The Quantification of Sampling Error in Coordinate Measurement

by
Joanne Coy

Submitted for the Degree of
Doctor of Philosophy

The University of Warwick Department of Engineering

May 1991

Table of Contents

Table of Figures

Table of Photographs

Abstract

This work was carried out between October 1986 and February 1989 at the School of Engineering, University of Warwick. The thesis begins with a review of the configurations of coordinate measuring machines in common use and an investigation into the types and magnitudes of the errors incurred due to various phenomena associated with the design, deformation or misalignment of the machine components. Some of the more significant of these errors are then measured and tabulated with a view to using them as a comparison to further work. Methods by which these errors can be rectified are then briefly reviewed.

Chapter 2 is concerned with the inadequacies associated with current coordinate measuring machine software algorithm design. Measurement practices are reviewed and sources of inconsistency or potential misinterpretation are identified. Sampling error is singled out as being of particular significance.

Chapter 3 reviews geometric element fitting procedures and the errors that can result from ill advised measuring practice. Systematic and random error analyses of the errors incurred in the estimates of geometric parameters are reviewed and an original investigation is performed into the errors incurred in parameters due to not considering all possible data (sampling error.)

Chapter 4 presents an assessment of the nature of the problem of sampling error and outlines the way in which a robust algorithm for the formal quantification of these errors should be formulated.

Chapter 5 then identifies the criteria that would maximise the implementability of an algorithm of this type. An algorithm satisfying these particular requirements is duly developed.

Finally, chapter 6 consists of an investigation into the effect of probe geometry on the phenomenon of sampling errors. A method is then developed whereby the probe geometry that will minimise sampling error can be readily selected.

Introduction

Over the past decade, the use of 3-axis coordinate measuring machines (CMMs) has become an integral part of engineering inspection practice. The reasons for the success of coordinate based inspection are in its distinct advantages over traditional measurement techniques i.e. inspection times are dramatically reduced, relatively complex inspection procedures can be performed easily and the interrelationships between features on a component, which are often vital to function, are readily calculable.

The development of computer aided coordinate measurement techniques has lead to the formation of a fundamentally new philosophy of geometric measurement. Representation of surfaces is now mathematical, the workpiece being described by sets of equations in a defined coordinate system. Consequently components are regarded as consisting of several definable geometric elements, planes, circles, spheres, helical surfaces etc. The geometry of the workpiece is ascertained by measuring several surface points within a three dimensional coordinate system . These coordinates are stored as a digital image of the workpiece in the computer memory, giving no indication of the dimension or quality of the workpiece with respect to the original drawing. Specific software is then employed in order to achieve comparisons and geometric parameter evaluation.

However, the efficacy of coordinate measurement techniques is significantly impaired by uncertainties either due to machine borne errors and/or inadequacies associated with the geometric evaluation software. Measurement error resulting from the CMM itself can be broadly categorised into two distinct types; random error mainly due to vibration or electrical noise, and systematic error resulting from deformation and misalignments of the component parts of the machine. The contacting probe can also

induce random error through lack of repeatability and systematic error resulting from its construction and design.

Studies on the reliability of commonly used and supposedly advanced geometric evaluation algorithms have produced some worrying results; it has been found that the least squares estimates of parameters can produce answers in excess of five times the actual value, and that fewer than 3% of straightness estimates are correct in accordance with existing standards (1). These uncertainties can result from several phenomena; instabilities in estimates due to ill-conditioning of the data fitting routine , random measuring error, problems in the digital interpretation and implementation of standards and quantitative validation of software. The two latter problems are of particular importance as the present unsatisfactory situation regarding CMM software cannot be significantly improved until some progress is made in their solution. Until then, as different quantities and distributions of data will be used, definitions of geometric form will vary among practitioners, similarly the quantitative validation of software, as hitherto there has been no way of knowing the accuracy of the geometric estimates.

If rigorous definitions of geometric form are to be effectively implemented in coordinate measurement procedures, it is required to express surfaces in terms of best estimates of appropriate parameters calculated according to appropriate mathematical criteria (2). This stipulation effectively means that in order to apply standards it is necessary to refer to algorithms having the capability to calculate geometric parameters to within a specified accuracy. The question of software validation is also a question of knowing the likely accuracy of the estimates of the geometric entities. Both of these difficulties are associated with the same phenomenon, i.e. the quantification of the discrepancy between the estimates of the appropriate parameters using all possible surface data, and the similar estimates made using a particular number and distributions of data points, this phenomenon will be henceforth termed sampling error.

It is the intention of this thesis to identify phenomena, borne by either machine or

software, impairing the efficacy of coordinate measurement, reviewing where possible, established methods by which these faults can be rectified. Ultimately attention is focussed upon the unique task of the formulation of an effective implementable algorithm for the quantification of sampling error. To this end an investigation into the viability of a general solution is performed, followed by a rigorous definition of the problem and the subsequent formulation of an original theoretical expression for sampling error. Experiments are then performed systematically to quantify the behaviourof this theoretical expression. In this way two algorithms are developed. The first a tentative excercise examining thebehaviourof the theoretical expressions. Although the algorithm developed provides reliable estimates of sampling error, its applicability is limited. However, the second is fully implementable in normal CMM inspection practice, requiring no excess data. Most importantly, it is proved that the autocorrelation function, calculated using an extremely sparse data set provides an adequate measure of spatial variation for the quantification of sampling error, and therefore provides the way forward in the solution of these problems.

References

[1] F. EDSON & G. PARRY (1985) Integrity of Software Associated with Coordinate Measuring Machines. NPL Proc. Conf. Software for Coordinate Measuring Machines, 25/9/1985.

[2]: M.G. COX & K. JACKSON: "Algorithms and Software for Metrology: A Statement of Need";NPL Report MOM65, June 1983.

Errors Associated With Coordinate Measuring Machines Corrective Calibration Techniques

1. Introduction.

The use of coordinate measuring machines (CMMs) has the distinct advantage over other, more traditional forms of measurement procedure by allowing the measurement and data analysis processes to be performed within a single, self contained unit at speeds that compare favourably with alternative inspection methods (table 1.1)(1). As a result, CMMs are now used ubiquitously throughout the majority of engineering industries, and coordinate measurement with its associated inadequacies have become an important area of research in engineering Metrology.

The following chapter introduces CMMs, outling their characteristics and describing some common configurations. The errors induced by the CMM construction are also identified and measured for comparison to later work. Existing procedures for rectifying these errors are then reviewed.

1.1. The Integration of CMMs in Industry

The fact that the CMM has risen, since it's introduction some fifteen years ago, to become an integral piece of equipment in most engineering inspection departments, is largely due to the following contributory factors:

i) Costs are cut dramatically by a reduction in inspection time, the utilisation of less skilled operators, and the reduced need for many expensive gauges.

ii) The versitility that CMMs afford give flexibility to keep up with constantly changing market demands.

Typical Inspection Times		
Features Inspected	Conventional Techniques (mins)	CMM (mins)
Length	6	0.3
Squareness	15	2.5
Flatness	4	2
Angle between Faces	15	2
Diameter	3	0.5
Hole Centre Location	6	0.5

Table 1.1

iii) CMMs also supply readily quantified digital information which provides vital feed-back for a computer integrated manufacturing system.

iv) Traditional inspection methods do not readily permit the measurement of all features on complex components and their interrelationships, which is essential for function related inspection.

v) Set-up time is substantially reduced as traditional surface plate inspection techniques are particularly time consuming in the establishing of appropriate reference points. CMMs facilitate the virtual elimination of these procedures.

vi) Elimination of operator influence digital output eradicates subjective interpretation and automatic data recording prevents errors in transcribing readings, which leads to the situation of de-skilling relatively complex inspection procedures.

vii) CMMs have a degree of diagnostic capability. Through ongoing statistical and geometrical analysis, deterioration in the quality of particular features can be identified and consequently a fault in the manufacturing process diagnosed and corrected before parts containing 'out of tolerance' features are produced.

1.2. Configurations of CMMs.

Several configurations of CMM have evolved, each basically being a means by which the contacting probe may be traversed along three axes with a known position. Each arrangement possesses it's own particular advantages, depending upon the application. The most popular configurations are as follows:

i) Bridge

ii) Column

iii) Cantilever

iv) Horizontal arm

v) Gantry

Each of these configurations is depicted in Figure 1.1.

1.2.1. Bridge

This particular construction is by far the most popular, mainly due to the fact that it's rigid symmetrical construction results in great accuracy. It's configuration is as follows; the probe quill (Z-axis) is mounted upon a horizontal member (Y-axis), which in turn is supported by two vertical members which traverse the beam in the X-axis. However, it must be remembered that the accuracy afforded by the rigid construction may be reduced if the two vertical members are not tracked in perfect alignment.

1.2.2. Column

These machines are usually refered to as universal measuring machines rather than CMMs. They usually have a moving saddle arrangement designed for maximum geometric accuracy and rigidity

1.2.3. Cantilever

The cantilever configuration comprises a probe quill (Z-axis) traversing on a mutually perpendicular horizontal overhung beam (Y-axis), which in turn moves along a

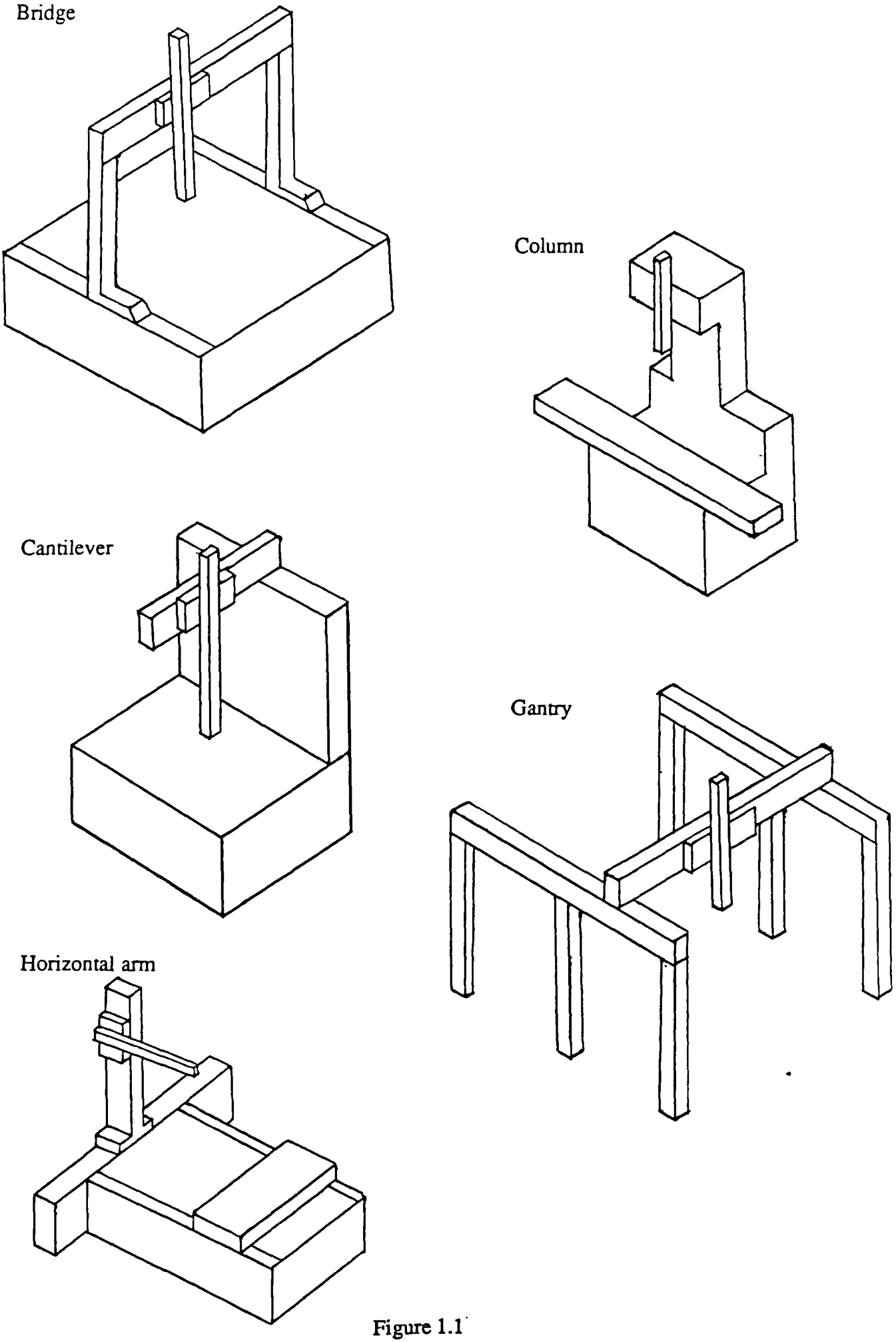


Figure 1.1

mutually perpendicular track (X-axis). This arrangement has the advantage of enabling parts larger than the machine table to be inspected without inhibiting machine travel.

1.2.4. Horizontal arm

The horizontal arm construction consists of a horizontal probe quill (Z-axis) mounted on a mutually perpendicular column (Y-axis) which in turn moves in a mutually perpendicular motion (X-axis). This open structure facilitates access sufficient to allow the inspection of components of complex geometry.

1.2.5. Gantry

The gantry configuration consists of a probe quill (Z-axis) moving on a mutually perpendicular beam (Y-axis) which, in turn, is supported by and traverses along two elevated rails (X-axis) which are usually supported by columns mounted directly on the shop floor or appropriate measuring surface. This arrangements particularly useful for the inspection of extremely large objects (2).

1.3. Inaccuracies Associated with CMMs

Like all numerically controlled machines, CMMs have errors inherent in either their construction and software that result in inaccuracies in positioning. The errors in the CMM hardware can be attributed to various sources either specific or non-specific

1.3.1. Non-Specific

(i) Ambient conditions.

(ii) Variable coefficients of expansion of machine parts.

(iii) Vibration via foundations.

(iv) Variable sensing force.

(v) Inappropriate probe diameter.

(vi) Surface finish and stiffness of workpiece.

(vii) Inappropriate measuring strategy.

1.3.2. Specific

(i) Interpolation error.

(ii) Errors in the machine scales.

(iii) Reversal error owing to play and friction in the slideways.

(iv) Geometry of machine, straightness of guideways etc..

(v) Infringement of Abbe principle.

The errors resulting from these sources can be divided into three catagories; repeatibility ,random and systematic. Errors of repeatability can be controlled to within acceptable limits by rigourous design of structure, alignment of slideways and enviroment control (3). Random errors can be induced by vibrational sources either within or external to the CMM, they can be quantified by repeated measurement of an artefact of known dimension (usually a slip or ring gauge)at a particular location in the measuring volume of the CMM, calculating the standard deviation of the measured dimensions and using this value as a measure of the random errors present in the CMM.

Systematic geometric errors such as linear displacement error, pitch ,roll and yaw, are a source of significant errors in the positioning of the CMM axes and are primarily caused by parasitic errors of machine elements due to manufacturing tolerances in the main spindle and the cross slides, and construction misalignments (4). These errors are effectively magnified by concentrated heat sources within the machine such as drive motors and spindle bearings which create thermal gradients along the machine structure. This phenomenon results in fairly significant machine deformation, spindle growth and lead screw expansion.

The aims of the following section are twofold. Firstly, to measure some of these machine errors with a view to comparing them with later work, in order to asses its significance, and secondly to review the calibration procedures used in CMM software to

improve the accuracy of the CMM by compensating for these machine errors.

2. Measurement of Geometric Error Components

2.1. Experimental Set-Up

The primary purpose of the experiment requires that the individual error components be known (for comparison with work in Chapter 5), it was therefore decided to mainly use laser interferometric methods as most of the error components can be assessed in this way using the same set of equipment.

A series of direct machine error measurement experiments has been carried out on the 'LK Micro-Four' coordinate measuring machine in the microengineering centre at Warwick University. At the time that the procedures were performed, the machine was approximately eight years old, and had undergone no significant corrective treatment. Throughout the experiments, the enviromental conditions remained constant at a temperature of 20 degrees Celsius and a humidity of 50%

The equipment used for the calibration procedures consisted of the 'LK Micro-Four' coordinate measuring machine, the 'Hewlett Packard HP5526A' laser interferometer system, a minicam containing a board for the laser with an RS 432 and RS 232 leading to a BBC microcomputer. The arrangement of the equipment is shown in Photograph 1.1 and schematically in Figure 1.2. Roll measurements were made using the Rank Taylor Hobson Talyvel electronic level.

2.1.1. The LK Micro-Four Coordinate Measuring Machine

This is a four axis measuring machine, consisting of three linear axes, and a rotary table. It consists of a cast iron base, mounted upon three supports, the front of which supports the rotary table. The measuring equipment, carried by the rear of the base, consists of a column moving on the longitudinal, or X-axis, a probe carrier (quill) moving vertically along the column on the Z-axis, and a probe bar moving transversely over the

