These changes in OPD can be obtained by moving the object or the scanning head or the reference mirror at a constant speed (Schmit and Olszak, 2002). A moving scanning head in the vertical direction is only considered here due to the fact that this is the typology of equipment used in the experimental activity carried out. The intensity of the light reaching the detector, namely $g(x, y, z)$, can be described by the following equation (cf. Larkin, 1996):

$$g(x, y, z) = a(x, y) + b(x, y)\, c[z - 2h(x, y)] \cos[2\pi\omega_0 z - \alpha_{WLI}(x, y)] \qquad (2.1)$$

In equation (2.1), x and y are the coordinates of the generic point in the interferogram, whereas z is the location of the scanning element for that particular interferogram whose light intensity is recorded. The coordinate z is also called *axial location* or *defocus of the object.* The quantity $a(x, y)$ depends on the intensities of both the

object and the reference beam, whereas the quantity $b(x, y)$ is only determined by the intensity of the reference beam. The function $c[z - 2h(x, y)]$ is called the *interferogram envelope function*, or simply *envelope*, and depends on the spectral profile of the utilised source of light. The exact shape of the envelope is considered as not critical by some authors and is commonly approximated by a Gaussian function (for instance, cf. Larkin, 1996).

The mean wavelength of the source of light influences ω_0, the *spatial frequency of the intensity* $g(x, y, z)$ *in the z direction.* The parameter $\alpha_{WLI}(x, y)$ is called the *phase change on reflection.* It can be used to take into account the fact that the variation of phase on reflection of a broadband light beam on certain type of test materials, e.g. metals, is not constant for all the wavelengths of the beam. A detailed investigation of the dispersion of the phase change on reflection has been proposed by Harasaki et al. (2001). The function $g(z)$, that describes the interference intensity distribution at a particular point (x, y) of the interferogram/test part when the scanner element is at the generic axial location $z = z(time)$, is commonly referred to as *correlogram* (Larkin, 1996). Usually, the WLI is used to determine the *profile* of a surface, $h(x, y)$. The profile can be defined as the graphic of the function $h(x, y)$ representing the signed distance of the surface under investigation at its generic point $P(x, y)$ from a pre-specified reference plan (Larkin, 1996).

In WLI, an oscillating correlogram $g(z)$, which unveils the presence of interference fringes in the axial direction z , appears only in a small area around the local focus on the test object (Schmit and Olszak, 2002). In fact, the wide spectrum of the white light source results in a short coherence length that provides interference fringes in $g(z)$ when the OPD is about zero (Harasaki et al., 2001). Further support for this can be found in the definition of coherence length reproduced verbatim from (cf. COHER): 'If a wave is combined with a delayed copy of itself (as in a Michelson interferometer), the duration of the delay over which it produces visible interference is known as the coherence time of the wave, $\triangle t_c$. From this, a corresponding coherence length can be calculated: $\triangle x_c = c \triangle t_c$'. For example, in order to give a quantitative measure, common light bulbs have a coherence length of about $1\,\mu m$,

whereas laser sources can have coherence length of some metres (cf. COHER).

The peak of the envelope $c = c(z) = c\,[z - 2h\,(x, y)]$, corresponding to the maximum contrast between the fringes, and the phase of the fringes, i.e. $(2\pi\omega_0 z - \alpha\,(x, y))$, give the height $h(x, y)$ of the point $P(x, y)$ on the object surface, provided that the OPD change rate is known (Larkin, 1996 and Schmit and Olszak, 2002).

Many algorithms have been developed in order to find the maximum of the envelope $c(z)$ from the finite and discrete set of intensity data points that a measurement yields (cf. Harasaki et al., 2000). Their analysis is, however, beyond the purposes of this research, due to the fact that the WLI microscope and the associated software utilised in the experimental activity carried out are based on inaccessible algorithms[2]. In addition, it should be noted that steps between two surfaces as high as a few millimetres are commonly measured with resolution of a few nanometers by commercially available WLI microscopes (for instance, cf. MICRO). On a macro scale, this corresponds to an instrument which would be able to distinguish between two points separated by a few millimetres while measuring length on a scale of kilometres.

2.2.2 Open research issues

The rate of change in the OPD is usually referred to as the *phase step* (Schmit and Olszak, 2002). Due to the fact that the knowledge of the phase step is essential for the assessment of $h(x, y)$, the motion characteristics of the scanner element are particularly critical. The most common method used to assess the phase step is to set the rate of change in the OPD to a nominal value by a calibration procedure (Schmit and Olszak, 2002). This is achieved by measuring a traceable length sample via the WLI microscope and by using a correction factor to match the measured and the nominal length. The correction factor is also often referred to as height correction (ADEP2). It is only after completing this calibration procedure that the

[2]Nor the source code neither a detailed documentation of the used algorithms has been provided by the manufacturer of the instrument.

instrument can be used for a measurement task (MICRO; ADEP2; NANOT). If the part to be measured is significantly different from the calibration sample used, the scanner element is travelling for a scan length different from that used during calibration. The implicit assumption that is made in this procedure is that the phase step is constant along a generic scan length (Schmit and Olszak, 2002). However, as Schmit and Olszak (2002). pointed out, this hypothesis is not always correct.

Improved technological solutions that reduce or eliminate the centrality of this assumption have been recently proposed (Schmit and Olszak, 2002). The white light interferometric microscope used in this experimental investigation does, however, require the type of calibration just mentioned. Consequently, the following interconnected issues need addressing:

- Firstly, a calibration sample with a feature of known length traceable to a national standard of length is necessary. The knowledge of the uncertainty of the traceable length sample is also a necessary prerequisite to the calibration procedure. The official national sample of length in the United Kingdom is the responsibility of the National Physical Laboratory (NPL). However, commercially available samples purpose built for calibrating white light interferometric microscope are most commonly traced to the United States official national sample, which is produced at the National Institute of Standards and Technology (NIST). It is anticipated that the availability of a large set of calibration samples covering a broad range of lengths, has a paramount role in the calibration procedures of measuring instruments (cf. BS ISO 11095).

- Secondly, in order to determine the range of measurement tasks that can be performed by the white light interferometric microscope in terms of variability, a gauge capability analysis is required (for instance cf. Montgomery and Runger, 1993).

- Thirdly, after calibrating the interferometric microscope using a specific traceable length, an assessment of the potential bias induced on the WLI measurements of a generic length is necessary. If a bias in the measurements

is quantitatively proven, then an alternative calibration procedure has to be developed in order to avoid obtaining distorted measurements.

In Chapter 3, an answer to the first need is given together with an analysis of the effects of two discretionary setup parameters on the variability of the measurement results. An answer to the second need is provided in Chapter 4, whereas calibration and bias related issues are investigated in Chapter 5.

In this section, focused on WLI microscopy, concepts in the domain of metrology have been deployed without a pertinent critical analysis. This is provided in the next section.

2.3 Metrological concepts deployed in assessing the WLI process

In order to deploy a measurement instrument or a measurement method, their performances have first to be assessed quantitatively (for instance cf. Montgomery and Runger, 1993). Although many might argue about the unavoidable necessity of dedicating resources to this sort of activity, strong evidence supporting such a need has been found in the literature (for instance cf. Feynman, 2000).

The analysis of the measurement instrument and of the volume measuring method developed in this investigation demands the introduction of some concepts that are outlined in this section.

The terminology used in this thesis relies on the 'International vocabulary of basic and general terms in metrology' (VIM) , published by ISO, the 'International Organisation for Standardisation' , and reproduced verbatim in BSI PD 6461-1 , a 'Published Document' (PD) issued by the 'British Standard Institution' (BSI) .

2.3.1 Uncertainty

The internationally recognised fundamental reference about uncertainty is the 'Guide to the expression of uncertainty in measurement' published by ISO. This guide is referred to as the GUM and it has been reproduced verbatim in BSI PD 6461-3 , a PD issued by the BSI.

In the GUM, the term 'uncertainty of measurement' is used in two different meanings. On one hand, it expresses the general concept of doubt about the validity of the result of a measurement. On the other hand, it denotes all the specific quantitative measures of this concept.

The formal definition of uncertainty from the GUM is as follows: 'uncertainty of measurement is a parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand'.

Examples of such a parameter are the following:

- the standard deviation of the result of a measurement. This parameter is called standard uncertainty.
- half-width of an interval about the measurement result that is expected to encompass a pre-specified large fraction (level of confidence) of the distribution of values that could reasonably be attributed to the measurand. This half-width is called expanded uncertainty. The level of confidence associated with the interval defined by the expanded uncertainty is very sensitive to the probability distribution of the measurement result. Such distribution has to be explicitly or implicitly assumed (cf. note 2 of section 2.3.5 in BSI PD 6461-3). Consequently, whenever it is possible, it appears sensible first to test the distributional hypothesis and then to calculate the expanded uncertainty.

This formal definition has given rise to some controversy among the experts (cf. the National forefront section in BSI PD 6461-3). These, however, are beyond the scope

of this investigation. In addition, in the GUM, section 2.2.4, it is stated that all the available concepts of uncertainty lead to the same manner of using the data and pertinent information.

Two main approaches to the evaluation of the uncertainty of measurement are proposed in the GUM: the type A and the Type B evaluation. Type A evaluation is defined as a 'method of evaluation of uncertainty by the statistical analysis of series of observations'. Type B is defined as a 'method of evaluation of uncertainty by means other than the statistical analysis of series of observations'. From a practical point of view, such definitions appear quite generic. They fail, in fact, to provide any specification of the conditions in which the series of measurement are meant to be taken. For example, neither reproducibility nor repeatability conditions are mentioned (cf. sections B.2.15 and B.2.16 in BSI PD 6461-3).

Whereas in the VIM and in the GUM the term uncertainty is applied solely to measurements, a phrase such as 'uncertainty of an instrument' has been found in common use (cf. page 14 in Cheok, 2005).

2.3.2 Error and accuracy

In the GUM, the term measurand means a 'particular quantity subject to measurement'. The definition of a measurand is in terms of specifying some physical states and conditions. Nevertheless, a measurand cannot be thoroughly described without providing an infinite amount of information. Thus, the provision of information is expected to be commensurate with the purpose of the measurement and its requested uncertainty. In fact, the less information present in the definition of the measurand, the larger the need for interpretation. A component of uncertainty is, therefore, added to the measurement result.

For example, 'the length of a nominally one-metre long steel bar' can be a suitable definition of measurand if a standard uncertainty in the millimetres range is tolerated. On the other hand, it cannot be accepted if a standard uncertainty in the

range of micrometres is demanded for the measurement result. In this case, 'the length of a nominally one-metre long steel bar at 25°C, and 101 325 Pa, between the two centre points of the two side surfaces' can be a more appropriate definition of the measurand.

Once provided with the definition of measurand, a measurement procedure is performed on a quantity that physically materialises the definition of measurand. This is called the realised quantity of the measurand (cf. sections D.2 inBSI PD 6461-3). The consistency between the definition and the realised quantity of the measurand is only ideally complete. A further component of uncertainty of the measurement result has, therefore, to be considered.

The value of the quantity intended to be measured, that is the value of the measurand, can be logically postulated, but, ultimately, it can never be known (cf. section D inBSI PD 6461-3). In the VIM, the value of a measurand is called a 'true value of a quantity', whereas in the GUM the attribute 'true' is omitted because is considered redundant (cf. note in section 3.1.1 and the section D.3.5 both in BSI PD 6461-3). Furthermore, the value of the measurand is an ideal concept, that cannot be uniquely determined due to the aforementioned uncertainties connected with the always partial information of the measurand definition and with the always imperfect measurand realisation.

In this investigation, the attribute 'true' referred to the value of a measurand is used when it is thought that clarity benefits from its introduction.

Dissatisfaction and controversy have been surveyed concerning the expression 'value' contrasted to 'true value' of a measurand (cf. the National forefront section in BSI PD 6461-3). On the one hand, it is believed that these arguments might appear more of a metaphysical rather than a practical nature. On the other hand, they highlight the effort made by the scientific community to clarify those widespread practical concepts that lean on the concept of 'value' or, equivalently, 'true value' of a measurand. These are the concepts of error and accuracy.

In the VIM, error of measurement, or error of the result of a measurement, is defined

as the 'result of a measurement minus a true value of the measurand'. The indefinite article in front of 'true value' underlines the previously mentioned multiplicity of values consistent with a given measurand. From its definition, it follows that the error of measurement cannot be known, so as the value of the measurand. Hence, it can only be estimated. The error of measurement is traditionally viewed as the sum of two additive components, namely a random error and a systematic error.

Random error is the 'result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions' (cf. section B.2.21 in BSI PD 6461-3). Due to the fact that only a finite number of measurements can be performed, the random error can never be known. It is thus modelled as a random variable (cf. section 5.1 in BS ISO 5725-1). Random error originates from unpredictable temporal and spatial variations of influence quantities, where an influence quantity is 'a quantity that is not the measurand but that affects the result of the measurement' (cf. section B.2.10 in BSI PD 6461-3). The random error component of a measurement result can never,therefore, be eliminated. Following the above definition, the expected value of a random error variate[3] is zero.

Systematic error is the 'mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand' (cf. section B.2.22 in BSI PD 6461-3). Similarly to the random error, the systematic error cannot be known. If the effect of an influence quantity is recognised as the originator of an estimated systematic error, then the effect can be quantified and a correction or a correction factor can be applied to the measurement result so as to compensate for the effect. In this manner, after compensating, the expected value of the systematic error is zero (cf. section B.3.2.3 in BSI PD 6461-3). Due to the fact that explicit reference to the concept of expected value is made, it is argued that also the systematic error is modelled as a random variable. The systematic error is also called bias in connection with the measuring instruments (cf. section 5.25 in BSI PD 6461-1). and with the properties of point

[3]variate is a synonym of random variable (cf. section C.2.2 in BSI PD 6461-3).

estimators in statistics (cf. page 293 in Mood et al., 1974).

In the GUM, the conceptual difference between the terms 'uncertainty' and 'error' is emphasised. The experimental standard deviation or sample standard deviation of the average of a series of observations is the uncertainty of the average due to unpredictable variations of influence quantities [4]. On the other hand, the random error in the average cannot be known. Nevertheless, according to its definition, the random error can be estimated for each average in a number of series of observations, by subtracting from each average the grand average of all the series. Similarly, the partial knowledge of a correction factor introduced to compensate a systematic error in a measurement result gives rise to an uncertainty component called "uncertainty of the correction". An explanatory graphical representation of the relationship between error and uncertainty is presented in Figure D.1 and D.2 in BSI PD 6461-3.

In BS ISO 5725-1, accuracy of measurement methods and results is defined as 'the closeness of agreement between a test result and the accepted reference value'. The term 'test result' appears to be the same concept as the 'measurement result' used in the GUM. Similarly, 'accepted reference value' seems to correspond to the 'true value (of a quantity)' used in the VIM and eschewed in the GUM. The difference in terminology might be due to the fact that BS ISO 5725-1 bases its vocabulary on BS ISO 3534-1 [5], rather than on the VIM and the GUM.

Furthermore, in BS ISO 5725-1, it is pointed out that accuracy for measurement methods and results, when applied to a set of measurement results, accounts both for random and systematic error (of each measurement result). This can also be inferred from the definitions provided above. In addition, it seems that the term 'error' is used in connection with a measurement result, whereas the term 'accuracy' is utilised to characterise a measurement method or procedure (cf. section 2.4 and 2.5 in BSI PD 6461-1 for the difference between measurement method and measurement procedure).

[4] These variations are also referred to as 'random effects'.

[5] At the present time, July 2006, BS ISO 3534-1 is a current standard, i.e. valid, upon which BSI signals some work in progress.

In BS ISO 5725-1, precision (of a measurement method) is defined as 'the closeness of agreement between independent test results obtained under stipulated conditions'. Thus, 'precision depends only on the distribution of random errors' (note 9 in BSI PD 6461-1) of the test results and not on the true value of the measurand. Precision is measured as the experimental (sample) standard deviation of the independent test results mentioned in its definition (note 10 in BSI PD 6461-1).

Although in the GUM the distinction between uncertainty and error is emphasised, the relationship between uncertainty and precision seems to be more difficult to describe. In fact, the experimental (sample) standard deviation of a series of independent measurement results taken in stipulated condition, estimates both the precision of the measurement method and the uncertainty of each measurement result. Hence, it is believed that the main difference between uncertainty and precision is the fact that uncertainty is referred to a measurement result, whereas precision is referred to a measurement method.

In BS ISO 5725-1, trueness (of a measurement method and results) is defined as 'the closeness of agreement between a test result and the accepted reference value'. The term 'trueness' has been invented in BS ISO 5725-1 in order to substitute the term 'bias', which originates from statistics, due to some controversies among some practitioners (medical and legal practitioners). Furthermore, in the same standard, it is pointed out that 'accuracy' was once used to express only the 'trueness' rather than both 'trueness' and the 'precision' of a measurement method, as it is proposed in the standard.

In addition, as a consequence of their definitions, it is believed that the distinction between 'trueness' (of a measurement method and results) and 'systematic error' (of measurement) is not completely clear. 'Trueness' is defined in more generic terms than the systematic error. Nevertheless, some degree of overlapping between the two concepts is apparent.

2.3.3 Calibration

The issues raised in section 2.2.2 about the WLI microscope calibration, lead to further insight into the concept of calibration.

BS ISO 11095 introduces a general framework for carrying out calibration of a generic measurement system. In fact, this standard provides general principles and a basic method for establishing calibration functions. Furthermore, it provides also a control method for assessing if a calibration function needs updating and for estimating the uncertainty of a measurement result due to the transformation using the calibration function. In addition, two further calibration methods are described for use when some specified particular conditions hold.

In BS ISO 11095, calibration is defined as a 'set of operations which establish, under specified conditions, the relationship between values of quantities indicated by a measurement system and the corresponding accepted values of some standards'.

In such a definition, the words 'values realised by standards' refer to the accepted values of reference materials. A reference material (RM) is defined in BS ISO 11095 as 'a substance or artifact for which one or more properties are established sufficiently well to be used to validate a measurement system' . The term 'system' describes not only a measuring instrument but also the measuring procedure, the operators and the environmental conditions connected with an instrument.

The focus of BS ISO 11095 is only on measurement systems for which not only RM's are available, but also whose accepted values are without error. However, an accepted value of an RM have an uncertainty. Hence, an error can always be present. Nevertheless, these uncertain accepted values of RM are still considered valid if their uncertainties are small compared with the uncertainty of their measured values.

The main result of a calibration procedure is a calibration function. In BS ISO 11095, the term 'calibration function' is not explicitly defined, although it is used. However, from the aforesaid definition and from the whole contents of BS ISO 11095,

it appears that 'calibration function' is a regression model fitted between the dependent variable measurement result and the independent variable accepted values of reference materials. More details about regression analysis can be found in Drapper and Smith (1966), which is also quoted in BS ISO 11095, in Sen and Srivastava (1990), which provides a detailed theoretical content, and in Faraway (2002), which also provides some practical guidance to the usage of R, a free statistical environment for statistical computing and graphics (cf. R Development Core Team, 2006).

In BS ISO 11095, the sole regression models investigated are those that lead to a linear relationship between measurement results and accepted values. In other words, only linear calibration functions are taken into account. On the other hand, if such a linearity is incompatible with the data, the same standard quite generically mentions the polynomial regression and gives some reference to Drapper and Smith (1966) (cf. note 2 in BS ISO 11095).

In setting up the experimental conditions to determine a calibration function, BS ISO 11095 advises the use of at least three RMs spanning the range of values that the system is to encounter during its normal operation. The larger the number of RM's, the easier the detection of departure from the assumed linearity of the calibration function.

In contrast, the calibration procedure described in section 2.2.2 for the WLI microscope, relies on a single RM. In chapter 5, it is shown that this procedure is a particular case of a one-point calibration method as described in section 8.2 of BS ISO 11095. In reality, the one-point calibration method is based on two points: the first provided by the RM and the second provided by requiring the calibration function to pass through the origin. Furthermore, it has to be noticed that BS ISO 11095 recommends not to use this procedure for calibration purposes, but only for the quick testing of an already existing linear calibration function.

Once a calibration function has been experimentally derived, i.e. once a regression model has been fitted to the experimental data, observed future measurement results are inserted in the regression model in place of the dependent variable and the

corresponding independent variable (regressor) is so determined. The calculated regressor constitutes the calibrated measurement result or the measurement result corrected for calibration.

2.4 Geometrical modelling in fitting clouds of points

The result of a measurement task performed using a WLI microscope can be represented by a text file containing the Cartesian orthogonal coordinates of the cloud of points describing the acquired profile of the measurand. A number of techniques, used not only in metrology but also in reverse engineering to obtain computer aided design (CAD) models of physical parts, have been developed to fit a digital model of the measurand to this cloud of points. An exhaustive overview and discussion of these techniques is presented in Várady et al. (1997). In particular, these authors first point out the distinction between three dimensional copiers and three dimensional scanners, then they identify the elements and the main problems encountered in each of the phases of a generic reverse engineering process. These encompasses:

- data acquisition;
- selection of data structures for representing shape from a hierarchy of shape models;
- segmentation of the cloud of points and surface fitting to each of the segments;
- combination of multiple scans (views) of the measurand, when they are needed to describe it completely;
- creation of a three dimensional geometric model (CAD model) of the measurand.

Várady et al. (1997) highlight also that the large number of decisions needed in a reverse engineering process is driven by the planned use of the model. For example,

the same cloud of points could be fitted by a set of surface patches (planar facets, quadrics, or higher order) each fitting a portion (segment) of the cloud or by a single parametric simple surface (plane, cylinder, sphere or others), where the parameters are estimated by some criterion of optimality such as least squares function, minimum deviation zone function or others (Nassef and ElMaraghy, 1999).

Moreover, Várady et al. (1997) point out the extreme importance of the availability of a priori information about the object, the measurand in this context, when performing reverse engineering tasks. The terms 'a priori information' or pre-experimental information is referred to the availability of geometrical knowledge about the measurand prior to the acquisition of the cloud of points. This knowledge assists and influences the decisions to be taken before and/or during the reverse engineering process, which can be either fully automatic or with some user interaction.

When performing a measuring task, sometimes also called an inspection task, most of the times the aim is to assess the compliance of some functional quality characteristic of a manufactured part (e.g. linear dimension, form, surface waviness, surface texture,etc.) to the specifications limits prescribed by the part designer. A priori information is therefore usually available in the form of drawings or in the form of a digital three dimensional geometrical model describing the 'nominal part'. A description of a procedure for assessing the compliance of these quality characteristics to their designed tolerances starting from the selection of the points to be measured on the actual part (inspection plan) is described in Capello and Semeraro (2001).[fig:5.4.2 Competing canditate calibration functions] These authors show how an obtained cloud of points is utilised in order to produce a 'substitute geometry'. A substitute geometry is the joint result of both the identification of a parametric family of elements (circle, plane, cylinder, etc.) from the available a priori knowledge of the part and of the estimation of the family parameters obtained by using some optimisation criterion (least squares method, minimum zone, etc.) with the experimental cloud of points. The evaluation of the deviation of the quality characteristics of interest from its nominal values is then performed on the basis of such a substitute geometry, thus enabling the comparison with the prescribed tolerances.

An acceptance or a rejection of the part can be therefore recommended. Some examples of this general framework can be found in the case of form error evaluations (cf. for instance Zhu et al., 2004 and Yang and Jackman, 2000) and in the case of angular error between two lines (cf. Huang, 2003). A specific applicative case is then presented in Shin and Kurfess (2004), where an alternative algorithm for identifying surfaces starting from a cloud of points is demonstrated on printed circuit boards. In a broader perspective, algorithms and methods used in inspection procedures based on computer vision are described in Marshall and Martin (1992).

On the other hand, when a measuring task is performed in order to gain an insight on the nature of some manufacturing process or on some natural object and phenomenon, the use of a priori knowledge in assisting the decision making process inherent in the reverse engineering procedure appears more questionable. For instance, if the aim of a measurement task is to evaluate the volume of a chip from a milling or turning operation, while fitting surfaces to the clouds of points the assumption of a particular measurand shape or of regions on the measurand should be supported by some pre-existent scientific evidence. Others examples where this situation is believed to hold are the measurement of the volume of craters of a single pulse during a laser operation, the measurement of the volume of a red cell in the blood of different animals, the measurement of the minimum and maximum cross section of a given tornado.

The main thrust of the method for measuring small volumes described in chapter 6 is to provide an investigative tool that matches both the needs mentioned above. However, priority has been given to the case where the shape of the measurand is irregular or complex and where pre-experimental information about the measurand is not available. In fact, if a method is provided for such a volume measuring task, then it is believed that it can also be applied if the measurand has boundaries of a simple shape (plane, sphere, cylinder, etc). In fact, in such cases, the deviation of this measurement results from those obtainable by using such pre-experimental information is negligible relatively to the estimate of repeatability standard deviation of the method itself. In Chapter 6, quantitative supporting evidence of this belief

and of the above statement is provided. The proposed method therefore, when fitting a surface to the cloud of points associated with a measurand, makes use of the simplest possibility which is the introduction of planar facets every three points. In this way, the method also benefit of the regular organisation in a grid of the projection of the data points on the $z = 0$ plane. However, it also makes use of a priori knowledge, although in a limited way, when the operator is asked to identify a segment in the cloud of points as a reference element and to select a parametric surface family (plane, sphere, etc.) for fitting it to such a segment. This use of pre-experimental knowledge is solely necessary to establish a reference element, so it affects the performances of the method in the same way as a selection of a reference element performed without using a priori knowledge would do.

2.5 Compendium

In the first section of this chapter, the practical utility of a more accurate measuring method of micro-volumes and of the volumetric wear ratio in $\mu - EDM$ has been shown. A measuring method based on the WLI microscope is considered as one promising approach to this issue.

Thus, in the second section, fundamental general concepts about the WLI microscope have been introduced. In addition, some particularly critical aspects in the deployment of the WLI microscope have been highlighted. Among them there is the calibration.

Hence, in the third section, some metrology concepts used in this chapter and throughout this study, have been introduced and critically analysed.

Finally, a number of fundamental concepts in the domain of reverse engineering and geometrical modelling are considered in order to put into a context the decisions taken when developing the proposed method for measuring small volumes.

Chapter 3

Building a traceable sample of length and investigating the effects of discretionary setup parameters

In this chapter fundamental issues connected with the practical usage of the WLI microscope are considered. In section one, a method for the realisation of traceable reference materials for measurements of lengths in the micrometre range is proposed. In section two, an estimate for the uncertainty associated with these reference materials is computed. In section three, the effects on the measurement results due to the scan length and to the initial position of the scanner element associated with a pre-specified stage orientation are investigated.

3.1 Building traceable reference samples of length

The most trusted sample of length available for this experimental study was a set of grade 1 (for the meaning of grade 1 and further details about the geometrical specifications of the blocks cf. BS EN ISO 3650) rectangular steel gauge blocks. The set had been previously provided with a certificate of calibration which testifies the compliance of each single block in the used set with the tolerances specified in

the standard BS 4311-3. The certificate of calibration is attached in Appendix A. This certificate shows the actual figures of the deviations of the central length from the nominal value. In addition, this document gives the values of uncertainty and the coefficient of thermal expansion that are utilised in this study. Particularly, for the gauge blocks that are used, i.e. for gauge blocks with nominal lengths, $l_{n,i}$, less than or equal to $10\,mm$, the deviation, Δ_i, of the central length $l_{c,i}$from the nominal length is in the limits of $\pm 0.25\,\mu m$ prescribed by the standard BS 4311-3 (cf. ibidem table 1). The variation in length[1], v, is also smaller than the prescribed limit of $0.16\,\mu m$. It must be noted that the limit for deviation of the central length from the nominal length quoted from BS 4311-3, i.e. $\pm 0.25\,\mu m$, might appear somehow in contradiction with the 'limit deviation of length at any point from nominal length', t_e. In fact, from table 5 of BS EN ISO 3650, it results $\pm t_e = \pm 0.2\,\mu m$. The reason could be that t_e is expressed with two figures, thus it represents the interval [0.16 ; 0.25]. Instead the limits for $l_{c,i}$are given by three figures, therefore they represents the intervals [0.246 ; 0.255] and [−0.246 ; −0.255]. Both of them still comply with the more general prescriptions of BS EN ISO 3650.

In order to have a reference length that can vary from micrometres up to a few hundred micrometres, the following procedure is proposed.

- Firstly, an auxiliary plate with a planar surface of the same texture as the measuring faces of the gauge blocks [2] is necessary to meet the definition of length of a gauge block (cf. BS EN ISO 3650). For this purpose, similarly to Malinosvsky et al. (1998), Malinovsky et al. (1999) and Decker and Pekelsky (1997), a quartz circular optical parallel of either $12.25\,mm$ or $12.37\,mm$ thickness with diameter of $30\,mm$, was used. The manufacturer claims that the measuring faces of the parallel were produced in compliance with parallelism and flatness tolerance specifications of $0.2\,\mu m$ and $0.1\,\mu m$, respectively. These tolerances appear to be of the same order of those imposed on the gauge

[1] for the definition of central length and of variation in length, cf. the standard BS EN ISO 3650.

[2] for more information on measuring faces cf. section 3.1 in BS EN ISO 3650.

blocks. In fact, for nominal lengths, $l_{n,i}$, in the range $[0.5\,mm;\ 150\,mm]$, which are to be used, BS EN ISO 3650 prescribes for grade 1 blocks a flatness tolerance $0.15\,\mu m$ (cf. table 3, in the mentioned standard). In addition, BS 4311-3 does not specify any larger tolerance for the flatness of gauge blocks in use (cf. section 7.3, ibidem).

- Secondly, two gauge blocks of nominal length $l_{n,1}$ and $l_{n,2}$, respectively, were wrung to the same auxiliary plate one beside the other, so they were in contact along one of the two long side surfaces. It can be argued, however, that the side faces were in contact in a limited number of points. In fact, For the side faces, BS EN ISO 3650 prescribes a flatness tolerance of $40\,\mu m$ for nominal length $l_{n,i} \leq 100\,mm$ and a perpendicularity tolerance of a side face with a measuring face as a datum of $50\,\mu m$ for $10\,mm \leq l_{n,i} \leq 100\,mm$. For the reference sample of length a nominal value $l_{n,s}$ can be established by the following equation:

$$l_{n,s} = l_{n,1} - l_{n,2} \quad (with\, l_{n,1} > l_{n,2}) \tag{3.1}$$

A photograph of a generic artefact obtained using the proposed method is shown in Figure 3.1. Moreover, a schematic representation of the arrangement for defining the reference sample of length is outlined in part (a) of Figure 3.2.

The nominal length of the step is just a conventional value. Let L_i be the actual length of the $i-th$ gauge block in a generic point on the unwrung measuring face (with $i = 1, 2$). It follows that the actual length of the step relative to two generic points on the two unwrung measuring faces of the two blocks, L_s, is given by the following equation:

$$L_s = L_1 - L_2 \tag{3.2}$$

The actual length of the step, L_s, is changing when being calculated relatively to different points on the two blocks. In fact, the length $L_i\,(i = 1, 2)$ is expected to vary from point to point on the unwrung measuring face of the $i-th$ gauge block.

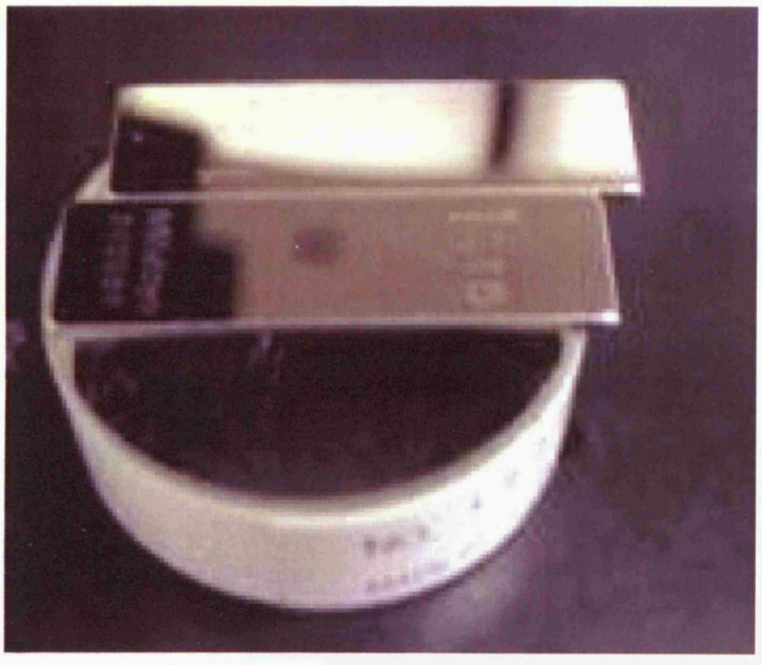


Figure 3.1: Photograph of a generic artefact obtained by means of the proposed method.

This circumstance is illustrated in part (b) of Figure 3.2, which has been adapted from Figure 3 in BS EN ISO 3650.

However, from the tolerance limits specified in BS EN ISO 3650, it is possible to derive the range of lengths in which L_s is expected to lie. Namely, the tolerance zone for L_i is completely defined by the following equation:

$$LSL_{L_i} \leq L_i \leq USL_{L_i} \tag{3.3}$$

Where LSL_Q and USL_Q are, respectively, the *lower specification limit* and the *upper specification limit* of the generic quantity 'Q'. For $Q = L_i$, BS EN ISO 3650 prescribes (Figure 3.3):

$$LSL_{L_i} = l_{n,i} - t_e \quad (with\, i = 1, 2) \tag{3.4}$$

$$USL_{L_i} = l_{n,i} + t_e \quad (with\, i = 1, 2) \tag{3.5}$$

Therefore, it follows:

$$LSL_{L_1} - USL_{L_2} \leq L_s \leq USL_{L_1} - LSL_{L_2} \tag{3.6}$$

From equations (3.5), (3.5) and (3.6) it results:

$$l_{n,1} - l_{n,2} - 2 \cdot t_e \leq L_s \leq l_{n,1} - l_{n,2} + 2 \cdot t_e \tag{3.7}$$

Equations (3.6) and (3.7) define, indirectly, specification limits for L_S. More formally, it holds:

$$LSL_{L_S} = LSL_{L_1} - USL_{L_2} = l_{n,1} - l_{n,2} - 2 \cdot t_e \tag{3.8}$$

$$USL_{L_S} = USL_{L_1} - LSL_{L_2} = l_{n,1} - l_{n,2} + 2 \cdot t_e \tag{3.9}$$

Equation (3.7) has to satisfy the physical constraint of non-negativity for a length. This results in the following inequality:

$$l_{n,1} - l_{n,2} \geq 2 \cdot t_e \tag{3.10}$$

Due to the fact that $min\,(2 \cdot t_e) = 0.4\,\mu m$ (cf. BS EN ISO 3650) and that $min\,(l_{n,1} - l_{n,2}) = 0.5\,\mu m$ (see the certificate of calibration in Appendix A), equation (3.10) is always satisfied.

A schema of the two blocks with the main quantities is illustrated in Figure 3.3.

In order to avoid significant bias affecting L_s, particular attention must be paid to handling the gauge blocks and to the wringing process. In fact, touching the block with bare hands puts them in contact with a body at a temperature between 34 and 35 °C[3]. Thus, in the very worst case of a long lasting contact with a bare hand, a gauge block of actual length L , expressed in mm, can present a bias given by the following equation

$$\Delta l = \alpha_{th} \cdot L \cdot \Delta T = 10.8 \cdot 10^{-6} \cdot 10^{6} \cdot L \cdot (34.5 - 20)\,nm = 156.6 \cdot L\ nm \tag{3.11}$$

where the coefficient of linear thermal expansion, α_{th}, is provided by the calibration certificate (cf. Appendix A) and where it has been assumed an average hand temperature of 34.5 °C. The gauge blocks were, therefore, handled wearing thermal insulating gloves and they were measured after waiting for some time in order

[3]The human body is only approximately isotherm. In fact, different parts have different temperatures, owing to different local blood flow. For instance, the knees has a temperature in the range [34 ; 35] °C, the cheek in the range [33 ; 34] °C, the ear in [31 ; 32] °C and the axilla in [36 ; 37] °C (cf. CIRDU).

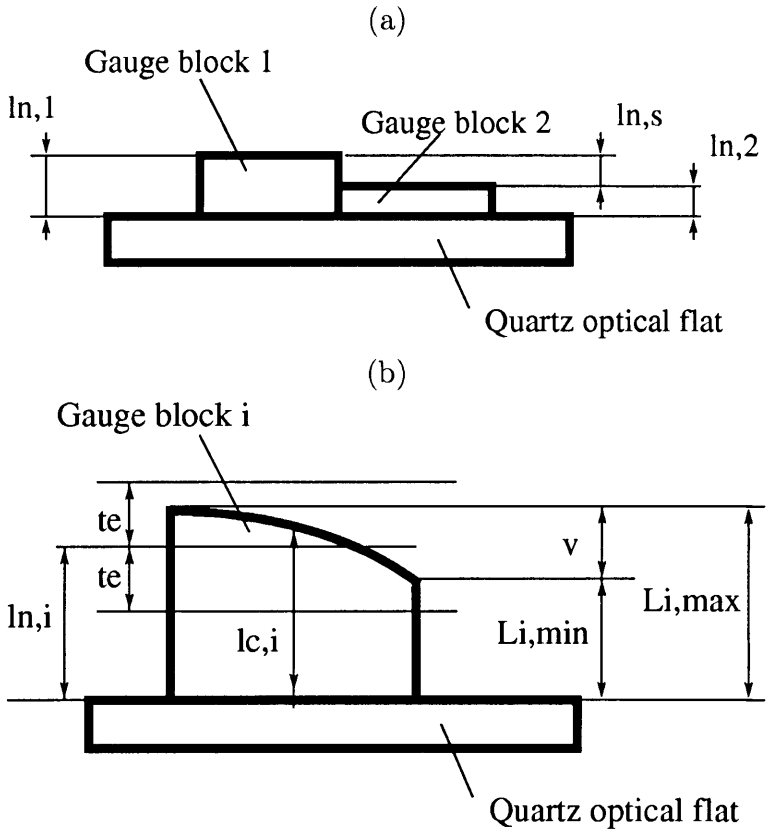


Figure 3.2: (a) Reference sample of length. (b) Main lengths and tolerances for the $i-th$ gauge block (After BS EN ISO 3650, with adjustments).

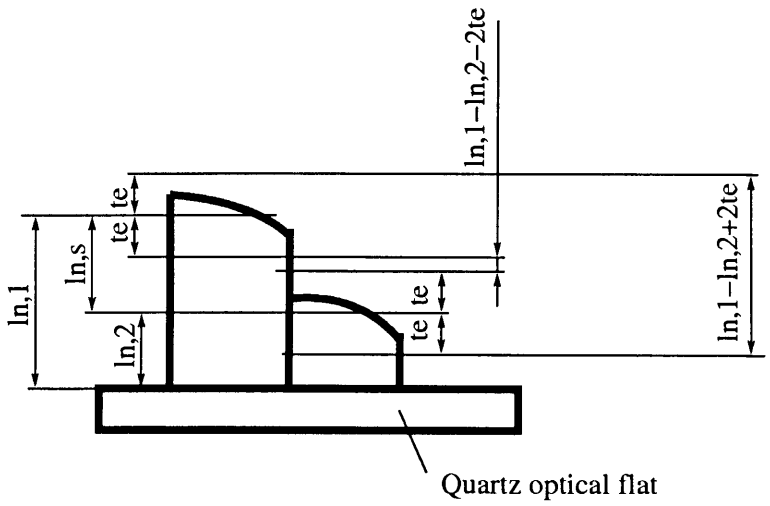


Figure 3.3: Exhaustive range of values for the length L_s of the reference step.

to allow the measurand blocks to reach a temperature close to that of the air (cf. section A.3 in BS 4311-3). Decker and Pekelsky (1997) advised leaving the blocks untouched overnight in order to stabilize them thermally. However, this research is constrained by time limits and, in addition, the particular aim of the measuring task under investigation is less demanding than in Decker and Pekelsky (1997). It is suggested, therefore, that fifteen minutes of thermal stabilisation time is sufficient. In fact, not only were the blocks stored in the same room as the interferometric microscope used for measuring them, but meticulous care was put in the thermal insulation of the hands while handling them in order to prevent the potential pitfalls of a relatively short thermal stabilisation time. The study from Scarr (1967)[4], in which it is shown the cooling curve for a $25.4\,mm$ slip gauge that was held in the hand for three minutes and then put in a stable ambient temperature[5], supports the fifteen minutes of thermal stabilisation. In particular, this cooling curve showed that the deviation from the nominal length went from $2.032\,\mu m$ to about $0.254\,\mu m$ after fifteen minutes and to about $0.127\,\mu m$ after thirty minutes.

Wringing is the procedure by which a gauge block is made to adhere tightly to the surface of the auxiliary plate, so that the block is fixed firmly in position. The two units, the block and the plate, when wrung, can also be handled as a single part (Decker and Pekelsky, 1997). Although Decker and Pekelsky (1997) suggested putting a very small amount of light textured oil on the surface of the plate and then cleaning it with a lint-free cloth, Malinovsky et al. (1999) showed that the interposition of oil between the plate and the block had the effect of increasing the deviation from the nominal length of the block by about $2\,nm$. They also showed that the space between the plate and the block increases with time due to the presence of oil. In addition, neither BS 4311-3 in section 8.2 nor BS EN ISO 3650 in section A.4 mentioned the usage of oil. Thus, no oil or any other substance was used in the wringing of the gauge blocks on the optical parallel surface. First the anti-oxidant liquid film that protects the blocks was removed using an alcoholic solution, then the blocks and the parallel were wiped free from dust as prescribed

[4] Cf. ibidem pages 11 and 12.

[5] No actual value for the temperature of the air was given.

in section A.4 of BS 4311-3[6].

The use of the quartz optical parallel, although with flatness tolerance insufficient for calibration of gauge blocks (cf. section 8.2 in BS EN ISO 3650), allows an assessment of the quality of the wringing by observing the wrung measuring face through the optical flat (cf. section 8.2 in BS EN ISO 3650 and section 4.1 in Decker and Pekelsky, 1997). A wringing is good if the look of the measuring face is uniformly wet. This indicates uniform contact between the gauge block and the parallel. Another sign of uniform contact is if interference bands, colours and bright spots are not visible (cf. section 8.2 in BS EN ISO 3650 and section 4.1 in Decker and Pekelsky, 1997). For blocks of grade one, BS EN ISO 3650 specifies that a wringing is still considered valid even if some bright spots or shades are visible. According to Decker and Pekelsky (1997), the space between the parallel and the gauge block, also known as wringing film, is 'very reproducible' and its thickness when the wringing is judged good is $10 \pm 5\,nm$ (cf. ibidem, section 4.1). [7]

An analysis of the uncertainty which characterises the reference length L_s is needed. In fact, unavoidable randomness affects both the WLI measurement process and the preparation of the reference sample of length. It is anticipated that the WLI measurements will not be a perfect match to L_s. Thus, the aforementioned analysis of uncertainty is required in order to establish to which of the following factors these deviations can be ascribed:

- the randomness affecting the measurement process;
- the nuisance random factors in the preparation of the reference sample of length;
- the lack of calibration of the instrument.

The analysis is presented in the next section.

[6] After usage the anti-oxidation film layer is restored as prescribed in section A.6 of BS EN ISO 3650.

[7] Neither the confidence level nor the coverage factor k is provided for such expanded uncertainty (cf. BSI PD 6461-4 for the definitions of the used terminology).

3.2 Investigating the uncertainty of the proposed reference length

The fact that the sample of length is used as a trusted length for the calibration of the white light interferometer microscope prevents an analysis of the uncertainty of the sample of length via a series of measurements taken by the same instrument. Furthermore, a measurement instrument significantly more accurate than that under investigation is not available for this investigation. In this context, the word 'accurate' is used with the meaning both of precision, i.e. variability between repeated measurements, and trueness, i.e. difference between the unknown true value or a reference value and the level of the measures given by the instrument (cf. sections from 0.4 to 0.6 in BS ISO 5725-1). The specification of the precision required to a measuring process in order for it to be considered adequate in an uncertainty study can be quantitatively set. In particular, similarly to Montgomery and Runger (1993), the ratio between the standard deviation of repeated measurements of the sample length in predefined experimental conditions and the expected standard deviation of the measuring process can be investigated. The standard deviation of the measuring process in the predefined experimental conditions must be known by previous studies. In analogy with the precision-to-tolerance ratio, a threshold for this ratio could be set to a lower limit of ten, which is frequently encountered (cf.Montgomery and Runger, 1993). Thus, on the basis of these considerations, this experimental approach to the assessment of the uncertainty, also known as type A evaluation of uncertainty (cf. BSI PD 6461-4 in section 8.1 and 8.2), does not appear to represent a practical viable solution for the particular case under study. As a consequence, the uncertainty assessment was based on information surveyed in published studies. This second approach is referred to as Type B uncertainty evaluation (cf. section 8.3 in BSI PD 6461-4).

The main sources of uncertainty that have been identified and that it has been possible to evaluate quantitatively are the following:

- the environmental conditions (air temperature, humidity and pressure);

- the wringing procedure, including the flatness deviation both of the auxiliary plate and of the gauge block surfaces.