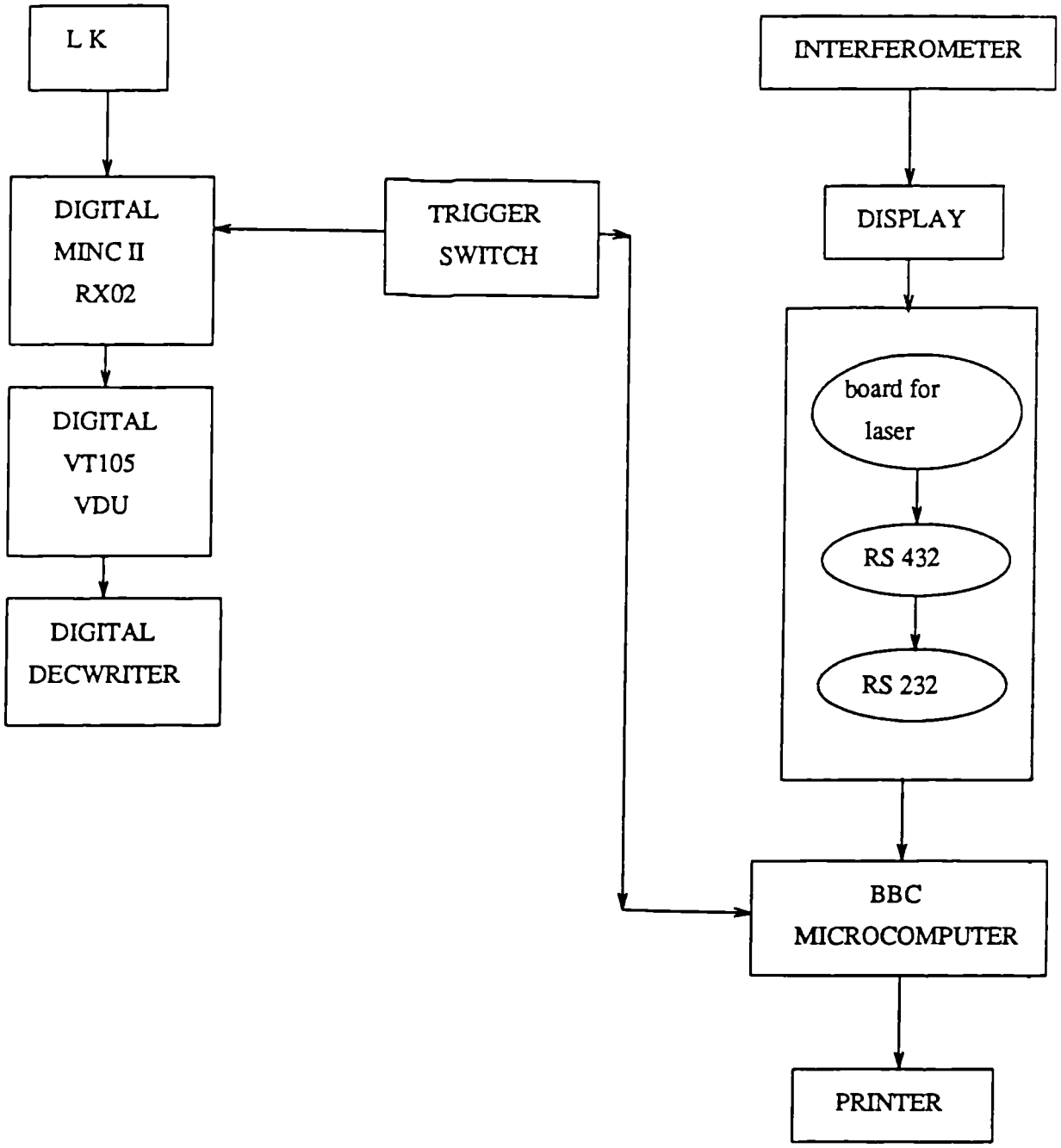
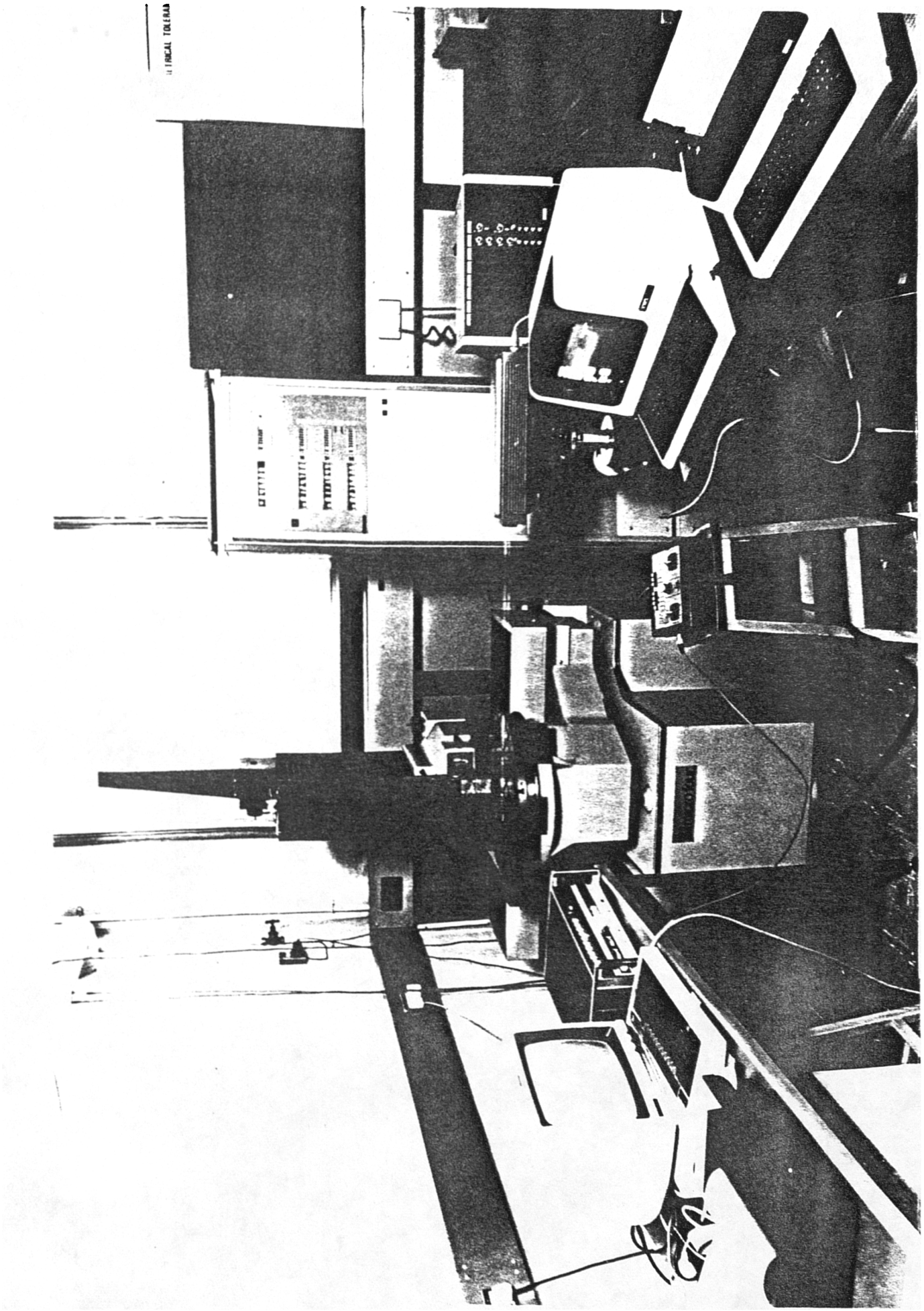

L K
INTERFEROMETER
DIGITAL
MINC II
RX02
TRIGGER
SWITCH
DISPLAY
board for
laser
DIGITAL
VT105
VDU
RS 432
RS 232
DIGITAL
DECWRITER
BBC
MICROCOMPUTER
PRINTER

Figure 1.2



Photograph 1.1

table on the Y-axis.

The measuring volume is a cuboid, 600mm x 280mm x 450mm. The manufacturer states that accuracy of the X,Y and Z axes is ±0.005mm, and 15 degrees ± 5' for the rotary table with a repeatibility in the order of ±0.005mm for the linear axes, and ±5' for the rotary table.

2.1.2. The Hewlett Packard HP5526A laser interfeometer system

The interferometer system was used to assess pitch, yaw straightness and linear displacement error all three linear axes and roll in the X and Y axes,

The accuracy of the interferometer in air is determined by the uncertainties of the source frequency and the refractive index of the air (5). He-Ne lasers generally operate very near a resolution of 633nm (6).) This value compares favourably with the best physical standards available, and, in most cases, is certainly acceptable for machine tool and CMM evaluation. Although manufacturers of laser interferometer equipment often stipulate that a warm up time of 30 to 40 minutes is sufficient time for the interferometer to stabilise, it is often advisable to allow at least an hour.

Instabilities can arise when the laser is not effectively isolated from the interferometer. This results in reflections from the optical system re-entering the laser cavity. These reflections are indistinguishable from the desired signal and are amplified within the resonator causing significant instablities. It must be pointed out that the Hewlett Packard interferometer facilitates this isolation within the laser head, and where great accuracy is required, there is some doubt as to whether this isolation is sufficient (6).

2.2. Experimental procedure

The readings from the CMM were obtained by means of its interface with the printer. Measurements from the laser interferometer set up were monitored through a 'Bede' minicam interface unit and a BBC microcomputer.

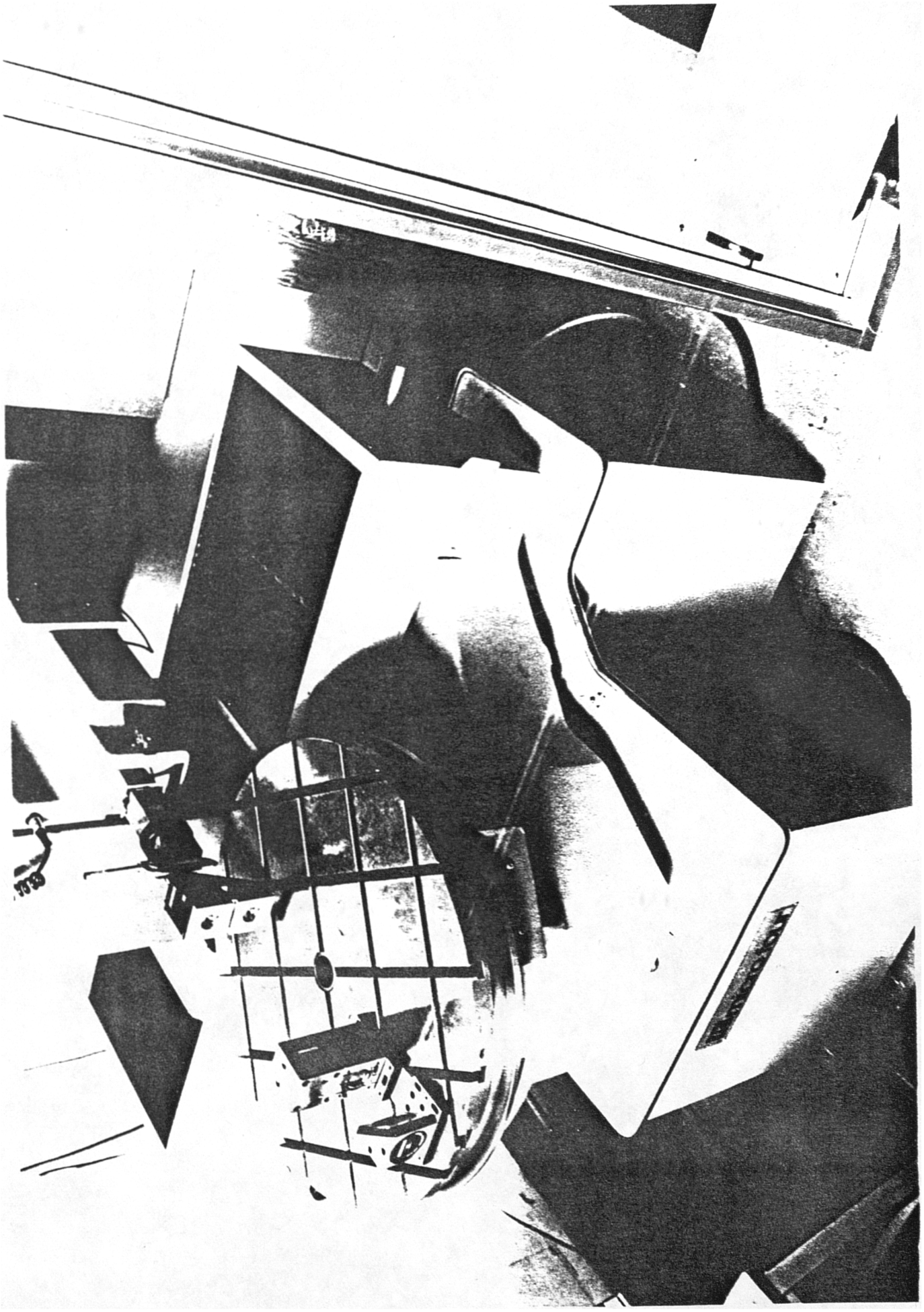
In order to ensure that the reading from both sets of apparatus were taken absolutely simultaneously, (to protect against the effects of machine and minute temperature drifts), a slight modification had to be made to the standard equipment to trigger the machines to render their respective results at the same time.

The standard CMM set up includes a handset for taking coordinate readings manually, behind the switch marked 'MANUAL' there were five pairs of contacts; one pair of which had remained hitherto unused. Wires were connected to these contacts and a new socket fitted to the side of the handset box. A cable connected this socket to the 'analogue in' I/O port of the BBC. This modification enabled the operator to trigger the two systems simultaneously by pressing the 'MANUAL' button. Readings were taken for each of the error components at intervals of 5 or 10mm along the respective axes. Figures 1.3 to 1.12 illustrate the way in which errors of linear displacement and angularity can be measured using the laser interferometer and the relevant accessories. The pictures were reproduced from the 'Hewlett Packard HP5526A Laser Interferometer System' handbook. Photograph 1.2 shows the pitch measurement procedure for monitoring the Y axis.

2.3. Results of Measurement of Machine Errors

Due to lack of equipment, roll in the Z-axis could not be assessed. The calibration of straightness was possible in the XY plane only as errors of straightness in the other planes was of such a degree that calibration for a length of traverse in excess of 15mm or so was not feasible as the laser could not reflect back into the appropriate cavity, and therefore repeated repositioning of the laser reflectors would be required in order to monitor a significant proportion of the axis. Five traverses measuring each error component were made and the mean measurements plotted and are shown by graphs 1.1 to 1.10.

These results are summarised in Table 1.2 for future reference, only the approximate range of values is included as all that is required is an idea of the general magnitude of errors that can be expected.