3.2.1 Environmental conditions

The contribution to the uncertainty of L_s given by variations of the barometric pressure and humidity from the standard conditions[8], were not taken into account. In fact, they are considered relevant only in the calibration of gauge blocks (cf. page 11 in Scarr, 1967, together with sections 5.3 and 8.3.3 in BS EN ISO 3650).

The room where the measurement tasks were carried out was thermostatically controlled at reference temperature of 20 °C. Nevertheless, variations in the air temperature are expected. In fact, a temperature sensor with a data logger was placed close to the white-light interferometric microscope. The time series of temperature values were automatically recorded every three minutes and they showed deviations from the set point temperature. Such deviations together with a potential residual thermal effect from the handling procedure on the temperature of the gauge blocks, represent a source of uncertainty for L_1 and L_2. In addition, as a precautionary measure, it has been preferred to potentially overestimate rather than underestimate the uncertainty of L_s. Therefore, it has been assumed that the temperature of a gauge block lies in the range $[a\,,\,b] = [18.5\,,\,21.5]$ °C with a probability of 100 % and that this temperature is distributed according to a triangular distribution of mode $c = 20$ °C.[9]. From Weisstein (2005), and from the coincidence of the second central moment with the variance, it follows[10]:

$$\sigma_{T_{actual}} = \sqrt{\frac{1}{18}\left(a^2 + b^2 + c^2 - ab - ac - bc\right)} = 2.5981 \;\; ^oC \qquad (3.12)$$

[8] In section 5.3, BS EN ISO 3650 specifies only the reference temperature of 20 °Cand the reference pressure of 101 325 Pa.

[9] The assumptions made are consistent with the methodology in sections 8.3.4 and 8.3.6 of BSI PD 6461-4.

[10] BSI PD 6461-4 presents analytically only the particular case $a = -A$, $b = A$, $c = 0$ with $A > 0$ and constant.

$$\mu_{T_{actual}} = \frac{1}{3}(a+b+c) = 20 \ \ ^{o}C \tag{3.13}$$

Where $\sigma_{T_{actual}}$ is the standard deviation of the air temperature under the assumptions previously set and $\mu_{T_{actual}}$ is its mean. Similarly to the method, the terminology and the symbols illustrated in BSI PD 6461-4, it holds:

$$\Delta l = l_{th} - L \tag{3.14}$$

The expression (3.14) defines the quantity l_{th} ,which is the length in a generic point of the block after a thermal expansion has taken place due to the deviation of the block temperature from its set reference (20 °C). From equation (3.11), therefore it follows:

$$l_{th} = L + \alpha_{th} L \left(T_{actual} - T_{reference}\right) \tag{3.15}$$

$$u\left(T_{actual}\right) = \sigma_{T_{actual}} = 2.5981 \ \ ^{o}C \tag{3.16}$$

In equation (3.16), $u\left(T_{actual}\right)$ is the standard uncertainty associated with the input estimate T_{actual}, which is the actual temperature of the block under the hypotheses previously introduced.

$$c_T = \frac{\partial l_{th}}{\partial T_{actual}} = \alpha L = 10.8 \cdot 10^{-6} L \ \ \frac{mm}{K} \quad (L \, in \ mm) \tag{3.17}$$

$$u_{T_{actual}}\left(L_i\right) = \frac{\partial l_{th}}{\partial T_{actual}} \cdot u\left(T_{actual}\right) = 2.8059 \cdot 10^{-5} L_i \ \ mm \ \ \ (i = 1,2; \ L \, in \, mm) \tag{3.18}$$

In equation (3.18), the quantity $u_{T_{actual}}(L_i)$, is the component of the combined standard uncertainty $u_c(L_i)$, of the output estimate L_i, due to the standard uncertainty $u(T_{actual})$ of the temperature T_{actual}, the input estimate.

Another contribution to the uncertainty of L_i , $u_c(L_i)$, is connected with variations in temperature of the gauge blocks and is given by the uncertainty of α_{th} in equation (3.15). In fact, in the calibration certificate the measure of α_{th} is reported with an uncertainty of $0.5 \cdot 10^{-6}\ K^{-1}$, without further details. According to BSI PD 6461-4, a Normal distribution is commonly assumed in such circumstances (cf. ibidem on page 11). The coverage factor reported in the calibration certificate is $k = 2$, which corresponds to a level of confidence of 95.45 % (cf. Appendix A). Thus, it follows:

$$u(\alpha_{th}) = \frac{U(\alpha_{th})}{k} = 0.25 \cdot 10^{-6}\ K^{-1} \tag{3.19}$$

$$c_{\alpha_{th}} = \frac{\partial l_{\alpha_{th}}}{\partial \alpha_{th}} = L(T_{actual} - T_{reference}) = L\,T_{actual} - 20\,L\ \frac{mm}{K} \tag{3.20}$$

$$u_{\alpha_{th}}(L_i) = \frac{\partial l_{\alpha_{th}}}{\partial \alpha_{th}} \cdot u(\alpha) = 0.25 \cdot 10^{-6} L_i\,T_{actual} - 5 \cdot 10^{-6} L_i\ mm \quad (i = 1, 2) \tag{3.21}$$

In equations (3.20) and (3.21) T_{actual} is expressed in degrees Celsius and $l_{\alpha_{th}}$, is the length of the block in a generic point after a variation in α_{th} has taken place. In equation (3.19), $U(\ldots)$, is the expanded uncertainty of the generic estimate in parentheses.

3.2.2 Wringing procedure

The possibility of assessing the quality of the wringing and of repeating the procedure when finding it unsatisfactory, can significantly increase the precision of the wringing

itself.[11] In detail, it is believed that factors affecting the variability of repeated measurements of L_i both in repeatability and reproducibility conditions are the following:

- The modification of the surface roughness of the blocks of a few nanometers in repeated wringing (Malinovsky et al., 1999).
- The position of the gauge block on the plate (cf. Malinovsky et al., 1999 and Decker and Pekelsky, 1997). This could be explained by the different local conditions caused by roughness, waviness and flatness of both the auxiliary plate and the blocks.
- The necessary cleaning operation after each measurements (Malinovsky et al., 1999).
- The wringing film is reported as an influential factor by different authors. Nevertheless, according to Malinovsky et al. (1999) the wringing layer is much more reproducible than in Decker and Pekelsky (1997).

In order to be conservative, the highest value of uncertainty found in the literature has been selected as indicative of the experimental conditions under investigation, namely:

$$u_w(L_i) = 8 \cdot 10^{-6}\, mm \quad (with\, i = 1, 2) \tag{3.22}$$

The figure in equation (3.22) is obtained by combining the contributions to the combined standard uncertainty $u_c(L_i)$ due to the factors 'wringing' and 'gauge block flatness and parallelism' so as they appear in table 2 of Decker and Pekelsky (1997). In fact, both factors seem to be necessary fully to cover the list above and not the sole factor 'wringing film'. In order to be on the conservative side, the combination has been made under the assumption of perfect positive correlation between the two

[11] Cf. page 1 in BS ISO 5725-1 for the meaning of 'precision' in this context.

factors rather than under the hypothesis of independence. In fact, under such a hypothesis, $u_w(L_i)$ is given by the sum of the contributing uncertainties rather than by the square root of the sum of the squared contributing uncertainties (cf. sections 10.2 and 10.3 in BSI PD 6461-4).[12]

3.2.3 Results of the investigation

The calculations of the combined standard uncertainty $u_c(L_i)$ of the length L_i, with $i = 1, 2$, are summarised in table (3.1) . Table (3.1) is based on the layout named 'The uncertainty budget table' proposed in section B.2 of BSI PD 6461-4.

In order to be on the conservative side, so as the estimate of the uncertainty is larger rather than smaller than the true unknown uncertainty value, the temperature and the coefficient of thermal expansion have been assumed perfectly positively correlated, namely with correlation coefficient equal to one. The contribution of T_{actual} and α_{th} to the squared uncertainty is, therefore, equal to the square of the sum of their contribution to the uncertainty (cf. Mood et al., 1974 page 178). Furthermore, it is not possible to assess the degree of correlation between 'wringing' and the group of 'temperature' and 'coefficient of thermal expansion'. Therefore, the two extreme cases of non-correlation and perfect positive correlation have been analysed in table (3.1) as it is suggested in section 10.4 of BSI PD 6461-4. In the case of non-correlation, the contribution of the two aforementioned components to the squared uncertainty is the sum of their squared contribution to uncertainty (cf. Mood et al., 1974 page 178 and BSI PD 6461-4 section 10.2). Whereas, in the case of perfect correlation, their contribution to the combined standard uncertainty is the sum of their contribution to uncertainty (cf. Mood et al., 1974 page 178 and BSI PD 6461-4 section 10.3). The expression of the combined standard uncertainty, both in

[12] With the terminology of the theory of probability, the variance of a sum of two random variables with correlation coefficient of one is equal to the square of the sum of the standard deviations (cf. pages 155 and 179 in Mood et al., 1974). Such sum is bigger than the sum of the variances of the two random variables, i.e. it is bigger than the variance of the sum of random variables under the hypothesis of Independence.

	Input quantity, x_i		
	Temperature	Coefficient of thermal expansion	Wringing
Expanded uncertainty, $U(x_i)$		$0.5 \cdot 10^{-6}$ K^{-1}	
Probability distribution	Triangular	Normal	Normal
Divisor, k		2	
Standard uncertainty, $u(x_i)$	2.5981 °C	$0.25 \cdot 10^{-6}$ K^{-1}	8 nm
Sensitivity coefficient, c_{x_i}	$10.8 \cdot 10^{-6}$ mmK^{-1}	$L \cdot T_{actual} - 20 \cdot L$ mmK^{-1}	1
Contribution to uncertainty, $c_{x_i} \cdot u(x_i)/nm$	$28.059 \cdot L$	$0.25 \cdot L \cdot T_{actual} - 5 \cdot L$	8
independent group contribution, $(c_{x_i} \cdot u(x_i))^2/nm^2$			64
Correlated group contribution	$L^2 \cdot (0.0625 \cdot T_{actual}^2 + 11.530 \cdot T_{actual} + 531.72)$		
Squared uncertainty (case 'uncorrelated'), $u_c^2(L_i)/nm^2$ $L^2 \cdot (0.0625 \cdot T_{actual}^2 + 11.530 \cdot T_{actual} + 531.72) + 64$			
Combined standard uncertainty (case 'uncorrelated'), $u_c(L_i)/nm$ $\sqrt{L^2 \cdot (0.0625 \cdot T_{actual}^2 + 11.530 \cdot T_{actual} + 531.72) + 64}$			
Simplified combined standard uncertainty (case 'uncorrelated'), $u_c(L_i)/nm$ $\sqrt{808.51 \cdot L^2 + 64}$			
Combined standard uncertainty (case 'correlated'), $u_c(L_i)/nm$ $L \cdot (23.059 + 0.25 \cdot T_{actual}) + 8$			
Simplified combined standard uncertanty (case 'correlated'), $u_c(L_i)/nm$ $28.434 \cdot L + 8$			

Table 3.1: Uncertainty budget table for L_i(with $i = 1, 2$). L is in millimetres, T_{actual} in degrees Celsius.

the case of correlation and of non-correlation [13], results a function of T_{actual}. Thus, a more conservative and simpler expression has been proposed in table (3.1), where T_{actual} has been given its anticipated maximum value, i.e. 21.5 °C. It can be verified that the simplified combined standard uncertainty in the case of perfect correlation is always larger than in the case of uncorrelation, given that $L > 0$. Thus, following a conservative approach, in the following sections the sole expression of combined standard uncertainty that is utilised is one that is simplified and derived under the hypothesis of perfect correlation, that is:

$$u_c(L_i) = 28.434 \cdot L + 8 \ nm \ \ (i = 1,2;\ L \ in \ mm) \tag{3.23}$$

From equation (3.2) and adopting a conservative approach which leads to the consideration of L_1 and L_2 as independent random variables [14], it follows:

$$u_c(L_S) = \sqrt{u_c^2(L_1) + u_c^2(L_2)} \ nm \ \ (L \ in \ mm) \tag{3.24}$$

$$u_c(L_S) = \sqrt{(28.434 \cdot L_1 + 8)^2 + (28.434 \cdot L_2 + 8)^2} \ nm \ \ (L \ in \ mm) \tag{3.25}$$

In this investigation, the calculations of $u_c(L_S)$ have been made by neglecting the differences between the infinite number of lengths realised by a single reference sample. $L_i (i = 1,2)$ in equation 3.25, has therefore been approximated by the corresponding nominal length adjusted with the deviations from the nominal at the centre of each block. These deviations are provided by the certificate of calibration (cf. Appendix A).

[13] Non-correlation is less constraining than independence. In fact, independence implies non-correlation, but the contrary is not true (cf. Mood et al., 1974 page 161).

[14] In fact, excluding the possibility of negative correlation, which is not expected for physical reasons, the independence is the situation in which the difference of two random variable has maximum variance (cf. Mood et al., 1974 page 179 equation (10)).

3.3 Studying the effects of discretionary setup parameters

In the actual usage of the WLI microscope, the length of the distance covered by the scanner element, also called scan length, and the initial position of the scanner element together with the orientation of the stage, have to be set before starting the measurement process.

On the one hand, in appendix D of ADEP1 on page D-18, it is stated that the scan length should be set at a value between 10 % and 20 % larger than the nominal length of the reference material. On the other hand, in ADEP2 on page D-17, it is stated that the scan length should be set at a value approximately 30 % larger than the aforementioned nominal length. Although only ADEP2 is consistent with the actual version of the software installed on the instrument, such differences in the prescribed scan lengths appear unjustified on the basis of the software and hardware configuration. Thus, an experimental investigation was carried out to shed some light on the relationship between the scan length and the measurement result.

Once the orientation of the stage supporting the part to be measured has been set, a range of positions of the scanning element gives rise to a different number of fringes on the part to be measured. On the one hand, on page D-17 of ADEP1, it is prescribed an adjustment of the initial scanner element position and the stage orientation so as the number of fringes in the field of view of the instrument is no more than three. On the other hand, on page D-17 of ADEP2, it is stated that the scanner element position and the stage orientation should be set up so as the same number of fringes is three. In general, once the stage orientation has been set, a degree of freedom appears to be left to the operator while setting the initial scanner element position. Hence, an experimental investigation was carried out in order to assess the effect, if any, of the initial scanner element position on the measurement result.

3.3.1 Scan length effect

In order to investigate the relationship between the scan length and the measurement results, two reference lengths of nominal 40 μm and 200 μm were prepared according to the method described in section 3.1. For each of them, scan lengths exceeding the nominal reference length of 10%, 15%, 20%, 25% and 30% were experimentally investigated experimentally. The designed experiment was replicated eight times. Furthermore, the sequence in which the tests were carried out was randomly selected as suggested on page 61 of Montgomery (2001). In addition, each measurement was taken in the same area of the reference length sample, without any repositioning of the stage. The resulting experimental data are illustrated in Figure 3.4.

The measurements were carried out with the same setting parameters. In particular, the height correction was conventionally set to one. In fact, under investigation is the sole effect of the scan length on the measurement results, while keeping all the other setting parameters constant, regardless of the potential systematic error that may arise in adopting this experimental strategy. This resulted in measurement results always smaller than the corresponding nominal values for both the reference lengths. Thus, the deviations and the percentage deviations from the nominal were always negative. Hence, in Figure 3.4 the absolute value of the percentage deviations has been shown so as to make them easy to read.

In Figure 3.4, it can be seen qualitatively that, by increasing the scan length, the measurement results tend also to increase. Although it might be suspected that the measurement results would be different for the different scan lengths under investigation, such a conclusion could be due to randomness in the data rather than to the fact that the measurement results are indeed different for different scan lengths. In order to support quantitatively such a qualitative conclusion, an *analysis of variance model*, ANOVA , has been fitted to the data.

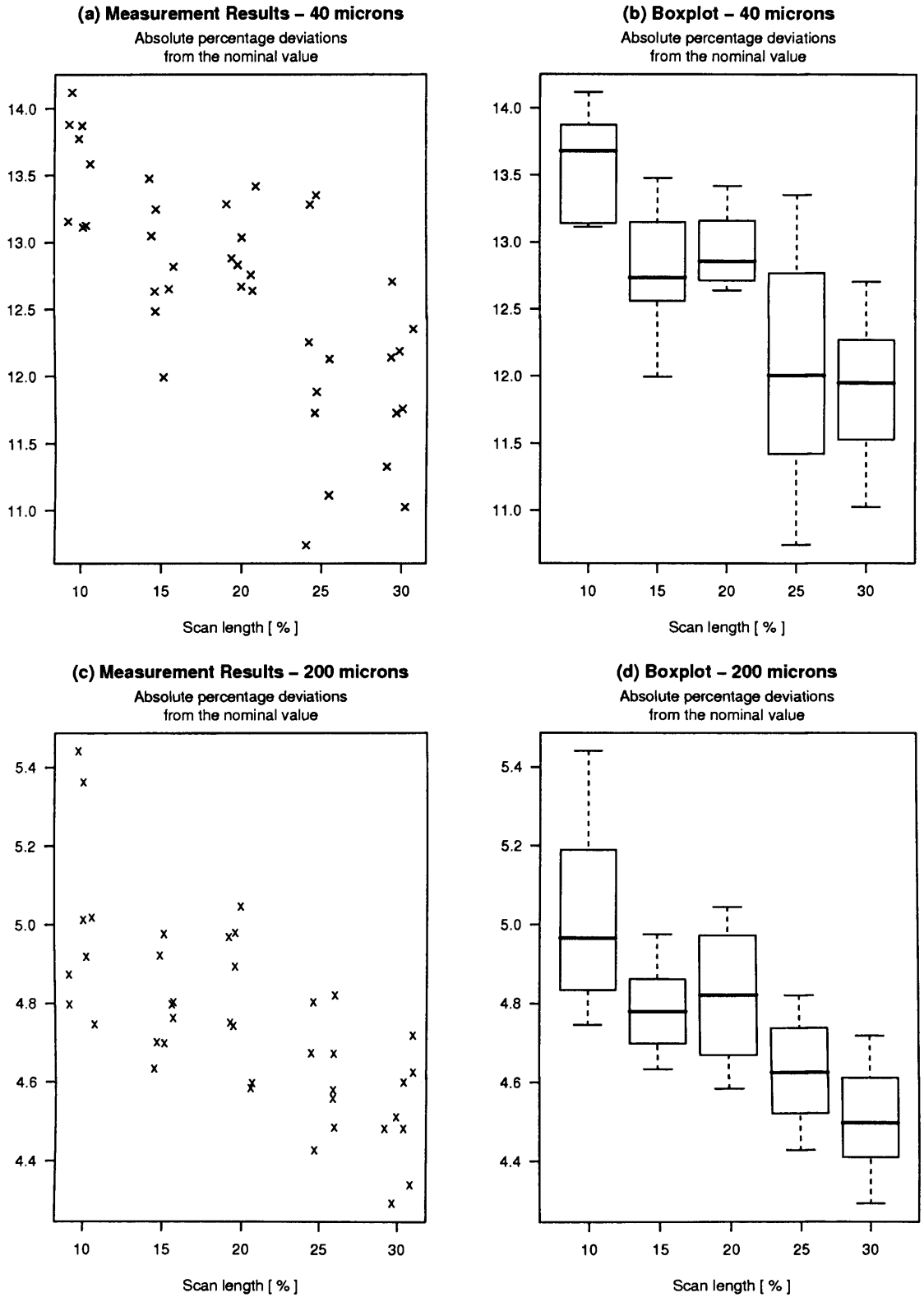


Figure 3.4: Measurement results for scan lengths varying in the range [10 %, 30 %]. (a) and (b) Nominal length 40 μm. (c) and (d) Nominal length 200 μm.

3.3.1.1 Analysis of the results

The percentage deviation of the $j-th$ measurement result from the nominal length of the reference sample, when measuring with the $i-th$ scan length, was set as the response variable, s_{ij} , of the ANOVA model. The scan lengths investigated were considered a random sample from a population of infinitive potential scan lengths that can be drawn in the interval [10%, 30%]. It is suggested that the validity of the conclusions that can be drawn from such a sample should not be limited to the specific five scan lengths investigated. In fact, in the case under study, it is preferable that the conclusions apply to the whole range [10%, 30%]. These considerations led to the selection of a one factor random effects ANOVA model for the analysis of the data (cf. Montgomery, 2001, chapter 12 for more details). Formally, it holds:

$$s_{ij} = \mu_s + t_{s,i} + e_{ij} \;\; with\, i = 1, \ldots, a \;;\; j = 1, \ldots, n \tag{3.26}$$

The model parameter μ_s is called the overall mean. The symbol $t_{s,i}$ denotes the random variable representing the effect of the i-th level of the factor scan length. The $\{t_{s,i}\}$ are assumed to be normally and independently distributed random variables with mean zero and constant variance $\sigma_{t_s}^2$, shortly $\{t_{s,i}\}$ $NIID\left(0, \sigma_{t_s}^2\right)$. The quantities e_{ij} are called the random error components. These quantities are random variables which account for all the sources of variability in s_{ij} that are not due to the scan length, if any variability due to the scan length is present. It is assumed that $\{e_{ij}\}$ $NIID\left(0, \sigma^2\right)$. This last assumption implicitly encompasses $E\left(e_{ij}\right) = 0$, $E\left(e_{ij}^2\right) = \sigma^2$ and $E\left(e_{ij} \cdot e_{i'j'}\right) = 0$, which in the literature are referred to as Gauss-Markov conditions (cf. section 2.5 in Sen and Srivastava, 1990). It is also required that $\{t_{s,i}\}$ and $\{e_{ij}\}$ are independent.

On the basis of this independence, from equation (3.26) it results (cf. also the corollary on page 178 in Mood et al., 1974) :

$$V\left(s_{ij}\right) = \sigma_{t_s}^2 + \sigma^2 \tag{3.27}$$

The symbol $V(...)$ indicates the variance of the random variable in parentheses. The quantities σ^2_{ts} and σ^2 are called variance components (Montgomery, 2001, page 518). The numerical calculations in the analysis of variance are the same as in the fixed effect model. Particularly, the following equations hold (cf. Montgomery, 2001, pages 66-69 and page 513):

$$MS_E = \frac{SS_E}{a \cdot (n-1)} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(s_{ij} - \overline{s}_{i.})^2}{a \cdot (n-1)} \tag{3.28}$$

$$MS_{factor} = \frac{SS_{factor}}{a-1} = \frac{n \cdot \sum_{i=1}^{a}(\overline{s}_{i.} - \overline{s}_{..})^2}{a-1} \tag{3.29}$$

$$E(MS_E) = \sigma^2 \tag{3.30}$$

$$E(MS_{factor}) = \sigma^2 + n\sigma^2_{t_s} \tag{3.31}$$

In equation (3.30) and (3.31), $\overline{s}_{i.}$is the average of the observations at the i-th level of the factor, i.e. i-th cell, whereas $\overline{s}_{..}$ is the grand average of all the observations. The ANOVA method of estimation of $\sigma^2$and $\sigma^2_{t_s}$ consists in equating the mean squares to their expected values so as to have a set of equations that are linear in respect of the unknown parameters. In the case under investigation, they are the equations (3.30) and (3.31). The ANOVA method of estimation of the variance components, can therefore be considered as a particular case of the method of the moments (cf. Mood et al., 1974, section 2.1). Consequently, from equation (3.30) and (3.31), it follows:

$$\hat{\sigma}^2 = MS_E \tag{3.32}$$

$$\hat{\sigma}^2_{t_s} = \frac{MS_{factor} - MS_E}{n} \tag{3.33}$$

where the symbol '$\hat{\xi}$' denotes an estimate of the unknown parameter ξ[15]. A more formal and complete description of this model is presented in chapter 1 and chapter

[15]In this study, Greek characters indicates unknown parameters, whereas Roman characters random variables.

3 of Searle and Casella (1992).

If the variance component $\sigma^2_{t_s}$is zero, then it can be argued that there is no effect of the scan length on the measurement results. In the analysis of variance, a test of hypothesis is carried out in order to draw some conclusion about $\sigma^2_{t_s}$ in the light of the experimental evidence.

The test procedure is as follows (more details are in Montgomery, 2001, Chapter 12):

$$H_0 : \sigma^2_{t_S} = 0 \quad H_1 : \sigma^2_{t_S} > 0;$$

$$F_0 = \frac{MS_{factor}}{MS_E}; \tag{3.34}$$

$$F_0 \quad F_{a-1,a(n-1)}. \tag{3.35}$$

Equation (3.34) describes the test statistic, whose interpretation is clarified by equations (3.30) and (3.31). In fact, according to these equations, F_0 is expected to be large when H_0 is not true. In equation (3.35), the probability density function of F_0 when H_0 is true, is displayed with the appropriate degrees of freedom.

The analysis conducted for the reference sample of nominal length 200 μm and 40 μm resulted in $F_0 = 10.331$ with a corresponding p-value $1.2 \cdot 10^{-5}$, and in $F_0 = 11.592$ with a corresponding p-value $4.253 \cdot 10^{-6}$, respectively. The p-value can be defined as the smallest significance level, or equivalently the smallest first type error, that would lead to a rejection of H_0 (cf. page 37 in Montgomery, 2001). In this case, H_0 is rejected for any reasonable pre-specified first type error, such as 1%, 5% and 10%. Although the mentioned levels are the most widespread in the community both of the academics and the practitioners, it appears appealing to select the first type error by minimising a function of the costs involved with the test. However, such kind of approaches are out of the scope of this investigation. A mention of them was, however, believed necessary for the sake of completeness. More detail about them, in the context of a practical application, can be found in Tagaras (1994) and in Tagaras (1996).

The analysis of variance supports the conclusion that varying the scan length of an amount included in [10%, 30%] in excess of the nominal length, leads to measurement results that are significantly different. That is, the differences between measurement results of the same part obtained by setting different scan lengths in the investigated interval cannot be accounted for only by random variations of some uncontrollable influence factor.

The application of the ANOVA method presented in equation (3.32) and (3.33), resulted in $\hat{\sigma}^2 = 0.3232$, $\hat{\sigma}^2_{t_s} = 0.4279$, $\hat{V}(s_{ij}) = 0.7511$ and $\hat{\sigma}^2 = 0.02991$, $\hat{\sigma}^2_{t_s} = 0.03488$, $\hat{V}(s_{ij}) = 0.06479$ for the nominal length $40\,\mu m$ and $200\,\mu m$, respectively.

These results are connected with the variability of the absolute percentage deviation from the nominal of the measurement results. Thus, by considering the definition of the estimates and the fact that all the deviations from the nominal are negative, the relevant estimates of variance for the measurement results, in μm^2 , are derived by multiplying the aforementioned estimates by the square of the correspondent nominal length. This gives to $S_{t_s,40} = 0.2616\,\mu m$, $S_{s,error,40} = 0.2274\,\mu m$ and $S_{t_s,200} = 0.3735\,\mu m$, $S_{s,error,200} = 0.3459\,\mu m$, for the nominal length $40\,\mu m$ and $200\,\mu m$, respectively.

The results from the analysis of variance fully holds if the assumptions of the ANOVA model are all satisfied. A discussion about the validity of the assumptions is, therefore, needed (cf. section 3.4 in Montgomery, 2001, chapter 3 in Drapper and Smith, 1966) and it is presented in the next section.

3.3.1.2 Checking the adequacy of the model

An analysis of the assumptions underlying the random effect ANOVA model is carried out in this section by considering the residuals. In this study, residuals are the predicted realisations of the random error component in equation 3.26. In particular, the random error component can be expressed as the difference between the response variable s_{ij} and its expected value for a given level of scan length $t_{s,i}$. The given level $t_{s,i}$ is a realisation of a random variable. In fact, it represents one of

the five scan lengths drawn from the infinite population associated with the range [10%, 30%]. More formally, it holds:

$$E\left(s_{ij}|t_{s,i}\right) = \mu_s + t_{s,i} \tag{3.36}$$

Then, from equation 3.36, it follows:

$$e_{ij} = s_{ij} - E\left(s_{ij}|t_{s,i}\right) \tag{3.37}$$

Thus, from equation 3.37 and the definition of residual, it follows:

$$\hat{e}_{ij} = s_{ij} - \hat{E}\left(s_{ij}|t_{s,i}\right) \tag{3.38}$$

In equation 3.38, the estimated expected value can be derived by applying the method of the moments to the $i-th$ sample, s_{i1}, s_{i2}, ..., s_{in}, thus obtaining :

$$\hat{E}\left(s_{ij}|t_{s,i}\right) = \bar{s}_i. \tag{3.39}$$

Alternatively, an estimate of the ANOVA model parameter μ_s and a prediction for the unobservable random effects $t_{s,i}$, can be obtained via the associated linear model [16] with four indicator variables, whose parameters are estimated by means of the ordinary least squares method (cf. chapter 4 in Sen and Srivastava, 1990). This more general second method leads to the same residuals given by equation 3.38 and equation 3.39. In particular, in section 2.3 of Sen and Srivastava (1990) the following equation is derived:

$$\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{H}) \cdot \mathbf{e} \tag{3.40}$$

In equation 3.40, $\hat{\mathbf{e}}$ is the vector of the residuals, $\mathbf{e}$ is the vector of the errors, $\mathbf{I}$ is the identity matrix and $\mathbf{H}$ is the hat matrix, which transforms the vector of the data

[16] The adjective linear refers to the parameters of the model and not to the dependent variables, that in this case are even not continuous.

s_{ij} into the vector of the values predicted by the model (also called fitted values). From equation 3.40, it follows that if the Gauss-Markov conditions hold, then the covariance matrix of $\hat{\mathbf{e}}$, namely, $var\,(\hat{\mathbf{e}})$, is given by the following equation:

$$var\,(\hat{\mathbf{e}}) = (\mathbf{I} - \mathbf{H}) \cdot \sigma^2 \tag{3.41}$$

The derivation of equation 3.41 is presented in Appendix B. From equation 3.41 it can be seen that although the errors are uncorrelated and with the same variance, the residuals are correlated and have different variances if the diagonal elements of $\mathbf{H}$, called leverages, are not all equal (cf. section 7.1 in Faraway, 2002). However, in Anscombe and Tukey (1963), it is pointed out that while on the one hand, the correlation affects the distribution of functions of the residuals, on the other hand the effects of the correlation of the residuals on diagnostic graphical procedures are usually negligible.

The k-th component of the residuals vector in equation 3.40, can be expressed as:

$$\hat{e}_k = e_k - \mathbf{h}_k^T \mathbf{e} \tag{3.42}$$

where $\mathbf{h}_k^T$ is the k-th row of the hat matrix, $\hat{e}_k$and e_k are the generic k-th components of the vector of the residuals and of the errors, respectively. Equation 3.42 is important because in section 5.2 of Sen and Srivastava (1990) it is shown that if the leverages go to zero, then $\mathbf{h}_k^T \mathbf{e}$ converges to zero in probability. Therefore, the smaller the leverages, the less the approximation in using the residuals in the place of the errors. It is shown in Appendix A.11 of Sen and Srivastava (1990) that the sum of the leverages equals the number of the indicator or dummy regression variables plus one, namely $\sum_k^{an} h_{kk} = a$. Thus, the average leverage is $\bar{h}_{..} = a/an$, where an denotes the total number of cases s_{ij}. In addition, in section 8.2.1 of Sen and Srivastava (1990), it is derived that $\frac{1}{na} \leq h_{kk} \leq 1$. Therefore, it is argued that this relationship permits the consideration of the leverages as always sufficiently close to zero to accept the approximation $\hat{e}_k = e_k$in 3.42. Consequently, "the residuals can be seen as the observed errors if the model is correct" (cf. section 3.0 in Drapper

and Smith, 1966). If the model is correct, therefore, the residuals should display properties that do not contradict the assumptions about the errors. Hence, it cannot be concluded that these assumptions are correct but only that the available experimental data do not support their denial. Therefore, this approach has similarities to the test of hypotheses when rejecting or failing to reject the null hypothesis.

The assumptions about the errors are described by $\{e_{ij}\}$ $NIID\,(0,\,\sigma^2)$. In section 3.1 of Drapper and Smith (1966), from the first normal equation of the least squares estimation, it is derived that the mean of the residuals is always zero. In addition, in section 7.1 of Sen and Srivastava (1990), it is highlighted the importance of collecting the data so as the error can be considered uncorrelated[17]. In fact, not only are data taken over a short time period most likely to be correlated (serial correlation), but also observations taken in the same zone of space are usually more similar to each other than to observations taken in other zones (spatial correlation).

In the light of these considerations, a plot of the $\hat{e}_k$'s versus the sequence order should display a central tendency equal to zero and should contain no relevant pattern if no serial correlation is present. The plot of the realisations of the residuals against sequence order is illustrated in Figure 3.5. It can be noticed that the realisations of the residuals appear quite structureless, although a vague pattern decreasing with the test order sequence seems present, particularly for the nominal length $200\,\mu m$. This suggests that, at least for the case $200\,\mu m$, the fitted model should include a dependent variable accounting for this run order influence. However, this pattern is believed to be too vague in order for its exclusion from the model to affect greatly the conclusions of this particular investigation. Consequently, at this stage, for the sake of simplicity, no further models have been fitted to the data.

If the model is correct and the assumptions hold, the errors should account for the variability due to all the uncontrolled and unidentified sources. Therefore, the residuals, which are considered as realisations of the errors within the above limits,

[17]Uncorrelation between random variables implies independence only if the random variables are normally distributed. If they are not, uncorrelation is just a necessary condition for the independence (more details are in Mood et al., 1974).

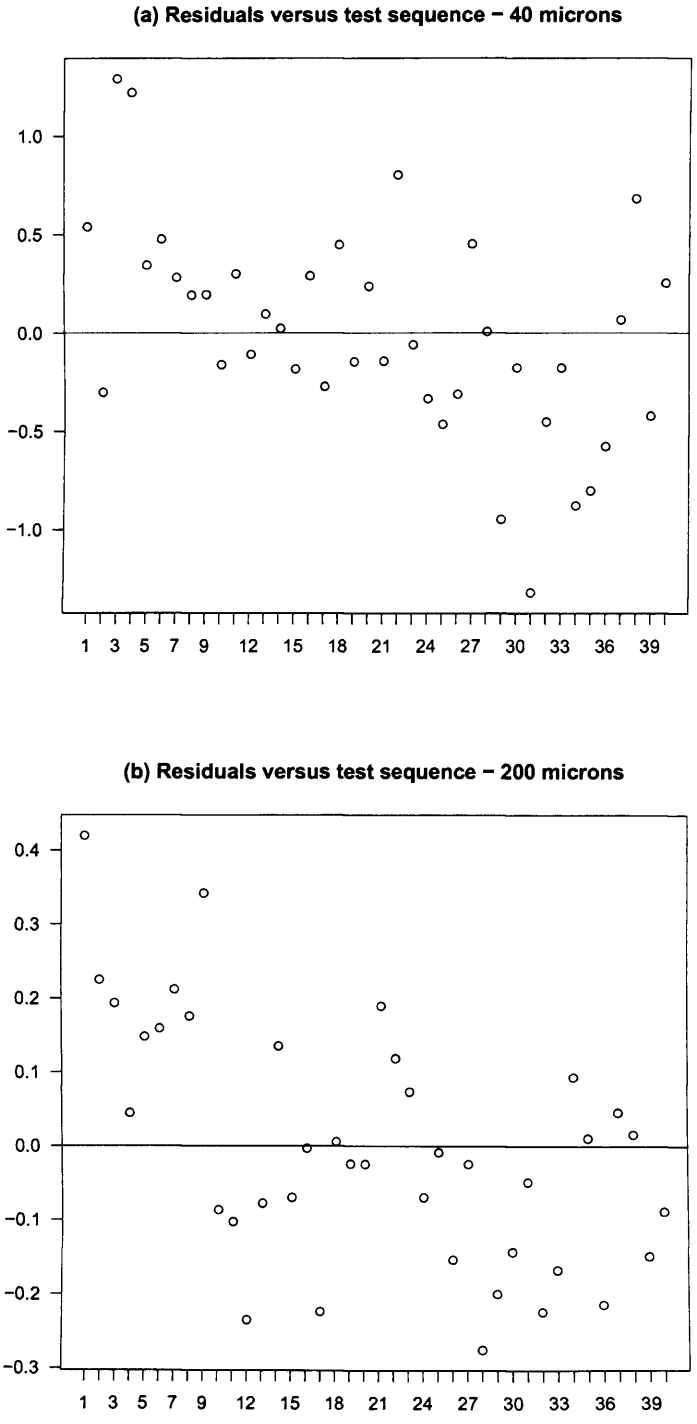


Figure 3.5: Residuals versus test sequence. (a) Nominal length 40 μm (b) Nominal length 200 μm.

should not exhibit any relationship with any possible quantity, including the fitted values. If this were not the case, such a quantity should be included in the independent variables of the model. In particular, a plot of the residuals against the fitted values displays patterns that indicate the need for adjusting the functional relationship between the dependent and the independent variables, perhaps involving also a pre-analysis transformation of the response. In addition, this plot can also highlight a violation of the assumption of equal variance of the errors (for more details cf. section 3.3 in Drapper and Smith, 1966 and section 3.4.3 in Montgomery, 2001). In fact, from equation 3.42, the variance of the generic k-th residual is given by:

$$V(\hat{e}_k) = (1 - h_{kk}) \cdot \sigma_s^2 \tag{3.43}$$

In the ANOVA model under investigation, the leverage are constants. Thus, the plot of the residuals against the fitted values provides a particularly good insight into the variance of the errors. In fact, as already pointed out and shown in equation 3.43, only if the leverages are constant do the residuals then have constant variance under the assumption that the errors are homoscedastic. Therefore, if the plot of the residuals against the fitted values exhibits the variability of the residuals as a function of the fitted values, then this circumstance can only be ascribed to a violation of the assumed homoschedasticity of the errors.

In part (a) and (c) of Figure 3.6, the plot of the residuals against the fitted values is presented for the $40\,\mu m$ and $200\,\mu m$ nominal lengths, respectively. In both cases, no significant patterns in the variability of the errors are apparent. In addition, it can be pointed out that in both cases there is a fitted value whose correspondent group of residuals displays a relatively larger variability. However, this observation does not appear sufficient to draw further conclusions. Therefore, there is no significant graphical evidence supporting the rejection of the errors' homoscedasticity.

Following the suggestions in section 3.4.3 of Montgomery (2001), a Levene's test was also conducted in order to assess the assumption of constant variance of the errors. In particular, this Levene's procedure tests the hypothesis of equal variance between each group of residuals (cell) identified by the same levels of the independent

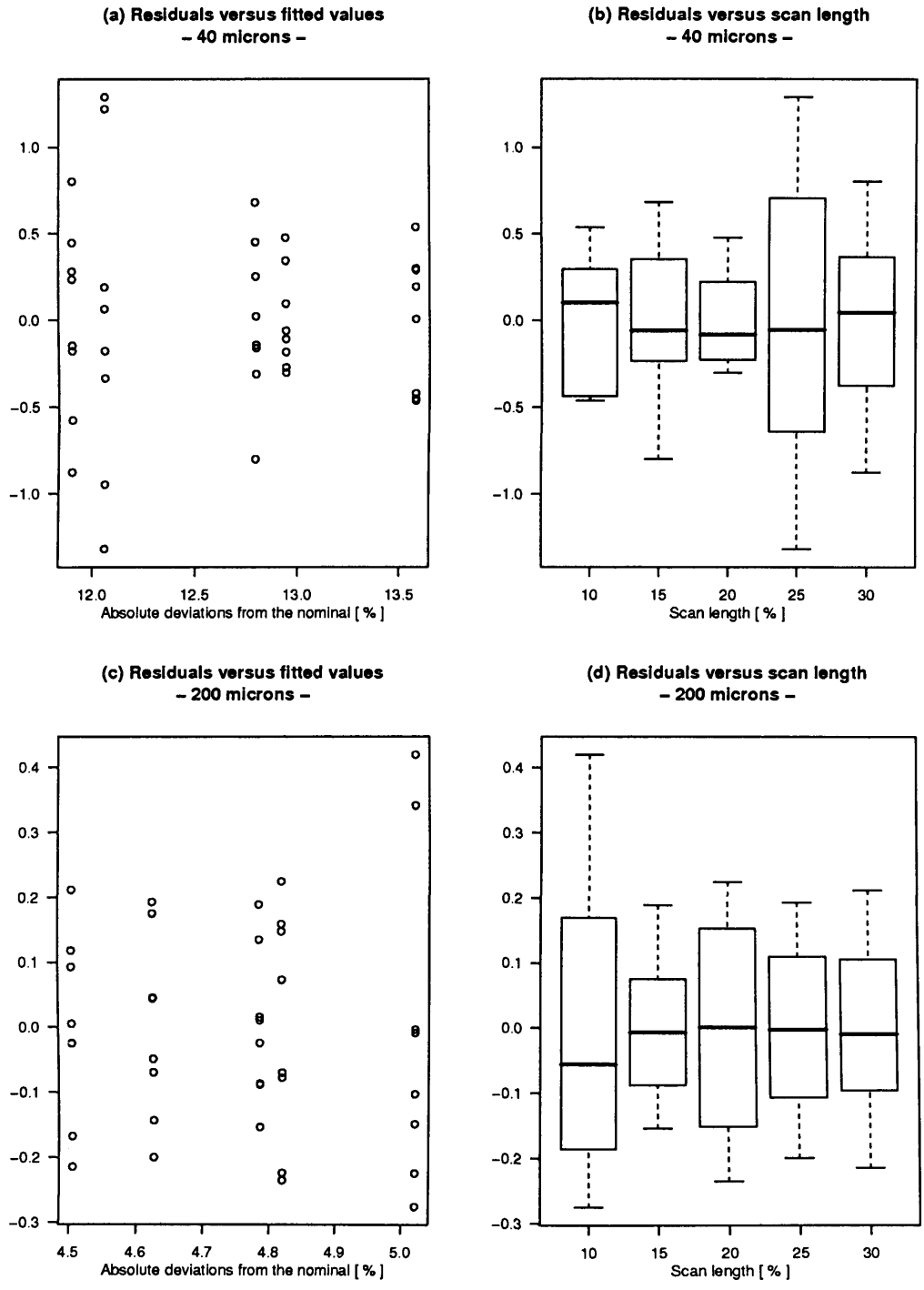


Figure 3.6: Residuals versus predicted values and versus scan length for 40 μm, (a) and (b), and for 200 μm, (c) and (d).

parameters (treatment or treatment combination). If the errors are homoschedastic, then these variances should be the same. Formally, the hypotheses are $H_0 : \sigma_1 = \sigma_2 = \ldots = \sigma_a$ and H_1 : 'H_0 : is not true for at least one σ_i'. The test statistic deployed in this procedure is the absolute deviation of the observations s_{ij}in each cell from the cell median, $\tilde{s}_{i\cdot}$, more formally: $d_{ij} = |s_{ij} - \tilde{s}_{i\cdot}|$, with $i = 1 \ldots a$ and $j = 1 \ldots n$. Then, in Montgomery (2001), it is stated that if the a cell means of d_{ij} are the same, the investigated variances are the same. Thus, the equality of means of these absolute deviations was tested by means of an ANOVA model and the corresponding F test. The test resulted in $F_0 = 2.34$ with $P_{value} = 0.074$ for the $40\,\mu m$ nominal length and $F_0 = 1.26$ with $P_{value} = 0.30$ for the $200\,\mu m$ nominal length. Thus, for both the lengths H_0 cannot be rejected. Hence, the test does not display any evidence for rejecting the assumed constant variance of the errors. Nonetheless, the $40\,\mu m$ nominal length case appears to be more problematic, due to the fact that the P_{value} is close to rejection values. The same conclusion is qualitatively provided by the graphical analysis of part (b) and (d) of Figure 3.6. In fact, variations in the interquantile ranges of the residuals are more evident for the $40\,\mu m$ nominal length than for the $200\,\mu m$ nominal length. However, the Levene's test does ensure that such variations are in both cases due to uncontrollable random influence factors affecting the data and not to differences in the residuals' variance for different scan lengths.

If the Gauss-Markov conditions hold, then equations 3.41 and 3.42 hold. From these equations and from what has been mentioned above in this section, it is argued that the residuals can be used, with some approximation[18], as a random sample extracted from the errors' population when assessing violations of the errors' normality assumption.

The graphical procedure adopted for assessing violations of the hypothesis of normality of the errors is as follows. The ordered residuals' realisations, $\tilde{e}_{[k]}$, are deployed in evaluating the corresponding empirical cumulative distribution func-

[18] The approximation are due to neglecting the effect of the h_{kk}in 3.42 and the correlations between the residuals in 3.41, so as they can be considered as random sample.

tion, F_k. Slightly different methods have been found in the literature for such a purpose, but the principle is the same (for more details cf. section 5.2.1 in Sen and Srivastava, 1990 and section 2.4.1 in Montgomery, 2001). Then, z_k, the quantile of the normal standard distribution corresponding to the evaluated empirical cumulative distribution function, is calculated using tables or some statistical software. If the errors are normal, the residuals' realisations should not exhibit any non-normal tendency. Thus, when plotted, the computed theoretical quantiles z_k and the corresponding residual realisations, $\tilde{e}_{[k]}$, should approximately lie along a straight line, if the distributional hypothesis is correct. In fact, it holds:

$$F_k = P\left(\tilde{e} \leq \tilde{e}_{[k]}\right) = P\left(\frac{\tilde{e} - E(\tilde{e})}{\sqrt{V(\tilde{e})}} \leq \frac{\tilde{e}_{[k]} - E(\tilde{e})}{\sqrt{V(\tilde{e})}}\right) \tag{3.44}$$

Hence, if the assumed distributional hypothesis is correct, from equation 3.44 it follows that:

$$z_k = \frac{\tilde{e}_{[k]} - E(\tilde{e})}{\sqrt{V(\tilde{e})}} \tag{3.45}$$

Equation 3.45 represents a straight line with intercept $-E(\tilde{e})/\sqrt{V(\tilde{e})}$ and slope $1/\sqrt{V(\tilde{e})}$. This procedure is often identified in the literature with the names 'Normal probability plot' (cf. Montgomery, 2001), 'quantile-quantile plot' and 'Q-Q plot' (cf. Faraway, 2002 and R Development Core Team, 2006). More generally, it can be used to assess violations of any distributional assumption of a sample, and not only of normality.