Photograph 1.2

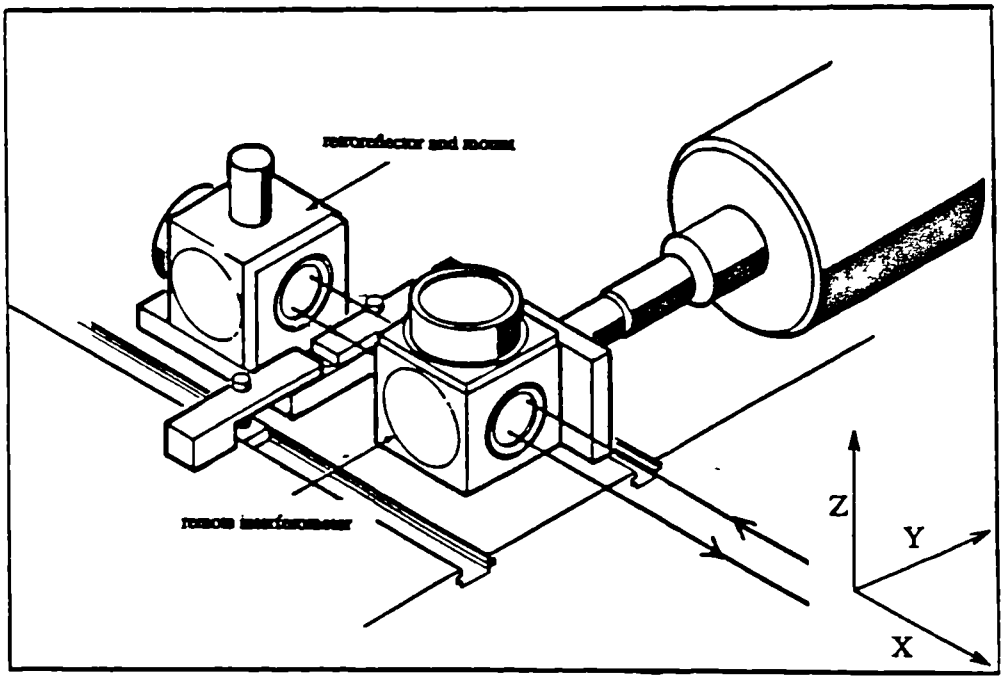


Figure 1.3

Linear Displacement Error in the X-Axis

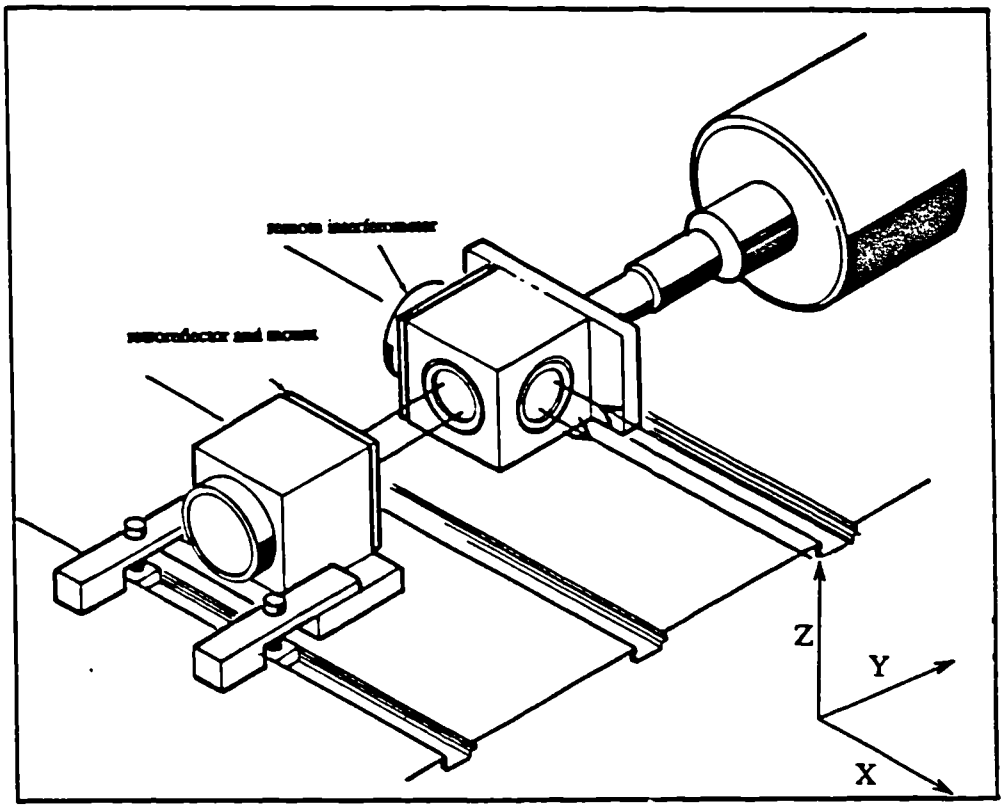


Figure 1.4

Linear Displacement Error in the Y-Axis

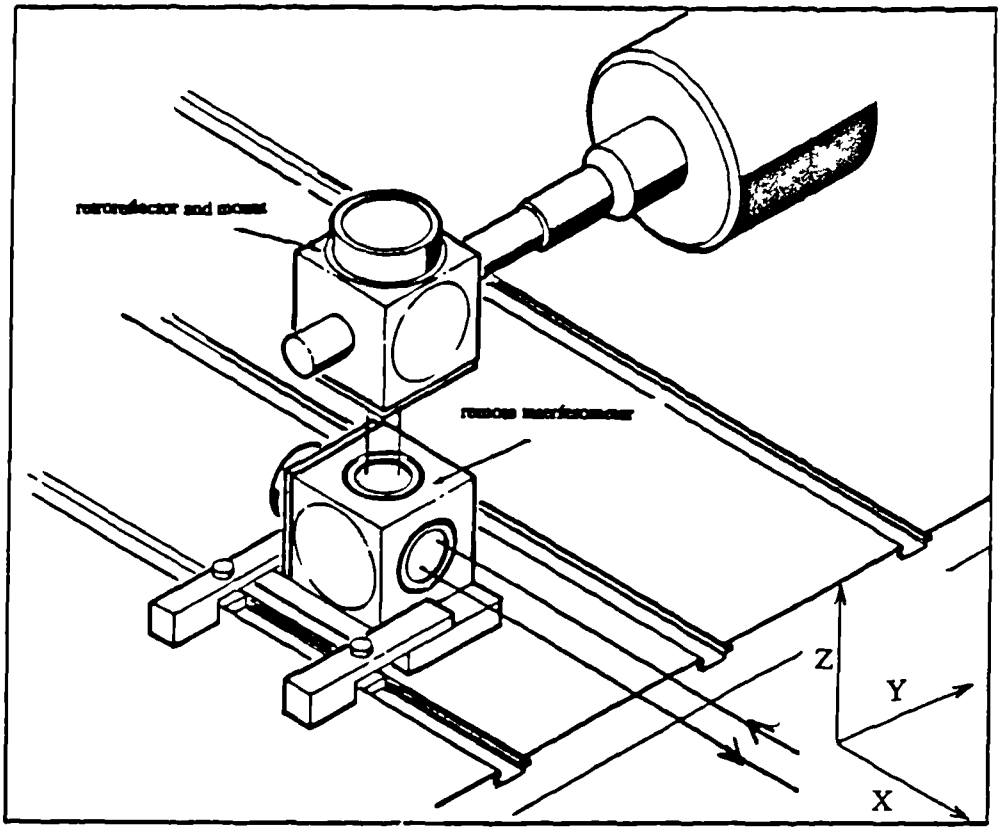


Figure 1.5

Linear Displacement Error in the Z-Axis

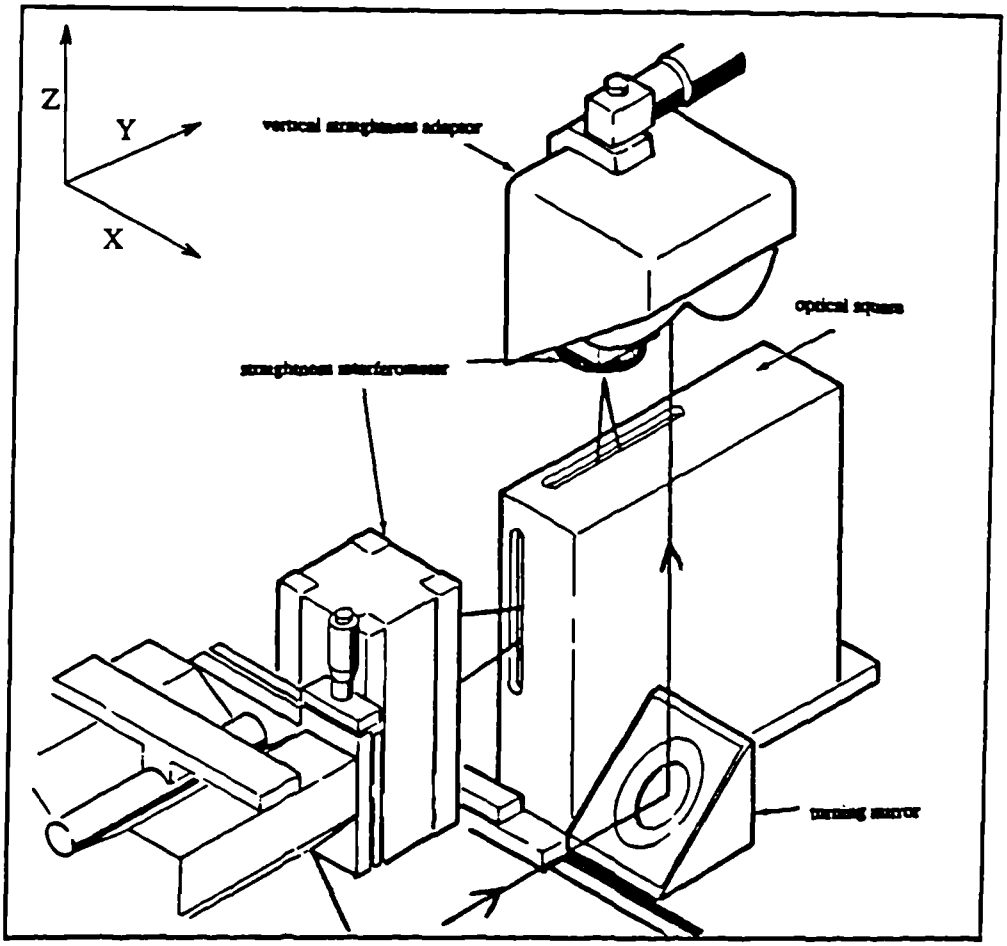


Figure 1.6

Measurement of vertical straightness error in the Z-axis

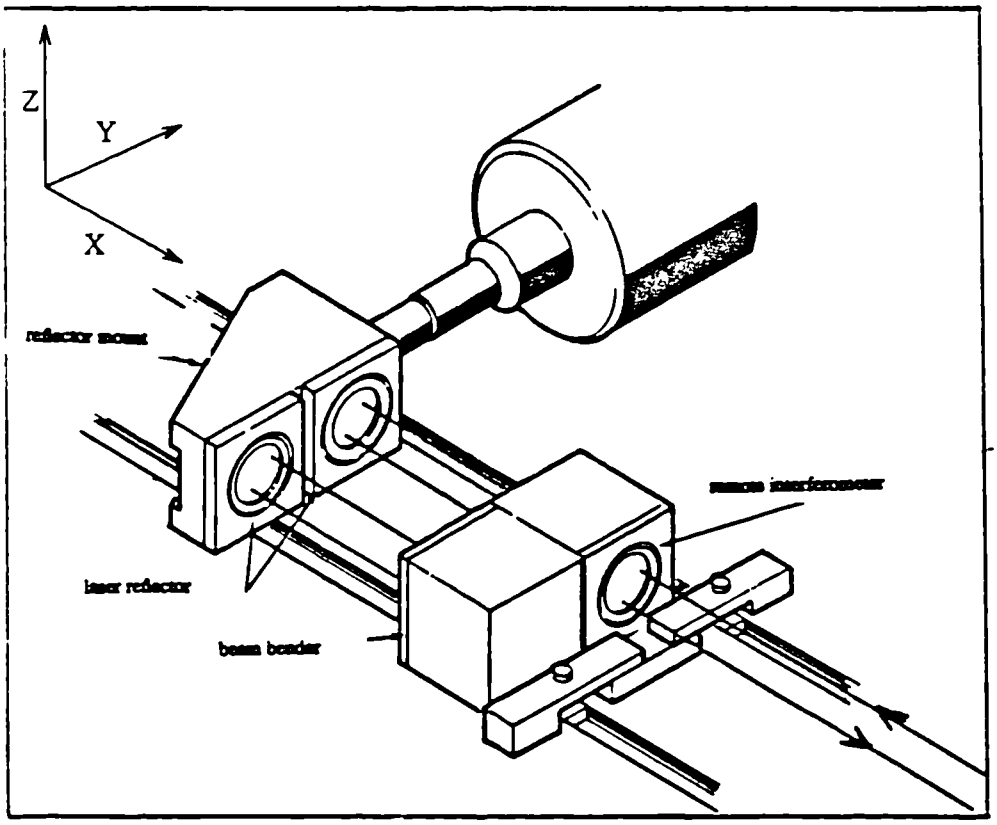


Figure 1.7

Yaw in the X-Axis

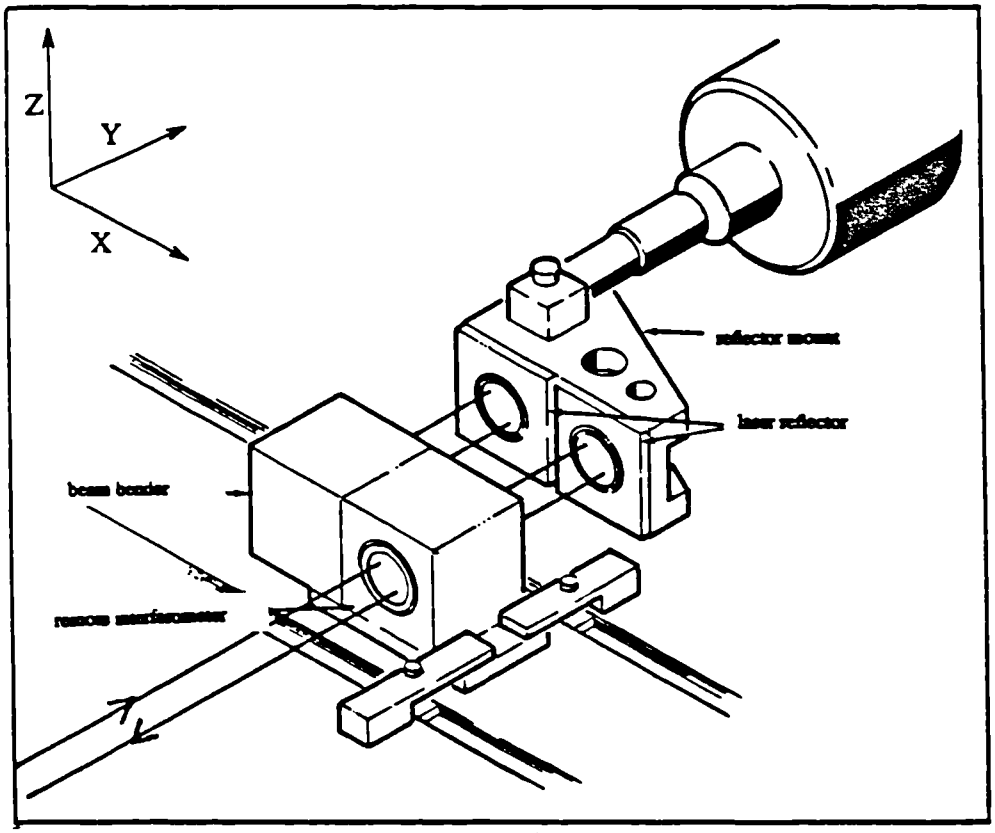


Figure 1.8

Yaw in the Y-Axis

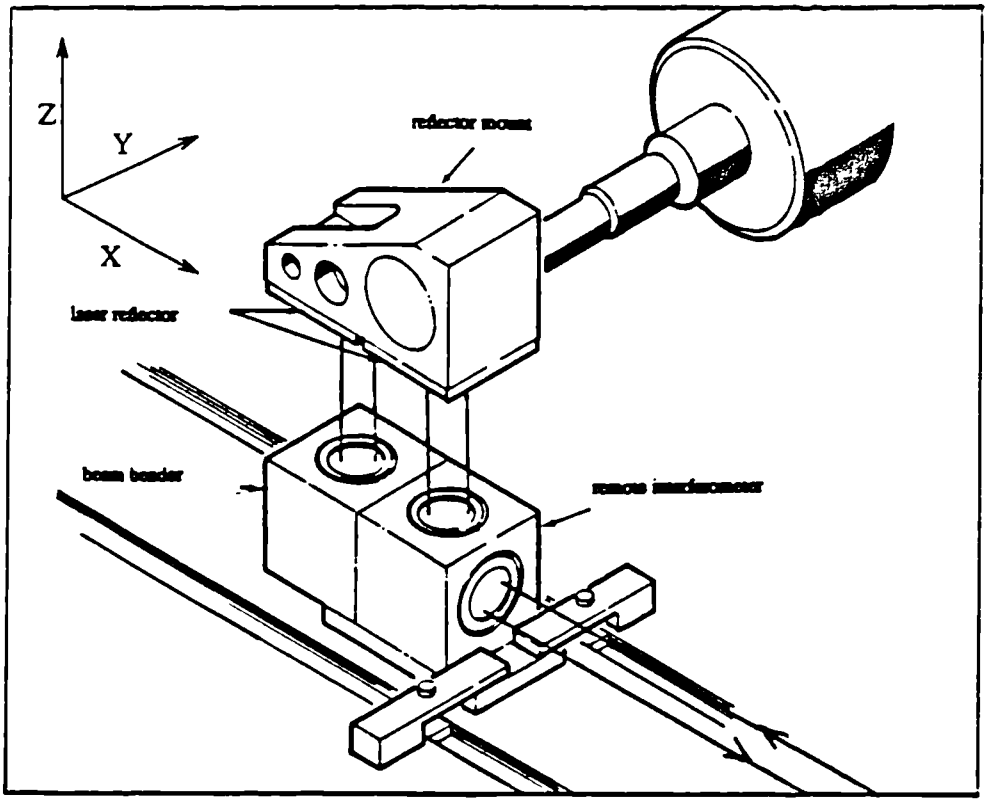


Figure 1.9
Yaw in the Z-Axis

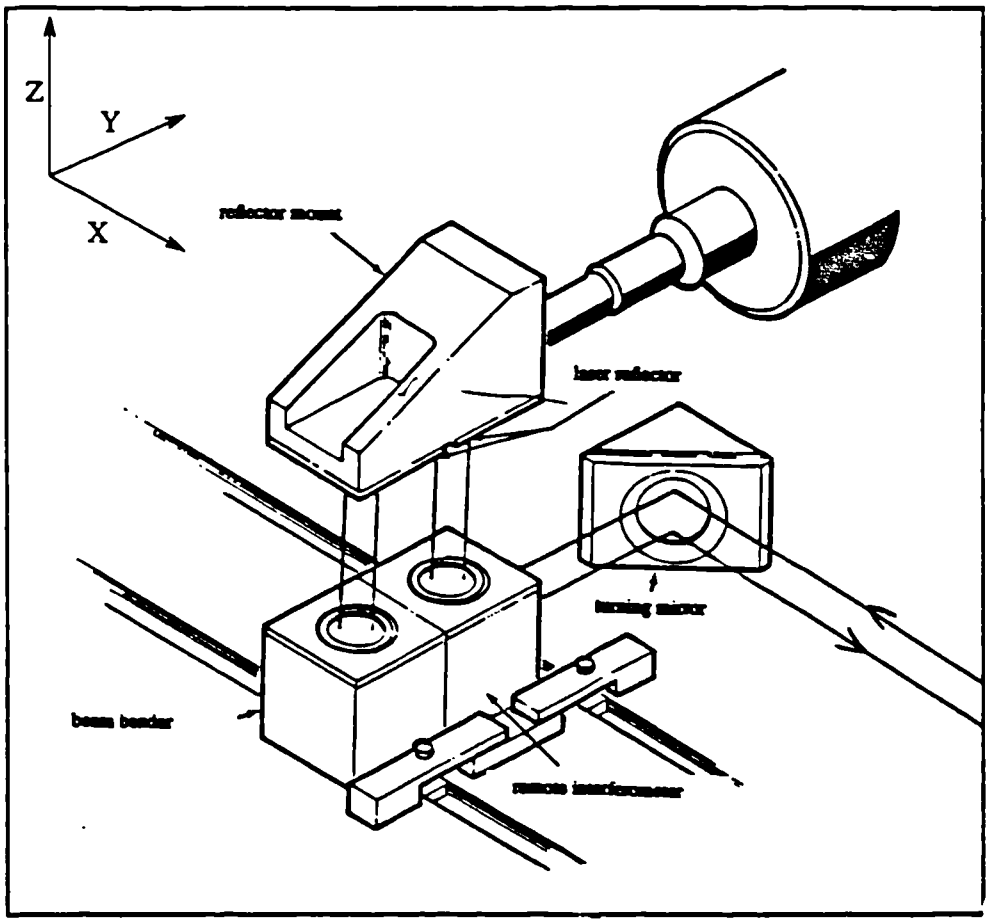


Figure 1.10
Pitch in the Z-Axis

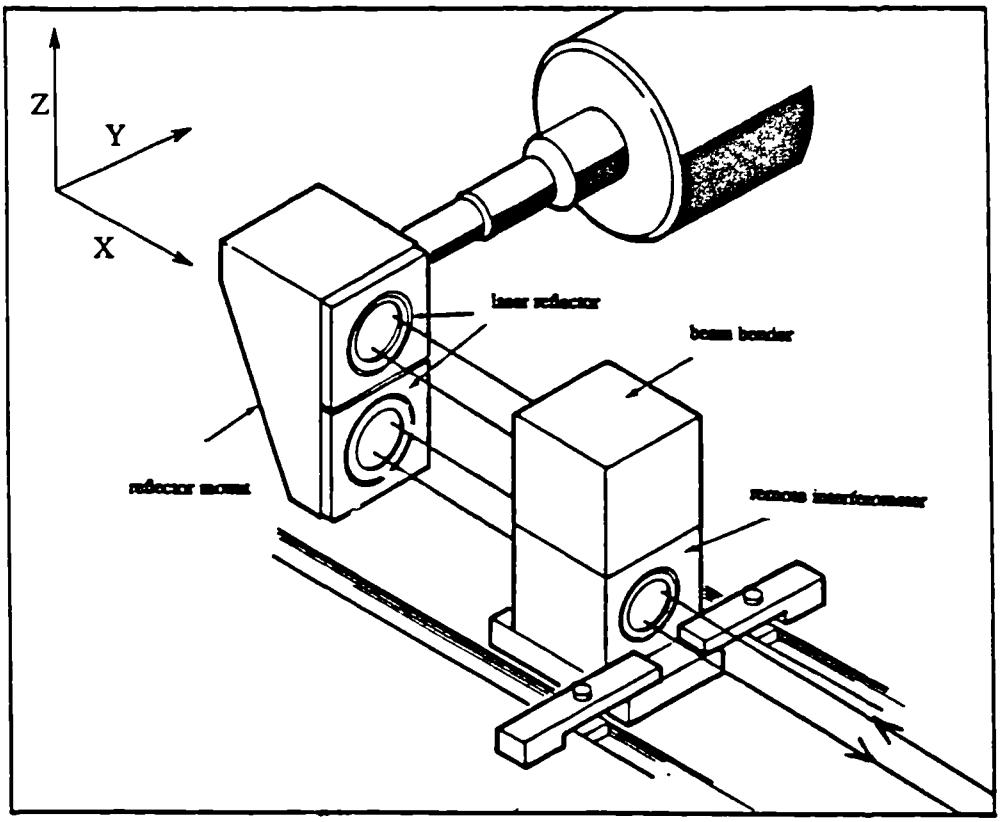


Figure 1.11
Pitch in the X-Axis

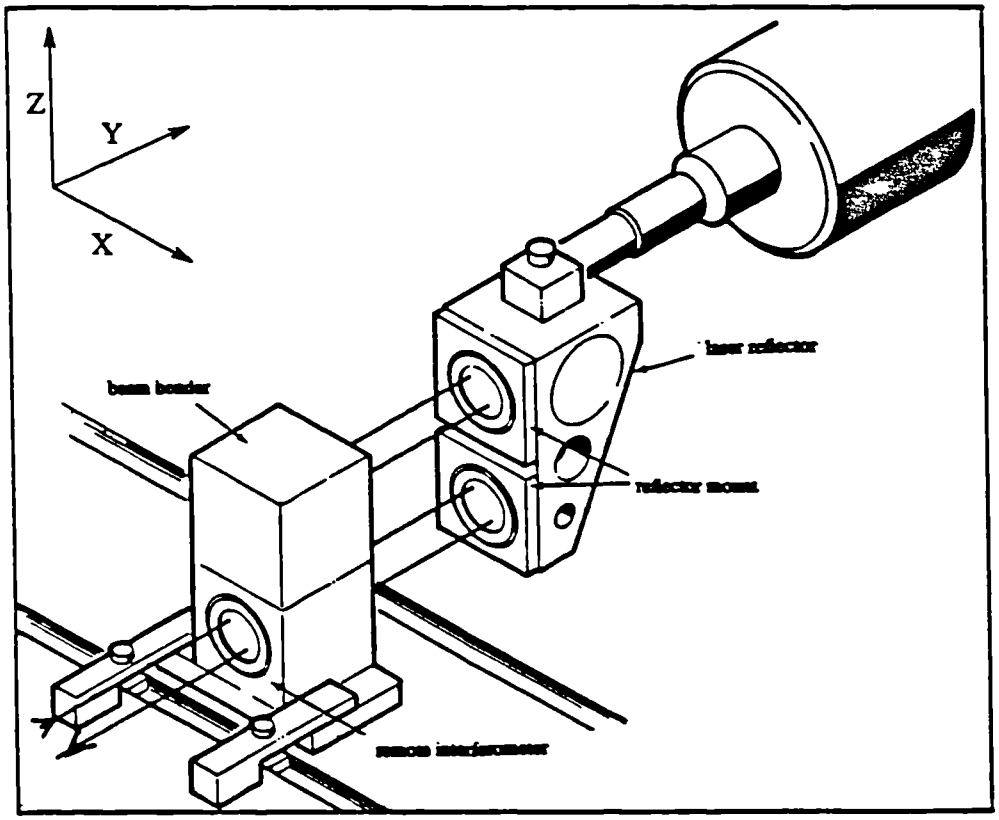
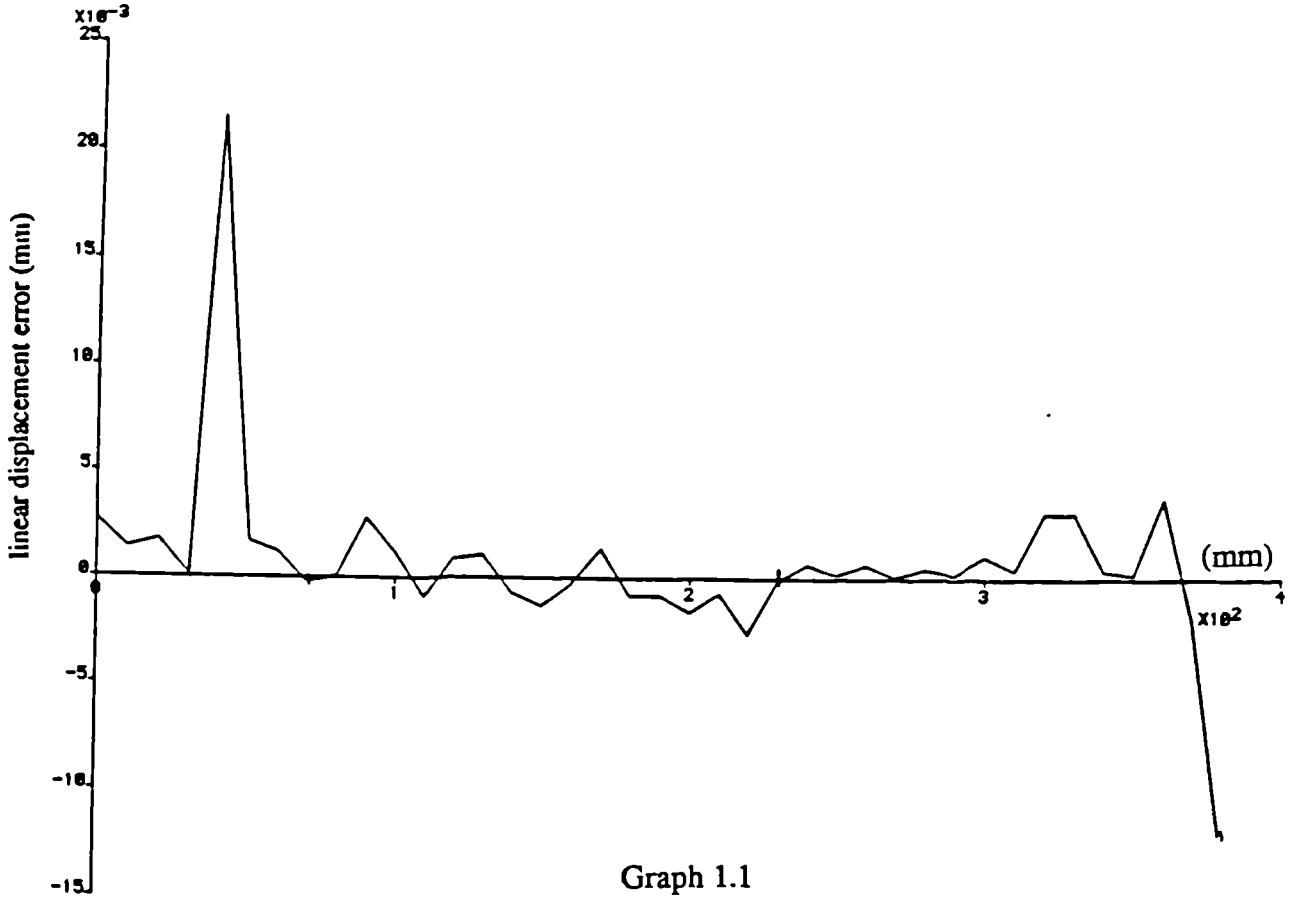