In this study, the normality plot of the residuals are presented in Figure 3.7, where the straight line is generated so as to pass through the first and the third quartiles. The residuals' realisations do not appear to violate dramatically the assumption of normality of the errors. In fact, apart from some departures on the tails, the residuals are quite close to the straight line. Although the deviations from a linear pattern near the tails may affect the levels of significance of the F test in the ANOVA procedure, they do not appear so large as to reject the assumption of normality of the errors.

In addition, a formal Anderson-Darling test of hypotheses has been also run on

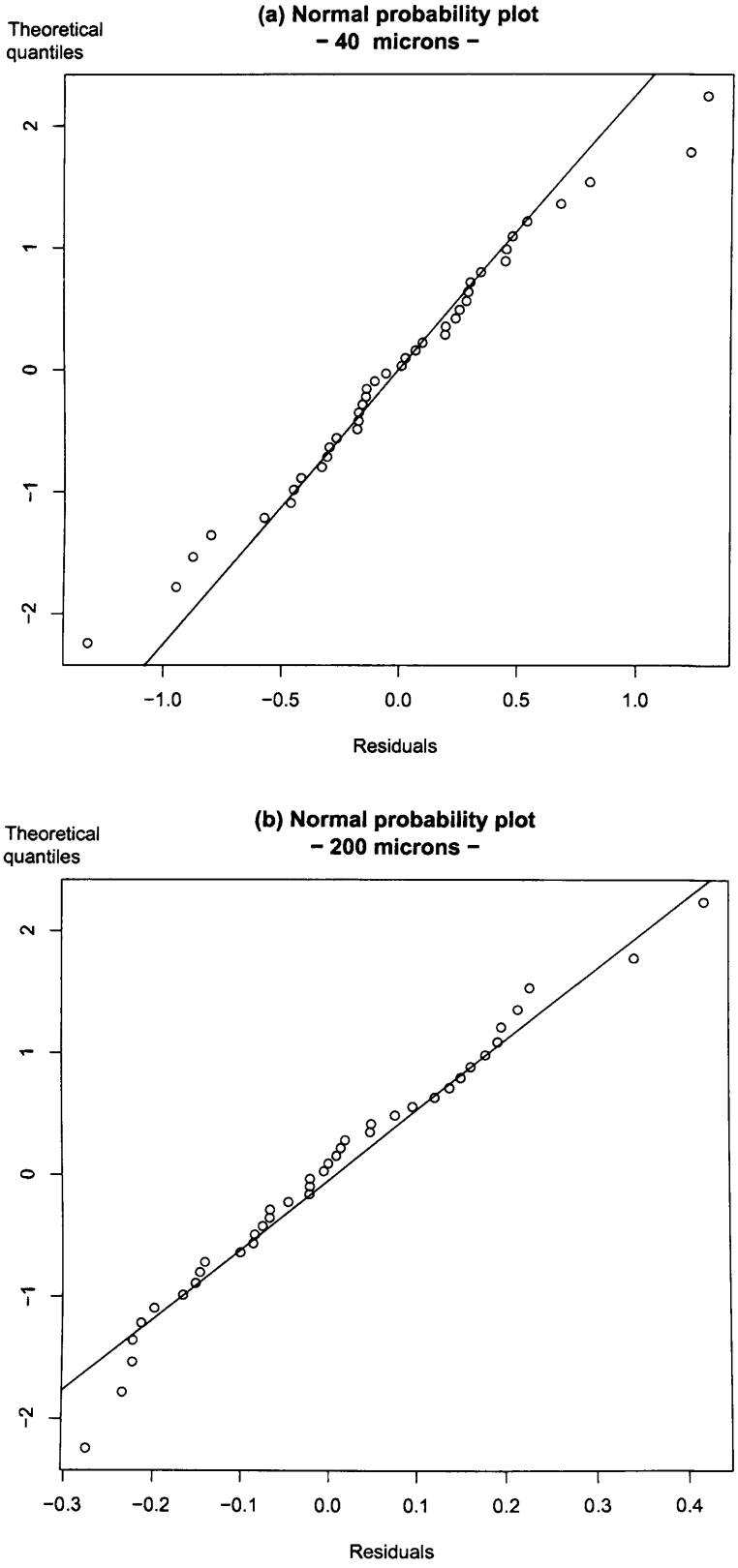


Figure 3.7: Normal probability plot for the residuals. (a) 40 μm nominal length (b) 200 μm.

the residuals' realisations. In this procedure, the null hypothesis of normality was tested by making use of the Anderson-Darling test statistic. This test is one sided. This means that values of the test statistics larger than the critical value lead to the rejection of the null hypothesis of normality. Alternatively to specify a critical value, a P_{value} approach can be used in rejecting or failing to reject the null hypothesis (for more details about the Anderson-Darling test cf. 1.3.5.14 in NISTS). In the case under investigation, the values of test statistic A and of the p-values, calculated using a software package (R Development Core Team, 2006). , were $A = 0.2777$ with $P_{value} = 0.6338$ and $A = 0.2857$ with $P_{value} = 0.6078$ for the 40 μm and 200 μm nominal lengths, respectively. There is no experimental evidence, therefore, for rejecting the normality of the residuals, hence of the errors.

3.3.2 Effect of the initial position of the scanner element

Once the stage orientation has been set, a degree of freedom in setting the initial scanner position appears to be left to the operator. Hence, an experimental investigation was carried out in order to assess the effect, if any, of the scanner initial position on the measurement results. This section describes the experimental study undertaken.

After determining the stage orientation, two slightly different initial positions of the scanner element were considered. Each of them allowed some fringes to be visible. The first position was selected at the top of the range of positions of the scanning element so as to make some fringes appear. In contrast, the second was selected at the middle of the same range.

As already observed, Figure 3.5 suggests that the measurement results may be affected by the sequence order of the experiments. The run order of the tests was therefore not randomly selected. In fact, if the sequence order has an effect on the measurement results, then the randomisation, by increasing the variability of the data, would make it more difficult to identify the potential effect on the measurement results due to the initial scanner position. Differently stated, it would be more

difficult to establish if differences in the measurement results are due to the factor under investigation, the initial scanner position, or to the potential nuisance due to the run order. Hence, the two scanner positions were tested in a sequence of pairs, where each pair contains both of them (for more details cf. chapter 4 in Box et al., 1978 and section 2.5 in Montgomery, 2001).

Seven and five pairs of measurement results were considered for the $40\,\mu m$ and the $200\,\mu m$ nominal lengths, respectively. The experimental data are displayed in Figure 3.8. From this figure, it can be noticed qualitatively that the initial positions of the scanner element further apart from the measurand (top) produce measurement results that are in general smaller than those obtained by scanner positions closer to the measurand (mid). In fact, in Figure 3.8, the ordinate represents the response variable, namely the absolute values of the deviations from the nominal lengths. Without the absolute value operator, these deviations are all negative.

In order to establish whether this behaviour is substantial or is due only to experimental variability, a randomisation test was run. In this procedure, the 14 and 10 response variable data are grouped in all the possible subsets of cardinality 7 and 5, for the two nominal lengths investigated respectively. This procedure resulted in examining $\binom{14}{7} = 3432$ and in $\binom{10}{5} = 525$ pairs of subsets for the $40\,\mu m$ and the $200\,\mu m$ nominal lengths respectively. For each of these pairs of subsets, the differences between their response variable means, alias average values, were computed. This allowed the building of an empirical density function of these differences (cf. Figure 3.9). In the literature, this density function is often referred to as the reference distribution (for more details cf. part I in Box et al., 1978). If the initial position of the scanner element does not have any effect on the response variable, then the actual value of the differences in the mean response variable for the two investigated positions, does not have to appear extreme when compared with the reference distribution. Alternatively, it can be concluded that the experimental data lead to the rejection of the hypothesis of no effect of the initial scanner position. It is worth noticing that this permutation method does not require any underlying hypothesis. More information about permutation methods can be found in Ernst

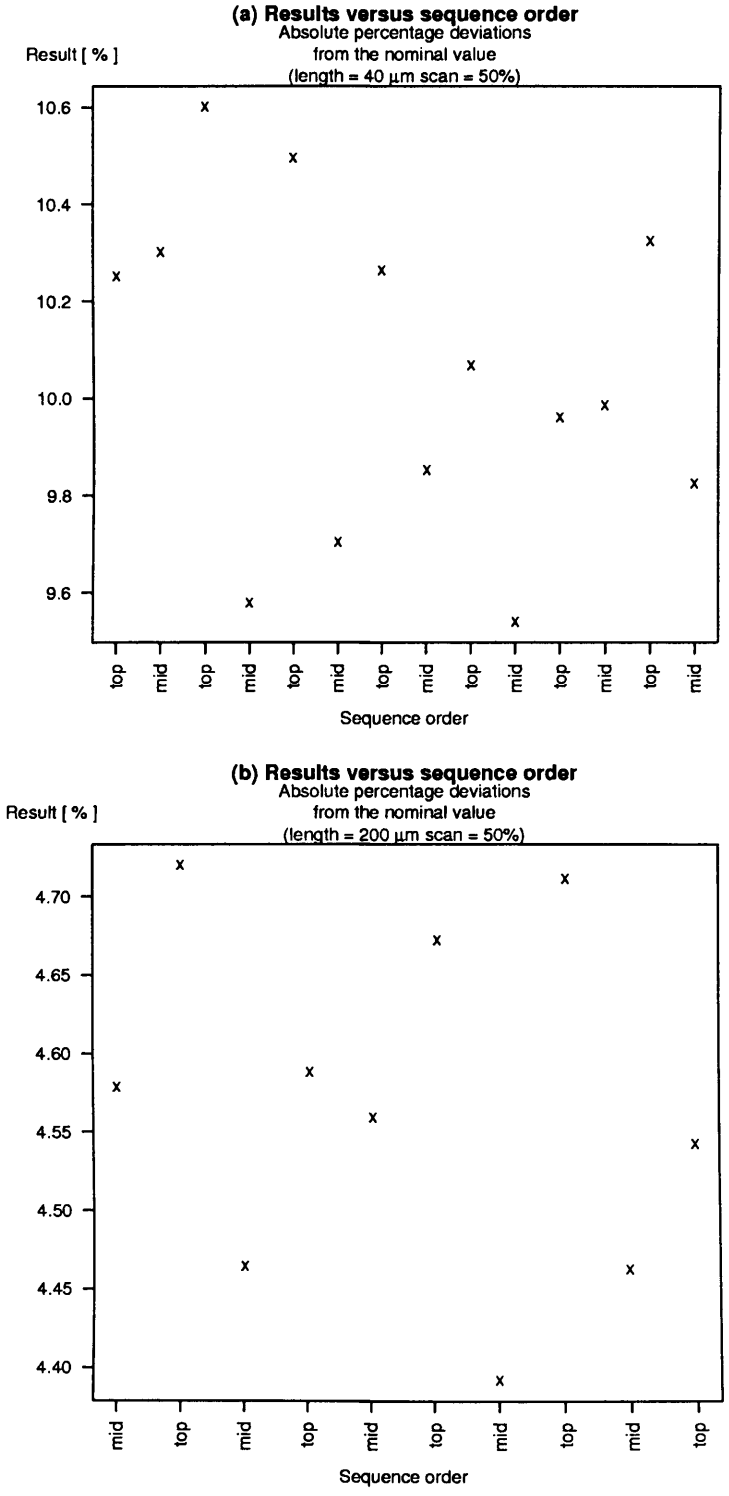


Figure 3.8: Effect of the initial position of the scanner. (a) 40 μm nominal length (b) 200 μm nominal length.

(2004) and Byrne (1993).

In this study, the computations were carried out by developing two programs in the interpreted software R (cf. R Development Core Team, 2006). The R source code of the programs is listed in Appendix C.

For the $40\,\mu m$ nominal length, the actual mean difference in the response variable due to the two different initial scanner positions investigated resulted in 0.4539%. However, in the reference distribution, only 11 cases out of $\binom{14}{7} = 3432$ provided a larger difference. Thus, the hypothesis of no effect of the scanner initial position can be rejected, due to the fact that $P_{value} = 11/3432 = 0.0032$.

Similarly, for the $200\,\mu m$ nominal length, the actual difference resulted in 0.1553. However, in the reference distribution, only 3 cases out of $\binom{10}{5} = 252$ provided a larger difference. $P_{value} = 3/252 = 0.0119$ and the hypothesis of no effect of the scanner initial position can therefore also be rejected.

The sample standard deviation of the differences of the measurement results when considering the two initial scanner positions under investigation, provides an estimate of the contribution to the overall measurement process variability that can be accounted for by different initial scanner positions. In particular, this sample standard deviation resulted in $S_{ip,40} = 0.1574\,\mu m$ and $S_{ip,200} = 0.1882\,\mu m$, for nominal scan length $40\,\mu m$ and $200\,\mu m$, respectively.

3.3.3 Interpretation

After setting up the scan length and the initial scanner position, it was noticed that the scanner element always moved vertically upwards by an amount that does not appear to vary significantly for different scan lengths. Furthermore, it was also observed that increments or reductions in the scan lengths result solely in increments or reductions of the non-initial part of the scanner's travel. These facts have two implications.

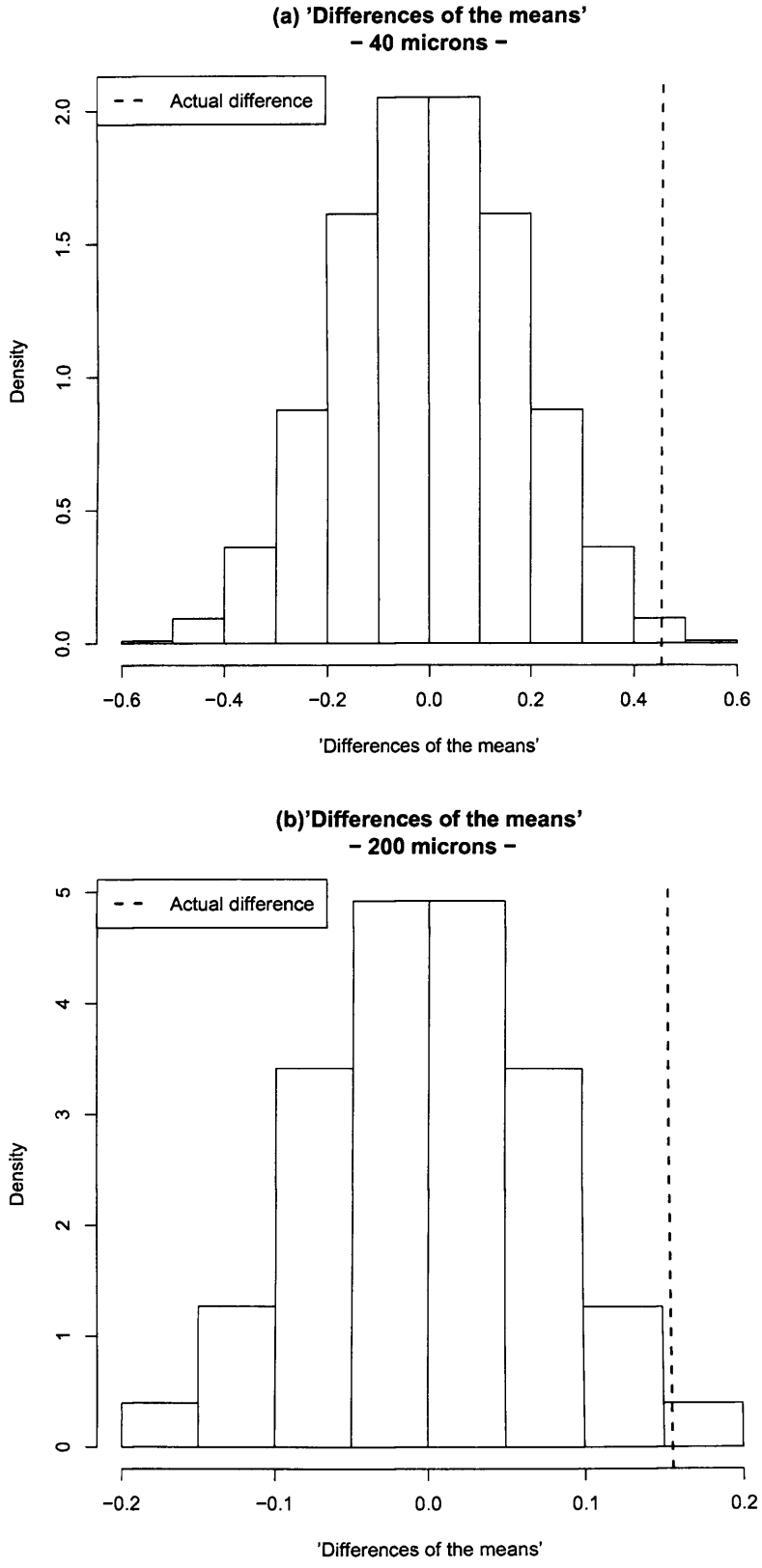


Figure 3.9: Calculated reference distribution of the differences of the means.

Firstly, the scanner element has to accelerate upwards, then it has to decelerate, stop and reverse the direction of the motion downwards, approaching the measurand. Therefore, these transient states of motion are potentially capable of perturbing the necessary condition of constant scanner speed while measuring. In fact, as Schmit and Olszak (2002) highlighted, the assumption of constant rate of change of the optical path difference is an extremely critical point in all the interferometric techniques for shape measurement. In the instrument used in this investigation, the rate of change of the optical path difference is solely connected with the speed of the scanner element.

Hence, it is argued that the initial scanner's position farther away from the measurand, but still allowing a few fringes to be visible, appears to be the most appropriate configuration. In fact, in this manner the accelerations and decelerations of the scanner element are likely to be kept farther away from the measurand, where the actual measurement data are collected. Consequently, it seems that the number of fringes that characterise the correct initial configuration of the stage orientation and the scanner's initial position according to ADEP1 and ADEP2 has to be considered with some flexibility rather than as a strict prescription.

Secondly, varying the scan length in the range [10%, 30%] in excess of the measurand's nominal length, has an effect on the measuring results, as has been experimentally proved. Notwithstanding, while measuring the same part with increasing scan lengths, the transient of motion associated with the final point of the scanner's travelling distance is farther away from the measurand. Hence, the larger the scan length, the less the potential effect on the measurement results due to transients in the motion status close to the end of the scanner's travel. Consequently, if the effects of the scan length on the measurement results are solely due to the deceleration of the scanner near the end of its scanning travel, then it may be expected that by increasing the scan length, the measurement results will display an asymptotic behaviour.

A broader range of scan lengths has, therefore, been experimentally investigated. In particular, for both the $40\,\mu m$ and the $200\,\mu m$ nominal lengths, scan distances in the

range [5%, 100%] were experimentally tested. Twenty evenly spaced levels inside this interval were considered in steps of 5%. For each level, three measurements were carried out. The results are displayed in Figure 3.10. In this figure, it can be noticed that there is a very visible increasing pattern [19]. There is, however, no trace of the predicted asymptotic behaviour.

It can be concluded, therefore, that transients in the scanner motion at the beginning and at the end of the scanning distance do not account for the effect of the scan length on the measurement results. Further investigation is needed in order to establish the origin of the scan length effect, that has been highlighted in this study.

A further conclusion that can be drawn from the analysis carried out is connected with the manner of establishing the scan length on the WLI microscope under investigation. While using the instrument, it was observed that the minimum scan length that can be set on the instrument is $3\,\mu m$. Consequently, when measuring lengths less than $2\,\mu m$, it is not possible to set the scan length as a percentage in excess of the length of the measurand. It is therefore not possible to follow the prescription of ADEP1 and ADEP2.

This circumstance induces the suspicion that the scan length should perhaps be set as a constant excess of the length of the measurand rather than an excess percentage of this. From the point of view of the laws of the motion, it is difficult to imagine how the length of the measurand can affect the part of the scanner's travelling distance that has to be added to the length of the measurand in order to obtain the scan length. The scan length should perhaps be calculated as the length of the measurand added to a fixed extra length necessary for the scanner's motion transients rather than added to a length variable with the measurand. Nonetheless, the calculation of a suitable fixed extra length does not seem a viable possibility without the provision of further technological information that was not possible to obtain from the instrument's manufacturer. The suggestions listed in ADEP1 and

[19]The variable on the y axis is an absolute value of the percentage deviations from the nominal lengths. The deviations from the nominal, i.e. the measurement results minus the nominal lengths, are all negative.

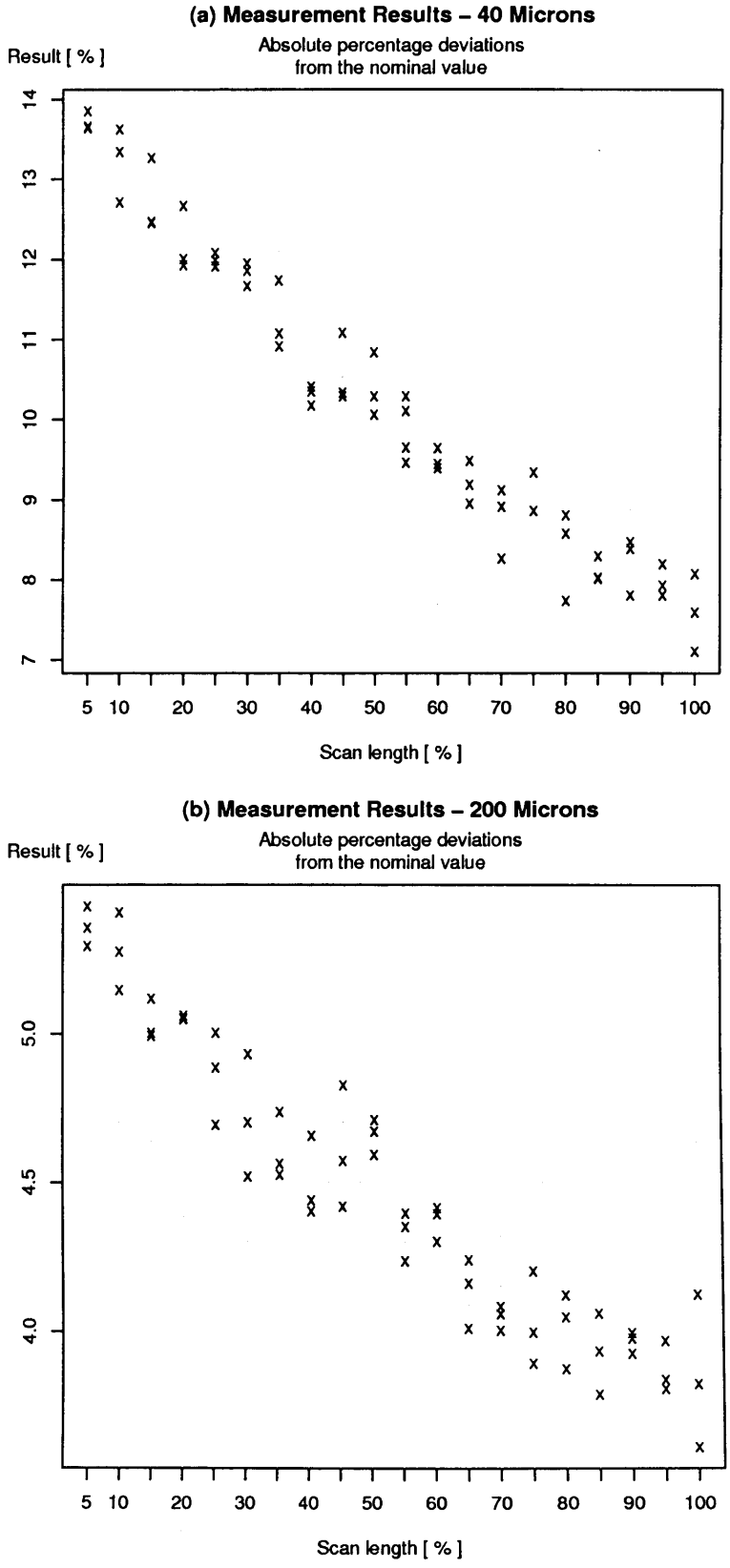


Figure 3.10: Measurement results for scan lengths varying in the range [5%, 100%]. (a) Nominal length 40 μm (b) Nominal length 200 μm.

ADEP2 have, therefore, been followed.

Conventionally, it has been decided to set scan lengths in a range of [10%, 30%] in excess of the nominal length. In doing so, it has been implicitly accepted that there is a source of variability of the measurement results accounted for by the variation of the scan length. This contribution to the variability of the overall measurement process has been estimated in section 3.3.1.1 as $S_{t_s,40} = 0.2616\,\mu m$ and $S_{t_s,200} = 0.3735\,\mu m$, for the nominal length $40\,\mu m$ and $200\,\mu m$, respectively.

Similarly, as reported in section 3.3.2, neglecting the effect of the initial scanner position corresponds to a contribution to the variability of the measurement results in the order of $S_{ip,40} = 0.1574\,\mu m$ and $S_{ip,200} = 0.1882\,\mu m$, for the nominal lengths $40\,\mu m$ and $200\,\mu m$, respectively.

Therefore, further investigation of the scan length determination can contribute to a reduction in the variability of measurement results up to about 70% and 73%, when measuring nominal lengths of $40\,\mu m$ and $200\,\mu m$, respectively. These figures highlight how critical is the set up configuration when using interferometric techniques.

3.4 Compendium

In this chapter, an affordable and versatile method for building traceable reference samples of length in the micrometric range has been proposed. The method is based on the usage of gauge blocks. In addition, a method for estimating the uncertainty of the resulting samples has been provided by following a "type B" approach to the expression of the uncertainty in measurement.

The effect on the measurement results due to discretionary set up parameters, such as the scan length and the initial scanner position for a given stage orientation, has been experimentally investigated. The main findings were the following:

- Variations of the scan length in the interval suggested by the instrument's manual result in significant differences in the measurement results;

- The effect of the scan length on the measurement result is not due to the transients of the scanner motion near the beginning and the end of the scan length. Additional technological information is necessary in order to carry out further investigations in this direction;

- The contribution to the overall variability of the measurement results due to the setup parameters scan length and initial scanner element position, given a stage orientation appears to be remarkably significant. These contributions have been quantitatively estimated when measuring reference samples of nominal lengths $40\,\mu m$ and $200\,\mu m$.

Chapter 4

Assessing the repeatability of the white light interferometric microscope

Precision is defined as 'closeness of agreement between independent test results obtained under stipulated conditions' (BS ISO 5725-1, section 3.12). In this section the repeatability, that is the precision of the results of WLI measurements under repeatability conditions (cf. BS ISO 5725-1, section 3.14), is quantitatively assessed. Although repeatability conditions constitute a lower bound for precision, they represent the conditions closest to those in which the WLI measurement method has been used in the evaluation of micro-volumes. In fact, all the measurements were taken by the same operator, using the same equipment and within short intervals of time.

As Montgomery and Runger (1993) highlight, the necessity of quantifying the precision of a measurement process, as defined in (BS 5233, section 2.08), emerges from the need to distinguish its contribution from the overall variability of the quality characteristic under investigation. In this case, the quality characteristic is the unknown length L_S of the reference length, the unknown step height (cf. section3.1). The identification and the quantification of the sources of variability of a measure-

ment process are one of the goals of a gauge capability analysis (Montgomery and Runger, 1993).

In order to achieve these goals, a design of experiment is introduced in section 1, whereas the experimental set-up is described in section 2. The estimate of the repeatability standard deviation and the examination of the residuals are presented in section 3. This study of the precision has highlighted a dependence of the measurement results on the sequence order. This dependence has been referred to as drift. Therefore, a further investigation of this drift is carried out in section 4. Finally, a discussion of the results is presented in section 5.

4.1 Design of the experiment

The quantification of the precision was carried out according to the ANOVA method of estimation of the variance components (see Montgomery, 2001, chapter 12, Montgomery and Runger, 1993, BS ISO 5725-1 and BS ISO 5725-2). In the case under investigation, the level of the test is associated with the different nominal lengths of the selected reference samples. Furthermore, the nominal lengths were chosen so that they approximately span the region of lengths entailed in the successive measurements of micro volumes. Just one level for the factor laboratory was considered, however, due to the fact that the quantification of the reproducibility was outside the scope of this analysis (for the definition of reproducibility see sections 3.17 and 3.18 in BS ISO 5725-1). In this study, potential sources of variability of the measurement process in the specific measurement task have been identified in the nominal length of the reference samples. Thus, 'nominal length' has been defined as the sole factor in the analysis of variance model.

In the analysis of the difference between test results of a measurement method, the bias of the measurement method has no influence and can be neglected (cf. sections 5.1.1.2 in BS ISO 5725-1). A calibration procedure is not, therefore, needed in order to carry out a gauge capability analysis. Hence, the value of the height correction factor has been conventionally set at one.

The j-th measurement result of the i-th step height has been set as the response variable h_{ij} , of a random effect ANOVA model with one factor.

In fact, the deviation $\Delta h_{ij} = h_{ij} - \mu_h$ of the measurement result h_{ij} from its overall mean μ_h, is modelled as two independent random variables $t_{h,i}$ and e_{ij}. The overall mean μ_h is relative to all the measurements taken and to all the randomly selected levels of the factor.

The first variable $t_{h,i}$, represents the contribution to Δh_{ij} due to the fact that h_{ij} is a measurement of the i-th step height. This contribution is referred to as a random effect. The second variable e_{ij}, represents the contribution to Δh_{ij} that is due to all these sources of variability that are difficult or expensive to identify and thus to control (for instance, small vibrations and local small fluctuations of the air temperature). This second variable e_{ij}, is the 'random error occurring in every measurement under repeatability conditions' (cf. section 5.1 in BS ISO 5725-1). Therefore, it holds:

$$h_{ij} - \mu_h = t_{h,i} + e_{ij} \;\; with\, i = 1, \ldots, a \;;\; j = 1, \ldots, n \tag{4.1}$$

Equation (4.1), is formally and conceptually identical to equation (3.26). In addition, the hypotheses laid out in section 3.3.1.1 also apply to this case. In particular, it is assumed that $\{t_{h,i}\} \sim NIID\left(0, \sigma^2_{t_h}\right)$, and that $\{e_{ij}\} \sim NIID\,(0, \sigma^2)$. It is also required that $\{t_{h,i}\}$ and $\{e_{ij}\}$ are independent. Therefore, the discussion developed in section 3.3.1.1, holds also in this study. As a consequence of all the above circumstances, the procedure detailed in section 3.3.1.2 for assessing the adequacy of the model using the analysis of the residuals, has to be carried out also in this investigation. Further details about this experimental investigation can also be found in Ferri et al. (2006).

The main purpose of this ANOVA is to estimate the repeatability standard deviation σ, which in BS ISO 5725-1 is also indicated by the symbol σ_r. In fact, it represents the component of variability of the response variable which is not originated from varying the nominal length of the part. This second contribution is given by σ_{t_h}.

Therefore, σ_r can solely account for the variability due to measuring the same parts using the same equipment, by the same operator in a short time interval, which is the definition of repeatability in BS ISO 5725-1. Thus, the number of the levels for the factor 'nominal length', a, and the number of replicates, n, were not selected on the basis of the operating characteristic curves as is common practice in experiments with random factors (see Montgomery, 2001, section 12.4). The design parameters a and n were selected instead in order to have the narrowest confidence interval for σ_r, subject to the constraints provided by an affordable level of experimental burden (Montgomery and Runger, 1993). Furthermore, the number of levels should be representative of the range of lengths associated with the measuring tasks to be performed (Montgomery and Runger, 1993).

In this research, the precision of the WLI microscope while measuring steps in the range [150 , 250] μm is investigated. Thus, a minimum of $a = 4$ steps was considered in order to span that interval. Independently of its cardinality, the set of the step heights was selected so that they span almost evenly the investigated interval and are a representative selection of the infinite number of possible step heights belonging to it. In fact, this sample has the same probability of being drawn from this interval as any other potential sample. In this sense, the selected set of step heights qualifies as a random sample.

For each potential number of levels a the number of replicates n is changed and the effects on the confidence interval for σ_r and on the experimental burden are considered.

If the hypotheses of normality of the random effect ANOVA model are satisfied, then it holds (cf. Montgomery, 2001, page 516, 543):

$$a \cdot (n-1) \cdot \frac{MSE}{\sigma^2} \sim \chi^2 (a \cdot (n-1)) \tag{4.2}$$

The notation $\chi^2 (a \cdot (n-1))$ indicates a chi-square distribution with $(a \cdot (n-1))$ degrees of freedom.

From equations (4.2) and (3.32) it follows, after a few passages:

$$P\left(s\cdot\sqrt{\frac{a\cdot(n-1)}{\chi^2_{\frac{\alpha}{2}}(a\cdot(n-1))}}\leq\sigma<s\cdot\sqrt{\frac{a\cdot(n-1)}{\chi^2_{1-\frac{\alpha}{2}}(a\cdot(n-1))}}\right)=1-\alpha \tag{4.3}$$

where $\chi^2_\alpha(n)$ is the percentage point of the chi-square distribution with n degrees of freedom that has α% of the probability to the right of $\chi^2_\alpha(n)$. The amplitude of the confidence interval in equation (4.3) for a unit of the estimated repeatability standard deviation, is given by the following equation:

$$P=\frac{A}{s}=\sqrt{a\cdot(n-1)}\cdot\left(\frac{1}{\sqrt{\chi^2_{1-\frac{\alpha}{2}}(a\cdot(n-1))}}-\frac{1}{\sqrt{\chi^2_{\frac{\alpha}{2}}(a\cdot(n-1))}}\right) \tag{4.4}$$

The effect of the design parameters a and n on the performance parameter P and on the experimental effort measured by the total number of tests $a\cdot n$, is given in Table 4.1 for a pre-specified value of $\alpha=0.05$.

Montgomery and Runger (1993) recommended limiting the consumption of resources for initial studies of this kind to no more than 25% of the total resources. Following this suggestion led to the setting of the maximum number of measurements for the planned experiment in the range of [60 , 70] tests. The calculations shown in Table 4.1 do not highlight relevant differences between the alternative designs that lead to an experimental burden in the pre-specified range. Nevertheless, the design $a=4$ and $n=16$ appears optimum and it was, therefore, selected.

4.2 The experimental set-up

The experimental activity was carried out with a $20x$ Mirau interferometric objective, a magnification lens $0.62x$ and a noise threshold of 0.024[1]. The choice of such

[1]During the motion of the scanner, the intensity of the light incident on any pixel has an amplitude. For each pixel, if this amplitude is less than the parameter noise threshold, then the pixel is labelled as a bad data point. Otherwise, the pixel contains a data value (cf. section 4.5 of ADEP2).

a Number of levels	n Replicates	P Amplitude of the confidence interval	$a \cdot n$ Experimental effort
4	10	0.486	40
	11	0.458	44
	12	0.435	48
	13	0.415	52
	14	0.398	56
	15	0.382	60
	16	0.369	64
5	9	0.458	45
	10	0.430	50
	11	0.406	55
	12	0.386	60
	13	0.369	65
6	8	0.446	48
	9	0.415	54
	10	0.390	60
	11	0.369	66

Table 4.1: Design of the precision experiment for $\alpha = 0.05$.

a low magnification originates from the need to broaden the field of view. In this manner, it encompasses a region of the gauge blocks with as many points as possible located away from the edges. In fact, the edges of the blocks are inadequate for establishing a standard of length, due to the presence of a round or a chamfer (cf. BS EN ISO 3650, section 7.3). Nevertheless, the drawback of a low magnification is poor lateral resolution. Therefore, it is suggested that the magnification should be as large as possible while accomplishing a measuring task.

As pointed out by Montgomery and Runger (1993), a complete replication of the measurement process must be performed while obtaining each test result. Otherwise, there is likelihood of neglecting significant sources of variability and of underestimating the variability of the measurement method.

In this study, varying the scan length in the range [10%, 30%] and setting slightly different initial scanner positions contributed to the overall spread of the results. In fact, these two sources of variability were not explicitly identified as factors in the proposed ANOVA model. Furthermore, their effects on the measurement results were approximately evenly distributed among the different step heights investigated. In fact, the manner in which the tests were carried out resulted in a de facto randomisation of these two influential factors.

When replicating the tests, the measuring position on each of the gauge blocks should not be changed. In fact, the variability of the WLI measurement procedure would be unduly increased by measuring different lengths L_S. Therefore, for each of the 4 steps 16 measurements were taken at 'virtually' the same x, y position on the steps. Every single measurement started and ended in the origin of the instrument coordinate system. The measuring position was reached automatically by programming the instrument. The word 'virtually' highlights the fact that the positioning system of the instrument was considered perfect and thus the contribution to the variability of the measurement results due to the positioning system has not been investigated. Notwithstanding, it is believed that this contribution is negligible relatively to the others.

Once a profile has been acquired, first it is 'levelled' and then the difference between

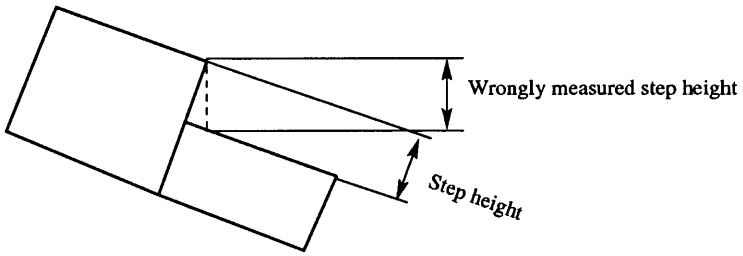


Figure 4.1: The rationale of levelling before measuring.

the averages of the measured z coordinates associated with the pixels in two areas of the field of view farthest from the edge is taken. Care has been taken in always selecting the same areas for calculating these averages when performing each measurement task. The word 'levelled' means that a software transformation is applied to the acquired profile so that all the pixels connected with a selected area on one block display the same measured z coordinates. The area for levelling was conventionally always chosen on the highest block. Figure 4.1 illustrates the reason why the levelling procedure was introduced.

The four levels of the factor nominal length, have been selected to cover, with some excess, the range of lengths [150; 250] μm. This yields the values 150 μm, 183 μm, 217 μm and 250 µm. The gauge blocks chosen for establishing each of these lengths had nominal lengths given by 1.30 mm and 1.15 mm, 1.19 mm and 1.007 mm, 1.22 mm and 1.003 mm, 1.50 mm and 1.25 mm, respectively.

Variations in the laboratory air temperature of about ±1 °C around the set point 20 °C are expected. The potential effects on the measurement results of these variations and of other uncontrollable potential nuisance factors are dealt with by randomising the sequence of the tests. In particular, to each measurement of the four steps an integer number was assigned from 1 to 64. The order of the measuring tests was then determined by selecting one of the possible 64! permutations. This strategy usually allows avoidance of serial correlation between the random errors and also between the random effects. Therefore, it usually allows the errors and the random effects to be treated as independently distributed random variables.

4.3 An estimate of repeatability and the analysis of the residuals

On the basis of the 64 measurement results, the point estimates of the standard deviation of the errors, $\hat{\sigma}$, and of the sample standard deviation of the response variable given the i-th step height, $\hat{\sigma}_i$ with $i = 1, \ldots, 4$, were computed in order to

assess the repeatability of the measurement method. The results are summarised in Table 4.2.

The variations of the interquantile ranges shown in Figure 4.2 lead to a suspicion that a dependence of the variability of the residuals on the nominal step height may be present. Thus a Levene's test was performed on the measurement results in order to reveal if the null hypothesis of constant variance of the errors should be rejected. More details about this procedure are provided in section 3.3.1.2. The results are $F_0 = 1.15$ and $P_{value} = 0.3345$. This indicates that the effect of the nominal step heights on the precision under repeatability conditions is not significant. The differences between the $\hat{\sigma}_i$'s are, therefore, solely due to randomness. Consequently, the ANOVA estimate of the precision given by $\hat{\sigma}$ is preferable to each of the four $\hat{\sigma}_i$'s. In fact, the estimator $\hat{\sigma}$ is proved to be equal to a weighted average of the $\hat{\sigma}_i$'s. Thus it has less variability than each single $\hat{\sigma}_i$ (cf. Montgomery, 2001, section 3.3.1).

From the normal probability plot[2] displayed in Figure 4.3, it appears that the hypothesis of Gaussian distribution of the residuals is not supported by the experimental evidence. In addition, an Anderson-Darling test (see section 3.3.1.2) was performed and resulted in a test statistic value $A_0 = 1.59$ and a corresponding $P_{value} = 3.813 \cdot 10^{-4}$. Hence, the assumption of Normality of the residuals has to be rejected. Departures from normality usually result in differences from the stated values both of the true significance level and of the power of the F test in the ANOVA procedure (cf. Montgomery, 2001, section 3.4.1).

In this investigation, the rejection of the normality assumption does not seem to have any consequence apart from the procedure for deriving a confidence interval for the repeatability standard deviation. In fact, due to the fact that the errors are not Gaussian, a confidence interval based on a chi-squared distribution does not appear appropriate.

Consequently, in this study, a Jackknife procedure was adopted for computing a confidence interval for σ. This procedure is robust to departures of the residuals from

[2]More details about normal probability plots are in section 3.3.1.2.

		Jackknife confidence interval (95%)
$\hat{\sigma}$ $[\mu m]$	$5.5 \cdot 10^{-1}$	(0.33 ; 0.79)
$\hat{\sigma}_1$ $[\mu m]$	$4.6 \cdot 10^{-1}$	
$\hat{\sigma}_2$ $[\mu m]$	$3.7 \cdot 10^{-1}$	
$\hat{\sigma}_3$ $[\mu m]$	$6.9 \cdot 10^{-1}$	
$\hat{\sigma}_4$ $[\mu m]$	$6.3 \cdot 10^{-1}$	

Table 4.2: ANOVA estimate of the repeatability standard deviation

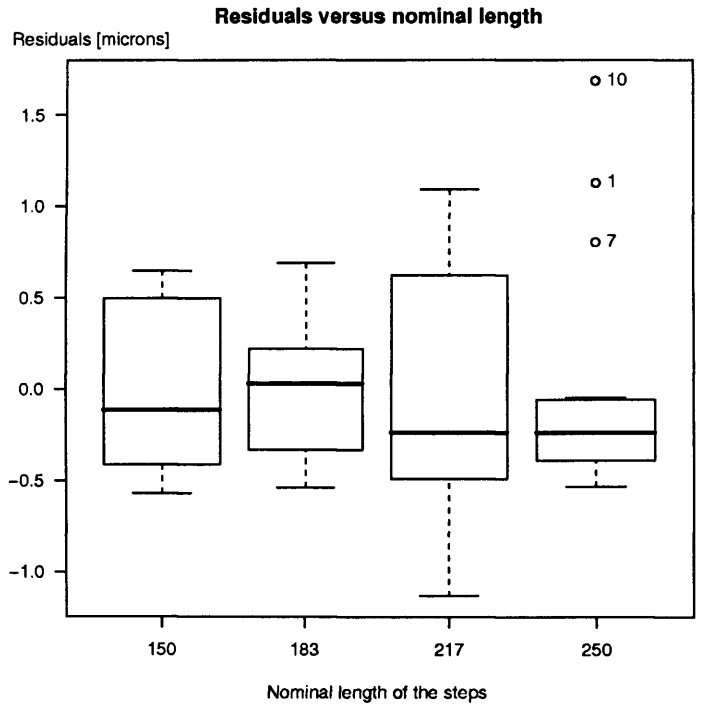


Figure 4.2: Boxplot of the residuals versus random sample of nominal lengths.

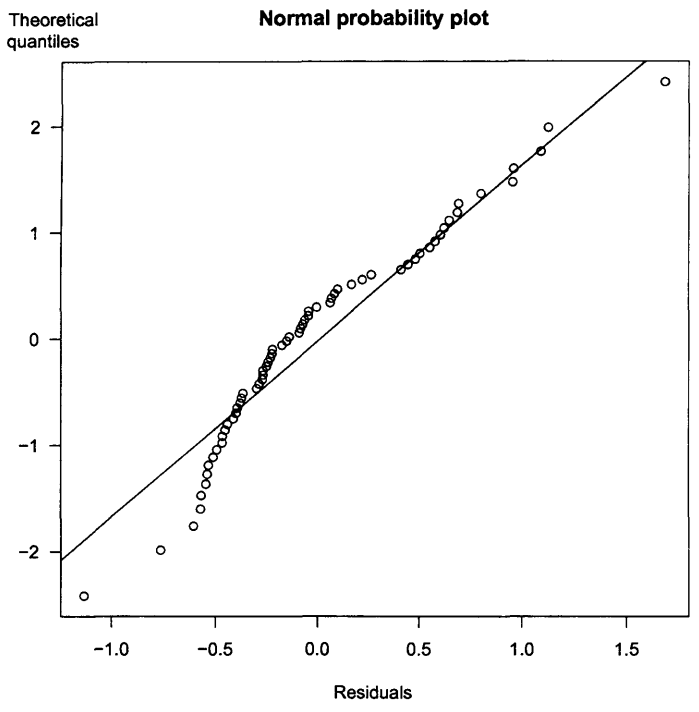


Figure 4.3: Normal probability plot of the ANOVA model residuals.

the normality hypothesis and consists in computing four 'partial estimates', $\hat{\sigma}_{-i}$, by leaving out the i-th factor level, with $i = 1, \ldots, 4$. Subsequently, four 'pseudovalues' $\tilde{\sigma}_i = a\,\hat{\sigma} - (a-1)\,\hat{\sigma}_{-i}$ are computed and their sample mean, standardised by the sample standard deviation, results asymptotically distributed as a Student's t due to the central limit theorem. Therefore, not only is this average used as a point estimator of the unknown parameter to be estimated, but it is also deployed in the construction of confidence intervals for the unknown parameter. More about jackknifing can be found in section 3.6.3 of Miller (1997). The confidence interval obtained, with 95% significance level, for σ is displayed in Table 4.2.

Figure 4.4 shows the realisations of the residuals, $\hat{e}_{ij}$, versus the sequence of the measurement tests. The relationship between residuals and errors has been analysed in section 3.3.1.2. A graphical examination of the plotted results highlights a possible violation of the assumption that the errors e_{ij}'s are independent and identically distributed random variables.

Therefore, a first order regression model of the residuals on the sequence order was fitted to the data in order to represent the dependence of the $\hat{e}_{ij}$'s on the run order, namely:

$$\hat{e}_{ij} = \beta_0 + \beta_1 O_{ij} + e_{ij,new} \tag{4.5}$$

where the symbol O_{ij} is the order in which the measurements are taken, β_0 and β_1 are the model parameters and $e_{ij,new} \sim NIID\,(0, \sigma^2_{new})$ are the errors. The parameters in equation 4.5, were estimated by applying the ordinary least square method. This yielded the estimates $\hat{\beta}_0 = 4.95 \cdot 10^{-1}\,\mu m$ and $\hat{\beta}_1 = -1.53 \cdot 10^{-2}\,\mu m$. In addition, for a generic regression model, it results from equation 3.40 that $\hat{\mathbf{e}}^T\hat{\mathbf{e}} = \mathbf{e}^T \cdot (\mathbf{I} - \mathbf{H}) \cdot \mathbf{e}$, due to the fact that the matrix $(\mathbf{I} - \mathbf{H})$ can be proved to be symmetric and idempotent. The statistics $\hat{\mathbf{e}}^T\hat{\mathbf{e}}$ is called the residual sum of squares (RSS). In section 2.7 of Sen and Srivastava (1990), it is proved that under Gauss-Markov conditions, $E\left(\hat{\mathbf{e}}^T\hat{\mathbf{e}}\right) = \sigma^2 \cdot (n-p)$, where p is the number of the unknown parameters in the regression model to be estimated, i.e. the number of the regressors incremented by one for the intercept term, and n is the overall number of experimental results. Therefore, by using the method of the moments, the following unbiased estimator

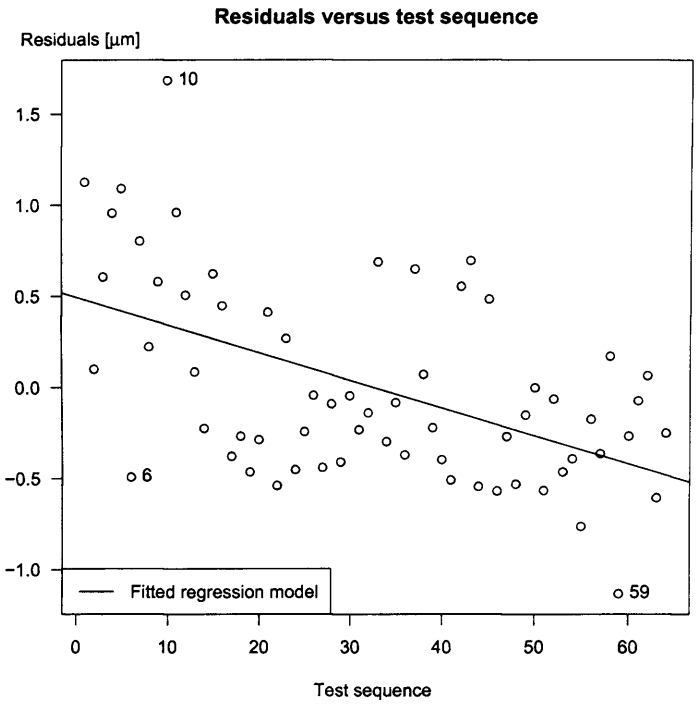


Figure 4.4: Realisations of the residuals versus run order .

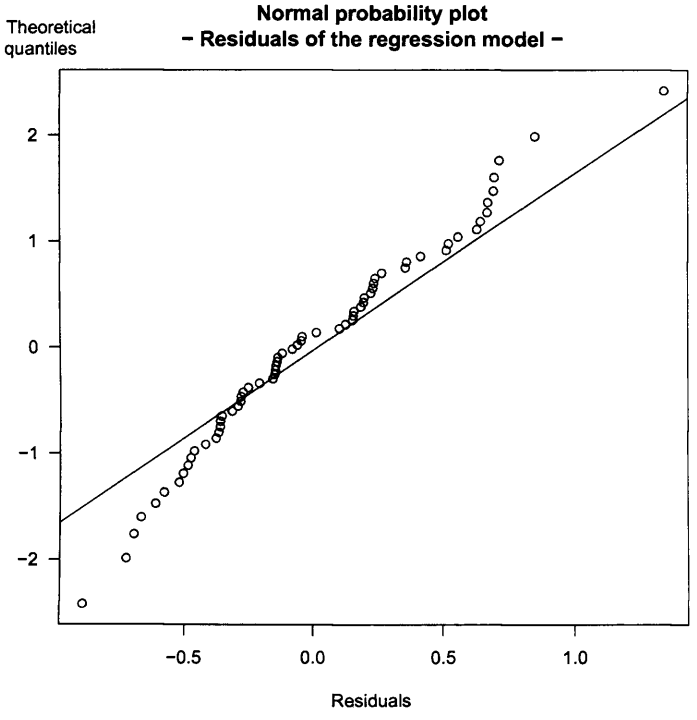


Figure 4.5: Normal probability plot of the residuals for the regression model.

of σ^2 is derived:

$$\hat{\sigma}^2 = \frac{RSS}{n-p} \tag{4.6}$$

Other properties of the estimator defined in equation 4.6 that make it preferable to other potential estimators are discussed in section 2.7 of Sen and Srivastava (1990). By expressing an ANOVA model as its equivalent regression model using indicator variables, it is derived that equation 4.6 is the same as the mean square error in the ANOVA. In this study, the same statistic is called estimate if the focus is on the numerical value provided when estimating a parameter, whereas it is called estimator if the focus is on the fact that a particular statistic is used to estimate a given unknown parameter. In the model of equation 4.5, the experimental investigation resulted in $\hat{\sigma}_{new} = 4.62 \cdot 10^{-1}\, \mu m$.

The F-test for significance of the regression (cf. section 1.4 in Drapper and Smith, 1966 and sections 3.1 and 3.2 in Faraway, 2002), resulted in F_0 =23.8 and $P_{value} = 7.9 \cdot 10^{-6}$. Consequently, the experimental evidence leads to the rejection of the null hypothesis $\beta_1 = 0$.

Moreover, the coefficient of determination resulted in $R^2 = 0.277$. This implies that the variability of the $\hat{e}_{ij}$, expressed as their sum of squares about their average, is approximately 28% accounted for by the sum of square due to the regression. This second term is given by the squares of the deviations of the predicted values from the average summed over all the predictions (cf. section 1.3 in Drapper and Smith, 1966). Alternatively, when predicting the value of a response variable, R^2 can be interpreted as the percentage reduction of the variability of the prediction error given by the availability of the predictors values (cf. section 2.1 in Faraway, 2002). For both interpretations, the higher $0 \leq R^2 \leq 1$, the better. Nevertheless, adjudging the appropriateness of a model on the basis of R^2 appears to be problematic. In fact, acceptable values of R^2 vary significantly with the area of application (cf. section 2.1 in Faraway, 2002). The purpose of this investigation is to ascertain whether a relationship between $\hat{e}_{ij}$ and O_{ij} is in place, rather than to establish the functional form of this relationship. No attempt at fitting more complex models has therefore been made, regardless of the fact that the quality of the fitting might be potentially

improved.

The normal probability plot of Figure 4.5 does not seem to exhibit significant departures from the assumed normality of the residuals and hence of the errors (cf. section 3.3.1.2 for more information). In fact, major deviations of the realised residuals from the plotted straight line are solely present on the left tail of the distribution. In addition, An Anderson-Darling test was performed and resulted in a test statistic value $A_0 = 0.533$ with associated $P_{value} = 0.166$. The experimental evidence does not therefore seem to be sufficient to reject the hypothesis of normality of the residuals.

The plot of the residuals against the fitted values, displayed in Figure 4.6, does not seem to exhibit any significant pattern and does not appear to show any violation of the assumed homoscedasticity of the residuals. In this regression model, oscillations in the variability of the residuals can be contributed to by non-constant leverages, as discussed in section 3.3.1.2.

Furthermore, in Figure 4.2, 4.5 and 4.6 it can be noticed that some values appear extreme in comparison with the majority of the data. For the sake of improved clarity, in these figures they have been labelled by their run order (1, 7, 10 in Figure 4.2; 6, 10, 59 in Figure and 4.5 and 4.6). The relatively extreme status of these points could be due either to the randomness of the measurement process or to unforeseen and unpredictable factors. In this study, these extreme values were considered to be caused by randomness and thus they were regarded as typical of the phenomenon under investigation. Hence, they were not removed from the data set. As a result, the estimate of the repeatability standard deviation is either correct or constitutes an upper bound for this unknown parameter.

In this section, not only has the quantitative analysis provided an estimate of the repeatability in pre-specified experimental conditions, but it has also ascertained the presence of a sequence-dependent pattern in the realised residuals of the ANOVA model. This pattern, called drift, is further investigated in the next section.

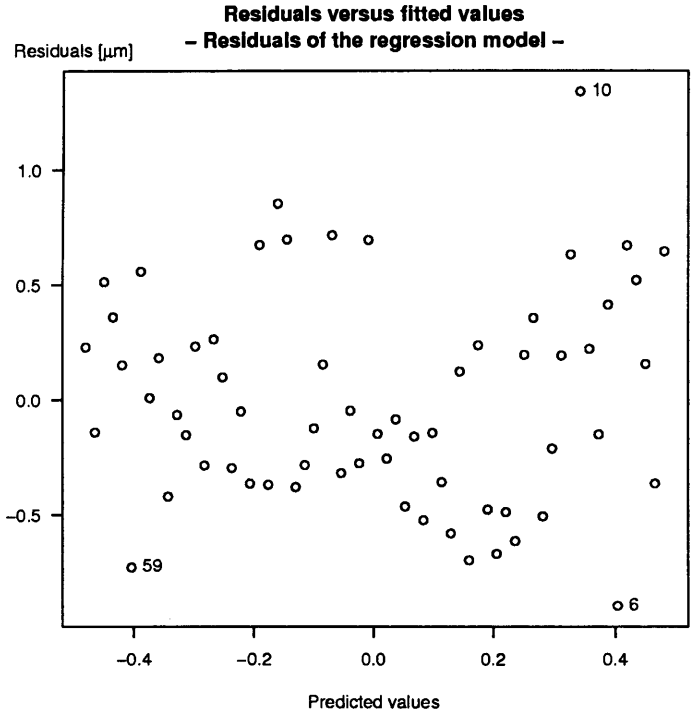


Figure 4.6: Realisations of the regression model residuals versus fitted values.

4.4 Further investigation into the drift of the measurement results

The existence of a linear dependence of the realisations of the residuals, $\hat{e}_{ij}$, on the test sequence can be accounted for by the effects of one or more factors that are not explicitly included in the model. In particular, this drift could be due to a continuous shift in the environmental conditions or to other factors directly inherent in the measurement process. For example, the characteristics of the motion during the scanning might contribute to this drift (cf. Schmit and Olszak, 2002).

In order to establish whether a contribution to this drift is given by some intrinsic component of the instrument, a further experimental investigation was carried out. In particular, 10 measurements test were performed on the same 40 μm nominal step height, with scan length = 44 μm and with the same initial scanner element position (the farther away from the measurand, cf. section 3.3.3). The measurements were taken in a time period of about 30 minutes. Therefore, given the laboratory thermal capacity and its thermostatic control system, it is believed that this time interval of 30 minutes is not sufficient for having changes in the environment so large as they possibly affect the measuring process. The results of this investigation are shown in Figure 4.7.

In this Figure, the predicted values given by two linear models that have been fitted to the data are displayed together with the measurement results. The model associated with the continuous line in Figure 4.7 is given by the following equation:

$$h_i = \beta_0 + \beta_1 i + e_i \quad (i = 1, \ldots, 10) \tag{4.7}$$

where h_i is the i-th measurement result, e_i's are the errors and β_0 , β_1 are the unknown parameters of the model. By applying the ordinary least squares method to equation 4.7, the experimental data lead to the equation

$$h_i = 33.6342 + 0.0659 \cdot i + e_i \quad (i = 1, \ldots, 10) \tag{4.8}$$

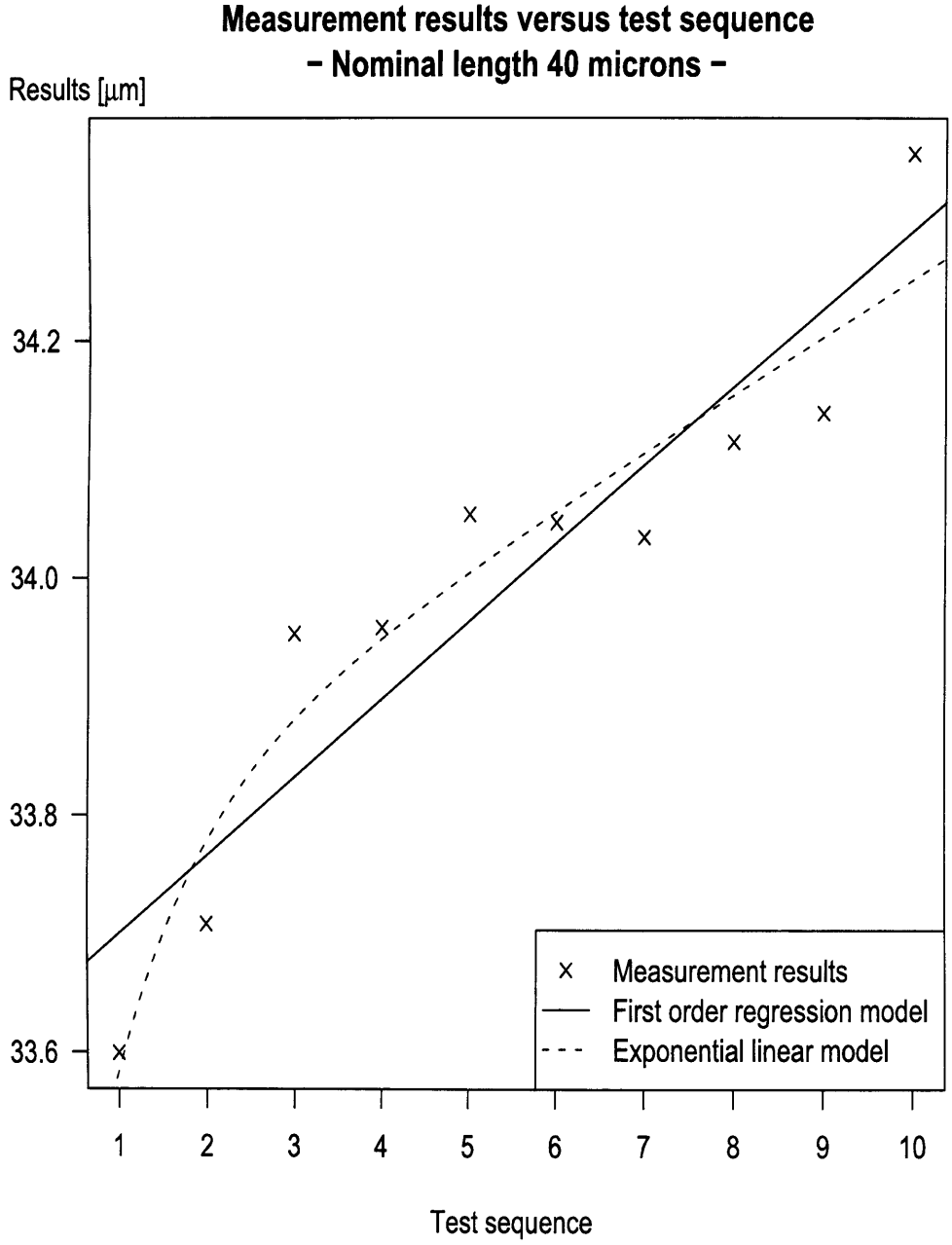


Figure 4.7: Measurement results versus run order.

with $R^2 = 86\%$, and a $P_{value} = 1.1 \cdot 10^{-4}$ while testing the significance of the regression ($F_0 = 49.32$). Thus the hypothesis $\beta_1 = 0$ is rejected. It also follows, from equation 4.6, that the estimated standard deviation of the errors resulted in $\hat{\sigma} = 0.0852\,\mu m$.

Two major conclusions can be drawn from equation 4.8. Firstly, the sequence in which the tests are carried out has an effect on the measurement results. Secondly, the positive value $\hat{\beta}_1 = 0.0659\,\mu m$ found in this investigation, contrasts with the corresponding negative value $\hat{\beta}_1 = -0.0153\,\mu m$ estimated in the regression of the residuals on run order in section 4.3. Furthermore, even without any additional derivation of the confidence interval for these parameters, from the already run F-test for the significance of the regression, it can be argued that their difference is significant and not due solely to experimental randomness. In fact, in this section β_1 is significantly larger than zero, whereas in section 4.3 it is significantly strictly negative. Consequently, albeit this investigation ascertains that some inherent component of the WLI microscope exerts a significance influence on the drift of the measurements with the run order, more investigation is required to establish the relationship between this drift and its potential causes. The data at hand cannot, in fact, support unambiguous conclusions.

It can also be noted that $\hat{\sigma}$ in this investigation is about five times smaller than the corresponding estimate in the regression of the residuals on the run order ($\hat{\sigma} = 0.462\,\mu m$, cf. the previous section). It is suggested that this can be due to the sources of variability included in the previous precision study and removed in this investigation. In particular, in the study presented in this section, the stage was never moved, the scan length was kept constant for all the tests and the initial scanner element position was allowed very limited variations.

From a graphic qualitative analysis of Figure 4.7, it emerges that the fitting of the first order linear model can be significantly improved. In fact, the data suggest that the model should also account for a curvature term in the dependent variable. Therefore, a linear model including an exponential relationship with the dependent

variable has been proposed, namely:

$$h_i = \beta_0 + \beta_1 \cdot i + \beta_2 \cdot e^{-i} + e_i \quad (i = 1, \ldots, 10) \tag{4.9}$$

where β_0, β_1, β_2 are the model parameters to be estimated, e_i's are the errors and e in the exponential term e^{-i} is the Napier's constant. Equation 4.9 represents at the same time a linear model in the unknown parameters and an exponential function in the regressor variable run order, i. Thus, in this study, it is conventionally referred to as exponential linear model, in the sense just explained.

By applying the ordinary least squares method to equation 4.7, the experimental data lead to the equation:

$$h_i = 33.7650 + 0.0487 \cdot i - 0.6225 \cdot e^{-i} + e_i \quad (i = 1, \ldots, 10) \tag{4.10}$$

with $R^2 = 91.6\%$, and a $P_{value} = 1.72 \cdot 10^{-4}$ while testing the significance of the regression ($F_0 = 38.16$). Thus the hypothesis $\beta_1 = 0$ is rejected. In addition, from equation 4.6, the estimated standard deviation of the errors resulted in $\hat{\sigma} = 0.0707\,\mu m$.

On the one hand, the exponential linear model is more complex than the first order linear model, due to the fact that it requires the estimation of three parameters rather than just two. Thus, following the principle of the Occam's razor, the simpler model should be preferred, unless it does not provide the same explanatory or predictive effectiveness as the more complex model. On the other hand, the larger models, i.e. the more numerous the parameters, the better the fitting to the data and, consequently, the smaller the RSS. Therefore, a trade-off between good fit and model size is needed. An objective and quantitative manner of establishing this balance while evaluating different competitor models for the same data, is by the use of criterion-based procedures (cf. section 10.3 in Faraway, 2002). In this study, the *Akaike information criterion* (AIC) is deployed. This criterion is defined as follows

(cf. section 6.8 page 185 in Venables and Ripley, 1999).:

$$AIC = -2 \cdot maximised\,log\_likelihood + 2 \cdot p \tag{4.11}$$

where p is the number of parameters in the model. If the Gauss-Markov conditions hold and the observed values[3] are normally distributed, i.e. if $e_i \sim NIID\ (0, \sigma^2)$, then, after a few passages from section 3.3 in Sen and Srivastava (1990) and page 185 in Venables and Ripley (1999), it can be shown that the $maximised\,log\_likelihood = constant - n/2 - n/2 \cdot \ln(RSS/n)$ when σ is unknown and it is estimated via its maximum likelihood estimator RSS/n. Hence, from equation 4.11, it results:

$$AIC = n/2 \cdot \ln(RSS/n) + 2 \cdot p + constant \tag{4.12}$$

Further information about AIC and a comparison with a very similar criterion, the Bayes Information Criterion (BIC), can be found in Akaike (1974) and Schwarz (1978), respectively.

From equation 4.12, it follows that the smaller the AIC, the 'better' the model. In this context, the adjective 'better' means that the model is parsimonious in the deployment of parameters (small p), while providing a good fit (small RSS) relatively to its complexity. In this study, the first order model and the exponential linear model yielded $AIC = -17.112$ and $AIC = -20.186$, respectively. Therefore, the increased complexity of the exponential model appears justified by the improved fit to the data, which is also apparent from the increased value of R^2. It can also be noticed that both the AIC values are negative as a consequence of the small values of RSS. In fact, $RSS = 0.05805$ and $RSS = 0.03495$ for the first order and the exponential linear models respectively.

The functional dependence of the measurement results on the test sequence that emerges from the better model highlights the presence of an asymptote. It is argued, therefore, that the instrument exhibits a possible transient or 'warming up' period.

[3] h_i in this particular case.

From a physical point of view, such a time interval might be associated with the need for some component of the WLI microscope to perform some measurement runs before reaching a regime condition. In particular, it is suspected that such a need might be associated with the thermal behaviour of the electrical motor driving the scanner element. In fact, Schmit and Olszak (2002) pointed out how critical are the WLI microscope quality characteristics of the scanner element's motion.

From a practical point of view, the stated potential presence of a transient period suggests that some idle measurements should be taken before accomplishing a measurement task. Figure 4.7 indicates that not less than 10 idle measurements should be considered.

The models presented in this section and the conclusions derived from them, only hold if the assumptions underlying them are not violated. Thus, an analysis of the residuals is necessary and it is presented in the next section.

4.4.1 Diagnostics of the fitted linear models

The analysis of the residuals for the first order linear model is summarised in Figure 4.8. In part (a) of this figure, the plot of the residuals against the test sequence seems to display a horizontal 'S' shape pattern, rather than being structureless. This might indicate that the functional relationship between the independent and the dependent variables is inappropriate. The same pattern appears in the plot of the internally studentised residuals versus the fitted values (part b of Figure 4.8). Internally studentised residuals are defined as the residuals divided by their corresponding estimated standard deviation $\hat{\sigma} \cdot \sqrt{(1 - h_{ii})}$, under the assumption of homoscedastic errors (cf. equation 3.43). This terminology is consistent with section 7.2 in Faraway (2002), even though elsewhere the term standardised residuals is also used (cf. section 6.3 in Venables and Ripley, 1999). From the fact that this plot presents a similar pattern to that of part (a) of the same figure, it can be argued that the assumed constant variance of the errors does not appear to be violated by the experimental evidence. Moreover, the absolute value of the studentised residuals in


CARDIFF UNIVERSITY

part (b) and (d) of Figure 4.8 is less or equal to 1.5. This means that no value appears to contradict the assumption of constant variance under the assumption of normality. In fact, under the normality assumption, there is only about 13.36% probability for these absolute values to be larger that 1.5. Furthermore, the normality probability plot of part (c) in the same figure does not highlight significant violation of the assumed normality of the errors. From part (d) of Figure 4.8, it can also be noticed that the maximum leverage is about 0.35, which is below the empirical threshold $2 \cdot p/n = 2 \cdot 2/10 = 0.4$ suggested in section 7.1 of Faraway (2002) and in section 10-7.2 of Montgomery (2001). This empirical threshold provides a rule of thumb for the detection of experimental results that exert a great influence on the fitted model. Hence, when fitting a model, the inclusion or exclusion of these experimental results with leverages larger than the threshold can significantly alter the conclusions drawn on the basis of the model.

Figure 4.9 displays the analysis of the residuals for the exponential linear model. In part (a) of this figure, a pattern is still present, although it seems attenuated in comparison with the corresponding plot for the first order model. Similarly to the analysis of the previous model, experimental evidence leading to a rejection of the hypothesis of constant variance of the errors is not apparent. Furthermore, from part (c) of Figure 4.9, the normality of the errors seems not to be disproved by the experimental results. In addition, in part (d) of the same figure, due to the fact that the empirical threshold for detecting influential points is set at $2 \cdot p/n = 2 \cdot 3/10 = 0.6$, it results that removing the first test results when fitting the model can potentially lead to significantly different conclusions. From a further examination of part (a) and (b) of Figure 4.9, it emerges that the tenth measurement result may have been contaminated by the presence of some nuisance factor. In this case, however, its exclusion from the fitting procedure is not expected to lead to any significantly different conclusion. In fact, its leverage is not too extreme and it is less or equal to 0.6 (cf. part d of the pertinent figure). No further investigation has therefore been carried out in relation to the experimental result number 10.

As a consequence of the above comments, the same exponential model was fitted to

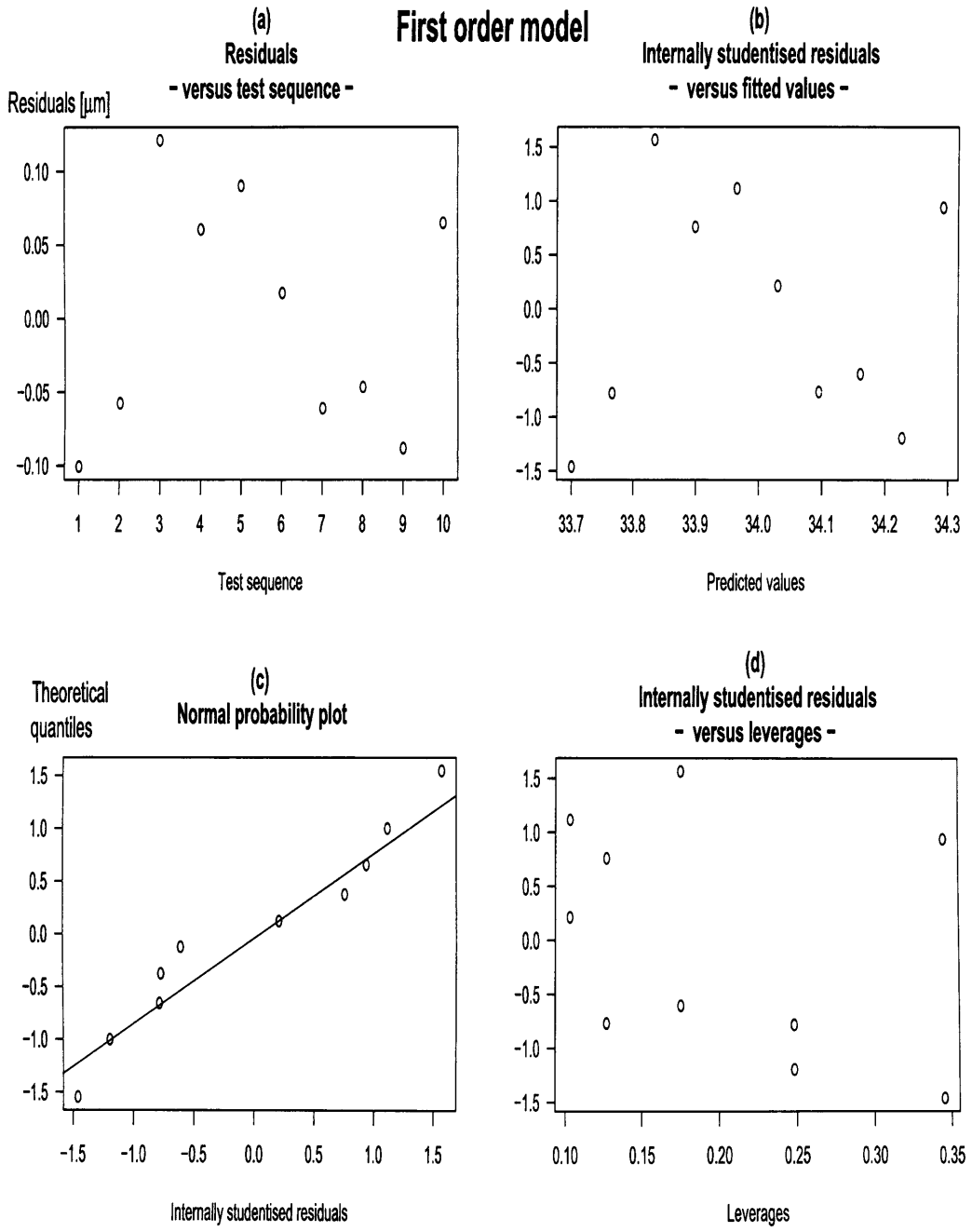


Figure 4.8: Analysis of the residuals for the first order linear model.

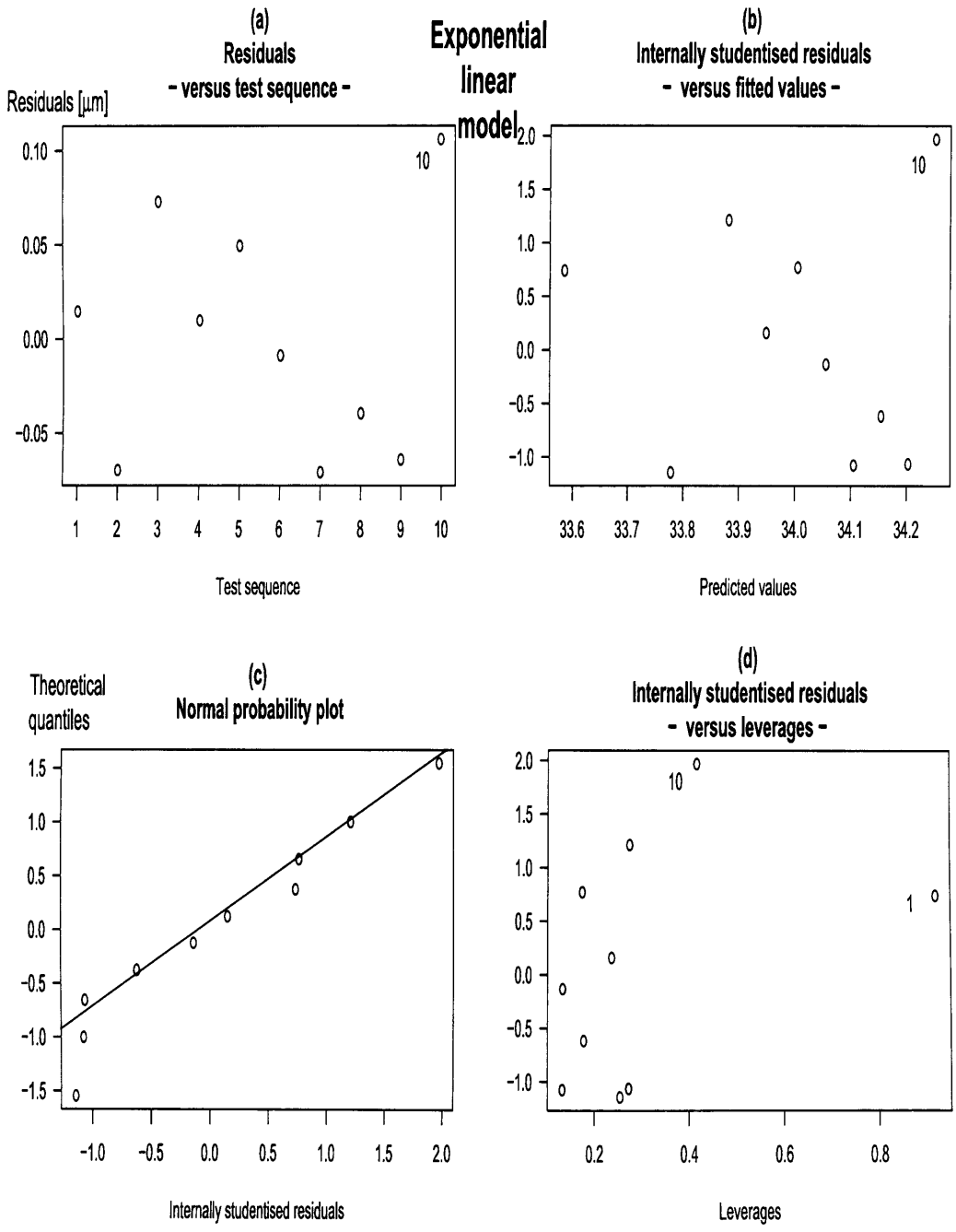


Figure 4.9: Analysis of the residuals for the exponential linear model.

the same data set but leaving out the first test. This yielded the following equation:

$$h_i = 33.8123 + 0.04280 \cdot i - 1.20276 \cdot e^{-i} + e_i \quad (i = 2, \ldots, 10) \tag{4.13}$$

with an estimated standard deviation of the errors $\hat{\sigma} = 0.07327\,\mu m$, a percentage of variance explained $R^2 = 86.62\%$, and a test for significance of the regression yielding $F_0 = 19.43$ with $P_{value} = 2.39 \cdot 10^{-3}$. The fitted model of equation 4.13 does not appear to change the conclusions already drawn on the basis of the previous fitting. Therefore, no further adequacy check of the model assumptions has been performed.

4.5 Discussion of the results

The experimental results have been analysed in this chapter by making use either of ANOVA models or of regression models. In reality, an alternative methodological approach of analysis could have been adopted once the presence of a drift in the measurement results during the precision study was discovered. In fact, the dependence of the measurement results both on the step height and on the run order, can be seamlessly considered in one single linear model. A linear model of such a nature, including both categorical variables (the step height) and quantitative regressors (the sequence order) are also referred to as analysis of covariance models (ANCOVA). More information about this particular application of a general linear model can be found in chapter 15 of Faraway (2002), section 6.1 of Venables and Ripley (1999) and section 14 – 3 of Montgomery (2001). In this investigation, an ANCOVA approach was not followed because of a preference for spending more resources on interpreting the results obtained from the performed analysis rather than carrying out further analyses.

On one hand, the estimate of the repeatability standard deviation for measurements of length in the range [150 , 250] μm was $\hat{\sigma} = 0.55\,\mu m$. On the other hand, estimates of the uncertainty of the reference samples of length can be obtained from equation

3.25. They range from 60.65 nm to 66.80 nm for the investigated samples of nominal lengths from 150 μm to 250 μm, respectively. Consequently, the ratio between the precision estimate and the uncertainty estimates is about 10 to 1. Therefore, it appears that the method for building traceable references of lengths is appropriate for carrying out experimental activities on the studied WLI microscope in the aforementioned range of lengths. From an economic point of view, this finding is quite significant. In fact, the sole alternative to the proposed method for building samples of a range of lengths appears to be the purchase of the corresponding ad hoc reference materials, whose cost is significantly greater (at the present time, a certified full set of gauge blocks has costs of the order of a single item ad hoc reference material).

The provided estimate for the repeatability standard deviation constitutes the main guideline when selecting an instrument for performing a pre-specified measurement task. In fact, a measurement instrument should have sufficient capability for performing the assigned task. In particular, the precision of an instrument has to be significantly lower than the variability of the quality characteristic that is meant to be measured. In fact, the measuring system should have enough capability to detect variations in the measurand that are considered critical. As an extreme case, if attention is not paid to this aspect, the measurement results will be characterised by a variability due to the sole instrument rather than to the measurands themselves. Montgomery and Runger (1993) report, as an example, a case where it is acceptable to have a measuring system contributing to the variability of the measured part for about 29%. Assuming the same situation, this study highlights that measurands with dimensions in the range [150, 250] μm and a variability not less than $\sigma_{part} = 0.55/0.29 = 1.90\,\mu m$ can be reliably measured in the same experimental conditions of this investigation. However, if the sources of variability due to the set up parameters (scan length and initial scanner position) and to the drift of the measurements with the run order are further analysed and eliminated, then more demanding measurement tasks can be performed and the spectrum of potential applications can be broadened.

4.6 Compendium

In this chapter, a precision study of the WLI microscope in repeatability conditions was carried out. The sole source of variability explicitly included in the analysis was the nominal length of the measurands. Other sources of variability, such as the setup parameters analysed in Chapter 3 (scan length and initial scanner element position) have been implicitly accounted for by randomisation techniques. In fact, it is believed that an operator approaching the instrument after a standard training course is unlikely to consider explicitly the effects of these set-up parameters on the precision. The main conclusions of this investigation are the following:

- The adequacy of the WLI microscope for performing a pre-specified measurement task in the range of lengths from 150 μm to 250 μm can be quantitatively assessed. Thus, this precision study constitutes a guideline for the selection of the WLI microscope among numerous potential alternative instruments for performing a particular measurement task (for instance, co-ordinate measuring machines equipped with optical probes).

- The method for building traceable samples of length presented in Chapter 3 provides lengths with an uncertainty that is negligible for a large number of applications, when compared with the repeatability of the WLI microscope. Thus, these samples of length can be deployed as reference materials when using the WLI microscope in the range [150 , 250] μm.

The precision study revealed the presence of a dependence of the measurement results on the sequence of the test. This dependence has been quantitatively analysed and modelled. The main conclusions that have been drawn are:

- A clear contribution to the dependence of the measurement results from the run order comes from some unidentified component of the WLI microscope and not from the surrounding environment. It is suspected that critical components to which these effects can be ascribed are the elements of motion and control of the scanner.

- The overall functional relationship between results and sequence of the test has a negative slope, whereas the sole unidentified component contributes to the same relationship with a positive slope. Therefore, the contribution of the unidentified component to this overall functional relationship is not unique. It is expected that other contributions exist.

- The existence of a transient and a steady state while operating the instrument, is suggested by the asymptotic function fitted to the experimental data. Thus, more stable performances are to be expected after running a few idle measurements prior to accomplishing the measurement task. The idle measurements should be not less than ten, according to the reported experimental evidence.

Chapter 5

Developing a method for the calibration of the white light interferometric microscope

In this chapter the calibration procedure built in the WLI microscope is analysed. Section 1 establishes a link between this procedure and the standard linear calibration. The derived interpretation is supported quantitatively by experimental evidence. The interpretation proposed also highlights the limitations of this built-in procedure. Among these limitations, the deriving potential bias is especially investigated. In fact, in section 2, quantitative experimental evidence and analysis are provided, displaying the presence of a bias induced by the built-in calibration procedure. Therefore, in section 3, a novel spline-based calibration procedure is presented in order to overcome this problem. Subsequently, in section 4, a novel control method for assessing the necessary stability of the measuring system, both during the calibration experiment and during its operational deployment, is described.

5.1 White light interferometer built-in calibration procedure

The calibration procedure for the WLI microscope deployed in this investigation is detailed in appendix D of ADEP1 and of ADEP2. From a practical point of view, it consists in measuring the length of a reference material first, and then setting a value for the parameter called 'height correction' in the software driving the instrument. This value is set automatically by the software, once the measured and the accepted values of the reference material have been provided as an input to the software. No further information about the exact meaning of this parameter and about the calibration procedure has been given by the equipment manuals. Consequently, an interpretative effort has been made, in order to establish the meaning of the 'height correction' parameter and of the calibration procedure. The resulting interpretation is presented in the next section.

5.1.1 Interpretation in the light of the standards

In this section, an interpretation of the 'height correction' parameter is provided by comparing the procedure presented in appendix D of both ADEP1 and ADEP2 with the one-point calibration method described in section 8.2 of BS ISO 11095.

In the one-point calibration method, a first order regression model of the measurement result on the accepted length of the reference material is assumed to constitute an adequate calibration curve. In this study, the accepted length of the reference material is its nominal length and the reference materials are those discussed in Chapter 3. Furthermore, the regression model has only one unknown regressor coefficient, β_1, due to the fact that the intercept parameter β_0 is set to zero. This is the same as assuming that the measuring system is yielding a zero value when measuring a reference material of accepted length equal to zero. In reality, therefore, this method is based on two points and not just one. The linear model is described

by the following equation:

$$h_k = \beta_1 \cdot l_{n,s} + e_k \qquad (with\ k = 1, \ldots, n) \tag{5.1}$$

where h_k is the k-th measurement result of the reference material of nominal length $l_{n,s}$ and $e_k \sim NIID(0, \sigma^2)$ are the errors. It is immediately verifiable that the ordinary least squares method results in the following estimate of β_1:

$$\hat{\beta}_1 = \frac{\bar{h}.}{l_{n,s}} \tag{5.2}$$

In equation 5.2, $\bar{h}.$ indicates the average over all the n measurement results. Furthermore, an estimate of σ^2 is obtained as described in section 4.3 and is given by

$$\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (h_k - \bar{h}.)^2 \tag{5.3}$$

It can be noticed that the estimate provided in equation 5.3 is identical to the sample standard deviation of the measurement results.

The main result of a one-point calibration experiment is the estimation of the calibration function, namely:

$$h = \hat{\beta}_1 \cdot h^* \tag{5.4}$$

with $\beta_1$given by equation 5.2, h representing the measured length and h^*representing the estimate of true value of the unknown length. Consequently, a future measurement result is transformed with the calibration function in order to obtain h^*. In particular, from equation 5.4, the transformation is:

$$h^* = \frac{h}{\hat{\beta}_1} \tag{5.5}$$

In this study, the 'height correction' parameter, HC , has been interpreted as the coefficient of the transformation, namely

$$HC = \frac{1}{\hat{\beta}_1} \tag{5.6}$$

Moreover, h^*represents the value displayed by the software of the instrument when performing a measurement task with a pre-specified HC, namely it holds $h^* = h^*_{displayed}$. Furthermore, h represents the raw results of the measurement, not displayed to the user, but a value that is expected to be stored internally by the instrument in some location of memory. Thus, from equation 5.5 and equation 5.6, it follows that if and only if $HC = 1$ then $h^*_{displayed} = h$.

In addition, it has been noticed that setting the HC by utilising the same experimental data, but starting from different pre-existing values of 'height correction', HC_0's, leads to different updated HC_1's values . This is explained as follows. On the one hand, from equation 5.5 and equation 5.6 the measurement results displayed by the software when $HC = HC_0$ are given by:

$$h^*_{displayed} = HC_0 \cdot h \tag{5.7}$$

On the other hand, h represents the measurement result that has to be associated with the reference material when building the new calibration curve leading to HC_1, namely:

$$l_{n,s} = HC_1 \cdot h \tag{5.8}$$

From equation 5.7 and equation 5.8 it therefore follows that:

$$HC_1 = HC_0 \cdot \frac{l_{n,s}}{h^*_{displayed}} \tag{5.9}$$

Figure 5.1 illustrates the logical process deployed in deriving the quantitative relationship between $HC_0$and HC_1.

The fact that equation 5.9 allows the user to predict the exact values of 'height correction' returned by the software constitutes strong supporting evidence for the correctness of this proposed interpretation.