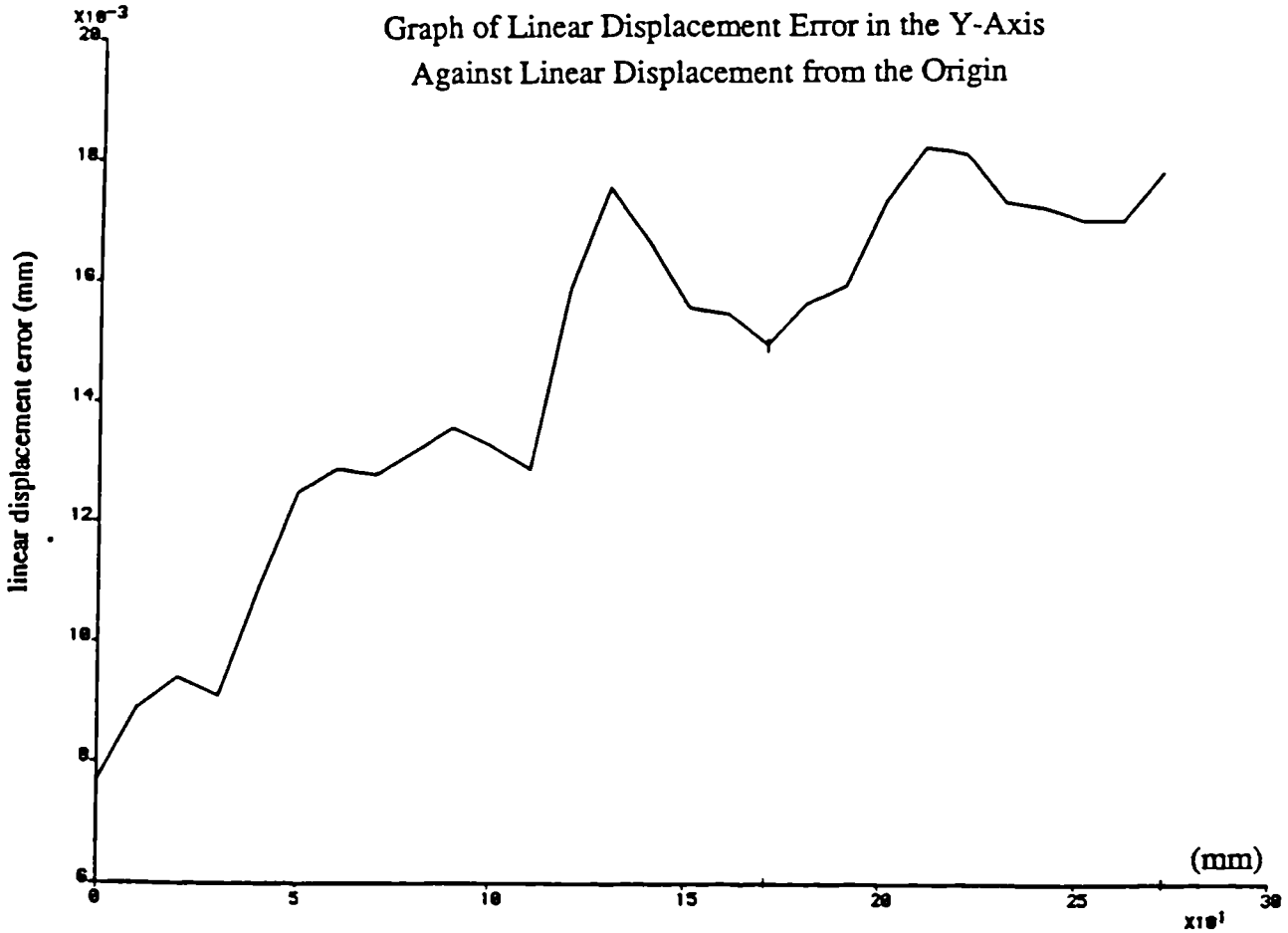
Figure 1.12
Pitch in the Y-Axis

Graph of Linear Displacement Error in the X-Axis
Against Linear Displacement from the Origin



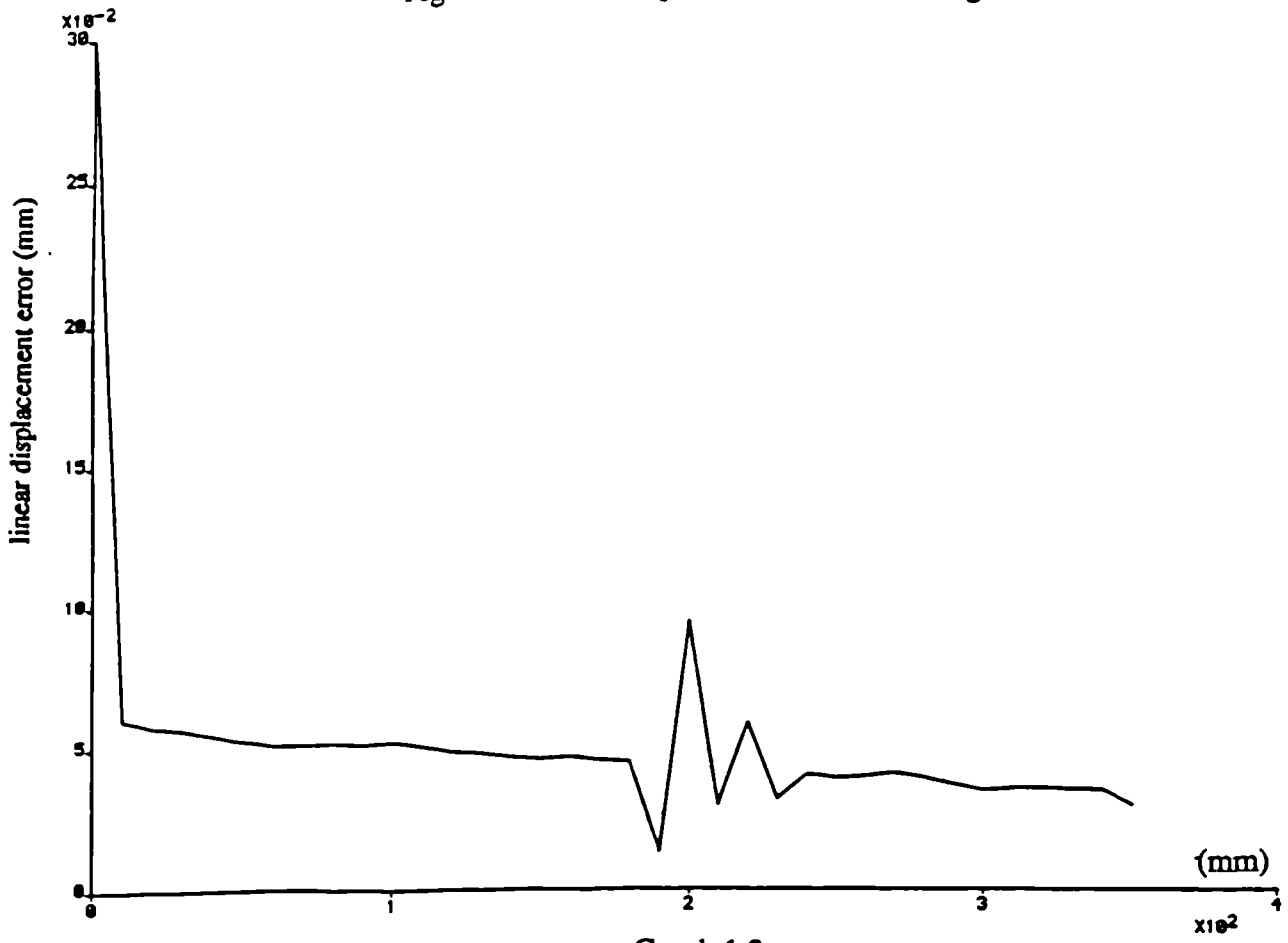
Graph 1.1

Graph of Linear Displacement Error in the Y-Axis
Against Linear Displacement from the Origin



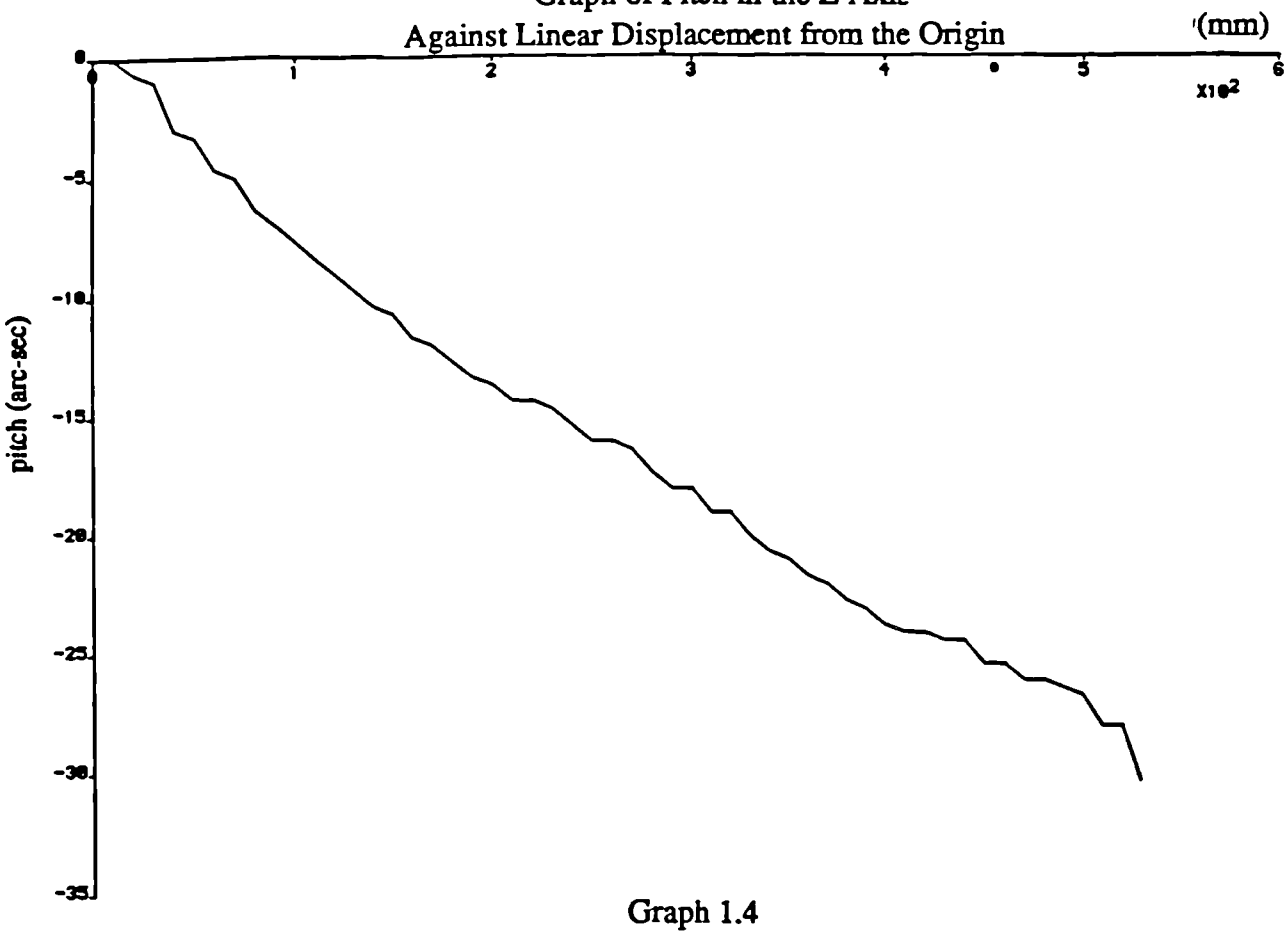
Graph 1.2

Graph of Linear Displacement Error in the Z-Axis
Against Linear Displacement from the Origin



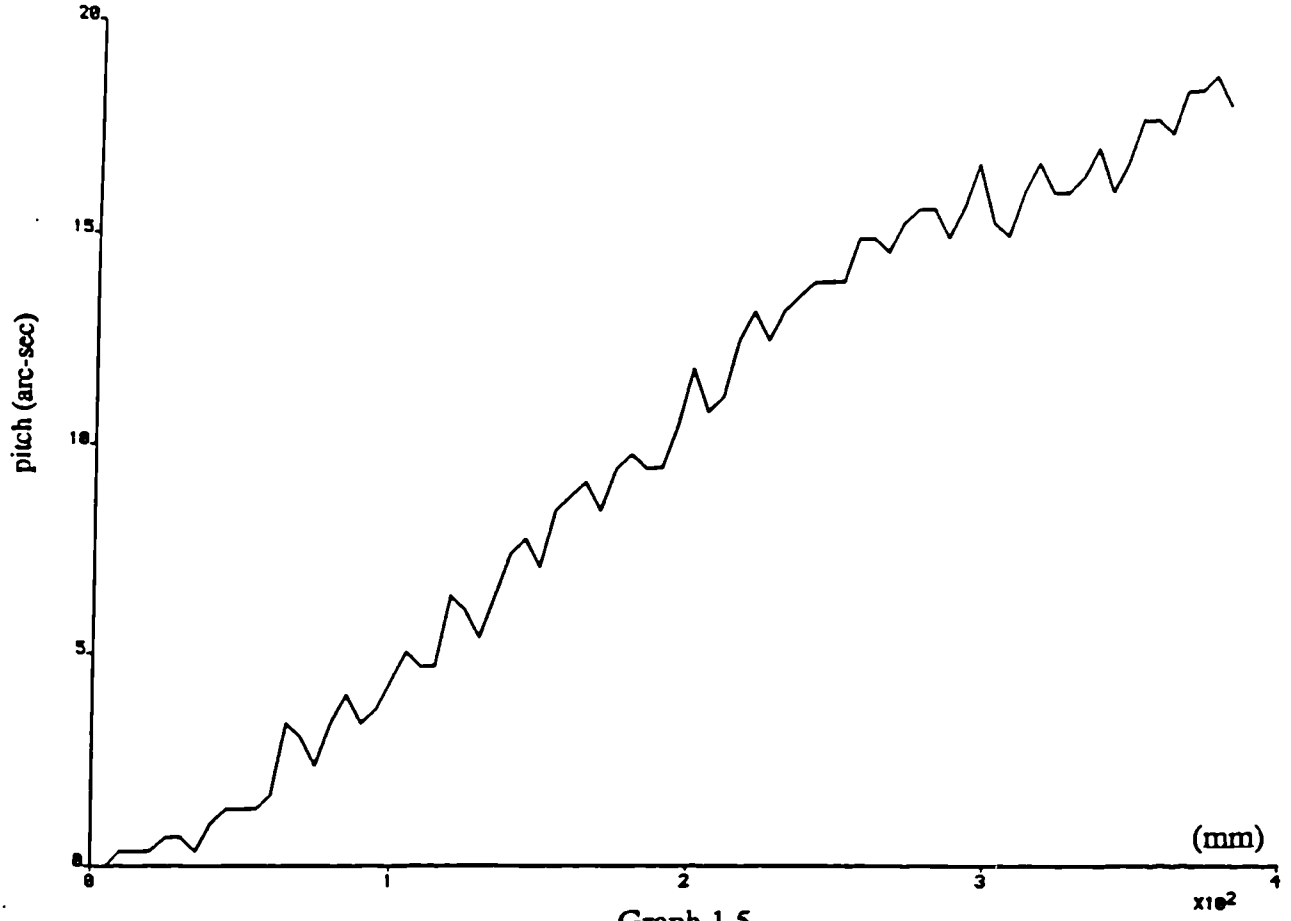
Graph 1.3

Graph of Pitch in the Z-Axis
Against Linear Displacement from the Origin



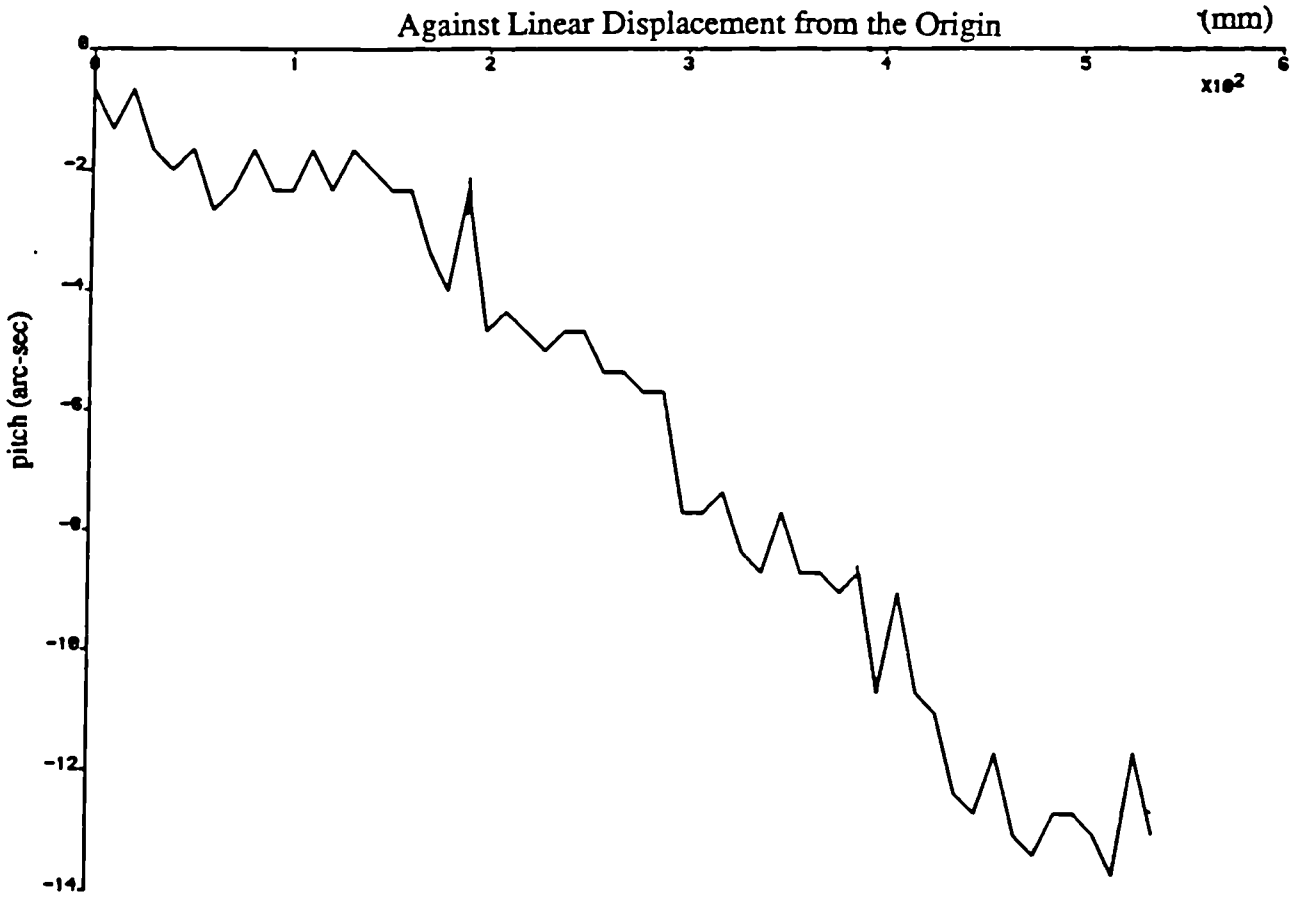
Graph 1.4

Graph of Pitch in the X-Axis
Against Linear Displacement from the Origin



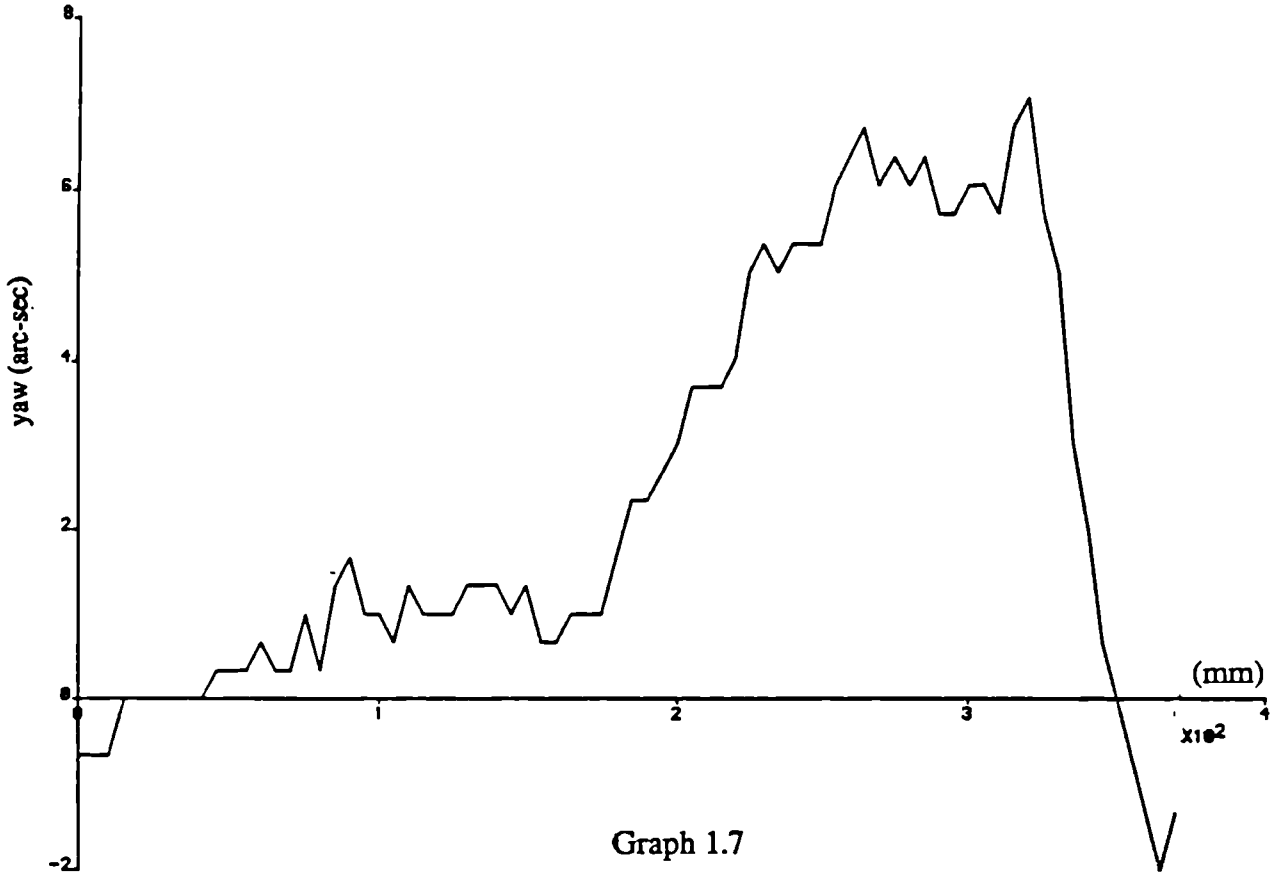
Graph 1.5

Graph of Pitch in the Y-Axis
Against Linear Displacement from the Origin



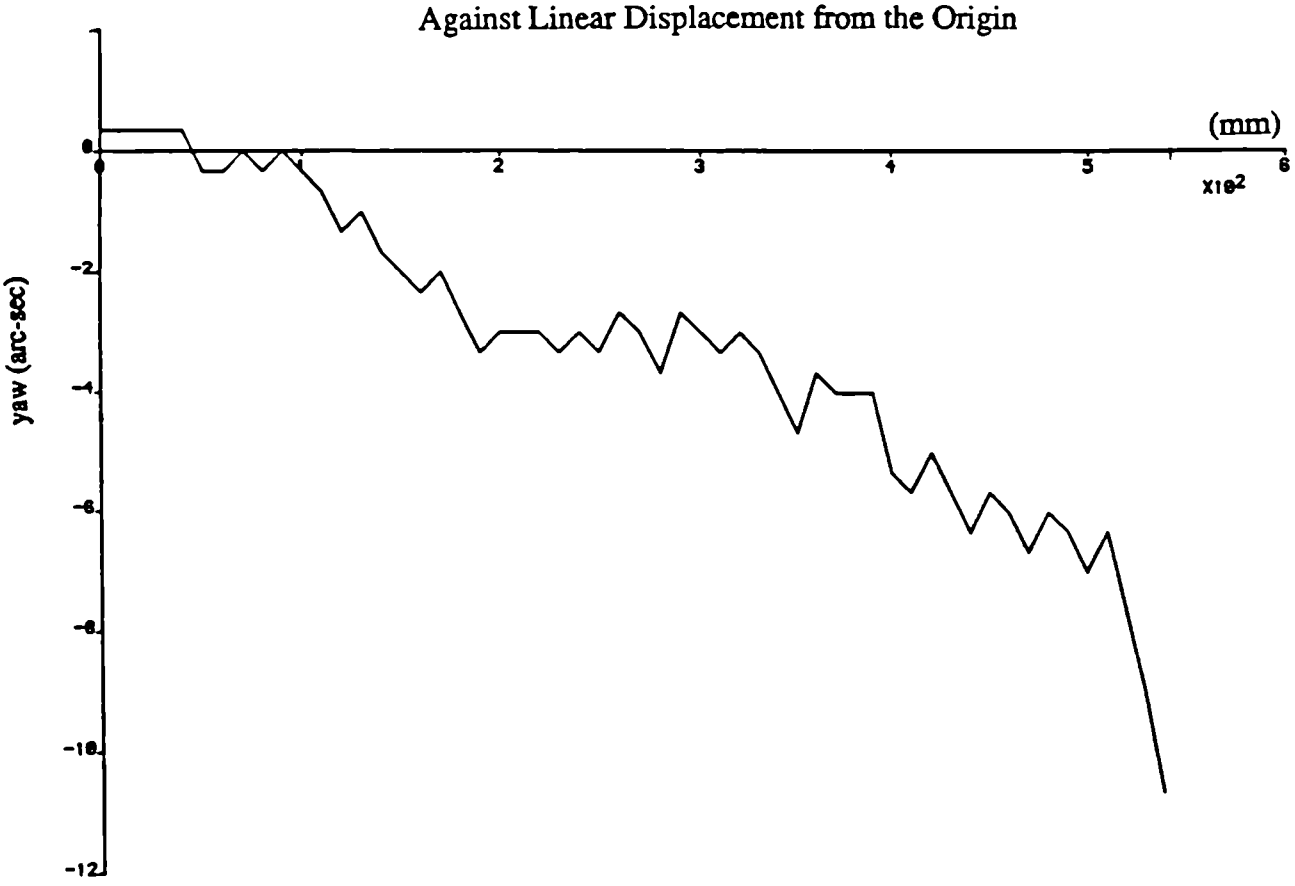
Graph 1.6

Graph of Yaw in the X-Axis
Against Linear Displacement from the Origin



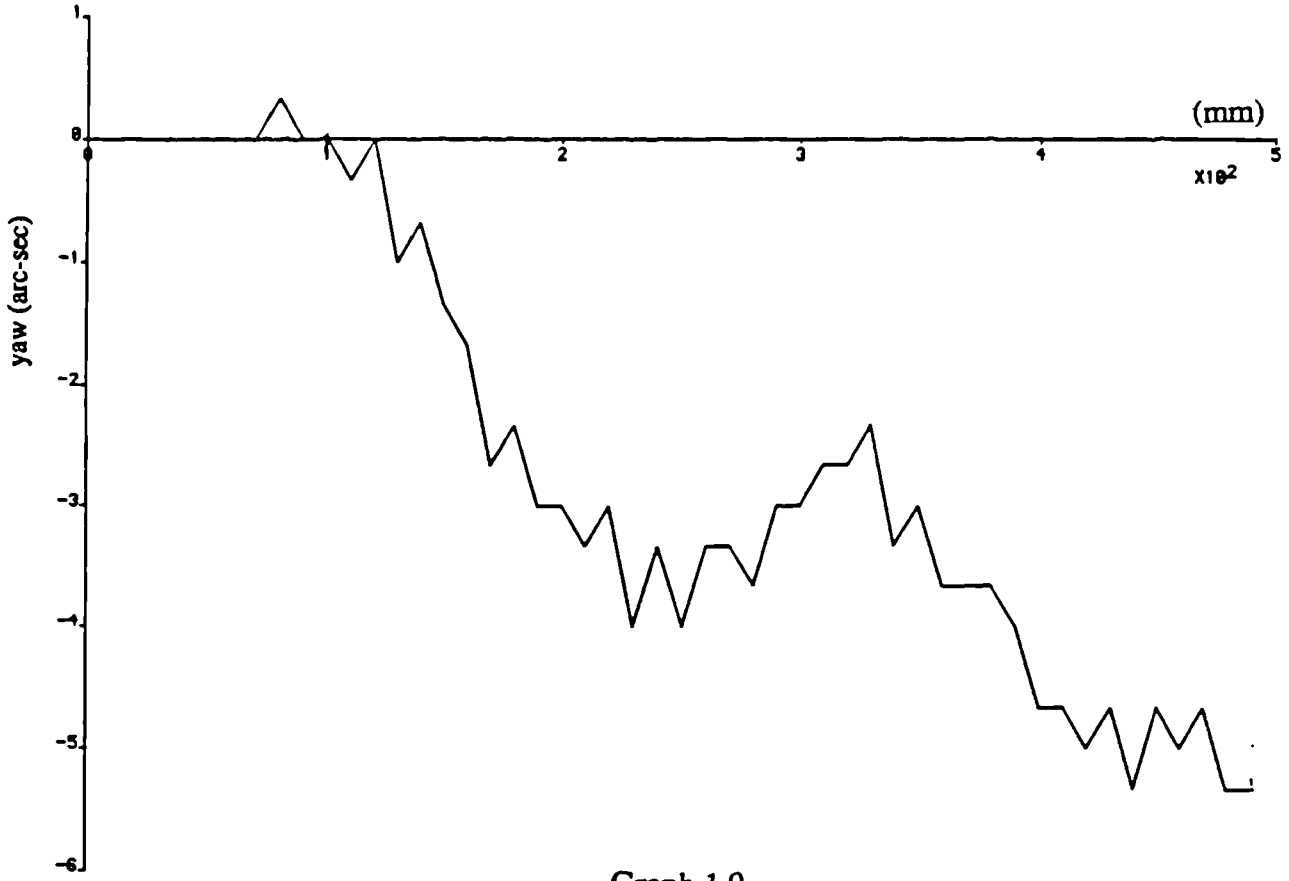
Graph 1.7

Graph of Yaw in the Y-Axis
Against Linear Displacement from the Origin



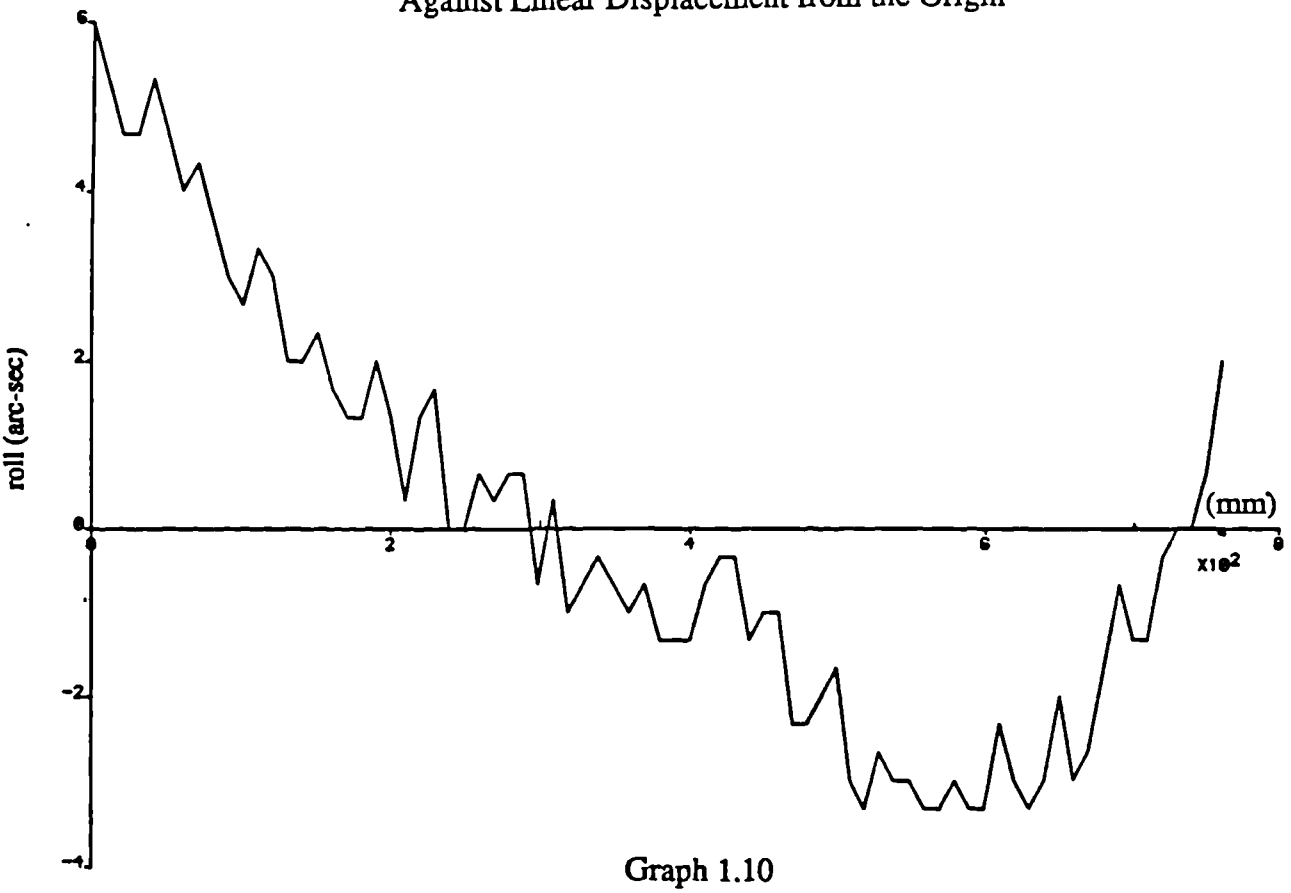
Graph 1.8

Graph of Yaw in the Z-Axis
Against Linear Displacement from the Origin



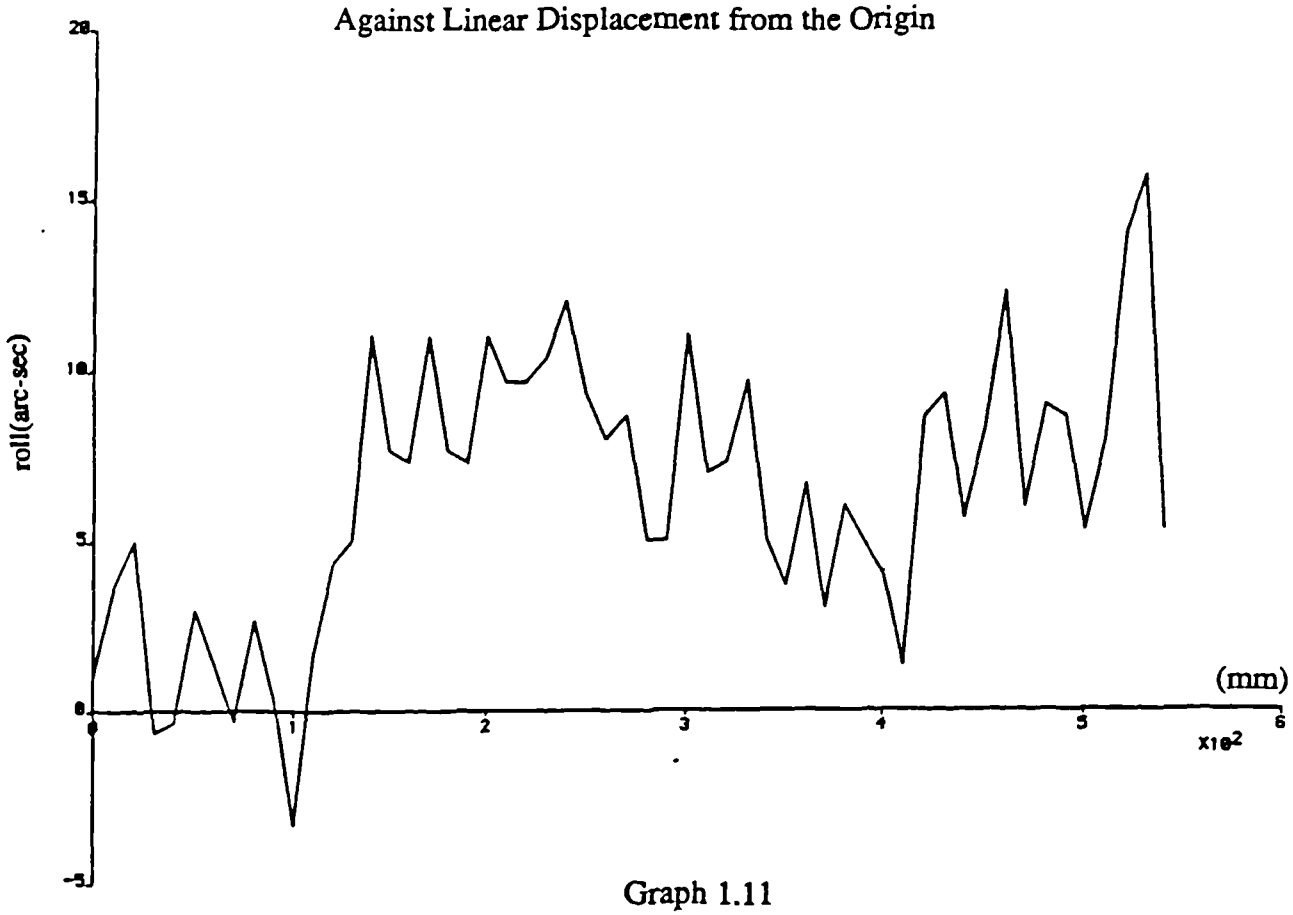
Graph 1.9

Graph of Roll in the X-Axis
Against Linear Displacement from the Origin



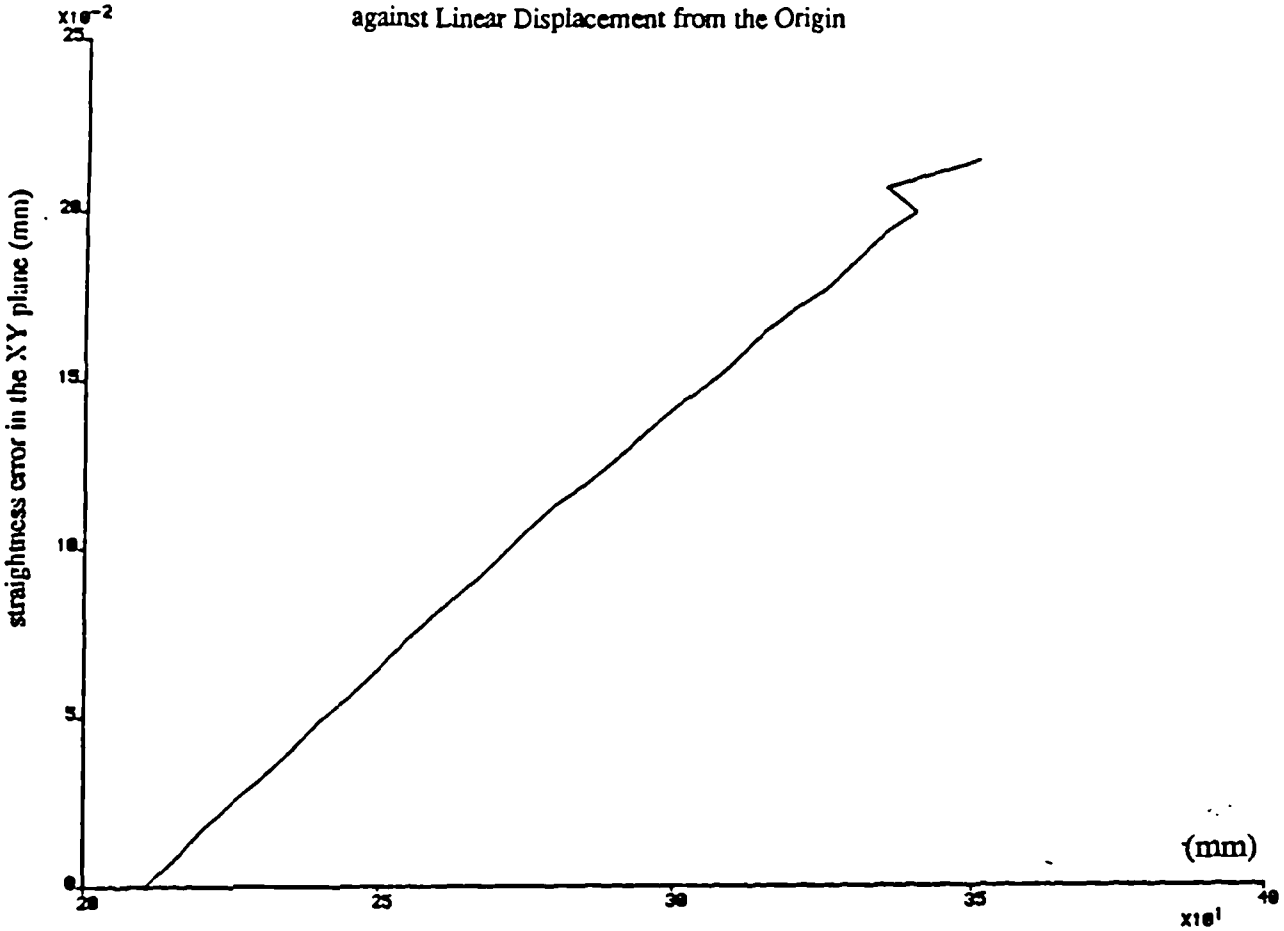
Graph 1.10

Graph of Roll in the Y-Axis
Against Linear Displacement from the Origin



Graph 1.11

Graph of Straightness Error in the XY-Plane
against Linear Displacement from the Origin