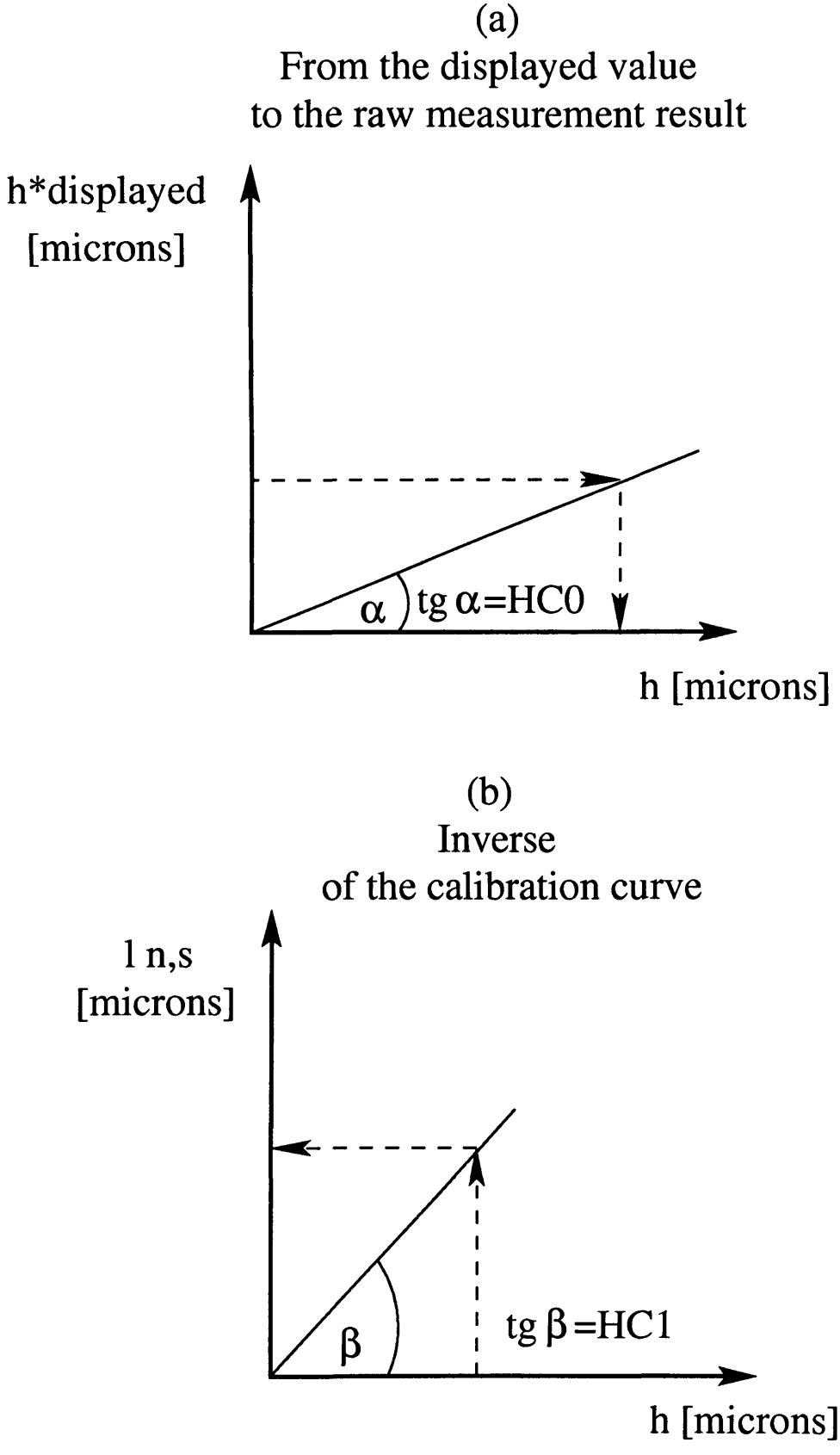


Figure 5.1: Rationale of the interpretation of the 'height correction' factor.

5.1.2 Limitations

The description of the calibration procedure reported in section D of ADEP1 and ADEP2 appears inadequate not only due to the fact that it does not provide any interpretation of the height correction factor, but also from a number of other points of view.

Firstly, in contrast with BS ISO 11095 (cf. ibidem section 5.3.4), the necessity of performing more than one measurement in order to set the height correction factor is completely neglected in ADEP1 and ADEP2. Hence, the means whereby the standard deviation of the errors can be estimated is completely overlooked in these manuals. More generally, the variability of the measurement results is not taken into any account. In addition, avoiding replicating the measurements prevents a check of the assumptions of the regression model using an analysis of the residuals. Only a single realisation of the residuals is, in fact, available. In particular, BS ISO 11095 in section 8.2.2 prescribes that, in a one-point calibration, the reference material should be measured at least twice.

Secondly, in these manuals no guidance has been given about the selection of the reference materials in connection with the measurement task that it is intended to carry out. However, BS ISO 11095 in section 8.2.3 prescribes that the accepted value of the reference material should be greater than the values encountered by the measuring system while accomplishing its measurement task. As an example, this means that it is inappropriate to use a $10\,nm$ traceable sample for calibrating the system, if the measurement task involves lengths in the micrometric range. The rationale behind this prescription may be the fact that the calibration curve, as every regression model, maintains its predictive and explanatory validity only in the experimental region used to estimate its parameters. Consequently, the 'one-point' calibration can be deployed only in the range of the corrected lengths ranging from zero to the accepted value of the reference material. The parameters of the model, β_1 and σ, were, in fact, estimated at these two points.

Thirdly, in the aforementioned manuals, no mention is made of the assumptions

underlying the one-point calibration model. According to section 8.2.2 of BS ISO 11095, they can be summarised by three main points:

- The accepted value of the reference material does not have any error. If this assumption does not hold, the model of analysis has an increased complexity (cf. chapter 5 in Faraway, 2002).
- The standard deviation of the residuals is constant.
- The relationship between the measuring system and the reference system, namely the calibration function, is linear in the interval from zero to the accepted value of the reference material.

This last assumption of linearity severely restricts the potential usage of the one-point calibration method. In fact, in section 8.2.1, BS ISO 11095 recommends that this method should not be deployed for calibration purposes. By contrast, it should be used essentially for the inspection of an already existing linear calibration function.

From the above considerations about the linearity and the selection of the reference material, it is argued that some room for using the one-point method in the calibration of the WLI measuring system does, however, appear to exist. In fact, from a mathematical perspective, every calibration function sufficiently regular to be expanded as a Maclaurin series (cf. Weisstein, 2006d), can be expressed by the sum of a first order polynomial function and an infinitesimal function of order at least equal to the amplitude of the interval where the function is defined (cf. Weisstein, 2006b). In this context, using the Landau symbols, the sentence ' function $f(x-a)$ of order equal to $(x-a)^n$ ' , with $a = constant = 0$, means $f(x-a) = o((x-a)^n)$ (cf. Weisstein, 2002).

Hence, the closer to zero the accepted value of the reference material, the more negligible the error in approximating the calibration function with a linear function. It may therefore be inferred that this WLI measuring system is designed for 'small'

accepted lengths of the reference material and thus for even smaller lengths of the measurands. The quantification of the adjective 'small' is out of the scope of this investigation. In fact, what has been addressed in this study is the usage of WLI microscopic techniques in the micrometric range, which represents the middle-upper range of the measuring tasks that these techniques can perform.

In the light of the aforesaid characteristics of the built-in calibration procedure, a main technological choice made by the manufacturer appears to be technically unconvincing. In fact, on the basis of an original design allowing the user to set a restricted interval of small scan lengths, the WLI system under investigation was subsequently provided with a scanner element extending significantly this range of scan lengths towards greater values (up to a few millimetres). Notwithstanding, the one-point calibration procedure has not been modified accordingly. Therefore, even though the scanner can travel significantly longer distances, the trueness of the measurement results in the micrometric range and above appears to be seriously compromised. In order to ascertain quantitatively the presence of bias induced on the measurements by the inadequate calibration procedure, an experimental investigation was carried out and is presented in the next section.

5.2 Bias assessment of the white light interferometric microscope

The built-in calibration procedure has been identified in the previous section with the one-point calibration method defined in section 8.2 of BS ISO 11095. This procedure is founded on the hypothesis of a linear relationship between the measurement results and the accepted values of the reference materials.Otherwise stated, it is based on the hypothesis of linearity of the calibration function. This assumption is experimentally investigated in this section.

For the sake of simplicity, the height correction parameter has been set at 1, namely $HC = 1$. Hence, as previously mentioned in section 5.1.1, the measurement results

displayed by the instrument are the same as those internally stored, i.e. $h^*_{displayed} = h$. Therefore, if the hypothesis of linear calibration curve is correct, then the difference between the measurement results taken on a set of reference materials and the true lengths of these reference materials should exhibit a linear relationship against these true lengths. This circumstance holds as long as the calibration function is linear and in spite of the fact that $HC = 1$ can be an inappropriate value for the height correction. In fact, if $HC = 1$ is inappropriate but the calibration function is linear, the aforementioned differences $h_{ij} - L_{s,i}$ have still to display a linear relationship with $L_{s,i}$. The symbol h_{ij} denotes the j-th measurement taken on the i-th reference material.

Figure 5.2 illustrates these concepts by means of the simulation of an example. In its part (a), simulated measurement results have been generated according to model $h^*_{display,ij} = h_{ij} = 3 \cdot L_{s,i} + e_{ij}$ with $i = 1, \ldots, 5$, $j = 1, \ldots, 10$, $\{L_{s,i}\} = \{0, 1, 2, 3, 4\}$ μm and $e_{ij} \sim NIID\,(\mu = 0, \sigma = 0.5)$. The data have been generated using R on a GNU/Debian system. 'R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories by John Chambers and colleagues. R can be considered as a different implementation of S' (cf. R Development Core Team, 2006). The abscissae of the data in both the parts of the figure instead of being just 5, i.e. $\{0, 1, 2, 3, 4\}$ μm, are as many as the data. In fact, for the sake of clarity of the figure they have been randomly offset by using a function of the R language (jitter()). In part (b) of Figure 5.2, the structural part of the theoretical model that generated the data, i.e. $h^*_{display,ij} = h_{ij} = 3 \cdot L_{s,i}$ cannot be distinguished from the first order linear model fitted to the same data. This represents a simulation of a perfectly linear measuring system. In real situations, had the fitted model in this figure departured from a first order linear model, then it could be concluded that the measuring system is not linear.

Furthermore, the true lengths $L_{s,i}$'s cannot be known. In spite of this fact, it can be derived from the specification limits LSL_{L_s} and USL_{L_s} presented in equation 3.8 and equation 3.9 and from BS EN ISO 3650 (cf. ibidem table 5), that the $L_{s,i}$'s

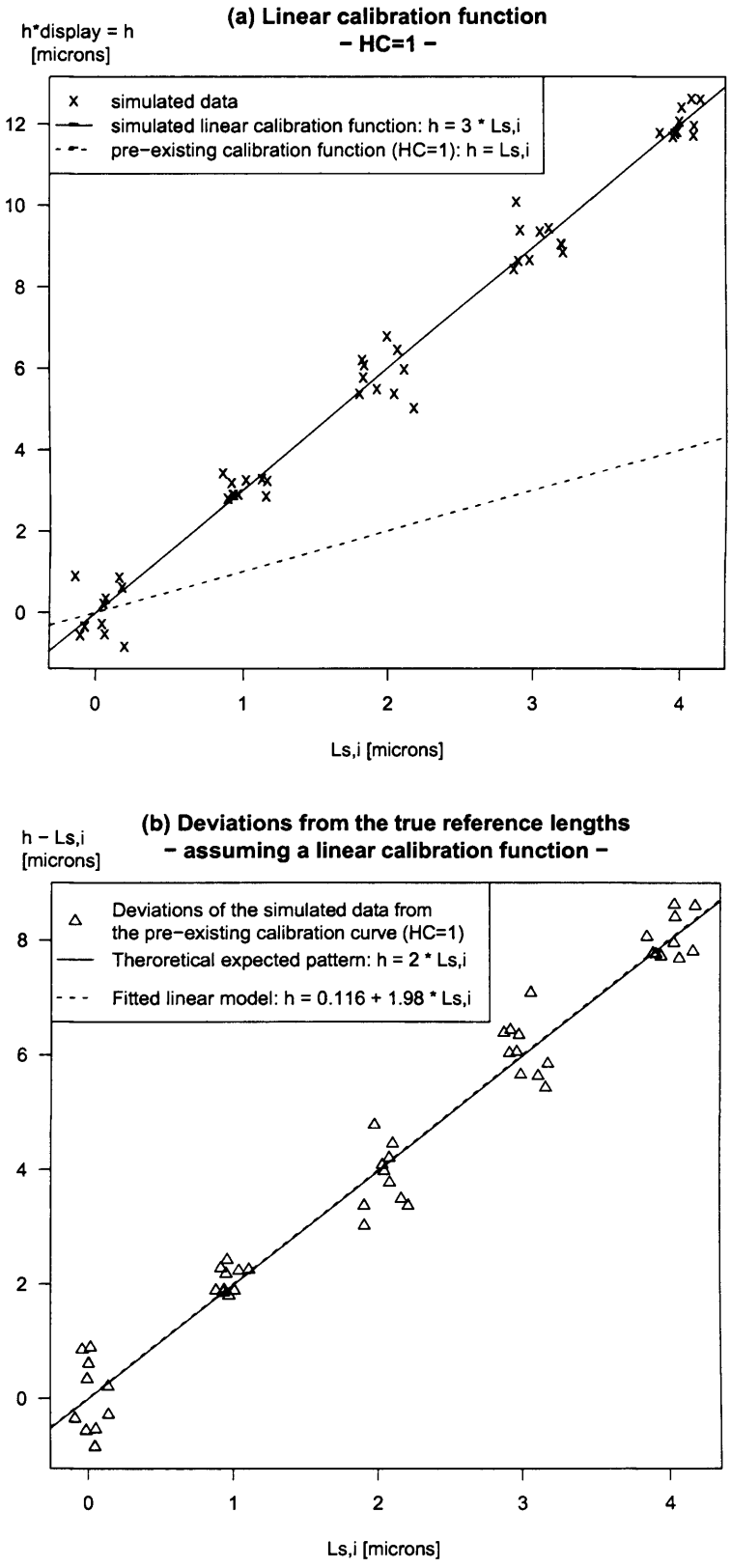


Figure 5.2: Simulation of a measuring system exhibiting a linear behaviour.(a) Recalibration based on 5 virtual accepted values (0, 1, 2, 3, 4) (b) Linearity of the deviations from the reference systems.

have to lie in the interval $l_{n,s} \pm 0.4\,\mu m$ for the gauge blocks used in this study.

Thus, if the accepted values of the reference materials are set at their nominal value $l_{n,s}$'s, then the expected maximum error $L_s - l_{n,s}$ is less than or equal to $0.4\,\mu m$. Hence, estimating the difference $h_{ij} - L_{s,i}$ by means of $h_{ij} - l_{n,s,i}$ involves a potential error that is limited in absolute value to $0.4\,\mu m$. Therefore, the deviations

$$d_{ij} = |h_{ij} - l_{n,s,i}| - 0.4\,\mu m \quad i = 1, \ldots, a\,;\, j = 1, \ldots, n \tag{5.10}$$

can be accounted for only by the measuring system and not by the reference materials. Alternatively, these deviations could have been calculated using the semi-amplitude of the expanded uncertainty of the reference materials as can be computed from the discussion in section 3.2.3. Nevertheless, the usage of the semi-interval of the specifications of the references is significantly larger than the corresponding semi-interval of the expanded uncertainty (for instance at 95% significance level). Thus, it was preferred, in order to have a simpler approach and to be more conservative.

Furthermore, if the relationship between the d_{ij}'s and the $l_{n,s,i}$'s exhibits evident departures from linearity, then it can be concluded that the assumption of a linear calibration function for the WLI microscope is not supported by the experimental data.

Figure 5.3 displays the plot of the d_{ij}'s and the $l_{n,s,i}$'s. The hypothesis of linearity does not appear to be supported. In fact, first and second order linear models were fitted to the data of Figure 5.3. On the one hand, the first order linear model resulted in $R^2 = 57.8\%$, $\hat{\sigma} = 1.08\,\mu m$ and $AIC = 195.45$. Furthermore, the test for significance of the regression yielded $F_0 = 84.91\%$ on 1and 62 degrees of freedom, with $P_{value} = 3.19 \cdot 10^{-13}$. On the other hand, the second order linear model resulted in $R^2 = 79.08\%$, $\hat{\sigma} = 0.767\,\mu m$ and $AIC = 152.54$. In addition, the test for significance of the regression gave $F_0 = 115.3$ on 2 and 61 degrees of freedom with $P_{value} = 2.2 \cdot 10^{-13}$.Thus, on the basis of the AIC values, the second order model appears to provide a more adequate interpretation of the data than the first order model (cf. section 4.4).

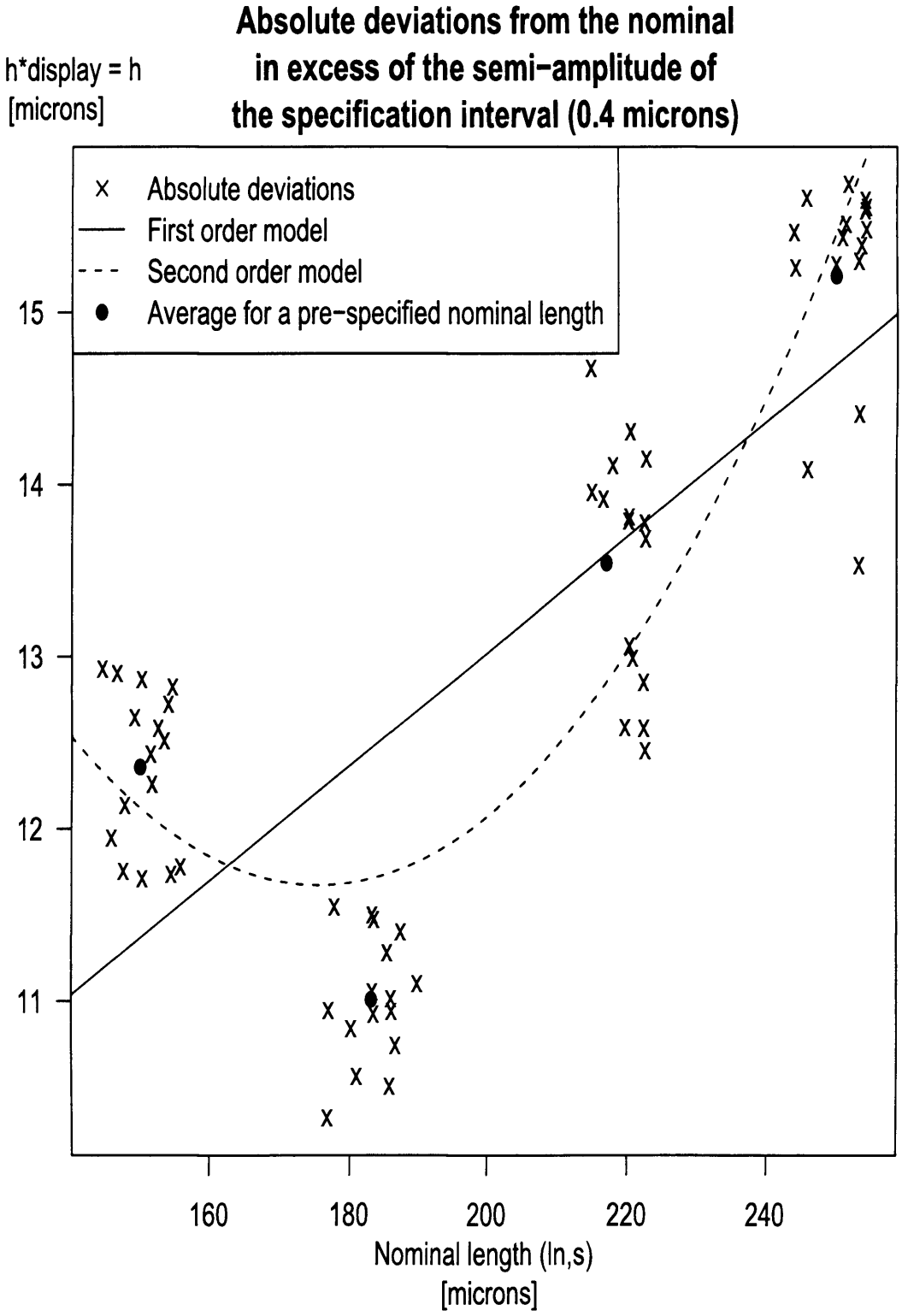


Figure 5.3: Absolute deviations d_{ij} and fitted linear models of first and second order.

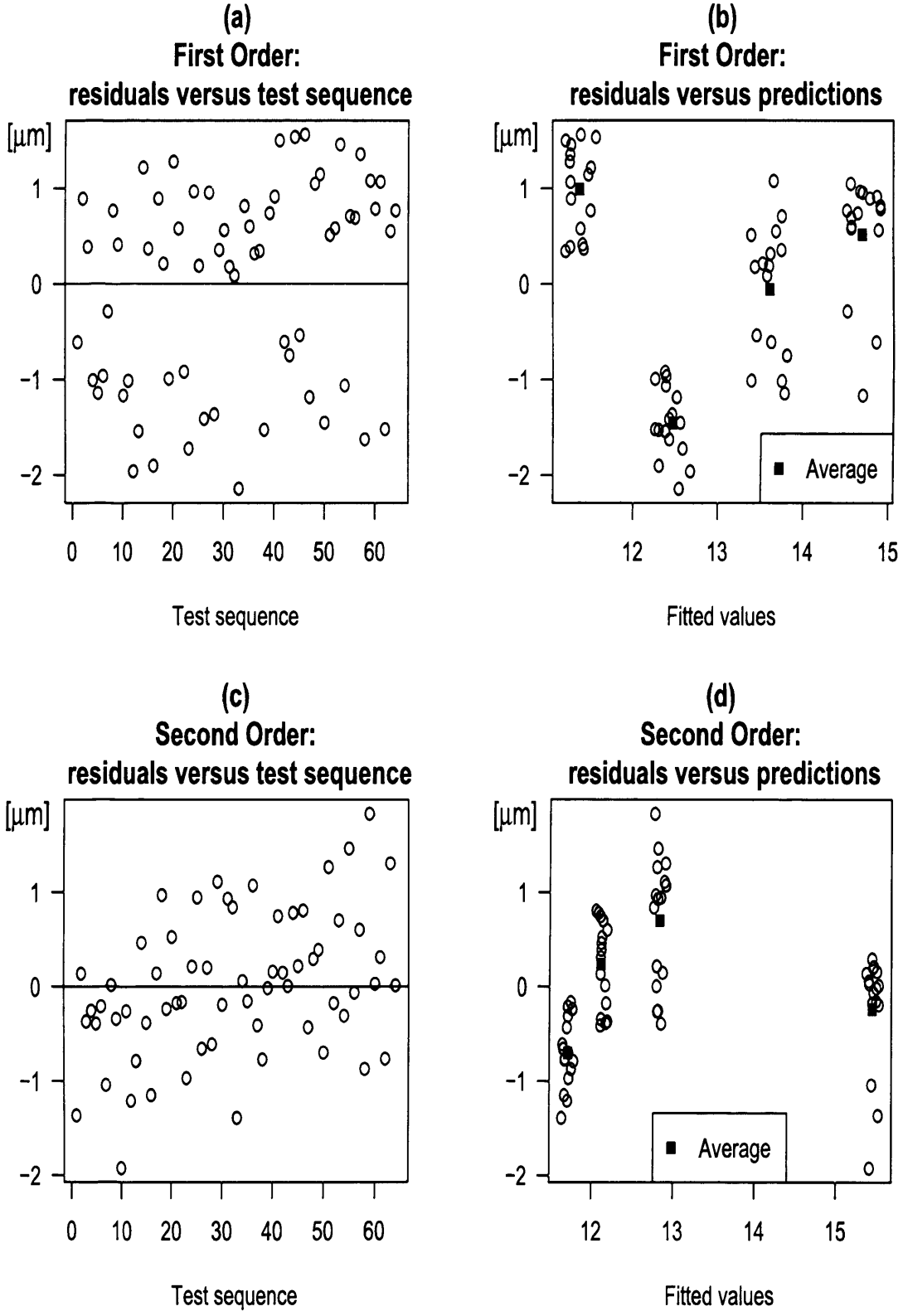


Figure 5.4: Residuals versus test sequence and versus fitted values for the first and second order models.

The plots of the residuals against the run order and the fitted values for both the first and the second order linear models are displayed in Figure 5.4. In this figure, the fitted values have been randomly displaced from their actual values, in order not to have overlapping data (cf. the function jitter() in the R language, R Development Core Team, 2006). No significant patterns are apparent in the plots of the residuals against the run order, although the presence of a mild ascending pattern might be suspected in the second order model. For both the models, however, no major violations of the assumed independence of the errors are supported by the experimental evidence. From the plots of the residuals versus the predicted values, two main conclusions may be drawn. First, the spread of the realised residuals for each fitted value does not appear to vary significantly with the predicted values. Thus, no violation of the assumed constant variance of the errors seems to be supported by the experimental evidence. Second, it can be argued that both the models display some degree of relationship between the residuals and the fitted values as these plots do not appear structureless. It is inferred, therefore, that models including other predictors and/or characterised by alternative functional relationships with the predictors can provide a better insight into the experimental data. Notwithstanding, this investigation is beyond the scope of this study. The purpose of this study is, in fact, to show how the assumption of linearity of the built-in calibration is most questionable.

In Figure 5.3 it can also be noticed that a large contribution to departures from linearity of the calibration function comes from the residuals associated with the nominal reference length $l_{n,s} = 183\,\mu m$. In fact, in connection with this nominal length, the figure displays the maximum deviation of the measurement results from the prediction of the first order model. Such a deviation is $1.45\,\mu m$, due to the fact that the prediction is $12.46\,\mu m$, whereas the average measurement result is $11.01\,\mu m$. From a conservative point of view therefore, the possibility that some unforeseen nuisance factor has affected the build of this particular reference sample cannot be excluded. If that were the case, the analysis presented in this section might have given different outcome regarding the linearity of the calibration function of the measuring system. However, even in these hypothetical circumstances, the built-in

calibration procedure would still be inadequate, in spite of the compliance of the measuring system to the linearity assumption. In fact, the first order regression model in Figure 5.3 has an intercept $\hat{\beta}_0 = 6.37\,\mu m$, where this estimate has been obtained by means of the ordinary least squares method. Furthermore, from a qualitative graphical inspection of Figure 5.3, it is inferred that this estimate would increase if the measurements of the sample with $l_{n,s} = 183\,\mu m$ were compliant to the linearity assumption. The presence of an intercept different from zero is also incompatible with the one-point calibration method and so, with the built-in calibration procedure.

On the basis of the experimental evidence, therefore, and on the basis of this analysis, it can be concluded that the calibration procedure built in the WLI microscope does not appear do be satisfactory. It is, in fact, based on assumptions that are unrealistic in the range of lengths investigated, i.e. in the range $[150\,,\,250]\ \mu m$. These unrealistic hypotheses are either the linearity of the calibration function or the absence of an intercept term in the calibration function. Consequently, a new calibration procedure taking into account these findings is introduced in the next section.

5.3 Proposed calibration procedure

5.3.1 Description

In the previous section it has been pointed out that there are two main critical points in the calibration procedure built in the WLI microscope used in this study when measuring lengths in the micrometric range. Namely, they are: the potential departures from linearity of the calibration function and the potential presence of a non-null intercept term in the calibration function.

In order to overcome these problems a new approach is introduced. It is founded upon the following elements:

Firstly, as strongly suggested in BS ISO 11095, more than a single reference material is considered. In particular, the reference materials are selected in such a way that they evenly span the region where the measuring system is to be deployed. The number of reference materials has been chosen in such a manner that each reference material is in the centre of a $3\,\mu m$ interval of lengths on the axis of the reference systems (reference and measuring systems are presented in section 4 of BS ISO 11095). In fact, in section 4.3 the repeatability standard deviation of the WLI microscope has been estimated as amounting to about $\hat{\sigma} = 0.5\,\mu m$ for lengths in the range from $150\,\mu m$ to $250\,\mu m$. Consequently, assuming that the result of a single measurement test is contributed by many independent random factors, each of them having a small, non-prevailing and finite effect, it can be argued, on the basis of the central-limit theorem (cf. sections 4.2 in chapter 5 and 3.3 in chapter 6 of Mood et al., 1974), that such a single result is expected to be normally distributed(similar arguments can be found also in sections 1.4, note 2 of Drapper and Smith, 1966). Hence, under this normality hypothesis, only about 0.27% of the measurement results are expected to fall outside the interval given by the accepted value of the RM $\pm 3 \cdot \hat{\sigma}$, i.e. given by the accepted value of the RM $\pm 1.5\,\mu m$. Thus, on the one hand, having references of length closer to each other of more than $3\,\mu m$ does not bring any benefit. It cannot in fact be resolved whether differences between measurement results of adjacent RM's are due to the difference in the accepted values of the RM's or are due to the precision of the measuring system. On the other hand, if the differences between nominal lengths of contiguous RM were larger than $3\,\mu m$, it would be difficult to support conclusions about the calibration function in areas of the reference system axis not covered by any RM. Therefore, a difference of $3\,\mu m$ represents an optimal balance between these two contrasting needs. The adjective 'optimal' refers to the expected reliability of the data analysis and of the subsequent conclusions drawn from it. However, it has to be noticed that BS ISO 11095 does not provide any detailed recommendation on how to select the number of the reference materials. In section 5.3.3, it merely points out that the RM's in the overall interval of the accepted values should be not less than 3 and that it should be increased in any subinterval where the linearity of the calibration function is questionable.

In section 9.2 the same standard also presents an example of linear calibration in a micrometric range. In this example, there is no degree of freedom left to the experimenter in selecting the accepted values. In fact, they are built in a single sample of standard reference material.

Moreover, as has already been pointed out in section 5.1.2, the likelihood of having a linear calibration function increases while reducing the amplitude of the interval of the accepted lengths. The overall range of accepted lengths over which a calibration function is demanded for accomplishing a pre-specified measuring task is, therefore, divided in subintervals of significantly smaller amplitude. A separate regression model is then fitted to the measurement results of each subinterval. In this manner, there are two main benefits. Firstly, the number of parameters potentially needed by the model is expected to be small, due to the expected closeness of the calibration function to a linear behaviour. Secondly, the possibility of having a calibration function that does not pass through the origin is anticipated to be significantly reduced. In fact, the data points in a subinterval only affect the model defined locally on the same subinterval and do not exert significant influence over the fitting procedure in regions distant from it. This undue global influence of data points could appear , for instance, in the case of polynomial regression (cf. section 8.3 of Faraway, 2002). At the junction point of two models pertinent to two adjacent subintervals however, the presence of any discontinuity is not expected. In fact, the observation of physical phenomena does not usually support discontinuity and sudden changes in the observed physical quantities, without the presence of identified causes. Consequently, in this proposed method, two adjacent models have been constrained to have the same value of the calibration function at the junction points between two adjacent subintervals. In the literature, this technique is referred to as 'broken stick regression' (cf. section 8.2.1 of Faraway, 2002), or 'broken line regression' (cf. sections 4.5 and 9.2.2 of Sen and Srivastava, 1990), when the models in each subinterval are of first order. However, when this approach is generalised so as to account for more complex functional relationships, it is referred to as 'broken curve regression' (cf. sections 4.5 and 9.2.2 of Sen and Srivastava, 1990). In particular, if the functional relationship is described by polynomials in each subinterval, then the resulting func-

tion defined on the overall interval is sometimes called a spline. Nevertheless, in a more strict sense that was not adopted in this study, the word 'spline' is used with an added limitation. This added limitation is that all the derivatives of order $s-1$ exist across the boundary of each subinterval, when the polynomials are of degree s.

Each point delimiting a subinterval is designated as knot. In this study, the selection of the knots has been driven by the experimental conditions. In fact, the knots were selected so as each resulting subinterval represented an homogeneous experimental unit. In particular, the data in each subinterval were gathered in a randomised sequence, whereas no randomisation was performed between data from different subintervals. Thus, the knots were selected so that the resulting experimental effort for each subinterval was affordable. Therefore, the knots constitute an input to the fitting procedure and do not offer augmented degrees of freedom in improving the fitting. Such a circumstance would make the problem of estimation of the model parameters no longer linear (cf. section 9.2.2 of Sen and Srivastava, 1990).

In this study, the proposed method for determining the calibration function represents an extension of the second order spline regression model presented in section 9.2.2 of Sen and Srivastava (1990), which does not allow a first order model to be fitted on the first subinterval. In particular, when one knot is present, the proposed model is given by the following equation, during the calibration experiment:

$$h_{ij} = \beta_0 + \sum_{l=1}^{2} \sum_{s=1}^{ord} \beta_{l,s} \left| k_1 - l_{n,s,i} \right|^s \cdot \delta_l + e_{ij} \quad i = 1, \ldots, a;\ j = 1, \ldots, n \qquad (5.11)$$

In equation 5.11, $k_1$represents a pre-selected knot in the overall interval $[a_{LB}, b_{UB}]$. , whereas *ord* is an integer as large as it is wanted and represents the maximum order of the polynomials that is taken into account. Thus, the proposed model is not limited to the second order as the model presented in section 9.2.2 of Sen and Srivastava (1990) is. During the calibration experiment, the overall number of RM's is a, whereas the number of replicates of the experiment is n. The independent variable is the nominal length of the RM's, namely $l_{n,s,i}$, whereas the

dependent variable $h_{ij}$is the j-th measurement result on the $i-th$ RM. The symbol δ_l represents the indicator function of the subintervals, i.e. $\delta_1 = 1$ ($\delta_2 = 1$) if $l_{n,s,i} \in [a_{LB}, k_1)$ ($l_{n,s,i} \in [k_1, b_{UB})$) and $\delta_1 = 0$ ($\delta_2 = 0$) elsewhere. Some of the unknown parameters $\beta_{l,s}$ can be set to zero independently during the ordinary least squares procedure. In this manner, simpler models can be build and lower AIC potentially be obtained in comparison with the second order model presented in section 9.2.2 of Sen and Srivastava (1990). The spline described in equation 5.11 is continuous in the knot. However, it is differentiable in the knots only if all the terms of first order are zero, namely $\beta_{1,1} = 0$ and $\beta_{2,1} = 0$.

In this investigation, a criterion-based approach to the selection of the order of each polynomial part of the spline was followed. Sections 10.3 and 10.4 of Faraway (2002) do, in fact, strongly recommend using a criterion-based method. However, orders greater than the second have not been examined. In fact, the selection of the RM's with the minimum acceptable difference between adjacent nominal lengths suggests that deviations from linearity are expected to be relatively small.

During the calibration experiment, the height correction was set at one and never changed. On obtaining the calibration function and when performing a generic measurement task, the measurement data from the instrument were exported in a coded text file. These data were decoded first and then they were transformed in the corrected measurement results by means of the calibration function. For the case investigated, the transformation of the data by the calibration function was implemented in a C program on a GNU/Debian system by making use of the GNU scientific library (GSL), (cf. Galassi et al., 2005).

A calibration experiment implementing the procedure described in this section is introduced in the next section for the overall interval of accepted lengths [180.5 , 219.5] μm.

5.3.2 Implementing the proposed procedure

The reference materials deployed in this investigation were the samples built by means of the gauge blocks as described in section 3.1. The procedures for positioning the stage of the WLI microscope and for determining the measurement results of the step heights were also identical to those utilised in the repeatability analysis (cf. section 4.2).

The number of replicates of the experimental design was limited to 3. The maximum affordable experimental burden and the possibility of frequent recalibrations led, in fact, to the selection of a very small number of replicates. The number of replicates chosen in this study is, however, very close to 4 replicates. This is the number selected in the example of section 9.2 of BS ISO 11095, where no supporting comments were provided.

In addition, the number of the reference materials was selected so as to maximise the use of the available surface on the auxiliary plate (quartz optical parallels) during the build of the samples. This resulted in a maximum of 7 RM's available at the same time on the stage of the WLI microscope. Consequently, on the basis of the discussion in the previous section, during each calibration experiment an interval of maximum amplitude $3 \cdot 7 = 21\,\mu m$ can be covered . Therefore, a knot in $k_1 = 201.5\,\mu m$ is sufficient to cover the range of nominal lengths ranging from $180.5\,\mu m$ to $219.5\,\mu m$. Hence, the overall interval results divided into two subintervals with common extreme in $k_1 = 201.5\,\mu m$. Furthermore, the first subinterval is spanned by 7 evenly spaced RM's , whereas the second by 6. Thus, the calibration function is a spline defined by two polynomial functions and it is continuous in k_1.

In the forward criterion-based approach to the selection of the model, the first step consists in calculating the AIC for the two first order polynomial functions that are fitted to the data by means of the least squares method. This yields a broken stick regression model (alias 'broken line' or 'hockey stick' or 'segmented regression'). The second step is fitting two calibration functions. Each of them is of the first order on one subinterval and of the second order on the other. The third step consists in

fitting a spline calibration function made of two second order models. Finally, the AIC of all the fitted models are compared with each other and the spline calibration function with the lowest AIC is selected.

The AICs resulting from this procedure are summarised in table 5.1. Consequently the spline with a second order polynomial function on both the subintervals is selected as a calibration function for the range of accepted lengths [180.5 , 219.5] μm. It can also be noticed that not only has the broken stick regression resulted in two straight lines with slopes significantly different (-0.931 and 1.00), but also that the straight line fitted over the global interval has yielded a $-6.46\,\mu m$ intercept. These remarks confirm how inadequate the assumption of linearity of the calibration function is for the measuring system under investigation. Consequently, the spline approach is preferred to the first order model fitted on the overall interval, even in the case of the broken line where the resulting spline model has higher AIC than this first order model.

The models investigated are displayed in Figure 5.5. It can be noticed that both the quantitative results shown in Table 5.1 and a qualitative visual analysis of Figure 5.5 support the idea that the different competing models are not much different from each other, as was expected from the discussion about the linearity, the Maclaurin series and the Landau symbols in section 5.1.2. However, The benefit of using an approach based on the splines is already apparent, due to the presence of a non null intercept when fitting the first order model on the overall interval.

In part (c) of Figure 5.5, the measurement results have their abscissae randomly distorted (cf. the function jitter() in the R language, R Development Core Team, 2006) in order to improve the clarity of the figure. The average of the measurement results for each RM and the measurement results predicted by the fitted model for each of the RM's deployed in the calibration experiment were calculated using R (cf. R Development Core Team, 2006). They are also shown in part (c) of Figure 5.5. Although from a graphical point of view the agreement between the predicted values and the averages appears qualitatively satisfactory, a test for lack of fit reveals that the fitting performances of the proposed model can be improved (cf. chapter

Model	First order model	Broken stick	First order on [180.5 , 201.5) μm Second order on [201.5 , 219.5) μm	Second order on [180.5 , 201.5) μm First order on [201.5 , 219.5) μm	Second order on both
AIC	132.02	133.11	131.39	134.28	129.40
RSS	57.80	56.47	51.33	55.28	46.34
R^2	98.76	98.79	98.90	98.81	99.01
β_0	−6.46	188.31	188.75	188.55	189.49
β_1	0.968	_	_	_	_
$\beta_{1,1}$	_	−0.938	−0.971	−1.05	−1.23
$\beta_{2,1}$	_	1.00	0.729	_	0.560
$\beta_{2,2}$	_	_	0.0178	0.986	0.0258
$\beta_{1,2}$	_	_	_	0.00596	0.0134

Table 5.1: Model selection for the spline calibration function and global first order model.

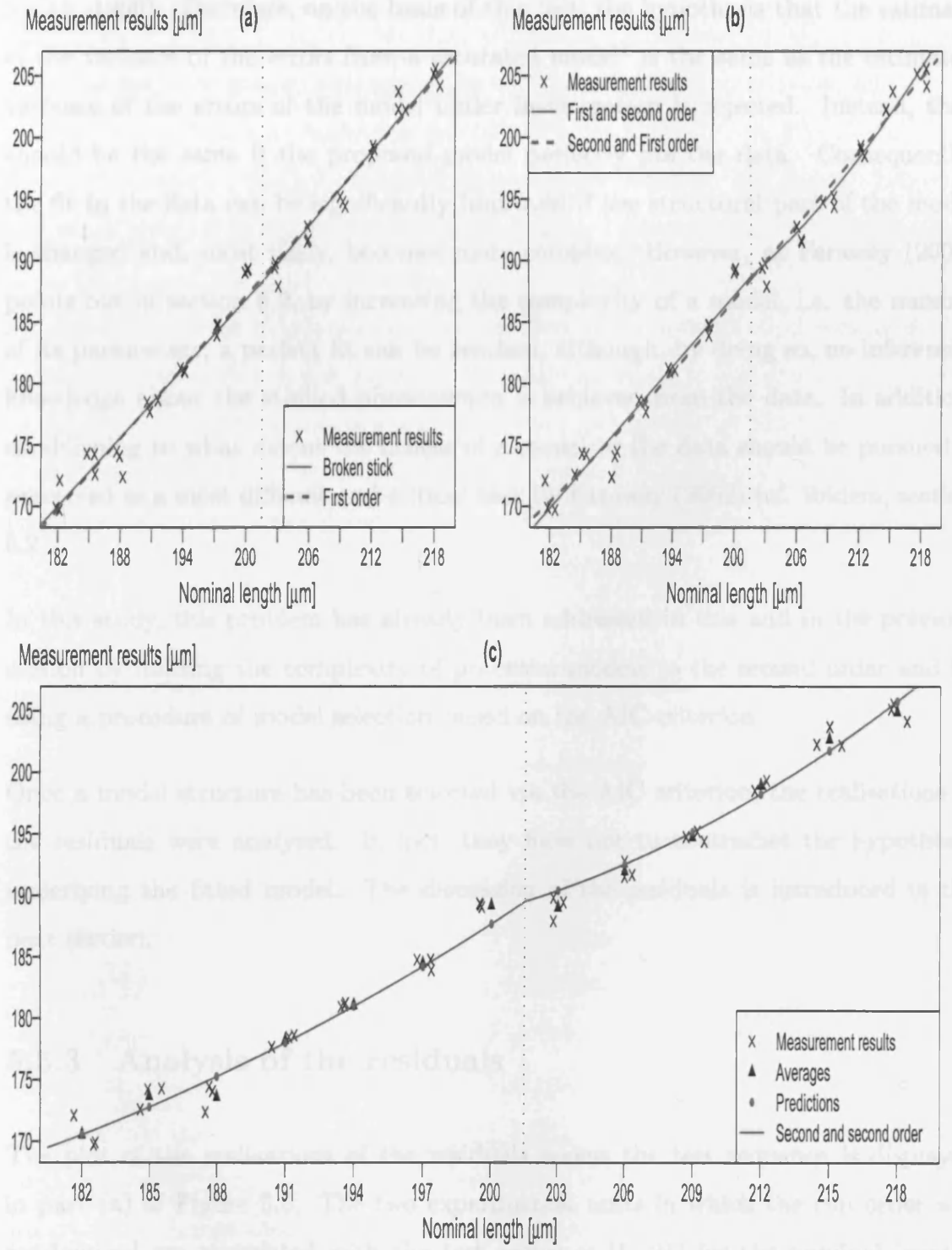


Figure 5.5: Competing candidate calibration functions. In part (c) the selected model is displayed.

6 in Faraway, 2002). In fact, an F test for lack of fit results in $F_0 = 6.715$ with a $P_{value} = 8.99 \cdot 10^{-5}$ (cf. chapter 6 in Faraway, 2002 and section 1.5 in Drapper and Smith, 1966). Therefore, on the basis of this test, the hypothesis that the estimate of the variance of the errors from a saturated model[1] is the same as the estimated variance of the errors of the model under investigation is rejected. Instead, they should be the same if the proposed model perfectly fits the data. Consequently, the fit to the data can be significantly improved if the structural part of the model is changed and, most likely, becomes more complex. However, as Faraway (2002) points out in section 6.2, by increasing the complexity of a model, i.e. the number of its parameters, a perfect fit can be reached, although, by doing so, no inferential knowledge about the studied phenomenon is achieved from the data. In addition, establishing to what extent the fitness of a model to the data should be pursued is perceived as a most difficult and critical task by Faraway (2002) (cf. ibidem, section 6.2).

In this study, this problem has already been addressed in this and in the previous section by limiting the complexity of potential models to the second order and by using a procedure of model selection based on the AIC criterion.

Once a model structure has been selected via the AIC criterion, the realisations of the residuals were analysed. In fact, they have not to contradict the hypotheses underlying the fitted model. The discussion of the residuals is introduced in the next section.

5.3.3 Analysis of the residuals

The plot of the realisations of the residuals versus the test sequence is displayed in part (a) of Figure 5.6. The two experimental units in which the run order was randomised are associated with the test sequence [1 , 21] for the nominal lengths

[1]This is also known as variance due to 'pure error'. A saturated model is a model which makes provision of a single parameter for each unique combination of the predictors, i.e. an ANOVA model (cf. section 6.2 in Faraway, 2002).

{182, 185, 188, 191, 194, 197, 200} μm and with the test sequence [22 , 39] for the nominal lengths {203, 206, 209, 212, 215, 218} μm, respectively. A dotted vertical line separates the two experimental units.

This plot appears not to show any evident pattern or serial correlation, apart from a short-range decreasing behaviour related with the cases 37, 38, 39. Therefore, it does not contradict the hypothesis of Independence between the errors. Although two cases (16 and 39) appear suspiciously extreme when compared with the other data, they are still at a close distance from the origin (less than $3\,\mu m$). Thus, it is difficult to establish if this is due to the presence of a uncontrolled contaminating nuisance factor or due to the natural randomness inherent in the measuring process. In this latter case, their exclusion from the analysis would be inappropriate. Part (b) of the same figure also shows that these cases do not exhibit any high leverage value nor large standardised residuals. Their exclusion from the analysis, therefore, is not expected to affect significantly the estimated values of the model parameters (cf. sections 8.5 about dfbeta and dffit and their relationship with the leverages and the standardised residuals in Sen and Srivastava, 1990. See also sections 7.2, 7.3, 7.4 in Faraway, 2002). Hence, their inclusion or exclusion from the analysis does not affect significantly the conclusions that are drawn.