Graph 1.12

Approximate Range of Geometric Errors		
Error	Axis	Range
Displ	X	$0 \rightarrow 22 \times 10^{-3}$ mm
	Y	$0 \rightarrow 18 \times 10^{-3}$ mm
	Z	$0 \rightarrow 30 \times 10^{-2}$ mm
Pitch	X	0.00→20.00 arc-sec
	Y	0.75→14.00 arc-sec
	Z	0.00→31.00 arc-sec
Yaw	X	−0.75→8.00 arc-sec
	Y	0.30→11.00 arc-sec
	Z	0.70→5.50 arc-sec
Roll	X	−3.50→6.00 arc-sec
	Y	−4.00→15.00 arc-sec

Table 1.2

The effects of these errors upon inspection routines may be substantialy reduced by the use of calibration data in the CMM software.

3. CMM Calibration

The calibration process should consist of the following steps:

i) Creating a mathematical model of the CMM to describe the interaction of the error components.

ii) Measuring the error components according to the model.

iii) Computing the multidimensional errors and their effects on single measurements and for variables of measuring tasks.

iv) Approving the results with the aid of calibrated measurement standards. (7)

There are basically two approaches to the calibration the CMM; either direct or indirect error correction methods.

The first method refers to the direct measurement of the geometric errors of a CMM by means external to the machine, commonly by the methods described in the above experimental procedure. The second method ascertains the geometric properties of the machine by measuring test bodies at different locations throughout the measuring volume, the CMM itself being the measuring instrument.

Although the second approach to calibration is relatively quick and readliy implementable, as no additional equipment needs to be set up, it is of limited diagnostic value. Consequently, although the first approach results in expensive 'down time', it is necessary when the sources of the geometric errors have to be established. However, indirect methods are invaluable for quick intermediate checking.

3.1. Direct Methods of Calibration.

The systematic error components of a CMM are established by taking individual calibrations of linear and angular error components by the means outlined in 1.3 (and shown in Figures 1.3 -1.12) throughout the measuring volume. There are two ways in which this type of calibration may be performed using a laser interferometer system:

i) The space grid method

ii) The synthesizing method

The space grid method involves length measurements in the form of axis calibrations in order to ascertain positioning errors along the three axes. Straightness, orthogonality and roll measurements enable axis positions to be combined in the form of an error map (8). This particular approach has an advantage over the synthesizing method in that it does not require the sometimes erroneous assumption that the CMM structure behaves as a rigid kinematic body, i.e. the errors in one carriage are entirely independent of the position of the other two carriages, so that effectively the effects of deformation and bending due to the moments induced by the carriages upon each other ignored as are errors due to reversal and random effects.

The synthesized approach combines individual calibrations in the form of a geometric structural model. The calibrated data required for this approach are; pitch, roll, yaw, linear displacement and straightness for each axis.

Although the laser interferometer is a valuable tool for machine calibration, other methods exist for the quantification of errors due to alignment of machine guideways i.e. straightness and squareness errors. Of particular importance for CMMs are methods using straightedges (9).

3.1.1. Processing the Calibration Data

In order to effectively utilise the information obtained by the calibration processes, it is desirable to be able to develop a mathematical model that will predict the value of the error component at any location along the traverses of the axes.

Where the error component describes a relatively simple shape, the fitting of a polynomial may suffice. Often, the degree of the polynomial required to reach any reasonable degree of accuracy may prove far too great for practical purposes. There are two approaches to solving this problem. The first is to transform the independent variable so that the new data is more readily fitted by a polynomial. Such transformations may include bi linear transformation, the taking of logarithms and reciprocals. The second approach is to represent the data as a series of polynomials. In a slightly different sense, transformations are also used in this method. The independent variable again undergoes linear transformation but with the purpose of avoiding ill-conditioning and subsequent loss of accuracy. (10)

Polynomials in data approximation applications usually employ the L_2 (least squares) norm,(the L_∞ , minimax norm is often favoured in function approximation) and the solutions are rarely derived directly, instead such methods as orthogonal triangularisation and singular value decomposition are used. Many ready made packages exist for polynomial and spline fitting, and it is generally advisable to utilise these routines. How-

ever when fitting splines it is vitally important that the knots be placed appropriately.

In general, more knots will be needed in regions where the behaviour of the curve is extreme, and fewer where it is relatively smooth, a sensible strategy for obtaining an approximation is as follows.

i) Position an initial set of knots in accordance with the above criterion.

ii) Obtain an approximating spline s(x) based on these knots.

iii) Examine key parameters such as residuals and the behaviour of s(x) in the regions where there are such features as inflexion points or extrema (also if regions have relatively few data points). At this stage, a graphical form of output, in which the data points, the approximating spline and the knot positions are displayed is particularly useful.

iv) In regions where the approximation is inadequate, introduce additional knots, prehaps after slightly adjusting the position of existing ones. There may be regions where the approximation follows the data values so exactly that the spline has oscillations with amplitudes of the order of the noise level of the data, or even greater, in which case a number of knots should be removed adjusting any as necessary.

v) Repeat as necessary from ii)

Two or three passes through this procedure is usually sufficient (10).

3.1.2. Analysis of the Calibration Data for Measuring Compensation

3.1.2.1. The Space Grid Approach.

Repeated positional, volumetric and angular readings are taken in forward and reverse directions, the mean value of the errors at each of several targets along the axis is calculated and represents the systematic error components. The difference between these two means is the hysteresis provided that there are no transient effects. Repeatability at each target position is evaluated in terms of the standard deviation about the mean for-

ward motion error. These values are combined in the form of an error map, each reference position being specified with respect to the origin and orientation of the axes.(7)

3.1.2.2. The Synthesized Approach.

As mentioned, all components in the CMM are regarded as rigid bodies, i.e. random deviations, reversal errors and elastic deformations are disregarded. The dispensation of elastic deformation effectively results in the criterion that the errors, either linear or rotational, on one carriage of the CMM are entirely independent of the positions of the remaining two carriages, and can therefore be measured seperately for each carriage. In total, twenty one error components are needed for a full calibration.

3.1.2.2.1. Interaction of Errors

When the individual error components of a CMM are known, they must be combined in such a manner that the position of the touch probe with respect to the origin of the instrument coordinate frame can be calculated. The following technique relies upon the assumption that the CMM behaves as a rigid kinematic body and is independent of machine configuration.

Consider a configuration of CMM consisting of three axes of measurement, X, Y and Z. The probing system is fixed upon the third carriage, which is mounted upon the second carriage, which in turn, is mounted upon the first.

Each of these traverses has an uncorrected travel vector, which when added describe the position of a reference point within the probing system. This reference point is displaced from the position of the touch probe by an additional vector; the probe vector. These vectors are regarded as being in a rectangular coordinate system X_R, Y_R and Z_R, which is aligned to the X axis and XY plane of the CMM.

The motion of each carriage is considered in turn. Begining with the motion of the third carriage, to its displacement vector is added its linear displacement error, the resultant is transformed by a matrix Φ containing the direction cosines of the X, Y and Z axes

with the rectangular coordinate frame. The probe vector is then transformed by multiplication of a matrix containing the rotational errors of that carriage, thus, in total, giving a resultant vector for the motion of carriage three. Movement of the second carriage is then considered, the resultant vector for carriage three is subsequently transformed by the rotational errors in carriage two, further, the vector sum of the displacement of carriage 2 and its linear displacement error (which is also transformed by the matrix Φ), giving a resultant vector for the movement of carriages two and three. Finally the motion of carriage one is considered. The resultant vector for the motion of carriages two and three is transformed by the rotational errors for carriage one, and the addition of a vector consisting of the sum of the displacement and linear displacement error of carriage one, (which is again transformed by Φ) gives the resultant equation for the errors in the total motion of the CMM traverses.

3.2. Indirect Methods of Calibration

Indirect methods of calibration do not generally require means of measurement extrinsic to the CMM, and can therefore be largely automated, and consequently reduce 'down time' of the CMM when compared to direct methods. Indirect methods evaluate the geometric errors of a CMM by measuring test bodies at positions throughout the measuring volume of the CMM. The systematic machine errors from these measurements depend upon the size and location of the test bodies and are derived by combining the results, with the aid of suitable mathematical models.

3.2.1. Principle of Indirect Methods

Geometric errors transform the ideal, perfectly square, straight lined coordinate system into a curvilinear coordinate system, which may be characterised, approximately, by a number of small local coordinate systems which are oblique angled, straight lined and have different scaling in each of its axes.

However, these local systems must be transformed into orthonormal coordinates, to

achieve this end, it is necessary to determine the squarenesses and the longitudinal errors of the axes. There are many ways by which to asses these errors using a variety of test bodies and mathematical models, two contrasting methods will be illustrated in order to give some idea of the diversity of methods that may be used.

3.2.2. Evaluation of a Local Reference System by means of a Kinematic Ball Bar

The relationship between a point P_{Ai} measured in the local coordinate system and the coordinate P_{Ri} in the orthonormal system may be represented by

$$P_{Ri} = M_i P_{Ai} \tag{1.8}$$

where

$$M_i = \begin{bmatrix} i+l & -\Phi_{xy} & -\Phi_{xz} \\ 0 & l+m & -\Phi_{yz} \\ 0 & 0 & l+n \end{bmatrix}_i \tag{1.9}$$

The subscript i denotes the number of the local coordinate system ($1,k$), and l,m and n are the linear scale corrections of the X, Y and Z axes respectively. Φ_{xy}, Φ_{xz} and Φ_{yz} are the squareness errors between the X and Y, X and Z, Y and Z axes respectively.

The kinematic ball bar equipment consists of two balls fastened to a bar with adhesive, these balls are mounted in sockets by means of permanent magnets. One magnetic socket assembly is mounted onto the rotary table of the CMM and the second replaces the probe on the machine quill (12) a nominally circular path. A self centring probe with three small balls picks up the coordinates of the probe centre. Deviations from circularity, when having undergone the appropriate manipulations, can give longitudinal and squareness errors.

If purely longitudinal errors are present, then the locus of the ball centre follows the form of an ellipse whose main axes are parallel to the corresponding axes of the machine. If only errors of rectangularity are present, the main axes of the ellipse are inclined at 45 degrees to the axes of the machine. The presence of both types of error results in the angle of the main axes of the ellipse being variable in proportion to the ratio of the sizes

of the elements of the two error types (11).

3.2.3. Calibration of Coordinate Measuring Machines Using Space Frames

The objective of using space frames is to identify trends in the error of the measuring machine system (the combined movements of the machine and probe), and therefore indicate where rectification needs to be made. This process is achieved by taking measurements on an artefact (often a carbon fibre frame of some description supporting steel balls), and moving this artefact in a well defined sequence throughout the measuring volume of the machine. The advantage of such a process is that it requires no time consuming leveling or aligning procedures.

The faults in the machine system are located by comparing measurements taken from the frame at each of its locations throughout the operational volume with the frame calibration data by the following procedure:

i) the machine errors are deemed to be small (negligible) if both the differences of the means and the standard deviations of the measured points are small.

ii) If the differences of the means are large and the standard deviation small, the machine errors will be significant, in a similar sense (+ or -), and will probably be due to geometric or linearity errors.

iii) If the differences in the means are small, and the standard deviation large, there will be large machine errors of either sense (+ or -). This type of error often originates at the probe, although if found in one small area of the measuring volume, the error may indicate an error in the guide beam.

iv) If the differences of the mean and the standard deviation are large, a combination of errors are implied, or the whole system may require considerable attention. (13).

4. Probes

CMM probes are generally available in two main types. 'Hard' probes are often

shaped as either tapered plugs for establishing the location of the centre of a bore, or as a ball for establishing surface location, these probes have to be triggered manually. However, the alternative 'touch trigger' probes are far more popular, and operate by means of providing a voltage jump when contact with the workpiece is made.

4.1. Touch Trigger Probes

The touch trigger probe is still the most widely used contact mechanism on CMMs. The stylus is the part of the probing mechanism that contacts the workpiece under inspection, and is screwed directly into the probe. It is the deflection of this stylus that triggers the probe mechanism. The trigger mechanism may be considered as a very repeatable re-seating mechanism. In order to unseat this mechanism to produce the trigger signal, a certain force must be overcome, the force in question is commonly known as the trigger force.

As the probe approaches the surface of the workpiece under inspection, contact is made by the stylus tip, however, the probe continues to travel, deflecting the stylus and trigger mechanism until the trigger point is reached. The duration of the travel between initial contact and the trigger point is known as pre-travel. Pre-travel itself is not an error constituent as datum setting procedures can effectively eliminate it.

As a consequence of the mechanics of the trigger mechanism, the trigger force and pre-travel will often vary with the angle of deflection. The variation in trigger force reaches a maximum at three points spaced equiangularly (120 degrees) in the 360 degree probing range. This phenomenon is known as 'probe lobing'. The effects of pre-travel variation can be further compounded by stylus length, as shown by the following approximation (14):

$$Pre\text{-}travel\ variation \propto stylus\ length^2 + C \tag{1.10}$$

Where C is the pre-travel variation with a stylus of zero length.

The repeatability of the probe (the ability to trigger at the same point each time), is

affected by the degree of pre-travel variability in such a direct way, that repeatability is often expressed in terms of pre-travel variability.

For example, the measure of repeatability of the Renishaw touch trigger probe is, for convenience, quoted uni-directionaly with a 95% confidence interval that the repeatability will be within $\pm 2 \times$ the standard deviation of the pre-travel from its mean value.

From the premise that a significant proportion of the probe error is directly allied to the trigger force, a logical progression would seem to be to recommend that the trigger force be lowered, however, if this suggestion were effected, CMM acceleration or vibration may result in false triggering.

The trigger force combined with the stylus length can also contribute to probe hysteresis, which is a systematic error due to the direction of the proceeding trigger. However, this phenomenon may be compensated for by sequential datuming. The accuracy of the probe can be affected volumetrically by the sphericity of the stylus ball tip and the stiffness of the probe (15).

4.2. Dynamical Probe Errors

The velocity at which the probe stylus approaches and contacts the workpiece can also have a profound effect on the measuring error, more specifically, due to the deformation of the stylus and the degree of over-travel error due to the delay between the trigger point being reached and the opening of the microcontacts of the location recording device.

The deflection of the stylus can be calculated in terms of its length, the maximum impact force and its rigidity characteristics (15).

4.3. Laser Scanning Probe System

For applications in high speed scanning of yielding materials, measurements of complex surface forms, such as turbine blades, static and dynamic non-contact gauging in high speed production monitoring; the use of the laser scanning probe is usually far

more effective.

The laser scanning probe is based on a GaAlAs laser diode, infra-red radiation is focussed by a lens system to a spot approximately 25μm in diameter at a stand off distance of 20mm from the probe body.

Some of the scattered light from the component is collected and focussed by a second lens onto an analogue silicon based photodetector. The position of the focussed spot on the detector is directly proportional to the displacement of the component surface with respect to the probe axis. Approximately fifty readings are made per second. High intensity ratio detection permits the inspection of a wide range of colours and surface finishes, while trigger and synch outputs allow synchronisation with the CMM. Accurate edge detection is usually assisted by continuous light level feedback combined with a small spot size (16).

5. Conclusion

The accuracy and therefore the efficacy of CMMs can, indeed be greatly enhanced by quantifying the error sources detailed above, and providing relevant compensation routines for implementation within the CMM software. A substantial degree of research is currently being performed in order to refine and improve upon the procedures discussed.

Although the machine errors measured on the 'LK Micro-Four' are relatively large, the application of error compensation software can result in these errors being reduced to a quarter or one fifth of their original value (11). It is upon this criterion that subsequent comparisons will be made with other CMM errors.

References

[1]:J.A. BLAND & R. DAVIDSON (1985): "Coordinate Measuring Machines'; Quality Assurance, September 1985, Vol. 11 No. 3.

[2]:G. SCHAFFER (1982): "Taking the measure of CMMs"; American Machinist, October 1982.

[3]:J. CHEN & Y.F. CHEN (1987): "Estimation of Coordinate Measuring Machine Error Parameters"; IEEE CH2413-3

[4]:M.A. DONMEZ, D.S BLOMQUIST & R.J. HOCKEN (1986): "A General Methodology for Machine Tool Accuracy Enhancement by Error Compensation"; Precision Engineering, Vol 8, No. 4.

[5]:W.T. ESTLER (1985): "High Accuracy Displacement Interferometry in Air"; Applied Optics; vol.24, no. 6.

[6]:R.J. HOCKEN (1979): "Lasers for Dimensional Measurement"; Annals of the CIRP, Vol 28.

[7]:M.BURDEKIN, B. DI GIACOMO & Z.XIJING (1985) :"Calibration Software an Application to Coordinate Measuring Machines"; NPL Conference Proc. on Software for Coordinate Measuring Machines.

[8]:K. BUSCH, H. KUNZMANN & F. WALDELE (1985): "Calibration of Coordinate Measuring Machines"; Precision Engineering, Vol 7, No.3.

[9]: M. BURDEKIN & F. KOENIGSEBERGER: "Accuracy Assessment and Analysis of Machine Tools"; Proc. Machine Tool Industry Research Association, One Day Seminar.

[10]:M.G. COX: "A Survey of Numerical Methods in Data and Function Approximation"; The State of the Art in Numerical Analysis ; Jacobsen, pp 627-668.

[11]:H. KUNZMANN & F. WALDELE (1985): "Two Numerical Error Correction Methods for Coordinate Measuring Machines"; NPL Conference Proc. on Software for Coordinate Measuring Machines, 25th September 1985.

[12]:H. KUNZMANN & F. WALDELE (1983): "On Testing Coordinate Measuring Machines with Kinematic Reference Standards"; Annals of the CIRP, Vol 32.

[13]:A. HARVIE (1986): "Factors Affecting Component Measurement on Coordinate Measuring Machines"; Precision Engineering, Vol 8, No.1.

[14]:Renishaw Metrology Ltd. (Touch Trigger Probes), Wotton under Edge.

[15]:L. NAWARA (1984): "The Investigations on Selected Dynamical Phenomena in the Heads of Multi-Coordinate Measuring Devices"; Annals of the CIRP, Vol.33.

[16]:Renishaw Metrology Ltd. (OP2 Laser Scanning System), Wotton under Edge.

CMM Software.

1. Introduction.

One of the primary advantages of coordinate measurement is that the data is in a form which has great potential for subsequent processing. From this data it is possible to calculate the form and position of geometric features and their interrelationships. CMM manufacturers have siezed the opportunity to maximise this potential by equipping most CMMs with computers containing a variety of geometric fitting subroutines, capable of performing relatively complex inspection procedures that may be brought into action through a simple keyboard command or menu selection

This chapter introduces the concepts behind the algorithm design including the relatively new developments based upon functionally significant considerations. Some of the inadequacies, ambiguities and potential sources of inaccuracy are identified with a view to finding a way forward in the fundemental improvement of CMM software.

1.1. Key Terms

1.2. Coordinate Reference Frames

The following chapter will refer to different frames of reference used in Metrology. There are three frames in common use; component, instrument and chart coordinates.

Component coordinates describe the system in which the component exists, such that each point on the component can be related to another independently of the orientation of the component in space. However, in order to measure the component and calculate the relevant parameters it is necessary to express the component geometry in terms

of the coordinate system existing in the machine, i.e. instrument coordinates. The transition from component to instrument coordinate frames involves quantifying the alignment between the two frames, as yet it is impossible to achieve this with absolute accuracy, so at this stage an error due to alignment inaccuracy is often introduced.

The instrument may then perform further transformations on the component data, such as magnification and radius suppression to produce a readable 'hard-copy'; the result is said to exist in chart coordinates (1).

1.3. Residuals

Residuals may be defined as that value that expresses the difference between a data point and the reference figure. Several criteria exist for expressing residuals, the most usual being linear and normal.

2. Coordinate Measurement Procedure

The current basis for assessing components by coordinate measurement generally adheres to the following procedure;

(i) The instrument and component coordinate frames are aligned

(ii) The datum framework is established

(iii) A set of reference figures is built within this framework

(iv) Tolerance zones are set upon these reference figures

(v) The data are checked for containment within these zones.

The data are used for two purposes; to establish the datum structure according to a predefined optimisation strategy, and as the means of measuring deviation from nominal form, but ideally, a set of data should be collected in order to establish the datum framework from which to assess other data. The strategy for creating the datum should depend upon it's intended use, i.e the function of the component (2).

2.1. Alignment of Instrument and Component Coordinate Frames

When fitting coordinate data to a reference figure residuals are best expressed, in the absence of function related conditions, normal to the reference figure. In order to achieve this, the instrument and coordinate frames must be aligned.

To illustrate, consider the alignment process for straightness measurement by least squares analysis. With reference to fig 2.1, in instrument coordinates, the residuals from the straight line representing the nominally straight side of component A, can be expressed as follows:

$$e_i = y_i - (bx_i + b_o) \tag{2.1}$$

In order to transform the residuals so that they are normal to the reference, they must be expressed in the following manner:

$$s_i = (1 + b^2)^{-1/2}(y_i - bx_i - b_o) \tag{2.2}$$

So that the least squares criterion now requires the following minimisation:

$$S = \sum(s_i)^2 = (1 + b^2)^{-1}\sum(e_i)^2 \tag{2.3}$$

(3)

2.2. Establishing the Datum Reference Frame

A datum can be described as a point or feature that indicates the origin of a dimensional relationship between a toleranced feature and a designated feature or features on a part, the designated feature serves as a datum feature, whereas it's true (ideal)geometric counterpart establishes the datum. However, true geometric measurements cannot be made from this true geometric counterpart, as it is purely theoretical, but a datum is assumed to exist in, and be simulated by the CMM and its associated processing equipment.

To establish a datum reference frame, a number of datum features are chosen, (predominantly out of functional considerations of the design) in order to establish the component in relation to a system of three mutually perpendicular planes, which is

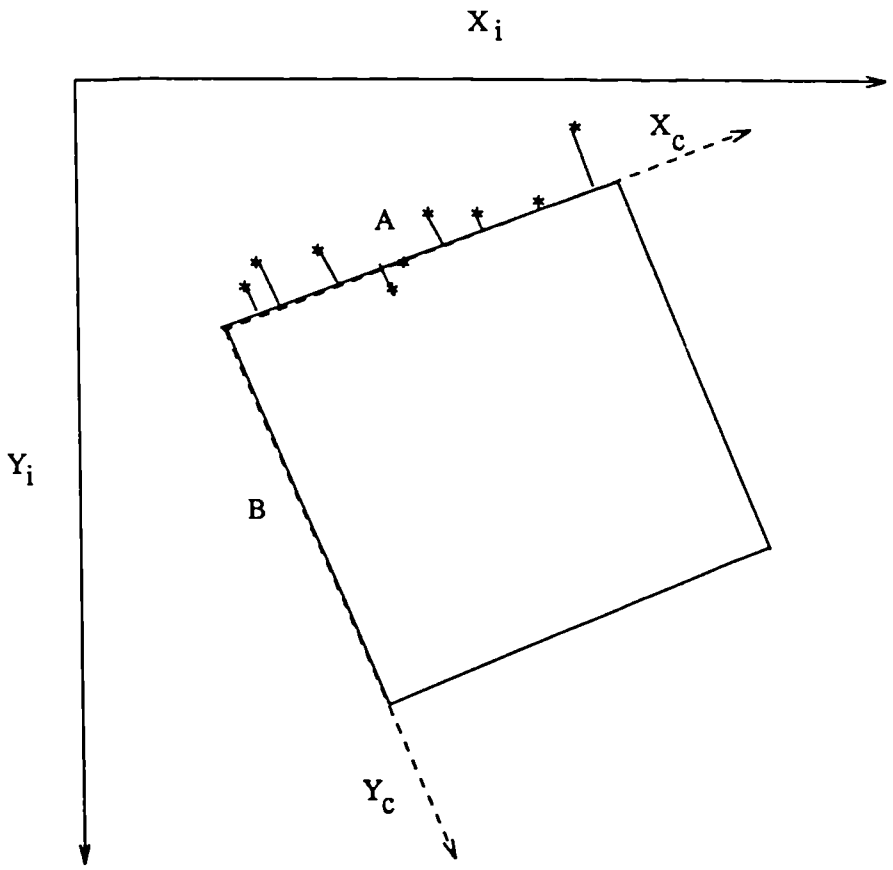


Figure 2.1

collectively referred to as the datum reference frame. This reference frame exists in theory only, consequently, there is a need to construct a method for simulating the theoretical reference frame from the existing features of the component. For implementation on CMMs, coordinate dimensioning must clearly indicate the features of the component that are used to establish the three mutually perpendicular planes in question.

In the majority of situations, a single datum reference frame will suffice, however, in some cases the physical separation or functional relationship between features will neccessitate additional reference frames, applied at specific locations on the component. Obviously, the feature control frame on the drawing should contain the appropriate datum feature references. The application of maximum material condition, MMC, or a change in the precedence of any of the referenced datums, may require different simulation methods, and, consequently establish a different datum reference frame.

Only the required number of datums should be referenced in a feature control frame when specifying geometric tolerances. However, the geometrical control afforded by these tolerances must be understood in order to effectively determine the number and type of of datum references required for a specific operation.

The concept of "datum targets", discussed in ANSI Y14.5M seems to be particularly useful for drawing relating to coordinate measurement as they designate specific coordinates, or systems of coordinates on a component that are used in establishing a datum reference frame. Datum targets are neccessary because irregularities of feature surfaces (usually due to manufacturing methods; moulding, casting, forging, machining etc.) often deem an entire surface an unsuitable medium from which to establish a datum. Each datum target point (coordinate) should be indicated by a symbol which may be located dimensionally, or alternatively, labelled and located tabularly, if possible on a direct view of the surface. If however, a direct view of the surface is not possible, the point location should be indicated on two alternate views (4).

The design of many components is such that there are three mutually perpendicular plane surfaces that are functionally appropriate for the construction of a datum reference frame. In this case, the reference frame would be established in the following manner: the primary datum feature relates the component to the datum reference frame by taking a minimum of three points on the first datum plane (these points being coordinates, ideally identified on the drawing by datum target points). The component is further related to the frame by taking two coordinates on the secondary datum feature. The relationship is then completed by taking at least one point on the tertiary datum feature.

If the diameter or centre of a cylinder is used as a datum, obviously it is subject to variation in both size and geometric form, therefore it is necessary to establish if maximum material condition (MMC) applies in each particular case.

In the case of the datum being the axis of the cylinder, the question of the most appropriate method by which its location is calculated still remains open, there are at least four reference figures that may be used: The least squares centre, the minimum radial zone centre, ring gauge centre and plug gauge centre. The choice of the most appropriate model to use should be made from functional considerations and to afford the best intermeasurement stability.

2.3. Tolerancing

The two basic types of tolerancing are of relevance to coordinate measurement, are those of form and position. Geometrical tolerance may be defined as 'The maximum permissable overall variation of form or position of a feature' (4). It defines the shape and size of a tolerance zone within which the surface or median plane or axis of the feature is to lie. The zone in which the feature is required to be contained is one of the following:

1) A circle or cylinder

2) The area area between two straight or curvilinear parallel lines.

3) The space between two parallel surfaces or planes

The geometric reference frame is a diagram composed of the constructional dimensions which serve to establish the true geometrical relationships between the positional features in any one group (fig 2.2).

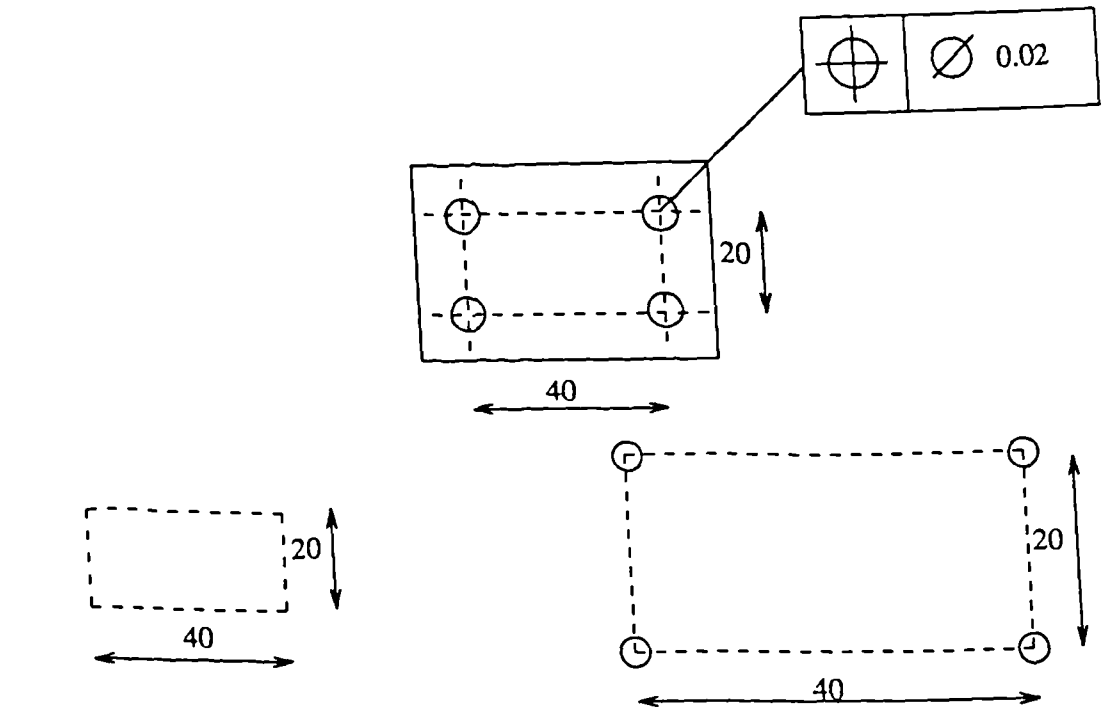
There are basically four distinct groups of tolerances, form, attitude, position and composite. Form tolerances are included to assess such errors as straightness, flatness, roundness, cylindricity, profile of line and profile of surface. Attitude type tolerances include parallelism, squareness and angularity, tolerances of position refer to criterion such as location of a particular feature, concentricity and symmetry. Composite tolerancing is usually expressed in the form of 'run out'.

The above definitions of tolerance zones are not directly implementable on CMMs for reasons discussed later.

In coordinate tolerancing there are generally certain problems as, inadvisable as it may seem, tolerances of position (their tolerance zones) are still often described or specified in relation to plane surfaces. This accuracy will be limited by the mutual accuracy of the reference plane surfaces. To illustrate, refering to fig 2.3, if sides a and b are not perpendicular, then the tolerance (of position) zone expressed relative to these surfaces, will not be centred at the nominal position of hole c and the component will be rejected although it is probably functionally sound. To avoid this situation, it is vital that tolerance zones should be built on a functionally significant framework.

2.4. Geometric Fitting Procedures

The choice of geometric fitting model primarily depends upon what is required of the measuring process. Geometric fitting merely involves the fitting of an equation to data by the minimisation of residuals according to various criteria (norms). Residuals are merely the diffence between the data predicted by a model and the data on the surface



geometric reference frame

geometric tolerance diagram

Figure 2.2

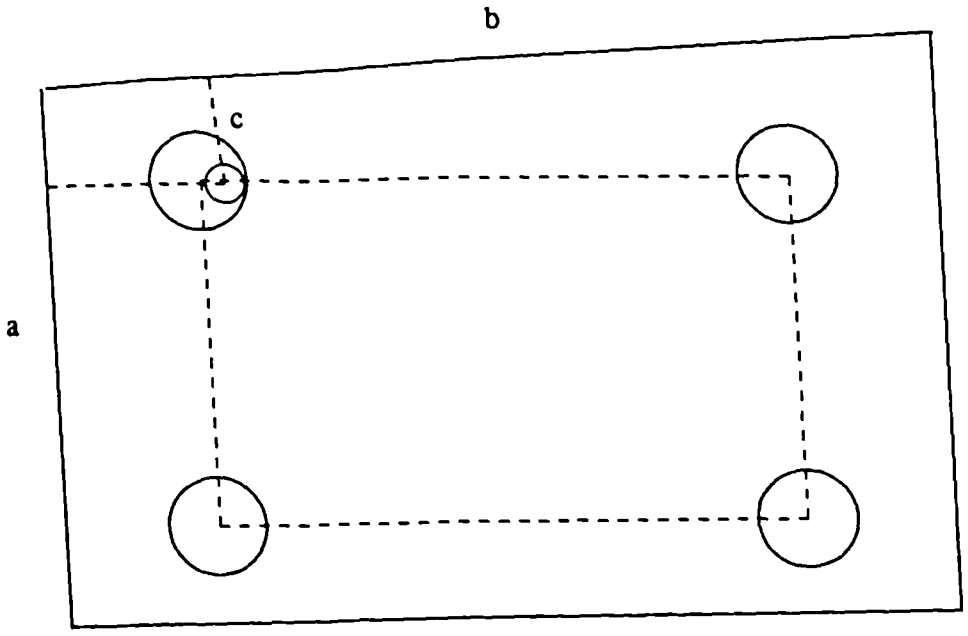


figure 2.3

under inspection. In the case of coordinate measurement, the measurements represent the surface of a manufactured component, and we need to know:

a) The geometry of the component

b) If the component is contained within tolerance limits.

However, there still remains the question of which of the three norms (L_1,L_2,L_∞) is the most suitable for these purposes (see appendix A2).

The L_2 fit tends to average out the errors and find the underlying parameters. If it is desired to ignore the odd wayward observation, L_1 norm is the most suitable choice. Perhaps of the three norms, L_∞ is potentially the most useful as it tends to minimise the maximum error by pulling the fit to outlying groups of data, which is potentially useful for applications in the design of functionally significant algorithms (discussed later.)

However, fitting L_1 and L_∞ norms is not satisfactory at present, as they are considerably less reliable and more computationally expensive than L_2 fitting algorithms. Much of this is due to the nature of the fits; for a least squares fit for a well posed problem, it is reasonably safe to assume that the local minima for the objective function are reasonably well seperated (5). For other norms this is not necessarily the case, and the local solutions may occur arbitrarily close to each other, indeed, even for linear L_1 fits, there can be a continuum of local solutions. The least squares formulation has further mathematical properties giving it considerable advantages over other norms in that the L_2 norm gives rise to differentiable objective functions, this is not true for other norms and it is therefore the L_2 norm that is most commonly used in coordinate measurement geometrical parameter estimation algorithms.

2.4.1. Element Fitting and the Estimation of Geometric Parameters

Least squares fit algorithms have been developed for virtually all geometric elements and are commonly used in CMM operations. If the residual error is defined to be the shortest distance to the curve or surface, then the models are deemed to be non linear.

Line and plane fits can be reduced to an eigenvalue problem, for which there exist stable algorithms, while other non linear fits can be tackled using a variant of the Gauss-Newton algorithm. However, linear models, using slightly different definition of residuals are also used for the plane sphere and cylinder fits and can be solved by the more computationally efficient linear least squares type algorithm. Even if the full non linear model is required, the linear approximations are important for obtaining good initial estimates of the optimisation parameters. There are also linearised models for the cone, ellipse and cylinder fits, but they often rely on the assumption that the residuals have a near Gaussian distributions (6).

If the data is accurate, and computational speed is important, which is certainly the case in coordinate measurement, then a linear model (linearised approximation) is appropriate. It is only when the data is less accurate and a true assessment is required, that the full non linear model should be used (5).

There is currently much work being done on L_∞ or 'minimum zone' fits for geometric elements and this is still much of a research topic. Algorithms for planes, lines, circles and to some extent spheres, are in a more developed and reliable state than cylinders and cones. Some of the algorithms use an exchange principle derived from a characterisation of the solution (7),(8), but for the more complex elements, no convenient characterisation is available, the algorithms have to concentrate on special cases which adds complication and reduces applicability. Often a linear least squares best fit would be found first so that the L_∞ search would have a good starting point and so reduce the number of iterations and consequently computation time.

Element fitting using the L_1 norm has not really been investigated to any great extent. More extensive use of L_1 fits will have to await further algorithmic research.

2.5. Functionally Significant Algorithms.

Many practitioners are starting to believe that the solution to many of the problems

associated with coordinate measurement is to be found in the designing of functionally significant algorithms particularly those adopting the basic principles of hard gauging. In addition to the advantages of functionally significant measurement, CMMs can overcome some of the problems associated with traditional gauging methods including:

i) Inflexibility to design changes

ii) Inability to provide process control information

iii) Long lead time needed to design, build and certify for use

iv) Cost of storing, recertification and maintenance

v) Reduced tolerance margins, due to gaugemakers tolerance

However, it is very difficult to simulate the go, no-go gauging action by corresponding profile according to Taylors principle, as the algorithms for these procedures is very complex and computationally expensive, being based on non-linear optimisation (9).

Existing data analysis software does not posses the ability to qualitatively determine the acceptability of a part utilising full MMC, specifically in the application of full MMC tolerancing to a number of component features and datums simultaneously in three dimensions. Such an inadequacy must be overcome if true functional significance is to be achieved.

A fundamental change of approach from calculating tolerances based on size and location (Taylor's Principle) to defining 'material free' boundary surfaces has produced some fruitful results (9). To illustrate the point, consider a positional tolerance applied at MMC; the change in approach is away from considering the problem in terms of the axis of the feature to considering it in terms of the surface of the feature. Traditionally in the example of a circular hole the axis of the hole must be contained within a cylindrical tolerance zone, with its axis located at the true position. Whereas the latter approach would merely specify that, while maintaining the specified size limits of the feature, no element of the feature surface shall be inside a theoretical boundary located at true posi-

tion. The new approach reduces the problem of inspecting individual points on the datum features in order to ascertain their containment within or violation of boundary surfaces. These boundary surfaces are controlled by the collective effect of tolerances such as size, form, orientation, location and the application of MMC. The net resultant boundary derived from these considerations is referred to in ANSI Y14.5M as the 'virtual condition', and it largely determines the gauge feature size. However, this approach may introduce some degree of difficulty in shifing the gauge definition with respect to the component, as discussed later.