Thus, although Figure 5.7 both in part (a) and (b) confirms that the cases 16 and 39 are extreme, they are not further analysed and are simply included in the analysis. In part (b) of Figure 5.6, all the measurement results pertaining to the nominal lengths 182 and 218 μm have leverage values close to the threshold suggested by several authors as a cut-off point (cf. sections 8.2 in Sen and Srivastava, 1990, sections 7.1 in Faraway, 2002 and section 10-7.2 in Montgomery, 2001). These authors do in fact conclude that, if a case has a leverage larger than this threshold, its dependent variables are far from the average point of the dependent variables over all the cases. Therefore, its remoteness causes such a case to exert an undue influence on the model predictions or explanations. A discussion of these results is presented in section 8.2 and 8.2.1 of Sen and Srivastava (1990). It is consequently inferred that conclusions less sensitive to the measurement tests of the RM's with nominal lengths 182 and

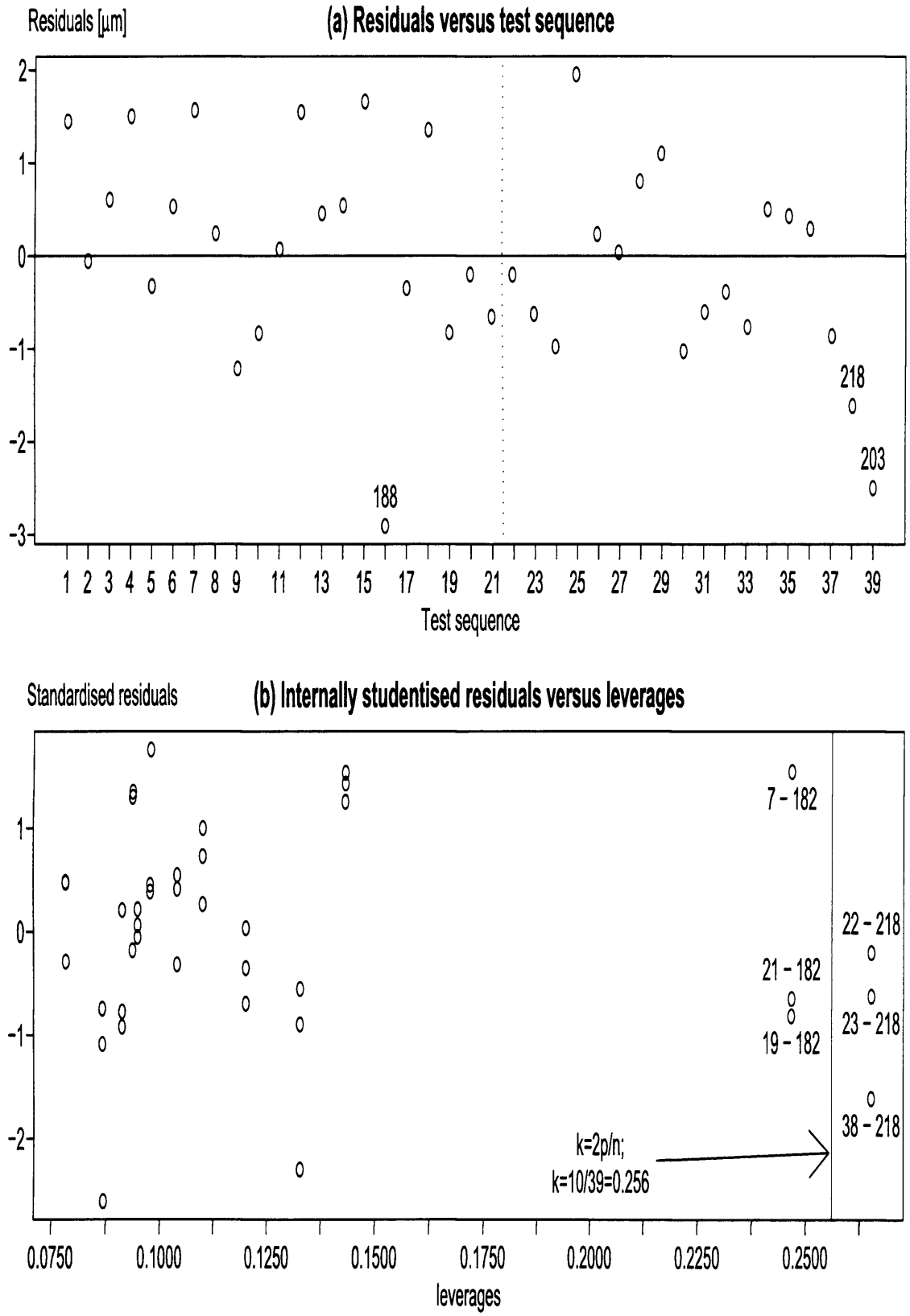


Figure 5.6: Residuals versus run order (a) and standardised residuals versus leverages (b).

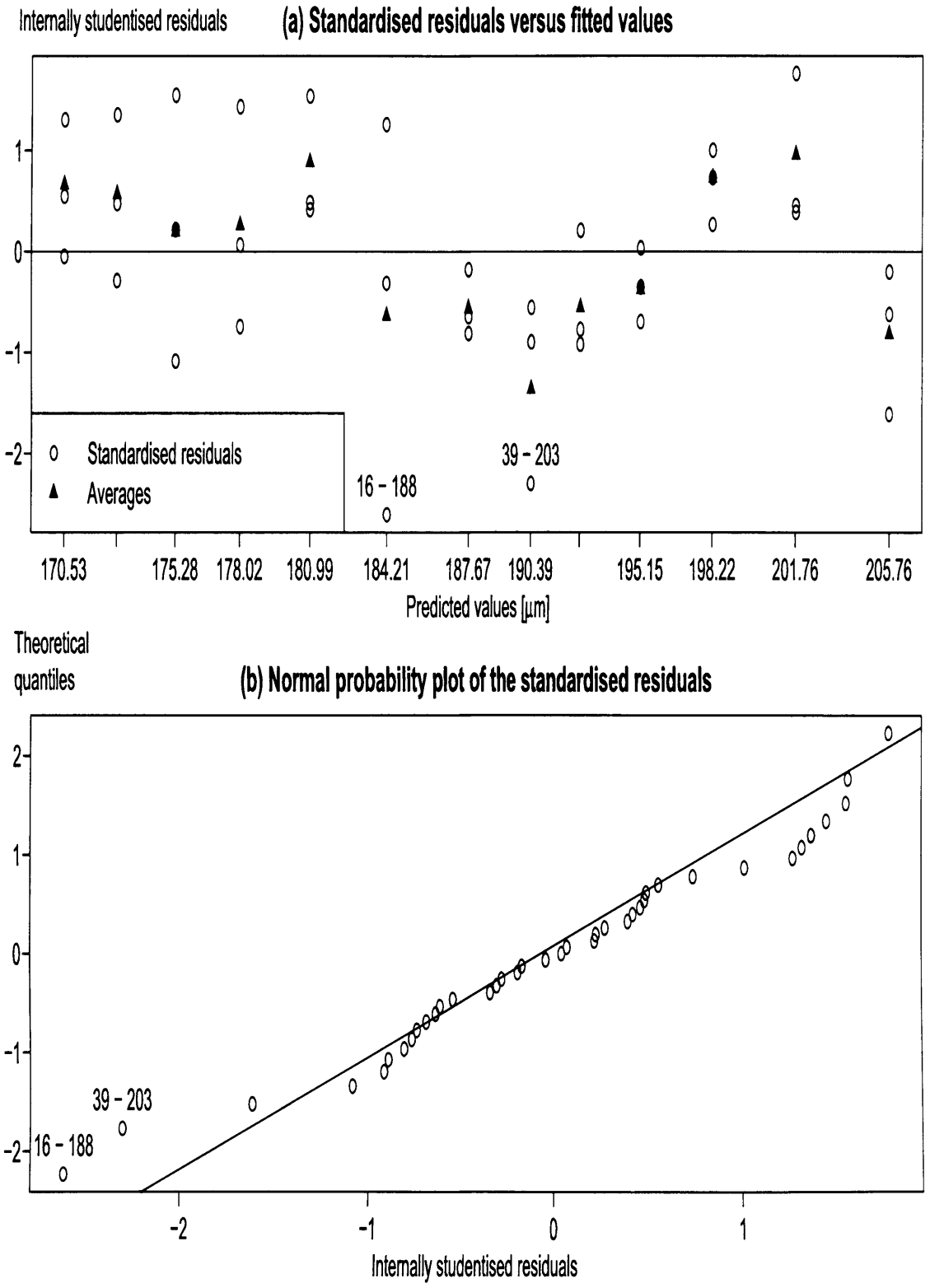


Figure 5.7: Internally studentised residuals versus fitted values (a) and normality probability plot of the residuals (b).

218 μm can be obtained by leaving them out of the defining procedure of the spline calibration function. Therefore, part (b) of Figure 5.7 suggests that the stability of the proposed calibration procedure would improve if each of its subintervals were to cover a narrower range of nominal lengths. The plot of the standardised residuals against the predicted values do not exhibit significant violation of the hypothesis of homoscedasticity of the errors, apart from the already mentioned effects due to the case 16 and 39. However, a pattern is appearing in this plot. It underlines the fact that some room is left for improving the fitting using a more complex structural functional relationship. The Normal probability plot of Figure 5.7, part (b), does not display significant departures from the assumed normality.

Once a calibration function is available, a control method is necessary in order to assess over the time whether the pre-existing calibration function can be safely deployed or a further calibration experiment is needed. A control method is presented in the next section.

5.4 Extended use of a calibration function

The two conditions needed for the calibration function to be valid are explicitly listed in section 4 of BS ISO 11095.

Firstly, the calibration experiment has to be designed to be representative of the normal operating conditions of the measuring system. However, the same standard fails to specify what the term 'representative' means from an operational point of view. The standard BS ISO 11095 does, in fact, prescribe the usage of more than one RM in its basic method for calibration. Nevertheless, nothing is said about the manner of choosing the number of RM's and the number of replicates of the calibration experiment in order for it to be representative of the normal operating conditions of the measuring system.

Secondly, the measuring system needs to be in a state of control. The state of the measuring system during the calibration experiment has to be non-significantly

different from the state of the same system in later periods of time. This means that the measuring system yields non-significantly different results when measuring the same quantities in later instants of time. In BS ISO 11095, this condition is also referred to as stability of the measuring process. Therefore, the calibration function fitted to the experimental results during a calibration experiment maintains its validity over a period of stability of the measuring process.

Consequently, the use of a calibration function, and more in general the use of a generic measuring system, cannot eschew assessing the period of time over which the measuring process is stable. Thus, the investigation of the stability and of the period of stability of a process requires the design of a control method.

On one hand, a control method provides a means of detecting whether changes (assignable causes) have occurred in the system that support an investigation and a recalibration. On the other hand, it makes provision of a method for estimating the uncertainty of the values that have been obtained by a given calibration function. However, this second aspect is beyond the scope of this investigation. Further details about it are presented in section 7.5 of BS ISO 11095.

The control method presented in section 7 of BS ISO 11095 appears difficult to extend to the proposed calibration method. In fact, it is based on the differences of the transformed measurement results from the accepted values of the RM's. And for the proposed calibration function, the probability density function of such differences assumes forms more difficult to evaluate than in the case of the linear calibration function presented in BS ISO 11095.

Therefore, in this study, it is proposed to assess the stability of the measuring system using the residuals of the calibration function. In fact, when the system is in control, the residuals account for all the sources of variability of the measuring process that are different from the accepted value of the RM's. Hence, if a non self-evident nuisance factor occurs and alters the stability of the measuring system, this has an effect on the mean or on the variability or on both the mean and the variability of the residuals. Examples of these not self-evident nuisance factors, also called

'assignable causes' in the terminology developed by Shewhart (cf. sections 4-2 in Montgomery, 1996), are a malfunction of one internal component of the measuring system, a change in the variability of the laboratory temperature and the occurrence of vibrations not sufficiently attenuated by the ad hoc stage. The assessment of the stability of the measuring system has been divided into two parts: the status of control while performing the calibration experiment and the status of control in subsequent periods of time. In fact, for these two parts, two distinct methods have been proposed and are presented in the next two subsections.

5.4.1 Control method during the calibration experiment

The residuals have been obtained on the fundamental hypothesis that the measuring system is in a state of control (cf. sections 4 in BS ISO 11095). Hence, their plot against test sequence, illustrated in part (a) of Figure 5.6, is expected not to exhibit any major violation of this assumption.

However, in order to provide quantitative evidence of the stability of the system, a further analysis is needed. In particular, the control method should provide a means of detecting small deviations from the status of control that part (a) of Figure 5.6 cannot highlight. The control method should also be suitable for sample sizes of the quality characteristic equal to one. In fact, it does not seem appropriate to aggregate residuals that are consecutive but are also most likely refering to different accepted lengths because of the randomisation procedure. Similarly, it does not seem appropriate to aggregate residuals refering to the same nominal length but not consecutive in the test sequence. Thus, in the light of these two considerations a cumulative-sum (CUSUM) control chart for the mean of the residuals and one for their variance has been designed. In particular, the internally studentised residuals are considered, so as they have the same variance, under the assumption of homoscedasticity of the errors. More details about CUSUM control charts are provided in section 7.1 of Montgomery (1996). The reference value $K_{CUSUM} = k_{CUSUM} \cdot \sigma_{CUSUM}$, ,and the decision interval $H_{CUSUM} = h_{CUSUM} \cdot \sigma_{CUSUM}$, , where σ_{CUSUM} , is the standard

deviation of the statistic whose realisations are plotted in the CUSUM chart. In this study, σ_{CUSUM} should be one, under the assumption of the calibration model (homoscedasticity of the errors) and if the estimates of the standard deviations of the errors used in standardising them were exactly equal to the unknown standard deviations. Therefore, it has been set $\sigma_{CUSUM} = 1$, independently from the sample standard deviation of the standardised residuals, which in every case is expected to be close to one. In addition, the state of out-of-control that the chart aims to detect has to be specified, namely:

$$\mu_{1,CUSUM} = \mu_{0,CUSUM} + \delta_{CUSUM} \cdot \sigma_{CUSUM} \tag{5.12}$$

$$H_{CUSUM} = K_{CUSUM} \cdot d_{CUSUM} = k_{CUSUM} \cdot \sigma_{CUSUM} \cdot d_{CUSUM} \tag{5.13}$$

In equation 5.12, $\mu_{0,CUSUM}$ and $\mu_{1,CUSUM}$ represent the mean of the plotted statistic when the process is in control and when it is out-of-control, respectively. Hence, $\mu_{1,CUSUM}$ is defined by assigning a value to δ_{CUSUM}, that is a specification input to the control chart design. Instead $\mu_{0,CUSUM}$ is set at zero, due to the definition of the standardised residuals and due to the first normal equation in the least squares procedure (cf. section 1.2 in Drapper and Smith, 1966). k_{CUSUM} is selected in accordance with the size of the shift, δ_{CUSUM}. The most used value is reported to be $k_{CUSUM} = \delta_{CUSUM}/2$ (cf. sections 7-1.3 in Montgomery, 1996). Once k_{CUSUM} has been determined, the value of h_{CUSUM} is selected so as the average run length when the process is out-of-control, $ARL(\delta = \bar{\delta})$, is minimised, when the wanted average run length for the in-control process, $ARL(\delta = 0)$, has already been given. This approach allows the determination of d_{CUSUM} and consequently h_{CUSUM} and H_{CUSUM}. In this study, it has been set $\delta_{CUSUM} = 1.5$ and $ARL(\delta = 0) = 500$. Thus, it follows that $k_{CUSUM} = 0.75$. Therefore, from Bowker and Lieberman (1972), it is derived $d_{CUSUM} = 4.7$ and $h_{CUSUM} = (\delta_{CUSUM} \cdot d_{CUSUM})/2 = 3.5$. From the discussion above, it also follows that $\hat{\mu}_{0,CUSUM} = 0$ and $\hat{\sigma}_{CUSUM} = 1$, thus $K_{CUSUM} = 0.75$ and $H_{CUSUM} = 3.5$. This design yields $ARL(\delta = 1.5) = 5.4$ (cf. sections 7-1.3 in Bowker and Lieberman, 1972). Therefore, on average, the chart takes 5.4 samples to signal an alarm, when a shift of the mean of 1.5 has occurred.

These figures appear compatible with the overall experimental effort (39 samples).

For controlling the variability the approach presented in section 7 – 1.9 of Montgomery (1996) was followed , where an ordinary CUSUM was designed for the following statistic:

$$s_{i,CUSUM} = \frac{\sqrt{|\hat{e}_{i,standardised}|} - 0.822}{0.349} \tag{5.14}$$

The random variable $s_{i,CUSUM}$ defined in equation 5.14 has an approximate Normal standard distribution when the process is in control. The design parameters $h_{CUSUM} = 3.5$ and $k_{CUSUM} = 0.75$ are the same both for the mean and the variance. The reference value K_{CUSUM} and the decision interval H_{CUSUM} are also the same, due to the fact that in both cases the standard deviation is equal to one and the in-control mean is equal to zero.

It has to be remarked that both the CUSUM's are designed under the assumption that the data are normally distributed and, in the second case, also standardised. However, while the internally studentised residuals are indeed standardised, on the other hand, it is believed that, in reality, they have a Student's t distribution with $n - p$ degrees of freedom, with n equals to the number of cases and p equals the number of β parameters in the calibration model. In fact, they are obtained as a ratio of a normal standard distribution, the residuals, and the square root of an independently chi-squared distributed random variable over its degrees of freedom, namely $\hat{\sigma}$ from equation 4.6 (for the distribution of the estimator of the variance of the residuals cf. section 3.4 in Faraway, 2002). Nevertheless, the Student's t tends asymptotically to a Normal when increasing its degrees of freedom, i.e. when n is sufficiently larger that p. Thus, in this context, as in the analysis of the residuals, this approximation is expected to hold. This observation accounts also for the fact that departures from normality most frequently occur on the tails in a normality probability plot of the standardised residuals. In fact, the difference between a Student's t (thicker tails) and a Normal (thinner tails) is most apparent on the tails of these probability density functions. Some further remarks about the relationship between the distributional proprieties of errors, residuals and standardised residuals

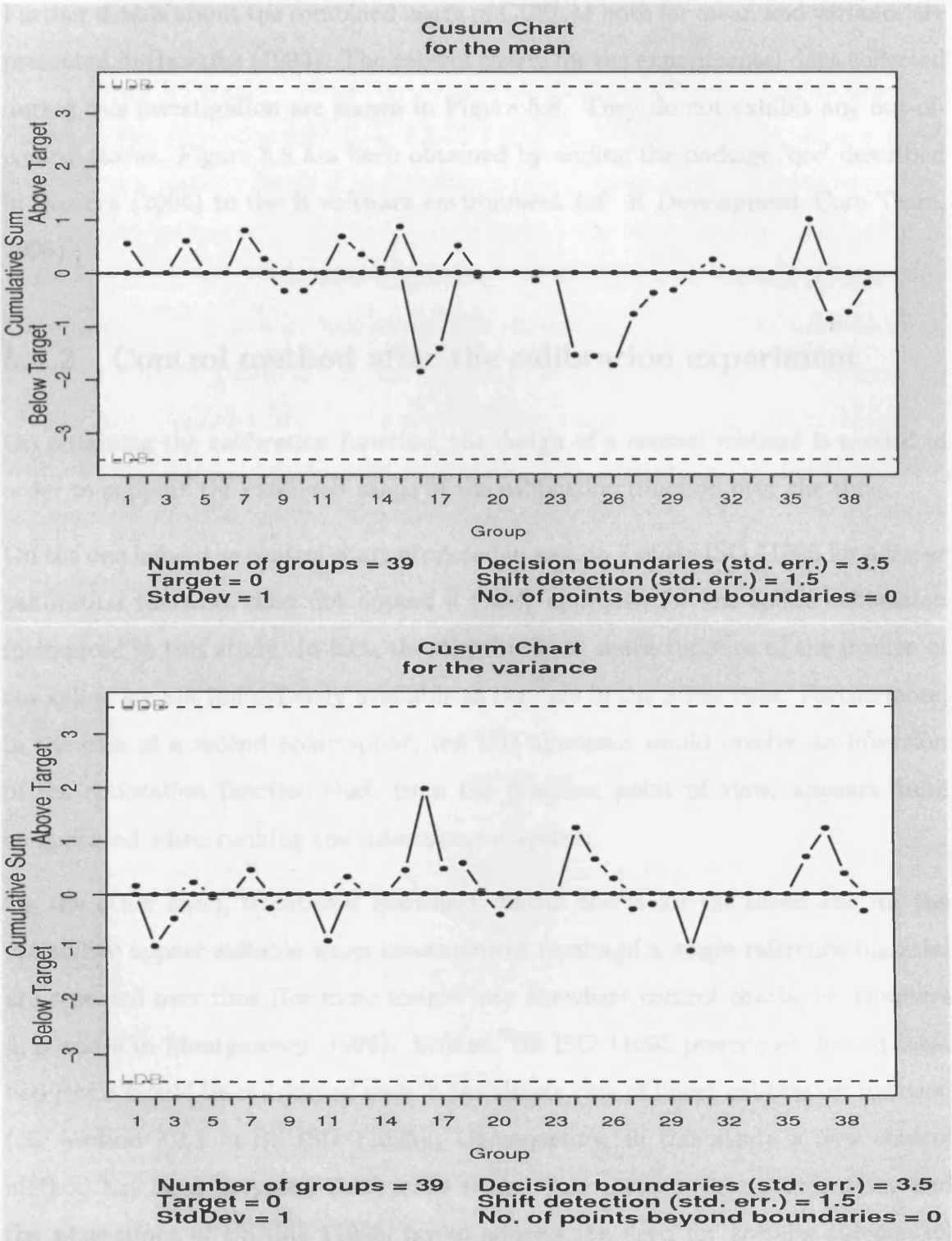


Figure 5.8: CUSUM control charts for the mean and for the variance of the standardised residuals.

can be found in sections 2.1.1 and 2.1.2 of Ryan (1997).

Further details about the combined usage of CUSUM both for mean and variance are presented in Hawkins (1993). The control charts for the experimental data collected during this investigation are shown in Figure 5.8. They do not exhibit any out-of-control status. Figure 5.8 has been obtained by adding the package 'qcc' described in Scrucca (2004) to the R software environment (cf. R Development Core Team, 2006).

5.4.2 Control method after the calibration experiment

On obtaining the calibration function, the design of a control method is needed in order to support the extended usage of the calibration function over the time.

On the one hand, the control chart proposed in section 7 of BS ISO 11095 for a linear calibration function, does not appear a viable approach for the spline calibration introduced in this study. In fact, the distributional characteristics of the inverse of the spline are not immediately available as they are in the linear case. Furthermore, in the case of a second order spline, the ISO approach would involve an inversion of the calibration function that, from the practical point of view, appears quite complicated when running the measurement system.

On the other hand, traditional Shewhart control charts for the mean and for the variability appear suitable when measurement results of a single reference material are repeated over time (for more insight into Shewhart control charts, cf. chapters 4, 5 and 6 in Montgomery, 1996). Instead, BS ISO 11095 prescribes that at least two RM's should be considered even in the simple case of linear calibration function (cf. section 7.3.1 in BS ISO 11095). Consequently, in this study a new control method has been proposed that, while taking into account the specifications and the suggestions of BS ISO 11095, try to address the need for keeping the control method affordable to run and as simple as possible.

Setting these targets resulted in selecting the minimum number of RM's as control points. Therefore, Two RM's for each subinterval were considered. Similarly, the

number of replicated measurements on each RM was set at 2. The frequency of control was set by analogy with the procedure adopted for similar measurement instruments in other laboratories. In fact the manual of the instrument provides only a generic indication, verbatim, 'calibration of the MicroXAM is necessary from time to time' (cf. appendix D in ADEP1 and in ADEP2). In the same manual it is also said that 'a check of calibration is advisable on at least a monthly basis' (cf. ibidem, on the same page). However, what the manufacturer means with 'a check of calibration' is not further specified. It was therefore decided to adopt the same frequency of control that is adopted at the Microfabrication Laboratory of the University of California at Berkley for a similar instrument (cf. section 9.5 in MICRO). Consequently, in the proposed procedure, the status of calibration is checked every time the instrument is powered. The resulting design therefore has a sampling interval that varies in a random manner.

If the measuring process is in control, then the residuals of the calibration function when measuring the RM's selected for control purposes should have the proprieties that they had during the calibration experiment. In particular, $\hat{e}_{i,j,t,control} \sim NID\,(0,\, \sigma_{control})$, where $\sigma_{control}$ is estimated by $\hat{\sigma}$, given by equation 4.6. In fact, the quantity $\hat{e}_{i,j,t\,control}$ is the difference between the j-th measurement (j =1, .., n_i; with $n_i = 2$) of the i-th RM (i =1, .., m; with $m = 4$) at the control instant t=1,2, ... and the value predicted by the spline calibration function for the corresponding RM, namely:

$$\hat{e}_{i,j,t,control} = h_{i,j,t} - \mathbf{x}_0^T \cdot \hat{\beta} \tag{5.15}$$

In equation 5.15 , $h_{i,j,t}$ is the j-th measurement of the i-th RM at the control instant t, and , using a terminology widespread in the literature, $\mathbf{x}_0^T$ is the vector of the regressor variables for which the prediction is made, whereas $\hat{\beta}$ is the vector of the parameters estimated via the least squares method during the calibration experiment (cf. sections 2.1, 2.2 and 2.3 in Sen and Srivastava, 1990). The prediction of the future value provided by the calibration function, i.e. $\mathbf{x}_0^T \cdot \hat{\beta}$, is a known deterministic quantity during the control period. Therefore, this prediction does not include the variability of the $\hat{\beta}$ as is the case when predicting future observations, considering

also the variability during the estimation experiment (cf. sections 3.5 in Faraway, 2002). In fact, the main purpose is to assess whether the status of the system has changed since the calibration experiment occurred. This purpose is different from providing a prediction of a future measurement result in x_0^T on the basis of the overall gathered data during the calibration experiment and during the test in x_0^T. This is also the reason why the $\hat{e}_{i,j,t,control}$'s are normally distributed rather than having a Student's t distribution.

Hence, the control limits for a probabilistic control chart for the mean of m RM's can be designed by setting the first type error at $\alpha_{test} = \alpha/m$, by making use of the Bonferroni inequality (cf. section 3-3.3 in Montgomery, 2001 and section 7.2.1 in BS ISO 11095). The symbol α is the family first type error, i.e. the probability that at least one of the m RM's signals an out-of-control condition at the instant t, when the process is in control. Thus it follows:

$$\bar{\hat{e}}_{i,\cdot,t\,control} = \frac{\sum_{j=1}^{n_i} \hat{e}_{i,j,t\,control}}{n_i} \tag{5.16}$$

$$i = 1, 2, \ldots, m; \;\; with \; m = 4; \;\; n_i = 2; \;\; t = generic \; control \; instant$$

$$UCL = +Z_{\frac{\alpha_{test}}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n_i}} \tag{5.17}$$

$$CL = 0 \tag{5.18}$$

$$LCL = -Z_{\frac{\alpha_{test}}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n_i}} \tag{5.19}$$

Where equation 5.16 is the statistic plotted on the control chart at each generic control instant t for each of the $m = 4$ RM's. UCL, CL and LCL are the upper control limit, the central line and the lower control limit, respectively. The symbol Z_ζ indicates the quantile of the Normal standard distribution that leaves ζ probability on its right.

In this study it is also proposed to control the variability of the measuring process by a Range control chart, whose limits have been probabilistically designed, and for which the simultaneous first type error has been accounted for by the Bonferroni

inequality as in the chart for the mean. In order to draw a probabilistic chart for the range, tables for the percentage points of the standardised range, $W = R/\sigma$, as reported in Harter (1960), have been used. As for the mean, the application of Bonferroni inequality results in $\alpha_{test} = \alpha/m$. Consequently, it holds:

$$P\left(W_{\frac{\alpha_{test}}{2}} < \frac{R}{\sigma} \leq W_{1-\frac{\alpha_{test}}{2}}\right) = 1 - \alpha_{test} \tag{5.20}$$

$$P\left(W_{\frac{\alpha_{test}}{2}} \cdot \sigma < R \leq W_{1-\frac{\alpha_{test}}{2}} \cdot \sigma\right) = 1 - \alpha_{test} \tag{5.21}$$

$$E\left(\frac{R}{\sigma}\right) = d_2(n_i) \quad with\ n_i = sample\ size\ of\ the\ i-th\ RM \tag{5.22}$$

where $d_2(n_i)$ is taken from table in Appendix VI of Montgomery (1996) and W_λ indicates the quantile of the distribution of the standardised range that leaves λ probability on its left. Hence, from equations 5.21, 5.22 and the estimation $\sigma_{control} = \hat{\sigma}$ the following control chart is obtained (cf. equation 4.6):

$$UCL = W_{1-\frac{\alpha_{test}}{2}} \cdot \hat{\sigma} \tag{5.23}$$

$$CL = d_2(n_i) \cdot \hat{\sigma} \tag{5.24}$$

$$LCL = W_{\frac{\alpha_{test}}{2}} \cdot \hat{\sigma} \tag{5.25}$$

The assumption underlying the equations 5.17, 5.24 and 5.25 is that the estimate $\hat{\sigma}$ obtained from the data of the calibration experiment is exactly equal to the unknown standard deviation, $\sigma_{control}$, of the the statistic under control, i.e. $\hat{e}_{i,j,t\,control}$.

For the collected experimental data it resulted $\hat{\sigma} = 1.167\,\mu m$ with $39 - 5 = 34$ degrees of freedom. By choosing $\alpha = 4\%$ for sake of convenience, it results $\alpha_{test} = 0.04/4 = 0.01$, $Z_{\alpha_{test}/2} = Z_{0.005} = 2.576$. Then, from table 1 in Harter (1960), it yields, for $n_i = 2$, $W_{\alpha_{test}/2} = W_{0.005} = 0.008862$ and $W_{1-\alpha_{test}/2} = W_{0.995} = 4.922533$. It may also be seen, from the table in Appendix VI of Montgomery (1996), that $d_2(2) = 1.128$. From this figures, the control chart for the mean is:

$$UCL_{mean} = 2.576 \cdot \frac{1.167}{\sqrt{2}} = 2.126 \tag{5.26}$$

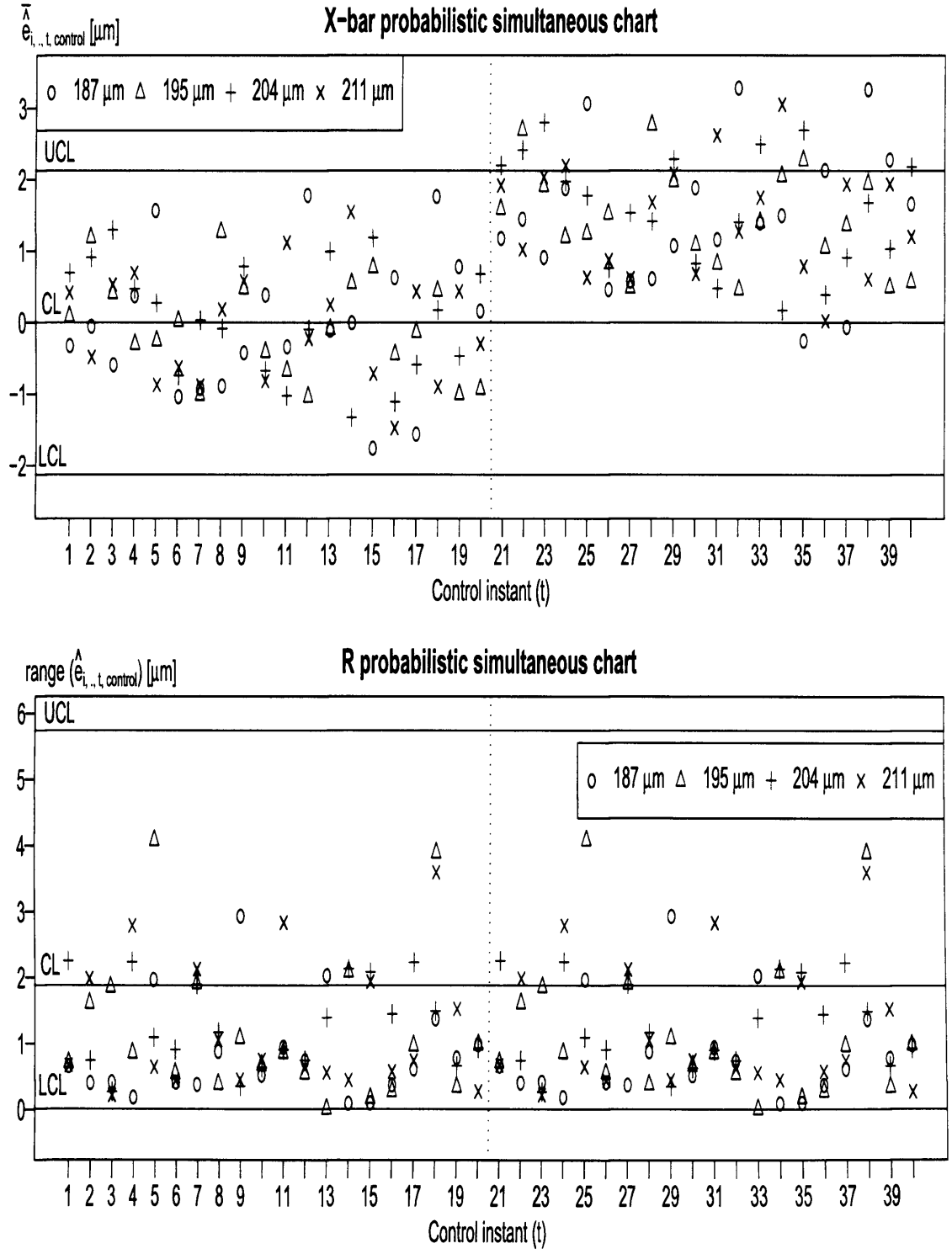


Figure 5.9: Simulated operation of the control charts for the mean and the standard deviation.

$$CL_{mean} = 0 \tag{5.27}$$

$$LCL_{mean} = -2.576 \cdot \frac{1.167}{\sqrt{2}} = -2.126 \tag{5.28}$$

Whereas, the control chart for the range is:

$$UCL_{range} = 4.922 \cdot 1.167 = 5.7440 \tag{5.29}$$

$$CL_{range} = 1.128 \cdot 1.1670 = 1.8838 \tag{5.30}$$

$$LCL_{range} = 0.008862 \cdot 1.167 = 0.01034 \tag{5.31}$$

The operation of the designed control chart is shown in Figure 5.9 using simulated data. The data have been pseudo-randomly generated in R (cf. R Development Core Team, 2006) from a $N\,(\mu = 0,\ \sigma = 1.167)$ the first 80 cases, i.e 80 pairs, and from an out-of-control distribution $N\,(\mu = 1.5,\ \sigma = 1.167)$ for the sole mean, the remaining cases from 81 to 160. In this simulation, no action has been taken on the detection of the out-of-control signal triggered by the chart for the mean, as it would have been in a real operation of the chart. It is for this reason that, several alarms occur in the range of cases from 81 to 160. In Figure 5.9 , the nominal lengths of the chosen RM's are also displayed. They are obtained by the following pairs of gauge blocks designated by their nominal length: $(1.190\,mm\,,\ 1.003\,mm)$, $(1.200\,mm\,,\ 1.005\,mm)$, $(1.210\,mm\,,\ 1.006\,mm)$and $(1.220\,mm\,,\ 1.009\,mm)$ for the $187\,\mu m$, $195\,\mu m$, $204\,\mu m$ and $211\,\mu m$ RM nominal values respectively. It can be noticed that they are also compatible with the recommendations of section 7.3.1 in BS ISO 11095. In fact, in this standard it is a recommendation that control RM's different from those used in the calibration experiment be selected. In addition, similarly to the calibration experiment, the selected RM's for the control method almost evenly span the investigated range of lengths. In Figure 5.9 the designed control limits (cf. equation from 5.26 to 5.31) are also shown.

5.5 Compendium

In this chapter, the calibration procedure built into the WLI microscope was investigated. This study resulted in an interpretation of the undocumented procedure in the light of standard BS ISO 11095. In particular, the main findings are:

- An identification of the built-in procedure with the one-point calibration procedure of BS ISO 11095 has been provided. This suggested interpretation was validated by the predictive method of the height correction factor that has been developed on the basis of this interpretation.
- Consequently, major resulting limitations of the built-in procedure were identified in the assumption of linearity of the calibration function and in the imposition of a intercept term equal to zero.

In addition, a comparison between simulated and actual measurement results has shown the inadequacy of the built-in calibration function. In particular, the linearity and the hypothesis of null intercept of the calibration function have been proved to be questionable. In fact, quantitative evidence of a bias induced by the built in calibration procedure is provided.

In order to overcome the identified problems generated by the built-in calibration procedure, therefore, a novel spline-based approach to the calibration of the investigated instrument was developed. The proposed calibration procedure has also been demonstrated by implementing it for reference materials with nominal lengths ranging from 182 to 218 μm. Furthermore, an analysis of the residuals has provided experimental evidence suggesting that each subinterval of the spline function should be less than 18 μm wide, in order to limit the undue influence on the model of particularly remote reference materials (high leverage values).

Finally, a novel control method for the practical use of the spline calibration function is proposed. It is different both from the method presented in BS ISO 11095, and from the Shewhart $\overline{X} - R$ control chart. In particular, the main results were as follows:

- A cumulative sum (CUSUM) control chart of the standardised residuals was designed and implemented for assessing the stability of the measuring system during the calibration experiment.

- A control procedure for assessing the stability of the measuring system after the completion of the calibration experiment was developed. It is based on evaluating the state of control from the residuals of the calibration function when applied to a new set of reference materials. Two control charts, one for the mean and one for the range, were probabilistically designed. In both cases, the control limits were designed by means of the Bonferroni inequality. Thus, they account for a pre-specified probability of a false alarm of the overall set of the selected control reference materials.

Chapter 6

Developing a method based on white light interferometry for measuring μ-volumes

A procedure for measuring micro-volumes of irregular and complex forms, both convex and concave, is described in this chapter. The proposed procedure is founded on identifying a reference element on the measurand itself. In this way, the accuracy (trueness and precision) of the measurement procedure relies on the possibility of creating masks in the software controlling the WLI microscope. In particular, the identification of the positional reference element on the part itself greatly facilitates the positioning of the measurand during repeated measurements. The contents of this chapter are as follows: In section one, the overall measuring procedure and inherent implemented programs are described. In section two, the algorithm for calculating the volume is presented. In sections three and four, the developed procedure is demonstrated on measurands of convex and concave form, respectively. In section five, an assessment of the precision in repeatability conditions when measuring a pre-specified convex part is estimated. In section six, the factors that affect the trueness of the measuring procedure are discussed, and in section seven, potential further developments are suggested.

6.1 Proposed micro volume measuring procedure

The profile $h(x, y)$ of the measurand is acquired with the height correction factor set at one. The instrument parameter 'noise threshold' is set in its lower range (less than 0.06) in order to take into account the description of such a parameter provided by ADEP2 in section 4.5. In fact, the lower the noise threshold, the smaller the number of the pixels in the acquired image that do not contain a data value and that are therefore tagged as 'bad data'. For sake of clarity, in this and in the following sections, it has been decided to call 'good data' all the points that are not tagged as bad data. The inclusion in the image of points extraneous to the part is the drawback of using a low noise threshold. This downside is, however, avoided by masking all the points not belonging to the part under investigation.

The proposed procedure is based on the absence of bad data in the acquired profile. The occurrences of bad data are processed using the functions offered by the software of the measuring instrument. Particularly, the 'fill bad data' function replaces the bad data with the average of the data contained in the adjacent pixels, whereas the 'interpolate bad data' function performs 'a localized low order polynomial fit to restore large regions of invalid data' (ADEP2, section 5.9.1).

The masking process is relevant to the developed procedure. In fact, it allows two different profiles from the measurand to be obtained: the feature profile and the reference profile. For instance, in measuring the volume of a pocket manufactured on a spherical surface two masks are used. The profile of the pocket (the feature profile) is acquired by the first mask, whereas the spherical polygonal surface limited by the pocket and the field of view (the reference profile) is obtained by means of the second mask. The case is illustrated in figure 6.1.

The masking procedure results in two different files containing the cloud of points representing the profiles of the feature and of the reference element, respectively. Each file, in ASCII format and with extension .sdf, contains the height information codified according to the protocol described in appendix E of ADEP2. When a point in the acquired data is covered by a mask, it is tagged as bad data. In order to trans-

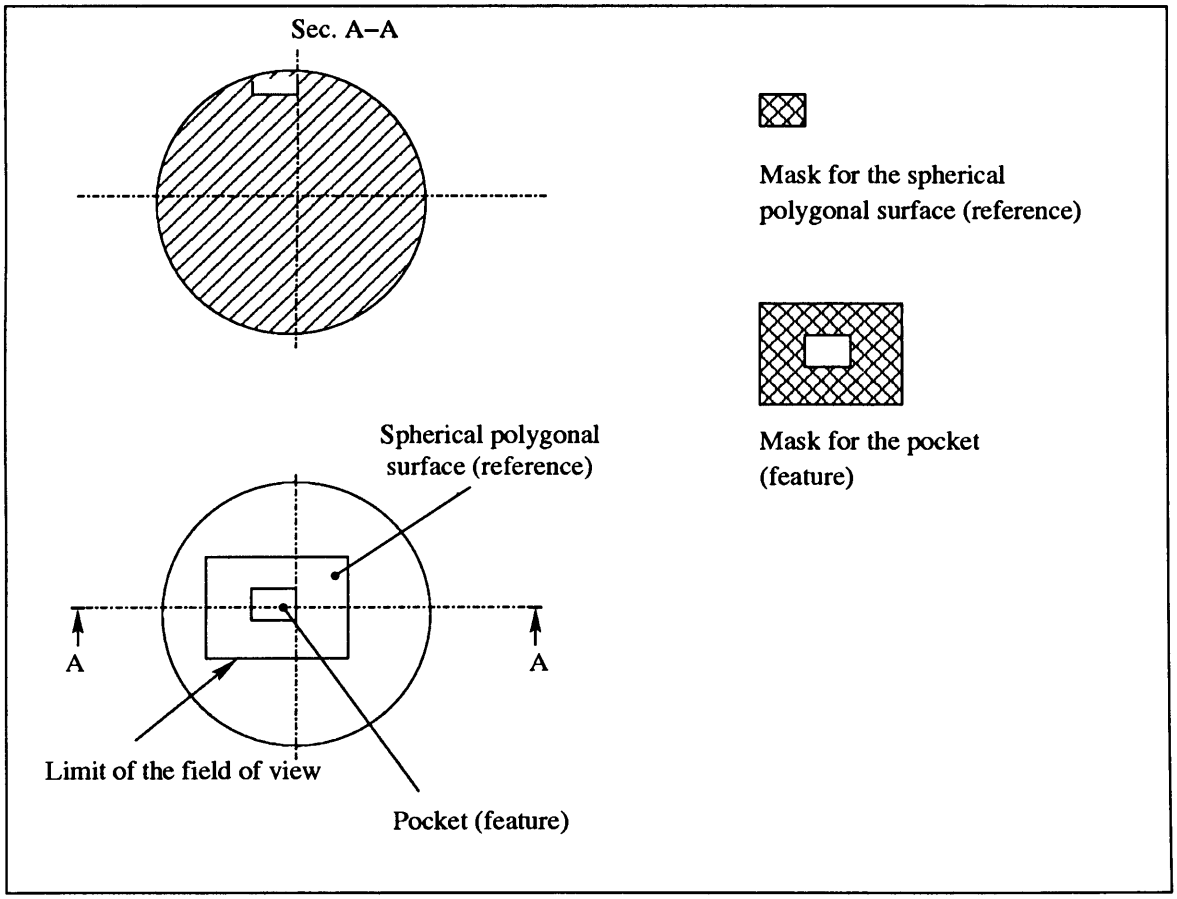


Figure 6.1: Application of the proposed volume measuring method to a pocket on a sphere.

late the codified information about the points into Cartesian coordinates $P(x, y, z)$, a program, named Coordinates, has been implemented in C . Coordinates provides the following output:

- the total number of points in the input profile;
- the number of points that are good data;
- the Cartesian coordinates of all the points;
- the Cartesian coordinates of the good data;
- the Cartesian coordinates of the good data transformed using the inverse of the calibration function.

The last three pieces of information are written in three separate ASCII files. In figure 6.2 a flowchart, drawn in compliance with BS 4058-ISO 5807, illustrates the relationship between the program Coordinates and the other programs described in this section.

On having the calibrated Cartesian coordinates of the reference profile, a least square algorithm fits a reference geometry to the cloud of points. In the above example of the pocket on a sphere, the least square algorithm provides the parameters of the sphere equation that minimises the sum of the squared difference between the acquired points and the corresponding points on the sphere. Given the specific measurement tasks under investigation, a least square algorithm for a planar reference surface, was implemented in C on a GNU/Linux system. The program was called LeastSquaresPlane and is based on the GNU Scientific Library (cf. Galassi et al., 2005).

The calibrated good data of the feature are then projected onto points of the reference geometry. The height difference between the calibrated good data of the feature and the corresponding projected points is then computed. The difference is expected to be always of the same sign. However, when this does not happen, such

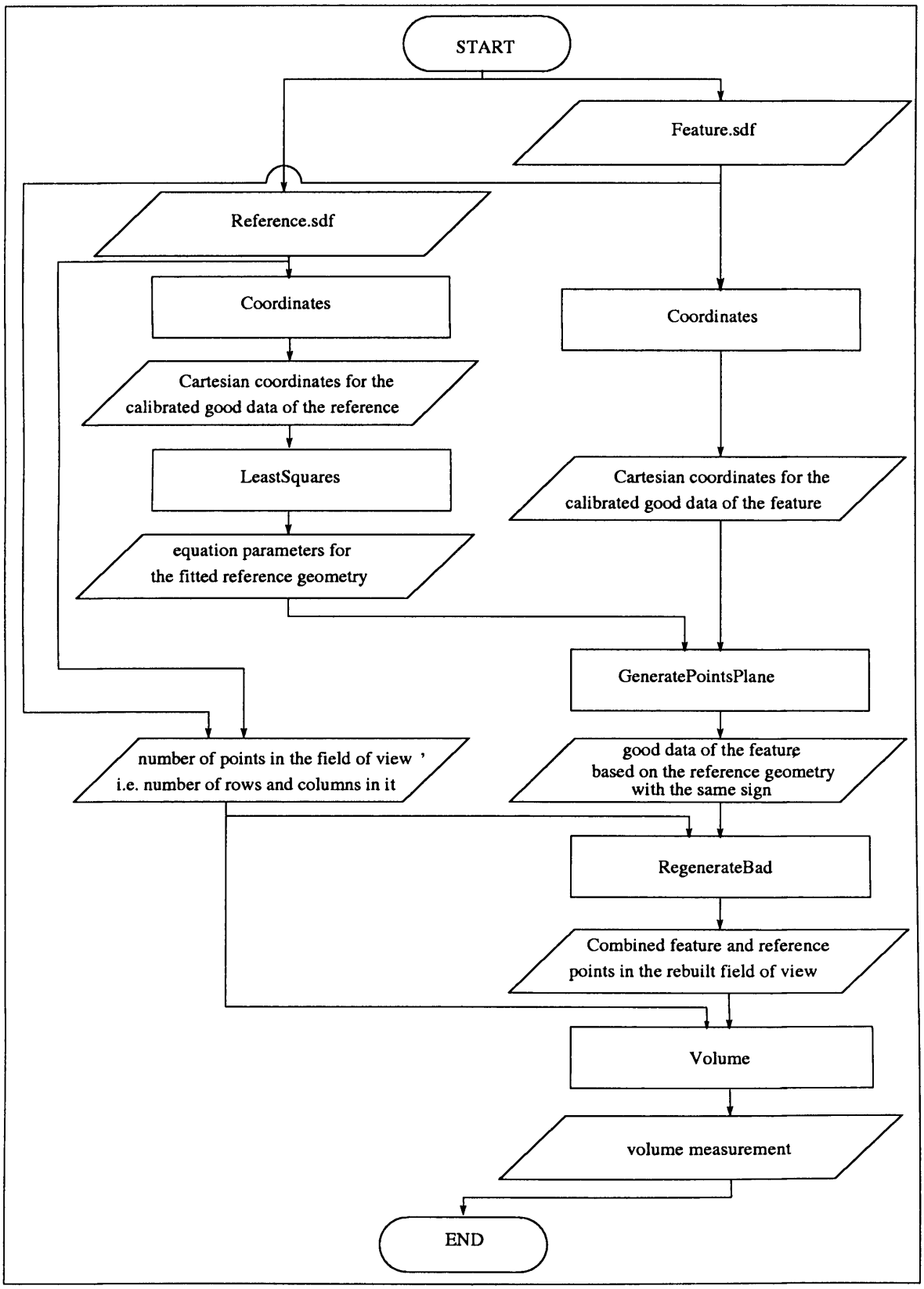


Figure 6.2: Proposed micro-volume measuring approach.

a difference has been set to zero, which is equivalent to putting the feature point on the reference geometry. It has, in fact, been assumed that a nuisance factor [1] has locally affected the measurement at that point, introducing a significant bias. As a result of this process, the zero for the heights of the feature points is on the reference geometry. In order to perform the projection and to calculate the above differences and their analyses, a program called GeneratePointsPlane has been implemented in C on a GNU/Linux system. GeneratePointsPlane needs an input of the Cartesian coordinates of the calibrated good points of the feature and the coefficients of the equation of the least squares plane resulting as an output of LeastSquaresPlane. GeneratePointsPlane produces the following output:

- the Cartesian coordinates of all the points of the feature projected on the reference surface;
- the differences between the heights of all points of the feature and their corresponding projection on the reference fitted geometry;
- the modification of the previous differences so as they have the same sign or are equal zero.

These three different pieces of information are stored in three separate ASCII files. Although the implemented program suits a planar reference surface, it can be modified for other surfaces, such as the polygonal spherical surface of the pocket in a sphere example.

The boundary and the number of points of the feature profile cannot be completely known a priori, before performing the measuring task. It can be different not only when measuring different pieces of the same feature, but also when measuring the same feature more than once. In fact, not only is every manufacturing process affected by a random inherent variability which makes the production of two identical features very unlikely, but also every measuring process is characterised by its own

[1] Scattering due to a local deep valley in the surface, for example.

variability. In order to compute the volume, therefore, it was considered convenient to merge the points of the feature profile referred to the reference geometry and the points of the reference profile itself. In this way, the program for calculating the volume processes a number of points spread on a rectangular area, the field of view and this number of points is deterministically known from the magnification settings of the WLI microscope. In order for the points of the reference profile not to give any biasing contribution to the volume measurement, however, all the heights of the reference profile must be set at zero. This result is self evident from the procedure for computing the volume that is illustrated in the next section. In the example of the pocket in a sphere, the good data of the pocket with all the heights of the same sign and referring to the spherical polygon, are combined with the points of the reference profile whose heights have been set to zero. The resulting set of points is spread across the whole rectangular field of view on the same number of rows and columns that would be used if the same feature had been acquired with the same instrument settings but without any mask. In order to combine the two profiles and to perform the adjustments mentioned above, a program called RegenerateBad was implemented in C on a GNU/Linux system. RegenerateBad needs input information from the .sdf files about the total number of points in the field of view. It also needs the feature profile with all the heights of the same sign as processed by GeneratePointsPlane. RegenerateBad produces an output of the Cartesian coordinates of the points of the whole field of view as described above and stores them in an ASCII file.

The measurement of the volume takes place using the output of RegenerateBad. The structure of the algorithm for calculating the volume is described in the next section.

6.2 The algorithm for measuring the volume

The algorithm developed for measuring the volume hinges on the regularity of the structure of the cloud of points that represents the output of the processed WLI

microscope measurement result (the output of RegenerateBad). An interpretation of the Appendix E of ADEP2 leads to the following conclusions: The projection of this cloud of points on the Cartesian plane $Z = 0$ constitutes a regular grid where these projections are evenly spaced by a distance 'Xscale' and 'Yscale' along the X and Y directions, respectively. 'NumProfiles' is the number of different Y-coordinates of this grid, whereas 'NumPoints' is the number of different X-coordinates for a given Y-coordinate. The plane $Z = 0$ is set at the point with the minimum z coordinate value in the reference system having the origin in the mean value of the heights. The internal parameters 'Xscale' and 'Yscale' are internally determined by the software of the WLI microscope on the basis of the magnification settings. However, the internal parameters 'NumProfiles' and 'NumPoints' are magnification-independent constants equal to 480 and 748, respectively (cf. Figure 6.3). It is suspected that these two last parameters depend only on the number of photo-sites of the CCD camera of the instrument that are associated with the pixels of the resulting image.

For the equipment used in this investigation, the coded information of the output text file (cf. Appendix E in ADEP2 about the UDF file header) was interpreted and resulted in the data summarised in Table 6.1. It is observed that the values of 'Xscale' and 'Yscale' shown in this table are equal to those reported in the corresponding coded text files in the output of the software controlling the WLI microscope. These result differed from those obtained by dividing the dimensions of the field of view by the corresponding 'NumProfiles' and 'NumPoints' . Although these discrepancies are not always negligible, it is believed that they are solely due to rounding procedures rather than to conceptual pitfalls in the proposed interpretation.

The volume to be determined can be identified by fitting triangles at every three points of the cloud, with similarities to the Stereolithography data format (STL), used in the manufacturing industry in particular for rapid prototyping applications (cf. section 1.2.2 in Pham and Dimov, 2001 and the description of the STL format at the tele-manufacturing facility project of the University of California, San Diego, namely TELEM).

This triangles fitting operation is effortless due to the regularity of the grid men-

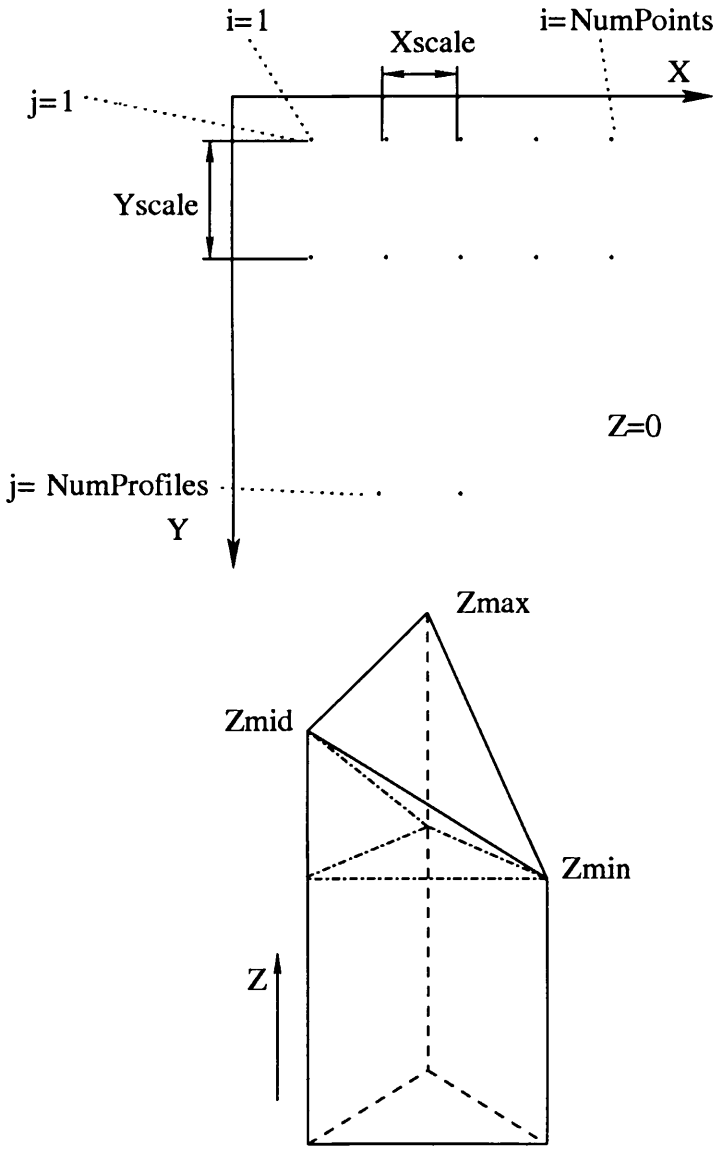


Figure 6.3: Grid and triangles.

Magnification	Field of view		Quantum of length	
	Horizontal [μm]	Vertical [μm]	'Xscale' [μm]	'Yscale' [μm]
$0.68 \cdot 20 = 13.6x$	656.48	498.09	0.880	1.0420
$2.00 \cdot 20 = 20x$	410.30	311.31	0.550	0.65127
$1.25 \cdot 20 = 25x$	328.24	249.05	0.440	0.52102
$2.00 \cdot 20 = 40x$	205.15	155.65	0.275	0.32564

Table 6.1: Field of view and image discretisation while varying the magnification set-up, but with the same $20x$ Mireau interferometric objective.

tioned above. Once the triangles have been fitted, then the volume to be determined is calculated by summing the contribution of the solid elements associated with each of the triangles. This sum operation is again facilitated by the regularity of the grid.

The volume of each of these solid elements is calculated as follows: first, the minimum of the three z-coordinates, z_{min},, is considered and the volume of the corresponding prism, V_{prism},, is calculated by multiplying this minimum by the area of the projection of the triangle on $Z = 0$. Such an area, S_{trg},, is computed from the coordinates of the vertices $P_1 = (x_1, y_1, 0)$, $P_2 = (x_2, y_2, 0)$ and $P_3 = (x_3, y_3, 0)$ from the following expression:

$$S_{trg} = \frac{1}{2!} \cdot abs \left(\begin{vmatrix} x_1 & y_1 & 0 \\ x_2 & y_2 & 0 \\ x_3 & y_3 & 0 \end{vmatrix} \right) \tag{6.1}$$

where $abs()$, is the absolute value operator. Consequently, it results:

$$V_{prism} = S_{trg} \cdot abs\,(z_{min}) \tag{6.2}$$

More details about equation 6.1 are in Weisstein (2006a). The three points are then ordered according to the values of their z-coordinates, namely, $P_1 = (x_1, y_1, z_{min})$, $P_2 = (x_2, y_2, z_{mid})$ and $P_3 = (x_3, y_3, z_{max})$, with $z_{min} \leq z_{mid} \leq z_{max}$,. The volume of the two tetrahedra with vertices $\{P_1,\ P_{2,min}\,(x_2, y_2, z_{min})\,,\ P_{3,min}\,(x_3, y_3, z_{min})\,,\ P_2\}$and $\{P_1,\ P_2,\ P_{3,min}\,(x_3, y_3, z_{min})\,,\ P_3\}$ is then computed using the following equations:

$$V_{tetrahedron,1} = \frac{1}{3!} \cdot abs \left(\begin{vmatrix} x_1 & y_1 & z_{min} & 1 \\ x_2 & y_2 & z_{min} & 1 \\ x_3 & y_3 & z_{min} & 1 \\ x_2 & y_2 & z_{mid} & 1 \end{vmatrix} \right) \tag{6.3}$$

$$V_{tetrahedron,2} = \frac{1}{3!} \cdot abs \left(\begin{vmatrix} x_1 & y_1 & z_{min} & 1 \\ x_2 & y_2 & z_{med} & 1 \\ x_3 & y_3 & z_{min} & 1 \\ x_3 & y_3 & z_{max} & 1 \end{vmatrix} \right) \tag{6.4}$$

More details about equation 6.3 and 6.4 can be found in Jackson and Weisstein (2006). Hence, the volume associated with the generic triangular element with vertices P_1, P_2 and P_3 is given by the following expression (cf. Figure 6.3) :

$$V = V_{prism} + V_{tetrahedron,1} + V_{tetrahedron,2} \tag{6.5}$$

This algorithm has been implemented on a Debian GNU/Linux system in ANSI C, making use of the ANSI C compliant library GSL (cf. Galassi et al., 2005) and following the recommendations and guidelines found in Lawlor (1995).

The implementation has been validated by calculating the volume of the eighth part of a sphere with a radius equal to one unit of length, for instance 1 μm. The theoretical volume was compared with the calculated volumes V while decreasing the grid pitch of the projection of the cloud of points on the plane $Z = 0$. The theoretical volume was computed with the constraint imposed by the 18 decimal digits of the 'long double' C type of variable used for the rational approximation of π . Grid pitches equal to 0.0010, 0.0005, 0.00025 and 0.0002 units of length were considered. For each of these cases, the percentage differences between V and V_{prism} and between V and $V_{prism} + V_{tetrahedron,1}$ were also computed. More details are shown in the display of the output of the implemented validation program shown in Appendix D. Figure 6.4 summarises the main results of the validation process.

From part (a) of this figure it is noticed that by reducing the grid pitch, that is by increasing the number of points in the method, which corresponds to using lenses with higher magnification, the difference of the programmatically computed volume from the calculated 'ideal' volume seems to display an asymptote. Moreover, this convergence to zero is quite fast due to the fact that, in the worst case, with

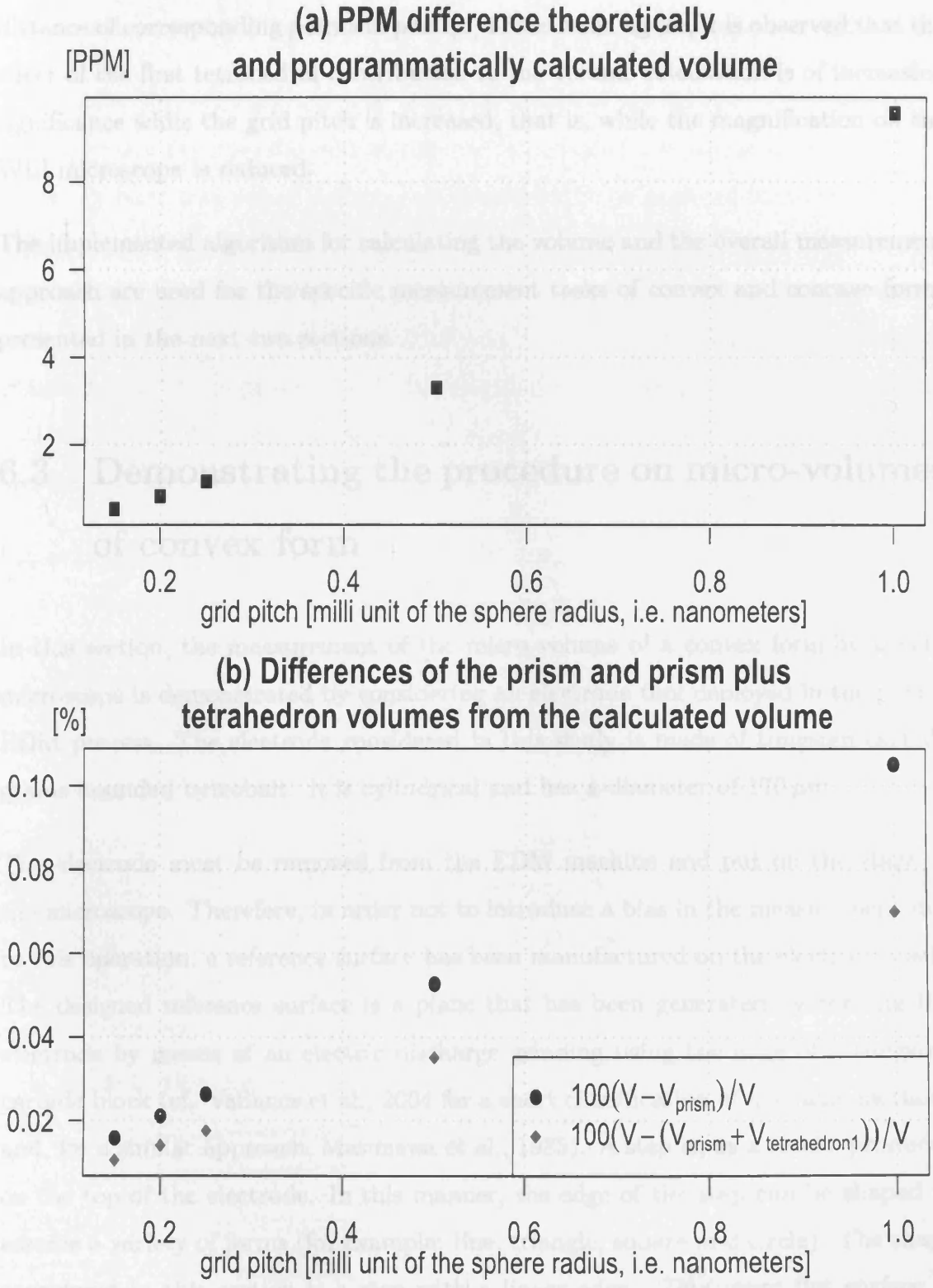


Figure 6.4: Summary of the validation results.

just one million points examined (grid pitch 0.001), the difference from the 'ideal' calculated volume is about 9.5 parts per million (PPM). From comparing the vertical distance of corresponding points in part (b) of the same figure, it is observed that the effect of the first tetrahedral contribution to the volume calculation is of increasing significance while the grid pitch is increased, that is, while the magnification on the WLI microscope is reduced.

The implemented algorithm for calculating the volume and the overall measurement approach are used for the specific measurement tasks of convex and concave forms presented in the next two sections.

6.3 Demonstrating the procedure on micro-volumes of convex form

In this section, the measurement of the micro-volume of a convex form by a WLI microscope is demonstrated by considering an electrode tool deployed in the micro-EDM process. The electrode considered in this study is made of tungsten carbide grains bounded by cobalt. It is cylindrical and has a diameter of 170 μm .

The electrode must be removed from the EDM machine and put on the stage of the microscope. Therefore, in order not to introduce a bias in the measurement due to this operation, a reference surface has been manufactured on the electrode itself. The designed reference surface is a plane that has been generated by shaping the electrode by means of an electric discharge grinding using the edge of a tungsten carbide block (cf. Vallance et al., 2004 for a short classification of available methods and, for a similar approach, Masuzawa et al., 1985). A step is, as a result, produced on the top of the electrode. In this manner, the edge of the step can be shaped to assume a variety of forms (for example: line, triangle, square and circle). The shape considered in this section is a step with a linear edge. The upper flat surface of the step corresponds to the feature in the general description of the procedure (cf. section 6.1), by contrast the lower flat surface of the step is selected as the reference.

The least squares method is used to fit a plane to the profile of the reference flat. Thereafter, the volume of the step included between the feature and its projection on the reference plan is measured in accordance with the proposed method.

This choice is connected with the potential practical application of the developed procedure in measuring the volumetric wear ratio in $\mu-$EDM, as mentioned in Chapter 2. In fact, this upper surface can subsequently be engaged in an on-purpose designed $\mu-$EDM operation, while the lower flat surface of the step is not. As a result, measuring the identified micro-volume before and after the $\mu-$EDM operation makes it possible to quantify in a direct way the volumetric wear of the electrode. This is possible independently of the irregularity and complexity of the shape of the feature profile after the machining operation. Figure 6.5 illustrates the case presented in this section.

In particular, part (a) of this figure shows the image of the electrode as it is displayed by the control software of the WLI microscope. The pixels that are not associated with the electrode have been masked out using a circular mask. Furthermore, the bad data have already been fitted or interpolated and the $Z = 0$ has been set at the lowest detected height. In part (b) of Figure 6.5, a selection of points from the feature and the reference profiles is made and displayed by Gnuplot (cf. GNUPL). The projection of the selected points of the feature profile on the reference geometry is also displayed. This selection, by reducing the number of displayed points contributes to making the figure clearer. This was the sole reason why a selection was made. The selection of points of the feature and the reference was made from the Cartesian coordinates of the good points that result in output from the program Coordinates, when the pertinent files were provided in input. The Cartesian coordinates of the projection of selected points of the feature on the reference geometry are as they result in output from the program GeneratePointsPlane. The example illustrated in Figure 6.5 is drawn from the first replicate of the precision experiment that is discussed in section 6.5, while the next section demonstrates the application of the volume measuring procedure to the case of concave forms.

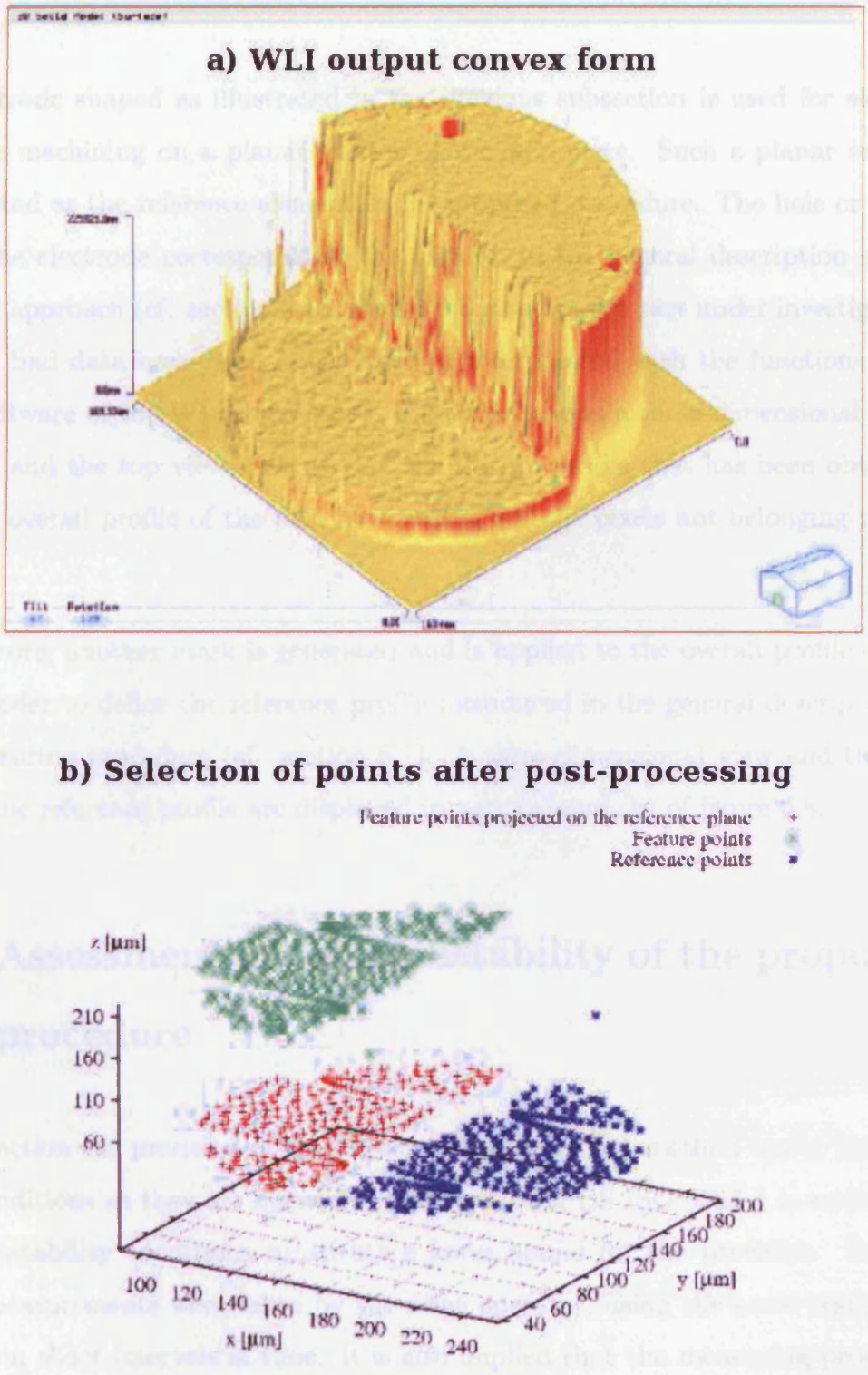


Figure 6.5: Volume of a convex form.

6.4 Demonstrating the procedure on micro volumes of concave form

The electrode shaped as illustrated in the previous subsection is used for electric discharge machining on a planar surface of the workpiece. Such a planar surface was selected as the reference element in the proposed procedure. The hole or mark left by the electrode corresponds to the feature, in the general description of the proposed approach (cf. section 6.1). Figure 6.6 displays the part under investigation after the bad data have been either filled or interpolated with the functions built in the software of the WLI microscope. Figure 6.7 shows a three-dimensional view, part (a), and the top view, part (b), of the feature profile that has been obtained from the overall profile of the part by masking out the pixels not belonging to the hole.

Furthermore, another mask is generated and is applied to the overall profile of the part in order to define the reference profile introduced in the general description of this measuring procedure (cf. section 6.1). A three-dimensional view and the top view of the reference profile are displayed in part (a) and (b) of figure 6.8.

6.5 Assessment of the repeatability of the proposed procedure

In this section the precision of the developed measurement method under repeatability conditions as they are defined in section 3.14 of BS ISO 5725-1 is estimated. The repeatability conditions constitute a lower bound for the precision. In fact, all the measurements were taken by the same operator, using the same equipment and within short intervals of time. It is also implied that the measuring process is under control (cf. section 1.1 in BS ISO 5725-1). However, in section 4.4.1 of BS ISO 5725-1, it is also stated that, during a repeatability experiment, 'the equipment should not be re-calibrated between the measurements unless this is an essential

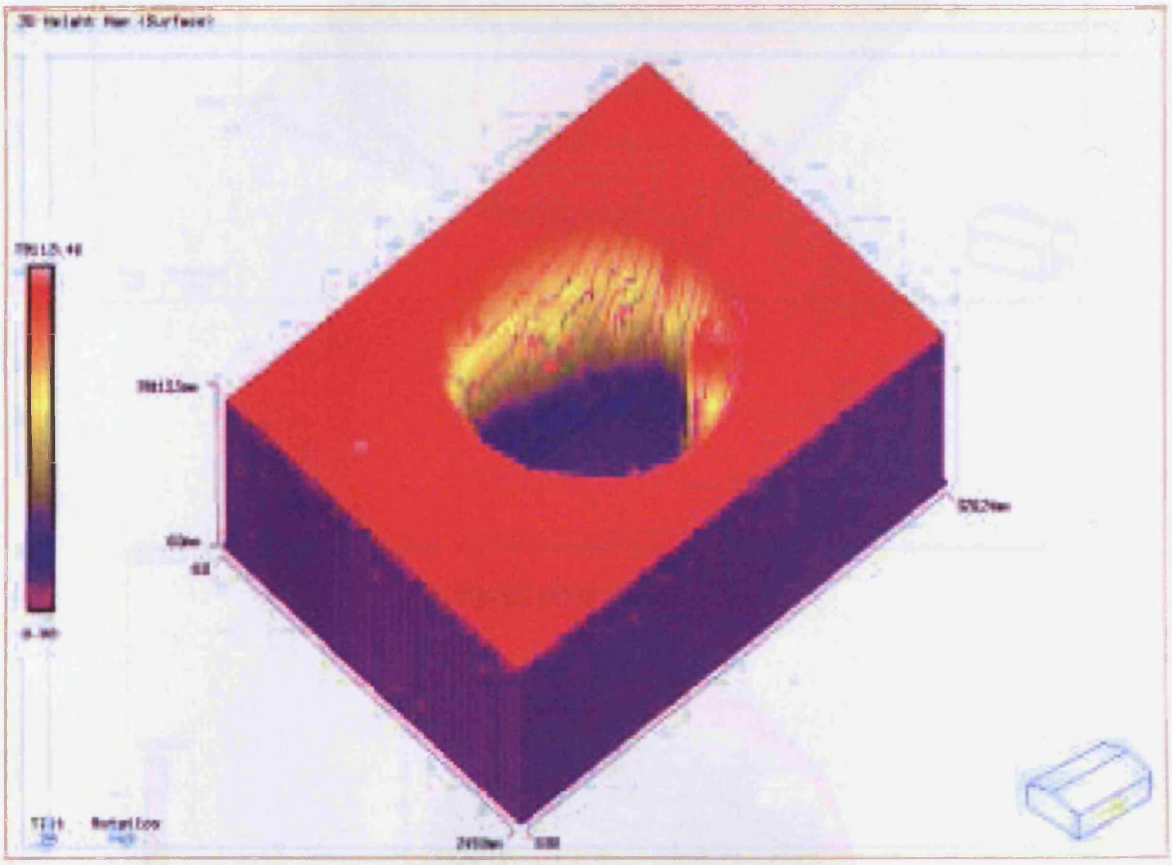
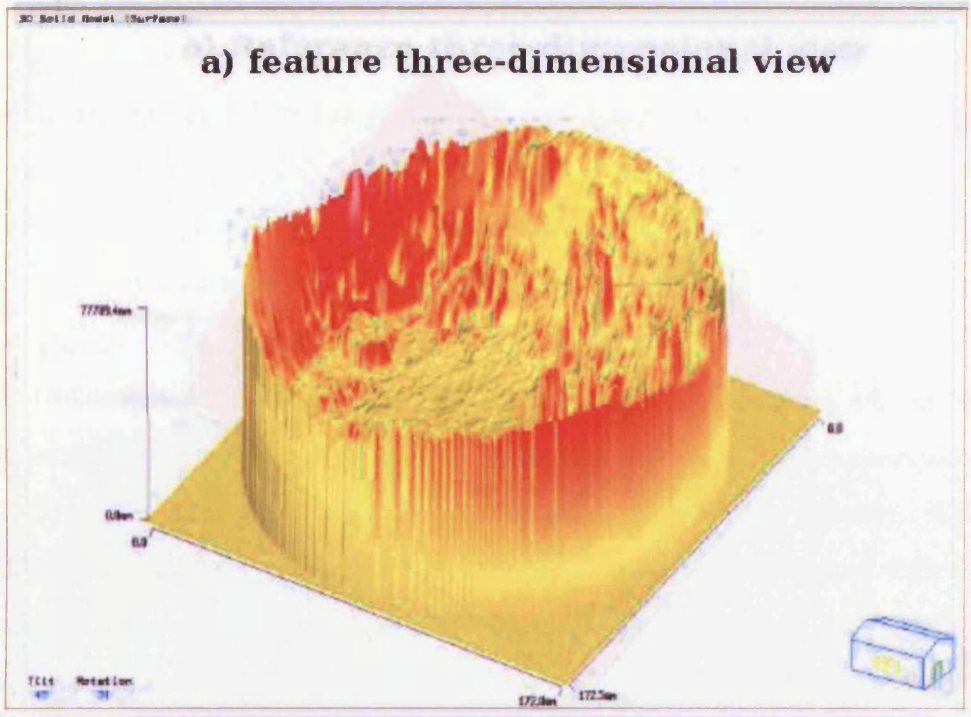


Figure 6.6: Profile of a concave form acquired by the WLI microscope.



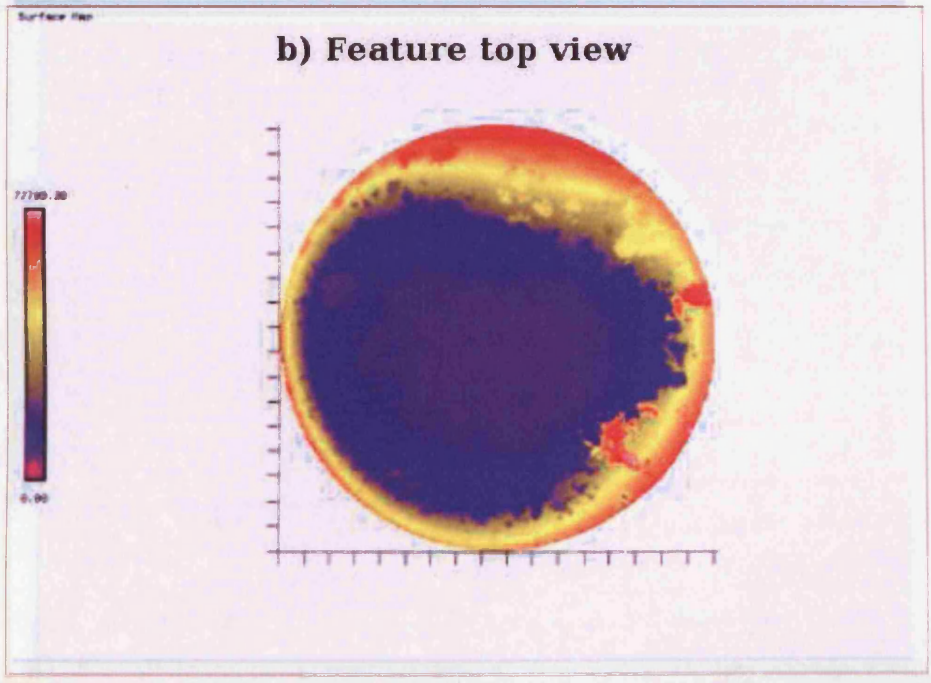
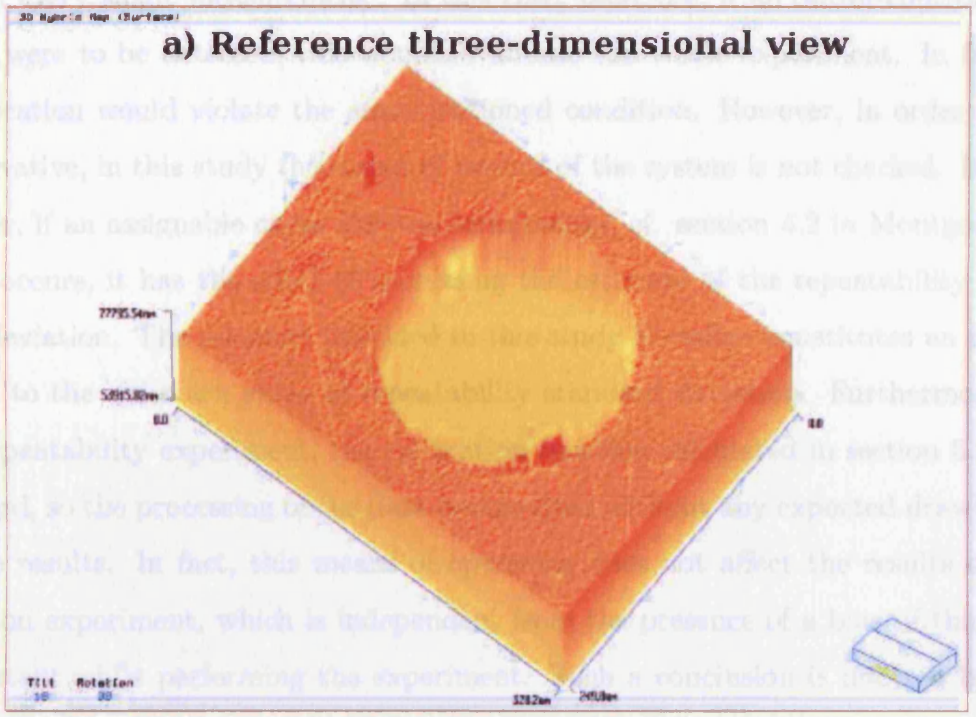


Figure 6.7: Feature profile while measuring the volume of a concave form.(a) Three-dimensional view. (b) Top view.



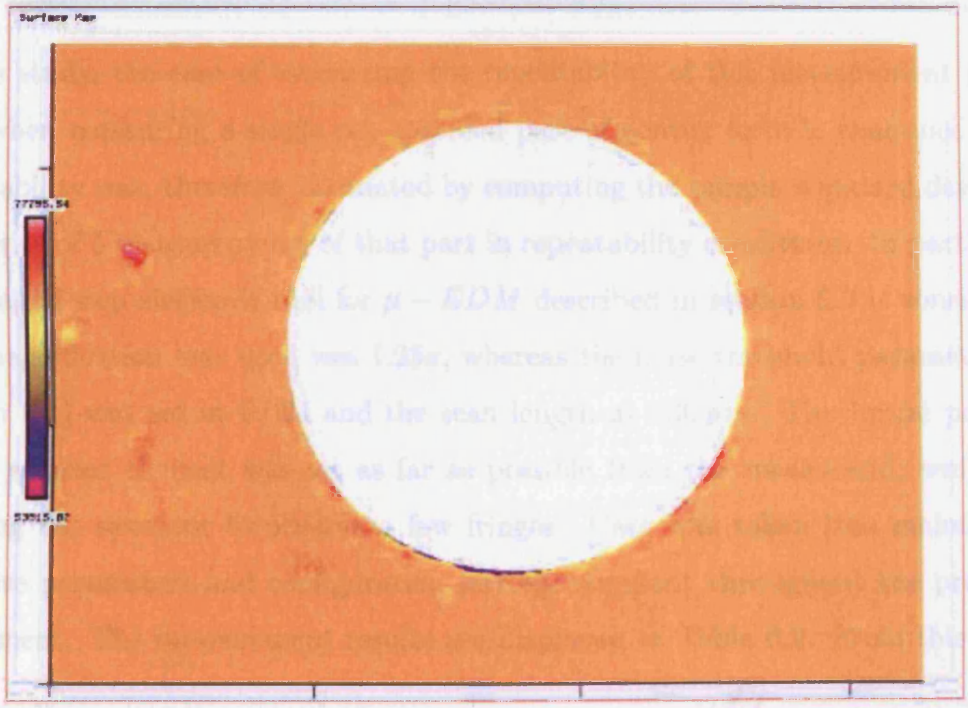


Figure 6.8: Reference while measuring the volume of a concave form.(a) Three-dimensional view. (b) Top view.

part of every single measurement'. In this case, therefore, if an out-of-control condition were to be detected, this would invalidate the whole experiment. In fact, a recalibration would violate the aforementioned condition. However, in order to be conservative, in this study the status of control of the system is not checked. In this manner, if an assignable cause (for the terminology, cf. section 4.2 in Montgomery, 1996) occurs, it has the effect of increasing the estimate of the repeatability standard deviation. The estimate provided in this study therefore constitutes an upper bound to the unknown value of repeatability standard deviation. Furthermore, in this repeatability experiment, the calibration function calculated in section 5.3.2 is not used, so the processing of the data is expedited without any expected drawbacks on the results. In fact, this means of operating does not affect the results of the precision experiment, which is independent from the presence of a bias, if this bias is constant while performing the experiment. Such a conclusion is inferred by the definition of sample standard deviation [2] and supported by section 5.1.1.2 of BS ISO 5725-1. Consequently, the measurement task was carried out by setting $HC = 1$ and neglecting the pertinent problems and proposed solution highlighted in Chapter 5 .

In this study, the case of estimating the repeatability of this measurement procedure when measuring a single pre-specified part of convex form is examined. The repeatability was, therefore, estimated by computing the sample standard deviation of a series of 5 measurements of that part in repeatability conditions. In particular, the shaped step electrode tool for $\mu - EDM$ described in section 6.3 is considered. The magnification lens used was $1.25x$, whereas the noise threshold parameter (cf. section 6.1) was set at 0.024 and the scan length at $230\,\mu m$. The initial position of the scanner element was set as far as possible from the measurand, while still allowing the operator to observe a few fringes. Care was taken into maintaining all these parameters and configuration settings constant throughout the precision experiment. The measurement results are displayed in Table 6.2. From this table,

[2] If $x_i = B + y_i$ with $i = 1, \ldots, n$ and $B = bias = constant$ then $V(X) = V(Y)$.

	Feature good points	Reference good points	Points forced	Volume V $[\mu m^3]$	A $[\mu m^3]$	ΔA [%]	B $[\mu m^3]$	ΔB [%]
1	43 095	43 871	6	1 646 319	1 474 078	10.5	1 531 093	7.00
2	43 243	43 560	1	1 653 740	1 475 994	10.7	1 534 785	7.19
3	44 051	42 186	0	1 684 918	1 510 293	10.4	1 568 045	6.94
4	43 591	41 535	2	1 679 259	1 507 049	10.2	1 564 003	6.86
5	42 592	43 679	0	1 646 228	1 483 732	9.9	1 537 482	6.60

Table 6.2: Volume measurement results in the repeatability experiment. $A = V_{prism}$, $B = V_{prism} + V_{tetrahedron,1}$, $\Delta A = 100 \cdot (V - A)/V$ and $\Delta B = 100 \cdot (V - B)/V$ ('Points forced' indicates the number of points adjusted to lie on the reference plane).

the repeatability standard deviation is estimated by the following quantity:

$$\hat{\sigma}_{volume} = S = \sqrt{\frac{\sum_{i=1}^{5} (V_i - \bar{V})^2}{5-1}} = 18\,614\,\mu m^3 \tag{6.6}$$

where $\bar{V} = 5^{-1} \cdot \sum_{i=1}^{5} V_i$. Due to the fact that $10^{12}\mu m^3 = 1\,cm^3 = 1\,cc = 1\,ml$, it follows that the result of equation 6.6 represents 0.018614 millionth of a millilitre. This estimated standard deviation is between 1.10% and 1.13% of the obtained measurement results. In Table 6.2, the number of good points on the feature and on the reference vary from one measurement to the other. It is therefore argued that such a number is affected by the natural variability of the measuring process. It is also observed that the good points represent about 12% of the overall number of points ('NumPoints'*'NumProfiles', $748 \cdot 480 = 359\,040$) in the field of view. This is due to the fact that most of the field of view is masked out and that the masked pixels are coded as bad data in the software controlling the measuring system.

From the data displayed in Table 6.2, the following two conclusions can be drawn. Firstly, the contributions of the two tetrahedra together account for about 10% of the measurement result and the sole volume $V_{tetrahedron,2}$ account for about 7%. Secondly, the maximum contribution to the variability of V is given by $\sqrt{V(V_{prism})} = 17\,257\,\mu m^3$, whereas $V_{tetrahedron,1}$ and $V_{tetrahedron,2}$ contribute only marginally with $\sqrt{V(V_{tetrahedron,1})} = 1\,886\,\mu m^3$ and $\sqrt{V(V_{tetrahedron,2})} = 3\,820\,\mu m^3$. These results hold for the particular case investigated and might significantly vary while varying the topology of the measurand. They qualitatively show, nevertheless, how beneficial the introduction of the two tetrahedron terms ($V_{tetrahedron,1}$ and $V_{tetrahedron,2}$) is. In fact, while contributing significantly to the measurement results (the mentioned 10% figure), they seem not to increase dramatically the overall variability of the measurement result.

After assessing the repeatability of the procedure, some considerations about its trueness are presented in the next section.

6.6 Biasing factors of the measurement procedure

In this section, the factors exerting a biasing effect on the measurement results of the proposed procedure are identified.

The proposed procedure demands the absence of bad data both in the feature and in the reference profile. However, while operating the instrument, it was found to be quite difficult to avoid completely the presence of bad data in these profiles. These bad data, therefore, are given a valid value on the basis of the good data close to them by means of the function 'fill bad data' and 'interpolate bad data' built in the software that controls the WLI microscope. This is an almost certain source of bias, due to the fact that the instrument does not provide any direct information about the z coordinates of the pixels classified as bad data. Furthermore, the more numerous the bad data that need to be filled in, the more the expected bias. It is anticipated, therefore, that the developed procedure appears difficult to apply in cases where the form of the measurand inherently generates bad data. Examples of these cases are:

- deep holes (or micro-holes) with aspect ratio sufficiently high to prevent the light to reaching the bottom or to be reflected back into the objective.
- Walls (or micro-micro-walls) so close to the vertical direction as to prevent the reflected light going back into the objective.

Moreover, the developed measuring procedure requires that the z coordinates of the feature profile have all the same sign when their projection on the reference geometry is subtracted from them in the GeneratePointPlane program. Hence, when this circumstance does not hold, such a difference is forced to zero. In performing such an operation, a bias is almost certainly introduced in the measuring procedure.

A further source of bias of the proposed measuring method is connected with the limited lateral resolution of the WLI microscope. In this study, with similarities to Weisstein (2006c), the word 'lateral resolution' is defined as the minimum distance

along the X or Y axis between distinguishable points in the recorded profile. Hence, it is given by 'Xscale' and 'Yscale' in Table 6.1, respectively. It is clear from this table that the lateral resolution can be significantly improved by increasing the magnification. Moreover, it is different for the two co-ordinate axes, it is therefore inferred that the positioning of the measurand on the stage of the WLI microscope can affect the bias induced by the lateral resolution. For example, in measuring the volume of a hole with an elliptical cross section or a pocket with a rhomboid cross section, the measurement results are expected to be different for different orientations of the part on the stage of the WLI microscope.

The bias induced by the resolution also provides a lower bound for the size of the parts whose volume can be measured by the proposed method, within a pre-specified trueness prescription. However, it is suspected that this lower bound can be lowered with the technical progress that makes available CCD cameras with an increasing number of photo-sites.

The identification of these limitations of the proposed measuring procedure leads to the suggestion of further direction of investigation. They are highlighted in the next section.

6.7 Potential sensitivity analyses

Sensitivity analyses that might be considered in a future study are those aimed to ascertain and quantify the limitations of the proposed measuring procedure that have been highlighted in the previous section. For example, the following investigations could be carried out:

- A sensitivity analysis of the procedure to the size of the feature, while considering also different forms, in order to detect where the proposed measuring method begins to be significantly biased. This study would provide the scale in which the method can be used, given a particular form of the measurand and the number of photo-sites of the CCD camera.

- A sensitivity analysis of the procedure to artificial changes in the number of bad data of a profile in order to discover the effect induced by the number of bad data and the subsequent use of the functions built in to the control software of the instrument. The number of bad data could, for instance, be artificially modified by introducing purposely built masks. Furthermore, the effects of different ways of filling bad data could also be considered.

- A sensitivity analysis of the measurement results to the changes in the orientation of the part on the stage, for different categories of forms. For instance, forms with an axial symmetric cross section and with a mirror symmetry could be considered (cf. Terr and Weisstein, 2003).

- A sensitivity analysis of the volume algorithm to the deployment of surfaces more complex than triangles when fitting the cloud of points obtained by the measuring process.

6.8 Compendium

In this chapter, a procedure for measuring volumes in the micrometre range and based on a WLI microscope was presented. The main advantages offered by the developed procedure are identified as follows:

- Volumes with irregular and complex boundary forms can be seamlessly measured, both in the case of convex and concave forms.

- The positioning of the measurand on the stage of the measuring instruments is not critical, due to the fact that the developed measurement procedure hinges on a reference built into the measurand itself.

- The joint deployment of prismatic and tetrahedral basic elements in the algorithm developed for measuring micro volumes proved to be a viable and effective approach both in terms of convergence and variability of the results.

Moreover, not only has the procedure been demonstrated in the case of a convex and a concave part, but its repeatability performance has also been assessed in the case of a convex part. In addition, some sources of bias in the practical use of the measuring procedure were identified and, consequently, further directions of investigation suggested.

Chapter 7

Contributions, conclusions and future work

This chapter summarises the contributions of this investigation to knowledge in the metrology domain and the conclusions reached. Possible directions for further investigation are also suggested.

7.1 Contributions

The main contribution of this investigation is the development of a procedure for measuring micro-volumes with a white light interferometric (WLI) microscope together with the necessary critical analyses of the sources of variability and of the calibration procedures of the WLI microscope measuring system. This work is a step forward in the direction of a full exploitation of measuring systems and endeavours to increase awareness of the centrality and the necessity of assessing the precision and trueness of measuring systems when on a micro- or nano-scale.

The specific contributions related to the task of measuring micro-volumes are:

- development of a novel procedure for measuring micro volumes of solids with irregular and complex boundaries of both convex and concave form. The pre-

cision of the proposed method greatly benefits from the fact that the positional reference is on the measurand itself, so that external positional references are not needed when taking a series of measurements. A quantitative repeatability study on a convex part confirmed this evidence;

- evaluation of the performances in terms of convergence and variability of the algorithm developed for measuring volumes. The contribution of the jointly deployed prismatic and tetrahedral solid elements to the measurement results is also quantitatively evaluated for the case examined;

- identification of the limitation of the proposed procedure for measuring micro volumes.

The contributions connected with the analysis of the variability of the measurement results obtained by means of a WLI microscope are:

- evaluation of the uncertainty contribution and of the effects of two discretionary setup parameters on the measurements of micro lengths taken by a WLI microscope. The two setup parameters are 'scan length' and 'initial scanner position for a given stage orientation'. It was found that variations of the 'scan length' in the interval specified by the instrument manufacturer significantly affect the measurement results. It also resulted that this effect does not appear to be caused by the transient of motion of the scanner element near the beginning and the end of the scan length. It was found, similarly, that variations in the 'initial scanner position for a given stage orientation' significantly affect the measurement results. The influence of both these setup parameters on the overall variability of the measurement results has also been quantitatively estimated in the two cases examined;

- estimate of the repeatability standard deviation of measurements of length taken by a WLI microscope along the z-axis in a pre-specified micrometric range;

- identification and characterisation of the dependence of the WLI length measurements on the sequence of the tests. Such a dependence was analysed by a regression analysis model. It is argued that this dependence is mainly due to some unknown critical component of the measuring system rather than to environmental conditions (a drift in the air temperature, for instance). On the basis of Schmit and Olszak (2002), it is also suspected that this unknown component might be the scanner motion subsystem of the WLI microscope. However, it is also argued that other contributions to this dependence exist;

- identification of a transient and a steady period in the operation of the WLI microscope. This finding is supported by experimental evidence and its subsequent analysis using a regression model. It is suggested that a few idle measurements be run prior to accomplishing a measurement task.

The detailed contributions pertinent to the calibration of the WLI microscpe are:

- identification of the calibration procedure built into the WLI microscope with the one-point calibration procedure described in BS ISO 11095. The consequent implications and limitations of this fact were highlighted and experimentally validated. The calibration procedure built into the particular WLI microscope used in this investigation is typically used in every WLI microscope, according to Schmit and Olszak (2002);

- provision of quantitative experimental evidence of the bias induced by the one-point calibration function;

- development of a novel spline-based calibration procedure in order to overcome the limitation of the one-point calibration;

- development of a novel control method, founded on two different sets of control charts, enabling the practical deployment of the developed calibration procedure.

Specific contribution connected with the availability of traceable reference materials is:

- provision of a cost-effective procedure for building traceable reference samples of length based on calibrated gauge blocks. The uncertainty of the developed procedure was also estimated using a 'type B' approach (BSI PD 6461-4). In this way, by comparing the uncertainty of a sample built by this procedure to the variability of the measuring process under investigation, it is possible to establish whether such a sample is suitable for the intended usage.

7.2 Discussion and Conclusions

The surveyed procedures for measuring micro volumes in connection with the determination of the volumetric wear in micro-EDM appear to offer room for significant improvement. In this context, the word improvement means that no surveyed micro volume measuring procedure endeavours to measure seamlessly micro volumes of solid with complex and irregular three-dimensional boundaries. Furthermore, none of the aforementioned procedures are provided with an analysis of their performance in terms of precision. Firstly, this study has aimed to fill this gap by providing a procedure for measuring micro volumes with irregular and complex boundaries and with either concave or convex form. Secondly, an estimate of the precision of the proposed procedure on a convex shape has also been provided when in repeatability conditions.

Moreover, in spite of the acclaimed key role of metrology in the current trend towards miniaturisation of products (cf. Leach et al., 2006, Dowling et al., 2004 and Masuzawa, 2000), in practice an underestimating attitude toward the necessity for a thorough assessment of the calibration procedures and of the precision of measuring systems has been experienced. The necessary investigations into the sources of variability and into the calibration procedures carried out in this work constitute supporting evidence for this underestimating attitude.

A large gap between declared intentions and current practice has not yet been bridged. This claim is supported by the fact itself that a very sophisticated and expensive piece of equipment, such as the WLI microscope investigated, relies on a calibration procedure recognised by the ISO as 'a weak and uncertain method because of the doubtfulness of the zero point. It should not be recommended for calibration purposes, but primarily for checking and existing linear calibration function.' (cf. section 8.2.1 in BS ISO 11095).

This work, by proposing a procedure for building traceable reference samples of length, also highlights how critical the availability of reference materials is. In fact, the unavailability of traceable reference materials in the same laboratory where a measuring system is in use inhibits any possibility of calibration. Moreover, it also prevents the establishment of any control system on the measuring system.

The main thrust of this investigation is the conviction that techniques for data analysis can provide a significantly large contribution in the exploitation of measuring apparatuses. This contribution appears to be especially beneficial in the light of the current trend toward miniaturisation. In particular, this study has highlighted the fact that gauge capability studies, the development of calibration techniques and the design of control charts should be key elements of this trend.

7.3 Future work

In this section, a few directions of potential further investigation are suggested. Two broad areas of further investigation have, in particular, been identified.

The first area is connected with the exploitation of the proposed procedure for measuring micro volumes in the analysis of micro manufacturing processes. Among them, the micro-electric discharge machining process, the machining and micromachining processes and, finally, the replications processes, such as micro injection moulding and hot embossing, could be considered. In this area, efficiency studies aimed at comparing competing machining processes could also be included.

In connection with the micro-electric discharge machining process, an investigation of the variability of the volumetric wear ratio, when constant setup parameters are considered in producing the same part, could be carried out. The outcome of such an investigation could provide a better understanding of the limitations of micro-EDM machining strategies, such as the uniform wear method (cf. Yu et al., 1998), that are based on a constant volumetric wear ratio.

Moreover, changes in the volumetric wear ratio when performing the same operation with changed technological parameters (for instance: maximum discharge current, open circuit voltage, duration of the pulse of voltage and time interval between adjacent pulse of voltage) could help to shed light on the nature of the material removal mechanism in micro-EDM. It is not entirely clear, in fact, whether the thermal model used for interpreting EDM operations on a macro scale is still applicable for EDM operations on a micro scale (cf. Singh and Ghosh, 1999). In this context, the term micro-EDM refers to EDM operations carried out with technological parameters that allow only low energy discharges to take place.

This procedure for measuring volumes could also be used to assess the effect of the metallic grain size in micro-EDM operations with tool steel. This investigation could enhance the knowledge of the removal mechanism in micro-EDM. In particular, it would allow a better insight into the effects, if any, of the occurrence of discharges predominantly in the internal part of a metallic grain or predominantly at its boundaries (mechanically more resistant at ambient temperatures).

Moreover, this procedure for measuring volumes could assist in establishing criteria for tool material selection in micro-EDM.

In connection with cutting processes, the procedure for measuring small volumes could provide quantitative information about flank and crater wear of the cutting tool (cf., for some similarities Dawson and Kurfess, 2005).

In connection with the replication processes, the proposed procedure for measuring volumes could be used in assessing the volumetric shrinkage on a micro scale for different families of geometric volumes (cylinders, spheres, cones, etc.) and for

different plastic materials, in order to build a database of coefficients for micro mould designing purposes. In addition, it could be deployed in establishing a quantitative relationship between volumetric wear on corners and edges of functional parts of the mould and the number of replicated micro parts produced.

In connection with efficiency studies, this procedure for measuring small volumes could be used to compare competing micro manufacturing processes on the basis of their material removal rate. In fact, the developed procedure allows the investigator to measure directly the volume removed by different processes in the unit of time. In this way, focused ion beam, laser machining, electro-chemical machining and electron-beam machining could be compared.

The second broad area is connected with perfecting and further analysing the proposed procedure for measuring small volumes. In particular, in the proposed procedure, the critical task of positioning the masks for obtaining the used profiles is left to the operator. This procedure could therefore be made less operator-dependent if some positioning and alignment algorithm were deployed, in a manner similar to that of Dawson and Kurfess (2005). In this manner, an improvement in precision is also expected.

The sensitivity analyses specified in section 6.7 could also be carried out. A comparative study between fitting the cloud of points extracted from the WLI data with triangles and with more complex surface elements could, in particular, be carried out.

References

ADEP1. *MapVue - Software User Manual Version 6.32.* ADE Phase Shift, Tucson, Arizona, 2000.

ADEP2. *MapVue - Software User Manual Version 6.42.* ADE Phase Shift, Tucson, Arizona, 2002.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6):716–723, 1974.

M. Alonso and J. E. Finn. *Fundamental University Physics - II fields and waves.* Addison-Wesley, Reading, Massachusetts, 1967.

F. Anscombe and J. W. Tukey. The examination and analysis of residuals. *Technometrics*, 5:141–160, 1963.

A. H. Bowker and G. J. Lieberman. *Engineering Statistics.* Prentice-Hall, London, second edition, 1972.

G. E. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters.* Wiley, New York, 1978. ISBN 0-471-09315-7.

BSI PD 6461-1. *General Metrology - Part 1: Basic and General terms (VIM).* BSI - British Standard Institution, London, 1995.

BSI PD 6461-3. *General Metrology - Part 3: Guide to the Expression of Uncertainty in Measurement (GUM).* BSI - British Standard Institution, London, 1995.

BSI PD 6461-4. *General Metrology - Part 4: Practical Guide to Measurement Uncertainty.* BSI - British Standard Institution, London, 2004.

BS 4058-ISO 5807. *Data Processing Flow Chart Symbols, Rules and Conventions -[ISO Title: Information Processing - Documentation Symbols and Conventions for Data, Program and System Flowcharts, Program Network Chart and System Resources Charts].* BSI - British Standard Institution, London, 1987.

BS 4311-3. *Specification for Gauge Blocks and Accessories - Part 3: Gauge Blocks in Use.* BSI - British Standard Institution, London, 1993.

BS 5233. *Glossary of Terms Used in Metrology (Incorporating BS 2643.* BSI - British Standard Institution, London, 1986.

BS EN 20286-1. *ISO System of Limits and Fits - Part 1: Bases of Tolerances, Deviations and Fits.* BSI - British Standard Institution, London, 1993.

BS EN ISO 3650. *Geometrical Product Specifications (GPS) - Length Standards - Gauge Blocks.* BSI - British Standard Institution, 1999.

BS ISO 11095. *Linear Calibration Using Reference Materials.* BSI - British Standard Institution, London, 1996.

BS ISO 3534-1. *Statistics- Vocabulary and Symbols - Part 1: Probability and General Statistical Terms.* BSI - British Standard Institution, London, 1993.

BS ISO 5725-1. *Accuracy (Trueness and Precision) of Measurement Methods and Results - Part 1: General Principles and Definitions.* BSI - British Standard Institution, London, 1994.

BS ISO 5725-2. *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.* BSI - British Standard Institution, London, 1994.

M. D. Byrne. A better tool for the cognitive scientist's toolbox: Randomization statistics. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pages 289–293, 1993. URL `http://chil.rice.edu/byrne/Pubs/cogsci93.pdf`. accessed on line on 7th August 2006.

E. Capello and Q. Semeraro. The harmonic fitting method for the assessment of the substitute geometry estimate error. part I:2D and 3D theory. *International Journal of Machine Tools & Manufacture*, 41(8):1071–1102, 2001.

G. S. Cheok, editor. *Proceedings of the 2nd NIST LADAR Performance Evaluation Workshop: March 15 -16,2005*, volume NISTIR 7266, 2005. National Institute of Standards and Technology.

S. S. Chim and G. Kino. Phase measurements using the mirau correlation microscope. *Applied Optics*, 30(16):2197–2201, 1991.

CIRDU. *The Temperature of the Human Body.* CIRD-Centro Interdipartimentale di Ricerca Didattica-Universita' di Udine, 2002. URL `http://web.uniud.it/cird/secif/termo/stati/es6.htm`. Accessed on 10th May 2005. This document is in Italian.

COHER. *Coherence (Physics).* URL `http://www.answers.com/topic/coherence-physics`. Accessed on 28th April 2005.

T. G. Dawson and T. R. Kurfess. Quantification of tool wear using white light interferometry and three-dimensional computational metrogy. *International journal of Machine Tools & Manufacture*, 45:591–596, 2005.

J. E. Decker and J. R. Pekelsky. Gauge block calibration by optical interferometry at the National Research Council of Canada. *Measurement Science Conference*, 1997. 22-24 Jannuary, Pasadena, California.

D. D. Dibitonto, P. T. Eubank, M. R. Patel, and M. A. Barrufet. Theoretical models of the electrical discharge maching process. I a simple cathode erosion model. *Journal of Applied Physics*, 66(9):4095–4103, 1989.

A. Dowling, R. Clift, N. Grober, D. Hutton, R. Oliver, O. O'Neill, J. Pethica, P. Nick, J. Porritt, J. Ryan, S. Anthony, S. Tendler, W. Mark, and R. Whatmore. *Nanoscience and Nanotechnologies: Opportunities and Uncertainties.* The Royal Society and The Royal Academy of Engineering, 2004. URL

http://www.nanotec.org.uk/finalReport.htm. Summary report accessed on 4th October 2006.

N. Drapper and H. Smith. *Applied Regression Analysis.* Wiley, New York, 1966.

M. D. Ernst. Permutation methods: a basis for exact inference. *Statistical Science*, 19(4):676–685, 2004.

J. J. Faraway. Practical regression and anova using R, 2002. URL http://www.stat.lsa.umich.edu/~faraway/book/. book freely available on line. Accessed on 17th July 2006.

C. Ferri, E. Brousseau, S. Dimov, and L. Mattsson. Repeatability analysis of two methods for height measurements in the micrometer range. In S. Dimov, W. Menz, and B. Fillon, editors, *4M2006 - Second International Conference on Multi-Material Micro Manufacture Proceedings*, Grenoble (France), 20th-22nd September 2006. Elsevier. ISBN 0-08-045263-9.

R. P. Feynman. Cargo cult science: Some remarks on science, pseudoscience, and learning how to not fool yourself - the 1974 Caltech commencement address. In J. Robbins, editor, *The Best Short Works of Richard P. Feynman - The Pleasure of Finding Things Out*, pages 205–216. Penguin books, London, 2000. ISBN 0-14-029034-6.

M. Galassi, J. Davies, T. J., B. Gough, G. Jungman, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual.* Network Theory Ltd, Bristol, second revised edition, 2005.

GNUPL. *Gnuplot homepage*, 2006. URL http://www.gnuplot.info. Accessed on 2nd October 2006.

A. Harasaki, J. Schmit, and J. C. Wyant. Improved vertical-scanning interferometry. *Applied Optics*, 39(13):2107–2115, 2000.

A. Harasaki, J. Schmit, and J. C. Wyant. Offset of coherent envelope position due to phase change on reflection. *Applied Optics*, 40(13):2102–2106, 2001.

H. L. Harter. Tables of range and studentized range. *The Annals of Mathematical Statistics*, 31(4):1122–1147, 1960.

D. M. Hawkins. Cumulative sum control charting: an underutilized SPC tool. *Quality Engineering*, 5, 1993.

J. Huang. Evaluation of angular error between two lines. *Precision Engineering*, 27 (3):304–310, 2003.

F. Jackson and E. W. Weisstein. *Tetrahedron.* From Mathworld - a Wolfram Web Resource, 2006. URL `http://mathworld.wolfram.com/Tetrahedron.html`. Accessed on 22nd September 2006.

K. G. Larkin. Efficient nonlinear algorithm for envelope detection in white light interferometry. *Journal of the Optical Society of America. A*, 13(4):832–843, 1996.

S. C. Lawlor. *ANSI C Programming.* West Publishing Company, Minneapolis-St.Paul, 1995. ISBN 0-314-02830-7.

R. Leach, D. Chetwynd, L. Blunt, J. Haycocks, P. Harris, K. Jackson, S. Oldfield, and S. Reilly. Recent advances in traceable nanoscale dimension and force metrology in the uk. *Measurement Science and Technology*, 17:467–476, 2006.

I. Malinosvsky, A. Titov, and C. A. Massone. High-precision method for gauge block measurements and comparison of gauge block interferometers. In *1998 Conference on Precision Electromagnetic Measurements*, pages 50–51, Washington,D.C., USA, July 1998.

I. Malinovsky, A. Titov, J. A. Dutra, H. Belaidi, R. d. S. Franca, and C. A. Massone. Toward subnanometer uncertainty in interferometric length measurements of short gauge blocks. *Applied Optics*, 38(1):101–112, 1999.

A. D. Marshall and R. R. Martin. *Computer Vision, Models and Inspection.* World Scientific Publishing, Singapore, 1992.

T. Masuzawa. State of the art of micromachining. *Annals of the CIRP*, 49/2: 473–488, 2000.

T. Masuzawa, M. Fujino, I. Kobayashi, and T. Suzuki. Wire electro-discharge grinding for micro-machining. *Annals of the CIRP*, 34/1:431–434, 1985.

MICRO. *Microlab Lab Manual.* Microfabrication Laboratory University of California at Berkeley, 2006. URL `http://microlab.berkeley.edu/labmanual/chap8/8.13.html`. Accessed on 9th September 2006.

R. G. Miller. *Beyond ANOVA Basics of Applied Statistics.* Chapman and Hall, London, 1997. ISBN 0-412-07011-1.

D. C. Montgomery. *Introduction to statistical quality control.* Wiley, New York, third edition, 1996. ISBN 0-471-30353-4.

D. C. Montgomery. *Design and Analysis of Experiments.* Wiley, New York, fifth edition, 2001. ISBN 0-471-31649-0.

D. C. Montgomery and G. C. Runger. Gauge capability and designed experiments. Part I: Basic methods. *Quality Enginerring*, 6(1):115–135, 1993.

A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics.* McGraw-Hill, New York, third edition, 1974. ISBN 0-07-042864-6.

NANOT. *Optical Profilometer.* Nanotechnology Core Facility-University of Illinois,Chicago, 2002. URL `http://www.ncf.uic.edu/manuals/OPTICAL_PROFILOMETER.pdf`. Accessed on 6th October 2005.

A. Nassef and H. ElMaraghy. Determination of best objective function for evaluating geometric deviations. *The international Journal of Advanced Manufacturing Technology*, 15(2):90–95, 1999.

NISTS. *NIST/SEMATECH e-Handbook of Statistical Methods*, 2006. URL `http://www.itl.nist.gov/div898/handbook/`. Handbook freely available on line. Accessed on 1st August 2006.

C. O'Mahony, M. Hill, M. Brunet, R. Duane, and A. Mathewson. Characterisation of micromechanical structures using white-light interferometry. *Measurement Science and Technology*, 14:1807–1814, 2003.

M. Patel, M. A. Barrufet, and P. T. Eubank. Theoretical models of the electrical discharge machining process. II the anode erosion model. *Journal of Applied Physics*, 66(9):4104–4111, 1989.

P. Pavlicek and J. Soubusta. Theoretical measurement uncertainty of white-light interferometry on rough surfaces. *Applied Optics*, 42(10):1809–1813, 2003.

D. Pham and S. Dimov. *Rapid Manufacturing:the Technologies and Applications of Rapid Prototyping and Rapid Tooling.* Springer-Verlag, London, 2001. ISBN 1-85233-360-X.

D. T. Pham, S. S. Dimov, S. Bigot, A. Ivanov, and K. Popov. Micro-EDM - recent developments and research issues. *Journal of Materials Processing Technology*, 149:50–57, 2004.

R Development Core Team. *R: a Language and Environment for Statistical Computing.* Vienna, Austria, 2006. ISBN 3-900051-07-0. URL `http://www.R-project.org`. Accessed in January-October 2006.

T. P. Ryan. *Modern Regression Methods.* Wiley, New York, 1997. ISBN 0-471-52912-5.

A. J. T. Scarr. *Metrology and Precision Engineering.* McGraw-Hill, New York, 1967.

J. Schmit and A. Olszak. High-precision shape measurement by white-light interferometry with real-time scanner error correction. *Applied Optics*, 41(28):5943–5950, 2002.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.

L. Scrucca. Qcc: an R package for quality control charting and statistical process control. *Rnews*, 4(1):11–17, 2004. ISSN 1609-3631. URL `http://cran.R-project.org/doc/Rnews/`.

S. R. Searle and G. Casella. *Variance components.* Wiley, New York, 1992. ISBN 0-471-62162-5.

A. Sen and M. Srivastava. *Regression Analysis - Theory, Methods, and Applications.* Springer-Verlag, New York, 1990. ISBN 0-387-97211-0.

D. Shin and T. R. Kurfess. Three-dimensional metrology of surface extracted from a cloud of measured points using a new point-to-surface assignment method: an application to pcb-mounted solder pastes. *Precision Engineering*, 28(3):302–313, 2004.

A. Singh and A. Ghosh. A thermo-electric model of material removal during electric discharge machining. *International Journal of Machine Tools & Manufacture*, 39: 669–682, 1999.

G. Tagaras. A dynamic programming approach to the economic design of X-charts. *IIE Transactions*, 26(3):48–56, 1994.

G. Tagaras. Dynamic control charts for finite production runs. *European Journal of Operational Research*, 91(3):38–55, 1996.

TELEM. *STL Format Description.* Tele Manufacturing Facility University of California, San Diego. URL `http://www.sdsc.edu/tmf/Stl-specs/stl.html`. Accessed on 21th September 2006.

D. Terr and E. W. Weisstein. *Mirror Symmetry.* From Mathworld - a Wolfram Web Resource, 2003. URL `http://mathworld.wolfram.com/MirrorSymmetry.html`. Accessed on 29th September 2006.

Y.-Y. Tsai and T. Masuzawa. An index to evaluate the wear resistance of the electrode in micro-edm. *Journal of Materials Processing Technology*, 149:304–309, 2004.

R. R. Vallance, C. J. Morgan, and S. M. Shreve. Micro-tool characterisation using scanning white light interferometry. *Journal of Micromechanics and Microengineering*, 14:1234–1243, 2004.

T. Várady, R. Martin, and J. Cox. Reverse engineering of geometric models – an introduction. *Computer-Aided Design*, 29(4):255–268, 1997.

W. Venables and B. Ripley. *Modern Applied Statistics with S-plus.* Springer, New York, third edition, 1999. ISBN 0-387-98825-4.

E. W. Weisstein. *Triangle Area.* From Mathworld - a Wolfram Web Resource, 2006a. URL `http://mathworld.wolfram.com/TriangularDistribution.html`. Accessed on 21th May 2006.

E. W. Weisstein. *Asymptotic Notation.* From Mathworld - a Wolfram Web Resource, 2006b. URL `http://mathworld.wolfram.com/AsymptoticNotation.html`. Accessed on 24th August 2006.

E. W. Weisstein. *Landau Symbols.* From Mathworld - a Wolfram Web Resource, 2002. URL `http://mathworld.wolfram.com/LandauSymbols.html`. Accessed on 24th August 2006.

E. W. Weisstein. *Resolution.* From Scienceworld - a Wolfram Web Resource, 2006c. URL `http://scienceworld.wolfram.com/physics/Resolution.html`. Accessed on 29th September 2006.

E. W. Weisstein. *Taylor series.* From Mathworld - a Wolfram Web Resource, 2006d. URL `http://mathworld.wolfram.com/TaylorSeries.html`. Accessed on 24th August 2006.

E. W. Weisstein. *Triangular Distribution.* From Mathworld - a Wolfram Web Resource, 2005. URL `http://mathworld.wolfram.com/TriangularDistribution.html`. Accessed on 14th May 2005.

E. W. Weisstein. *Optical Path Length.* From Scienceworld - a Wolfram Web Resource, 2006e. URL `http://scienceworld.wolfram.com/physics/OpticalPathLength.html`. Accessed on 3rd July 2006.

T.-H. Yang and J. Jackman. Form error estimation using spatial statistics. *Journal of Manufacturing Science and Engineering*, 122(1):262–272, 2000.

Z. Y. Yu, T. Masuzawa, and M. Fujino. Micro-EDM for three-dimensional cavities - development of uniform wear method -. *Annals of the CIRP*, 47/1:169 – 172, 1998.

Z. Y. Yu, J. Kozak, and K. P. Rajurkar. Modelling and simulation of micro EDM process. *Annals of the CIRP*, 52/1:143–146, 2003.

X. Zhu, H. Ding, and M. Y. Wang. Form error: An iterative reweighted least squares algorithm. *Journal of Manufacturing Science and Engineering*, 126(3):535–541, 2004.

Appendix A

Certificate of calibration of the gauge block set

CERTIFICATE OF CALIBRATION

ISSUED BY MITUTOYO (UK) LTD
CALIBRATION LABORATORY

DATE OF ISSUE 25-February-2005

CERTIFICATE NUMBER 133268

UKAS CALIBRATION 0332

Mitutoyo

Mitutoyo (UK) Ltd
Heathcote Industrial Estate
WARWICK
CV34 6TE

Tel: (01926) 450044
Fax: (01926) 887112
email: calibration@mitutoyo.co.uk

HEAD OFFICE
Mitutoyo (UK) Ltd
Joule Road
West Point Business Park
ANDOVER
SP10 3UX

Tel: (01264) 353123
Fax: (01264) 354883
email: enquiries@mitutoyo.co.uk

APPROVED SIGNATORY
L G Saundry

CUSTOMER	University of Wales Cardiff
MANUFACTURER	Mitutoyo Corp.
DESCRIPTION	Set of M88 steel Gauge Blocks
GRADE	Grade 1
IDENTIFICATION	0102048
CALIBRATION CONDITIONS	Ambient temperature 20 +/- 0.5 degrees Centigrade
BASIS OF CALIBRATION	Accuracy requirements of BS.4311:PART3:1993 specification for gauge blocks in use.
DATE OF CALIBRATION	10-December-2002

MASTER GAUGE BLOCK

Set I.D.	Certificate No.	Date of Calibration.
WCL510	200200541	17/01/2002
WCL510	200110931	04/01/2002

PARALLELISM: (VARIATION IN LENGTH)

All the gauges in this set have been examined for parallelism of the gauging surfaces and found to comply with the specified tolerances of the above standard.

MEASUREMENT OF LENGTH:

Gauges in this set have been measured for axial length by comparison using traceable calibration standards, at 20 degrees centigrade. The deviation from nominal size at the centres of the measuring faces complies with the specified tolerances of the above standard and is recorded in the table on the following page(s).

CERTIFIED BY

This certificate is issued in accordance with the laboratory accreditation requirements of the United Kingdom Accreditation Service. It provides traceability of measurement to recognised national standards, and to units of measurement realised at the National Physical Laboratory or other recognised national standards laboratories. This certificate may not be reproduced other than in full, except with the prior written approval of the issuing laboratory.

Figure A.1: Certificate of Calibration (first page).

CERTIFICATE OF CALIBRATION

Mitutoyo

Mitutoyo (UK) Ltd Heathcote Industrial Estate Warwick CV34 6TE

UKAS Accredited Calibration Labratory No. 0332

SERIAL NUMBER 133268

PAGE 2 OF 2 PAGES

DATE OF CALIBRATION 10-December-2002

SET ID No: 0102048

SPECIFICATION : Accuracy requirements of BS.4311:PART3:1993 specification for gauge blocks in use.

SET TYPE: M88

GRADE Grade 1

Parallelism and Deviation from nominal in units of 0.01µm at 20.0°C

Serial Number	Nominal Length (mm)	Dev	Par	Serial Number	Nominal Length (mm)	Dev	Par	Serial Number	Nominal Length (mm)	Dev	Par
010682	0.5	6	3	010596	1.23	5	1	010015	5.5	1	2
000331	1	7	7	010126	1.24	3	2	005843	6	-6	4
010539	1.0005	3	5	010649	1.25	1	2	010016	6.5	6	3
010083	1.001	6	4	010631	1.26	0	4	010176	7	4	1
010053	1.002	-1	3	010065	1.27	3	3	010239	7.5	-3	1
010095	1.003	5	2	010050	1.28	6	2	004677	8	6	2
010060	1.004	8	6	010556	1.29	1	4	002589	8.5	4	3
010687	1.005	-8	5	010020	1.3	4	3	010352	9	3	1
010039	1.006	6	3	010625	1.31	1	2	010005	9.5	-1	4
010607	1.007	1	1	010572	1.32	1	2	010513	10	-9	8
010579	1.008	-4	2	010602	1.33	4	3	007404	20	4	4
010060	1.009	4	2	010070	1.34	14	4	003723	30	2	1
010730	1.01	3	3	010560	1.35	13	4	001599	40	-5	4
010097	1.02	1	4	010086	1.36	3	2	004070	50	-3	4
010670	1.03	5	4	010568	1.37	3	3	001556	60	8	9
010091	1.04	11	5	010502	1.38	6	4	000604	70	-10	3
000328	1.05	6	4	000656	1.39	8	3	000366	80	16	5
010614	1.06	6	2	000988	1.4	3	2	000849	90	-19	3
010685	1.07	3	3	000906	1.41	6	3	004811	100	5	4
010163	1.08	-4	3	010046	1.42	2	4				
010714	1.09	2	5	010080	1.43	8	3				
010268	1.1	5	4	010606	1.44	1	5				
010578	1.11	1	4	010563	1.45	-4	4				
000843	1.12	1	1	010621	1.46	3	1				
010818	1.13	-4	6	010036	1.47	-1	3				
010167	1.14	4	4	010079	1.48	2	3				
010089	1.15	2	4	010613	1.49	-1	6				
010623	1.16	-1	3	010300	1.5	6	6				
010103	1.17	7	4	010604	2	6	3				
000586	1.18	8	6	010041	2.5	-4	4				
010582	1.19	4	2	010680	3	10	1				
010288	1.2	2	3	010611	3.5	8	6				
010015	1.21	-4	2	010639	4	-2	2				
010578	1.22	3	4	010074	4.5	5	2				
				010637	5	6	2				

Coefficient of thermal expansion :
(10.8 +/- 0.5) x 0.000001 /K

Comments and Notes:

Uncertainty of measurement:

>	0 up to	10mm	±0.08 µm
>	10 up to	25mm	±0.10 µm
>	25 up to	50mm	±0.12 µm
>	50 up to	75mm	±0.15 µm
>	75 up to	100mm	±0.18 µm

The reported expanded uncertainty is based on a standard uncertainty multiplied by a coverage factor k=2, providing a level of confidence of approximately 95%. The uncertainty evaluation has been carried out in accordance with UKAS requirements.

Figure A.2: Certificate of Calibration (second page).

Appendix B

Residuals covariance matrix

In this section it is proved that if the Gauss-Markov conditions hold, i.e. if $(i)\, E(e_{ij}) = 0\, (ii)\, E\left(e_{ij}^2\right) = \sigma^2\, (iii)\, E(e_{ij} \cdot e_{i'j'}) = 0$ for all $i = 1 \ldots a$, $j = 1 \ldots n$, $(ij) \neq (i'j')$, and given that equation 3.40 holds, i.e. given that $(iv)\, \hat{\mathbf{e}} = (\mathbf{I} - \mathbf{H}) \cdot \mathbf{e}$, then it follows that the covariance matrix of the residuals is given by the equation

$$var(\hat{\mathbf{e}}) = (\mathbf{I} - \mathbf{H}) \cdot \sigma_s^2 \tag{B.0.1}$$

Proof

By definition, it holds:

$$var(\hat{\mathbf{e}}) = E\left((\hat{\mathbf{e}} - E(\hat{\mathbf{e}}))(\hat{\mathbf{e}} - E(\hat{\mathbf{e}}))^T\right) \tag{B.0.2}$$

In equation B.0.2, the symbol $\mathbf{A}^T$ denotes the transpose of the matrix $\mathbf{A}$. From (i) and (iv) it follows that $E(\hat{\mathbf{e}}) = 0$. Hence, equation B.0.2 becomes

$$var(\hat{\mathbf{e}}) = E\left(\hat{\mathbf{e}}\hat{\mathbf{e}}^T\right) \tag{B.0.3}$$

From (iv) and equation B.0.3, it is derived

$$var(\hat{\mathbf{e}}) = E\left((\mathbf{I} - \mathbf{H})\, \mathbf{e}\mathbf{e}^T\, (\mathbf{I} - \mathbf{H})^T\right) \tag{B.0.4}$$

In Appendix B.2 of Sen and Srivastava (1990), it is reported that $E(\mathbf{AUB} + \mathbf{C}) = \mathbf{A}\,E(U)\,\mathbf{B} + \mathbf{C}$,where A, B and C are constant matrices, whereas U is a matrix of random variables. Therefore, from equation B.0.4, it follows:

$$var(\hat{\mathbf{e}}) = (\mathbf{I} - \mathbf{H})\,E\left(\mathbf{e}\mathbf{e}^T\right)(\mathbf{I} - \mathbf{H})^T \tag{B.0.5}$$

The definition of covariance matrix and (i) result in

$$var(\mathbf{e}) = E\left((\mathbf{e} - E(\mathbf{e}))(\mathbf{e} - E(\mathbf{e}))^T\right) = E\left(\mathbf{e}\mathbf{e}^T\right) \tag{B.0.6}$$

Hence, from (ii), (iii) and equation B.0.6, it is derived

$$var(\mathbf{e}) = E\left(\mathbf{e}\mathbf{e}^T\right) = \mathbf{I}\cdot\sigma_s^2 \tag{B.0.7}$$

Therefore, equations B.0.5 and B.0.7 yield B.0.8

$$var(\hat{\mathbf{e}}) = (\mathbf{I} - \mathbf{H})\,\mathbf{I}\cdot\sigma_s^2\,(\mathbf{I} - \mathbf{H})^T = (\mathbf{I} - \mathbf{H})\,(\mathbf{I} - \mathbf{H})^T\cdot\sigma_s^2 \tag{B.0.8}$$

As it is shown in sections 2.2 and 2.3 of Sen and Srivastava (1990), $H = X\left(X^TX\right)^{-1}X^T$, where X is the design matrix of the model. Thus, by substituting such expression in the equation $(\mathbf{I} - \mathbf{H})\,(\mathbf{I} - \mathbf{H})^T$ and semplying via matrix algebra, it results $(\mathbf{I} - \mathbf{H})\,(\mathbf{I} - \mathbf{H})^T = (\mathbf{I} - \mathbf{H})$. Consequentely, equation B.0.8 becomes

$$var(\hat{\mathbf{e}}) = (\mathbf{I} - \mathbf{H})\cdot\sigma_s^2 \tag{B.0.9}$$

Quod erat demonstrandum (i.e. which was to be proved).

Appendix C

Permutation tests in the initial scanner position

For the nominal length $40\,\mu m$ the R source code is as follows:

```
all.comb <- combinations(14,7)
diff.means <- numeric(3432)
top.index <-c(1,3,5,7,9,11,13)
top.values <-
  used.data.in.fringes.analysis.40$Pct.results [top.index]
middle.values <-
  used.data.in.fringes.analysis.40$Pct.results[-top.index]
real.diff.means <- mean(top.values)-mean(middle.values)
for (i in 1:3432) {
  samp <- all.comb[i,]
  pseudo.top <-
    used.data.in.fringes.analysis.40$Pct.results[samp]
  pseudo.middle <-
    used.data.in.fringes.analysis.40$Pct.results[-samp]
  diff.means[i] <- mean(pseudo.top)-mean(pseudo.middle)
}
```

```
no.greater <- sum(diff.means > real.diff.means)
p.value <- no.greater/3432
hist(diff.means, main="(a) 'Differences of the means'\n
  - 40 microns -",freq =F,xlab="'Differences of the means'")
abline(v=real.diff.means,col=2,lty=2,lwd=2)
legend(x="topleft","Actual difference",lty=2,lwd=2,col=2)
```

For the nominal length 200 μm, the R source code is as follows:

```
all.comb <- combinations(10,5)
diff.means <- numeric(252)
top.index <-c(2,4,6,8,10)
top.values <-
  used.data.in.fringes.analysis.200$Ptc.results[top.index]
middle.values <-
  used.data.in.fringes.analysis.200$Ptc.results[-top.index]
real.diff.means <- mean(top.values)-mean(middle.values)
for (i in 1:252) {
  samp <- all.comb[i,]
  pseudo.top <-
    used.data.in.fringes.analysis.200$Ptc.results[samp]
  pseudo.middle <-
    used.data.in.fringes.analysis.200$Ptc.results[-samp]
  diff.means[i] <- mean(pseudo.top)-mean(pseudo.middle)
}
no.greater <- sum(diff.means > real.diff.means)
p.value <- no.greater/252
hist(diff.means, main="(b)'Differences of the means'\n
   - 200 microns -", freq = F, xlab="'Differences of the means'")
abline(v=real.diff.means,col=2,lty=2,lwd=2)
legend(x="topleft","Actual difference",lty=2,lwd=2,col=2)
```

Appendix D

Validation of the algorithm for measuring volumes

The output of the implemented validation program for measuring volumes is shown in Figure D.1. It compares the theoretical volume of the eighth part of a sphere with the values computed by the implemented algorithm.

Printed by FERRI Carlo

Sep 23, 06 12:02 **SphereResultsComparison.txt** Page 1/1

```
----------------THEORETICAL VOLUME OF 1/8 OF A SPHERE--------------------
Long Double Decimal digit precision LDBL_DIG:  18
Double Decimal digit precision DBL_DIG:  15
The rational approximation of PI with 34 decimal digits
is provided by the constant M_PI1 in math.h:
3.14159265358979323846264338327950 29
Given the Long double decimal precision available,
the usable PI rational approximation of M_PI1 is of 18 decimal digits.
The 19th digit is shown to properly round the number to the 18th decimal digit:
3.141592653589793238 5
The theoretical Volume/8 of the sphere of radius 1.0000000000000000000 micrometre is:
0.5235987755982988731 cubic micrometres
(only 18 decimal digits are significant)
-----------------------------------------------
-----------------------------------------------
NumProfiles is 1001
NumPoints is 1001
NumLines is 1002001
Volume of the part with Prisms only is: 0.5230430467975 19 cubic micrometres
Volume of the part with the first Tetrahedron is: 0.5232269765797 36 cubic micrometres
Computed Volume of the part is: 0.523593765690266 cubic micrometres
Prism percentage deviation is 0.105181 %
First Prism Deviation is 0.070052 %
-----------------------------------------------
-----------------------------------------------
NumProfiles is 2001
NumPoints is 2001
NumLines is 4004001
Volume of the part with Prisms only is: 0.5233210337675 62 cubic micrometres
Volume of the part with the first Tetrahedron is: 0.5234131692040 18 cubic micrometres
Computed Volume of the part is: 0.523597046201650 cubic micrometres
Prism percentage deviation is 0.052715 %
First Prism Deviation is 0.035118 %
-----------------------------------------------
-----------------------------------------------
NumProfiles is 4001
NumPoints is 4001
NumLines is 16008001
Volume of the part with Prisms only is: 0.5234599134502 20 cubic micrometres
Volume of the part with the first Tetrahedron is: 0.5235060459921 70 cubic micrometres
Computed Volume of the part is: 0.523598167933154 cubic micrometres
Prism percentage deviation is 0.026405 %
First Prism Deviation is 0.017594 %
-----------------------------------------------
-----------------------------------------------
NumProfiles is 5001
NumPoints is 5001
NumLines is 25010001
Volume of the part with Prisms only is: 0.5234876915851 48 cubic micrometres
Volume of the part with the first Tetrahedron is: 0.5235246099035 35 cubic micrometres
Computed Volume of the part is: 0.523598343994974 cubic micrometres
Prism percentage deviation is 0.021133 %
First Prism Deviation is 0.014082 %
-----------------------------------------------
-----------------------------------------------
NumProfiles is 6668
NumPoints is 6668
NumLines is 44462224
Volume of the part with Prisms only is: 0.5235154703913 19 cubic micrometres
Volume of the part with the first Tetrahedron is: 0.5235431682622 16 cubic micrometres
Computed Volume of the part is: 0.523598496183203 cubic micrometres
Prism percentage deviation is 0.015857 %
First Prism Deviation is 0.010567 %
-----------------------------------------------


-------------------------------------------------------------------------------
                    Summary of the Results
                                                          Difference,%
theoretical Volume                    0.5235987755982 98  0.0000000000
Computed Volume (grid pitch 0.001)    0.5235937656902 66  0.0009568220
Computed Volume (grid pitch 0.0005)   0.5235970462016 50  0.0003302904
Computed Volume (grid pitch 0.00025)  0.5235981679331 54  0.0001160555
Computed Volume (grid pitch 0.0002)   0.5235983439949 74  0.0000824302
Computed Volume (grid pitch 0.00015)  0.5235984961832 03  0.0000533644

-------------------------------------------------------------------------------
```

Saturday September 23, 2006 1/1

Figure D.1: Output of the validating program for volume measurements .