Soft functional gauging involves duplicating a functional gauge on a CMM. This involves defining zones on a component within which there should be no material. To determine whether the part satisfies a 'go' condition, the measured points on the feature surfaces of the component can be compared to the mathematical definition of the gauge.

Another of the problems to be overcome in the implementation of this technique is the method by which the designer 'designs' the functional gauge on the CMM. The process would require the definition of perfect geometric figures in space; plane, cylinders, cones etc. The next problem is to overlay the measured points onto the gauge definition, and to consequently determine if the measured points interfere with the gauge definition. If so, the next stage is to determine if the measured points can be shifted with respect to the gauge definition so that the interference no longer exists. If a no-go condition is found, then the data is rotated and translated in order to get as many features as possible within tolerance, and to record in descending order the features violating tolerance, to achieve this, there are two main areas of concern in which there are radical differences in approach to that of standard coordinate measurement. Firstly, and most fundamentally, in the geometric fitting routines, 'best fits' i.e. unadulterated least squares techniques are not considered appropriate as the parameters do not represent the true functional surface. Instead, methods are found, usually based on traditional least squares routines, of manipulating the data being used in the fit so that only the functionally

significant data is finally used. The second difference is in aligning the workpiece for inspection. For truly functional inspection alignment is not merely a case of aligning the appropriate plane side of a workpiece with the instrument coordinate frame, as for straightness measurement mentioned earlier, but of overlaying the points onto the gauge definition, optimising the orientation of the workpiece i.e. the component coordinate frame within the instrument coordinate frame, with respect to the location of datum 'targets' within their respective tolerance zones.

2.5.1. Functional Fitting Routines

The philosophy behind functional fitting routines is clearly reflected in ANSI Y14.5M, which seeks to define geometric shapes by their function, and by unambiguous inspection techniques that simulate functional performance. For example, consider the plane, ANSI define it as being the locating surface on a surface plate, this definition means that the functional plane is that passing through the three high points on the plane, the remaining coordinate points are termed as having 'minus material deviation'(10).

A number of methods exist for fitting geometric elements to the functionally appropriate data, the following section merely gives examples of a few :

2.5.1.1. Plus Material Form Shift

This technique involves utilising both best-fit estimates of geometric parameters and form error in order to calculate the functional size of a feature.

Errors of form are expressed in terms of parallel or concentric defining surfaces, within which the feature surface is contained; the seperation of the defining surfaces being the error of form. The 'Plus Material Form Shift' technique assumes that the best fit element describes a surface half way between the defining surfaces. The functional dimensions of the features are then calculated by adjusting the appropriate parameters according to the maximum material condition. Such that, if the functional dimension of a circular pin is sought, its best fit diameter is increased by half the error of form.

Similarly, if the functional diameter of a hole is required, its diameter is reduced by half the form error.

2.5.1.2. Sigma Shift

This approach first calculates a best fit and then processes the residual definition statistically. A standard deviation is computed and from this, a statistical confidence factor is established. In Gaussian distributions, a sigma factor of one will contain 68% of the data. A two sigma confidence level will contain 95% of the data.

When a geometric element is sigma shifted, the modelled surface is moved so as to remain parallel to the best fit surface so that, by a statistical projection, the desired percentage of data points lie above or below a desired level. For example, as in the case of a plane the surface is moved normally upwards to the best fit plane, and for a hole, the radius is reduced by the standard deviation times the sigma factor. However, there are distinct disadvantages associated with this method in that it can lead to gross overcompensation and assumes that the surface is Gaussian distributed.

2.5.1.3. Successive Approximations

This method tries to deal with the problem of modelling surfaces based on functional fit criteria. 'Functional fit criteria' is not a mathematically pure concept but an ad-hoc model of reality in metrology. The successive approximation technique uses a procedure like best-fit. An approximation is made which discards the uninteresting data and re-evaluates using only the pertinent data. In the case of a plane, the three high points must be located and a plane subsequently fitted to them, so 'pertinent' data is all that above the best fit line and 'uninteresting' data is all that below. The amount of pertinent data is reduced with successive approximation, until the quantity of pertinent data falls below a certain level.

However, this approach requires a great amount of computation time, and the three highest points may lie very close together, effectively forming one or two points of

contact, and therefore the solution would not represent the true functional plane. Also the assumption is made that the data does in fact lie on the peaks. This latter problem may be overcome by isolating groups of points surrounding a peak, each grouping then being analysed with a three dimensional best fit curve that models a topographic peak, but, this is potentially very dangerous, as it can predict peaks that weren't really there.

2.5.2. Shifting the Gauge Definition with respect to the Component

The best fit between component and theoretical gauge definition can be achieved by moving one with respect to the other translationally (x,y) and rotationally θ, in such a way that the sum of the distances between some given point on each feature (most usually the centre of the feature) and its corresponding point on the gauge definition (E_i) relative to its relevant tolerance (T_i) are minimised. This would effectively involve an iterative search in x, y and θ to minimise $\sum\left[\frac{E_i}{T_i}\right]$ (11)

In order to apply a similar type of procedure to the concept of material free boundaries, changes would have to be made to the objective function, as it is no longer a case of minimising E_i/T_i, but of ensuring that the feature profile is contained within a boundary set by the gauge definition can be replaced by $\sum r_j$.

3. Limitations of Current Algorithm Design

For each aspect of measurement on a CMM, the results obtained using similar algorithms and sets of test data can vary greatly; the practical result of which is that an unacceptably high proportion of components inspected by coordinate measurement are not assembling with adequate precision, if at all.

The overall accuracy of the CMM may be tested by the measurement of an artefact, the dimensions of which have been verified by a primary measurement method (ring or bar gauges). The artefact in question is then inspected at various locations and orientations throughout the measuring volume of the machine and an error grid is constructed.

Many investigations into the integrity of CMM software have been carried out, usually by using a synthesized set of test data from a model of known error of form, and comparing the results obtained by the evaluation routines, an assessment of the software can then be made.

A series of tests were performed by Edson and Parry, (12), and the results showed that the use of least squares to calculate errors of form can produce answers which are in excess of five times the true error. Fewer than 3% of the answers produced for straightness were correct when compared with straightness calculated in accordance with BS 5204, ANSI Y14.5M or ISO/R1101.

Circular least squares fitting routines fared little better, with more than 90% of hole diameters calculated to be between 0.15% to 0.5% larger than the maximum diameter plug gauge size, and a similar proportion of centre position coordinates were found to contain errors between 0.02mm and 0.2mm. In practical terms, in addition to the case of out of tolerance parts being accepted, there arises the converse situation of 'good' parts being rejected,

The divergences of estimated parameters from their true values are due to several factors including:

Alignment error

Non optimum definition of geometric residuals

Lack of implementable data sampling techniques

Error due to not Considering all Possible Data

Lack of specification of inspection procedures

Lack of implementable definitions of geometric form

3.1. Alignment Error and the definition of geometric residuals.

These important factors are individually discussed in the following sections and form the basis of the following chapters The term residual may be defined as the amount

the observation deviates from the model, and ideally, a method of assessing residuals should be defined in standards. There is a need to build up a framework of concepts so that measurement procedures can be performed in a systematic manner. Providing either the component and instrument frames are aligned, or their misalignment is known and can be compensated for by calculation, the data may be regarded as being in component coordinates.

Even if the residuals were expressed in perfect component coordinates, there remains uncertainty in the measurements made from them as the direction of approach of the probe will probably be along the instrument coordinate frame, and therefore over-travel and other probe induced measuring uncertainties will not be in the same sense as these residuals.

From equation 2.2 it is obvious that the estimate of slope of the component surface is vital to the integrity of the subsequent analysis. It is imperative that the error in this estimate is quantified either to provide appropriate compensation or to be in a position to make some statement as to the quality or precision of the inspection procedure.

The complexity of the situation is exacerbated by the use of computer based error compensation techniques. Correcting systematic errors implies that the data obtained are chart coordinate representations involving a known reversible transformation. The correction process is an inverse transformation to ideal instrument coordinates with the errors fixed in that frame. After correction, the data is assumed suitable for fitting reference figures based on component coordinates. This process deals well with, for example the form errors of instrument translation because these do relate directly to instrument coordinates. Difficulties arise if the error depends on the particular measurement as it could with probe errors i.e interaction between the component and instrument system could make it virtually impossible to obtain a unique inverse transformation (2).

3.2. Non Optimum Residual Definition

The coordinate data taken from the component may possibly contain rather more useful information than is utilised by current fitting routines. Ultimately from such information, the best trend for residual definition can be identified. The optimum definition of the residual of course depends upon the purpose of measurement; if the geometry of the reference figure is to be ascertained, then the ideal residual direction would be that causing the least instability with respect to number and placement of data points, and the most suitable trend for checking for containment within tolerance zones would be that which most adequately describes the surface.

3.3. Optimum Data Sampling

For efficient measurement procedures, it is necessary to plan a measurement sequence such that there is a balance between the degree of computational effort and time expended, and the accuracy of the results obtained. Clearly, a method of assessing the statistical uncertainty in the estimation of parameters must be found. Ultimately, practitioners would then be in a position to be able to specify an optimum number and distribution of data points that would need to be taken in order to satisfy a predetermined tolerance or limit of accuracy. Such a development would hold subsequent advantages in that the reduced data aquisition time would lessen the effect of measurement instabilities due to machine and temperature drift (13).

However, certain conditions may hinder the implementation of an optimal measurement scheme, such conditions may include machine limitations , and impaired access.

Ultimately, truly effective coordinate measurement algorithms would be capable of deriving the optimum sampling procedure from surface information gathered by the CMM itself.

3.4. Error due to not Considering all Possible Data

When fitting geometric elements, errors are incurred in the estimates of the various parameters due to not utilising all the possible data, these errors are commonly known as sampling errors. In the case of coordinate measurement, the effects of sampling error are probably particularly severe, as relatively few data points are used. This problem is particularly important in alignment procedures, as discussed earlier, but it is also vitally important that sampling errors in the estimated parameters of circles, planes, cylinders etc. are also quantified, so that some degree of certainty can be associated with the accuracy of the inspection process. It is therefore important that in the formulation of a technique for quantifying these errors that provision should be made in order to ensure applicability in the fitting of most commonly used geometric elements can be maintained. Additionally, such a technique must be implementable on all engineering surfaces, and not rely upon the often erroneous assumption that the surface profile on the component under inspection consists of a Gaussian amplitude distribution.

3.5. Implementable Definitions of Geometric Form.

Departures from nominal shape are expressed in terms of surfaces, for example, BS 817 (1972) gives the following definition for flatness error on a working surface:

> The distance seperating two parallel planes between which that surface can just be contained.

Similarly, roundness is described in terms of two concentric circles (BS 3730, (1982)).

For the purposes of coordinate measurement and other digital techniques, these definitions are not implementable as ambiguities arise which consequently result in a divergence of practice. For definitions to be algorithmically implementable, the overall error of form should be related to a set of parallel or concentric features just containing best estimates of a discrete set of representative points on the working surface, the points

being derived through an appropriate mathematical model, from a finite set of measured values Therefore, in the case of the plane, an implementable definition would have related flatness to a pair of parallel planes just containing best estimates (13).

BS 3730 deals with roundness in a rather better manner as it relates to known measured profiles rather than unknown working surfaces; its definition for roundness is as follows:

> The departure from ideal roundness is assessed as the difference between the largest and the smallest radii of the measured profile of the workpiece measured from one of the following centres:

a) Least squares centre

b) Minimum zone centre

c) Minimum circumscribed centre

d) Maximum inscribed centre

However, no algorithm is specified for the calculation of the centres b, c and d. Therefore, although rather more implementable than the standard for planes, the definitions are still grossly ambiguous, and offer no indication of either random or systematic errors that may be incurred. Indeed, Cox and Jackson state that standards should ideally include references to algorithms for estimating the uncertainties in the assessment parameters, and any systematic uncertainties in the mathematical model (partial arcs etc.)

4. Conclusion

Among the many needs for development in CMM algorithm research, one specific problem seems to re-occur at almost every stage. This is the need to mathematically quantify the errors induced by not utilising all possible data. The resolution of this problem for a wide range of shapes and non Gaussian surfaces is presented in the following chapters of this thesis.

When setting up the datum reference frame, the instrument coordinate frame must be aligned to the component coordinate frame, this process involves the measurement of slope of one of the plane sides of the component. The datum frame is set relative to this value, so, for the integrity of the subsequent measurement process, it is vital that the error in this estimate of slope is either minimised or quantified so that a rigid body transformation correction process can be effected. In order to perform either procedure, it is neccessary to know the true slope, i.e. that estimated from virtually all data on the surface profile, and to know how the error of the slope estimated from the particular number and distribution of data points relates to it.

The geometric element fitting procedures also have a large degree of uncertainty associated with them. In addition to random and systematic inaccuracies due to machine error, the algorithms themselves induce their own instabilities. A significant degree of the instabilities is due to ill-conditioning, which for lines and planes usually implies that the data point being used in the fitting routine lie very close together, and for circular components the implication is that the data points have only formed a partial arc around the circular profile. However, due to the inevitable irregularities inherent in all engineering surfaces, uncertainties will again result from not considering all available surface data. If coordinate inspection procedures are to be entirely reliable, then these errors must be quantified, or it will not be possible to garantee the accuracy of the inspection process.

The need has also been expressed for the design of a data sampling scheme (13) , whereby the amount of data collection and procesing is minimised with respect to the accuracy required of the procedure, i.e. an optimum sampling scheme. This is important in speeding both the total measurement and computation time. The formulation of such a scheme can only be achieved by knowing the accuracy of the process for varying permutations of number and distributions of data points. One of the primary factors in this error assessment is sampling error. This has hitherto been neglected by the CMM com-

munity, whereas error induced by ill-conditioning have been widely investigated ((14),(15)). When these two error components have been combined with a random error analysis, it should then be possible to formulate an optimum sampling scheme for all geometric parameter evaluation procedures.

As mentioned earlier, standards must include references to specific algorithms for quantifying errors associated with measuring uncertainty, surely such references must include an algorithm for quantifying sampling error. All that remains is to formulate a reliable, accurate and implementable algorithm.

References.

(1): D.G. CHETWYND & P.H. PHILLIPSON (1980): "An Investigation of Reference Criteria Used in Roundness Measurement"; J. Phys. E, Sci. Instrum, Vol. 13.

(2): D.G. CHETWYND (1986): "The Definition of Geometric Residuals"; Microengineering Report, University of Warwick.

(3): D.G CHETWYND (1985):"The Definition of Geometric Residuals"; Proc. NPL Conf. on Software for CMMs, Sept 1985.

(4):ANSI Y14.5M

(5):P. FORBES: NPL report unpublished.

(6):T.S.R. MURTHY (1983): "A Method for the Evaluation of Elliptical Profiles"; Precision Engineering, Vol 5, pt. 2, pp 77-81.

(7): D.G. CHETWYND (1985): "Linearized Exchange Algorithms in Metrology" ; Proc. NPL Conf. on Software for CMMs, Sept. 1985.

(8):G.T. ANTHONY & M.G. COX (1985): "Reliable Algorithms for Roundness Assessment According to BS 3730"; Proc. NPL Conf. on Software for CMMs, Sept. 1985.

(9):W.H. RASNICK & N.E. ZURCHER (1986):"Soft Functional Gauging on Coordinate Measuring Machines"; Pre-print for submission to Prec-Eng Conf.

"Thresholds for Precision Engineering", Dallas, Texas, November 1986.

(10):K. ULBRICH (1987): Internal report , Electronic Measuring Devices, Flanders New Jersey, U.S.A..

(11):A. SUTHERLAND (1987): Internal report, Ferranti Industrial Electronics, Dalkeith.

(12): F. EDSON & G. PARRY (1985): "Integrity of Software Associated with Coordinate Measuring Machines"; Proc. NPL Conf. on Software for CMMs, Sept. 1985.

(13): M.G. COX & K. JACKSON (1983): "Algorithms and Software for Metrology: A Statement of Need";NPL Report MOM65, June 1983.

(14):J.C. MC. COOL (1979): "Systematic and Random Errors in Least Squares Estimation of Circular Contours"; Precision Engineering, Vol 1, Pt. 4, pp 215-20.

(15):D.J. WHITEHOUSE (1973): "A Best Fit Reference Line Used in Partial Arcs"; J. of Phys. E: Sci: Instruments, Vol 6.

Geometric Element Fitting Procedures

1. Introduction

This chapter introduces geometric element fitting procedures and the errors that may be incurred in their application to coordinate measurement. Some commonly used fitting procedures and their underlying principles are outlined. Where available, existing error analyses for parameter estimates are reviewed. An original investigation is then conducted into the behaviour of sampling error incurred in the estimation of some least squares parameters. From the results of this investigation, trends in the behaviour of sampling error are observed, and recommendations are made for general coordinate measurement practice so that excessive error may be avoided. More importantly, it is desired to ascertain wether similarities in the behaviour of the sampling errors exist between the various geometric elements so that the possibility of finding a general, formal solution for the quantification for these errors can be established.

2. Geometric Evaluation

2.1. Background

The inspection of components by coordinate measurement involves manipulation of coordinate data in order to ascertain information about the component geometry. In applications such as geodesy, it is required to represent the underlying geometry of the component surface usually by the fitting of polynomials or splines of varying degree, i.e. no fixed model is used. Alternatively, the coordinate data may represent the surface of a component nominally manufactured to within specific tolerances, and it may be required

to know if its dimensions are indeed within the required tolerance. Also, the general form of the component may be known, but its precise dimensions and shape may be required. In both the latter two situations, a fixed model is fitted to the data. It is with this type of fitting procedure that this chapter is concerned.

2.2. Conditioning of Geometric Fitting Routines

The conditioning of a problem can be thought of as a measure of how sensitive the estimated parameters are to changes in the data used in their calculation. A well conditioned model is one where a small change in the data will result in a correspondingly small change in the solution, and conversely, an ill conditioned problem is that where a relatively small change in the data can result in a large alteration of the values of the calculated parameters.

Ill conditioning affects each algorithm differently, but always adversely. In linear least squares problems, ill conditioning is manifest in a loss of precision in the parameters. Iterative algorithms such as Gauss-Newton also suffers a loss in accuracy but have the additional problem of a slowing down in the rate of convergence. As a consequence unless good initial estimates are available to begin the search, the solution may even diverge (1).

2.3. Mathematical Techniques

The majority of data fitting excersises can ultimately be considered as the solution of an overdetermined system of n equations :

$$a_i(\mathbf{u}) = 0, \qquad i=1,....,n$$

In the model parameters $u = (u_1,.....,u_n)$. In the linear case this problem may be formulated as a minimisation thus:

$$\min \| A\,\mathbf{u} - \mathbf{b} \|_N$$

Where $\|.\|_N$ represents a norm, either L_1 ,L_2 or L_∞ (see appendix A2). If the L_2 norm is chosen, then the problem is a linear least squares type which can be solved using

a QR type orthogonal transformation algorithm employing Householder or Givens transformations (1). The choice of L_1 or L_∞ norms results in a problem that is best solved iteratively using linear programming techniques (1,2).

If the equations are non linear in the parameters $\mathbf{u}$ then the most common approach is to use an iterative algorithm that linearises the problem at each stage. The Gauss-Newton algorithm for the solution of non linear least squares problems is of this type and is very reliable provided the problem is well conditioned and good estimates of the parameters are available to begin the procedure (1).

At present, non linear solutions using L_1 and L_∞ norms are unreliable and computationally expensive. The problem seems to be in the nature of the fits; for a least squares problem the local minima for the objective function are well seperated, but this is not neccessarily true for other norms, indeed, there may even be a continuum of local minima. These fitting routines are still in the research stage and are not therefore as yet available for CMMs.

2.4. Element Fitting

Least squares algorithms have been developed for fitting most geometric elements and are used extensively. If the residual error is defined as the shortest distance between the data point and the surface being fitted then the problem is non-linear. Line and plane fits can be expressed in terms of an eigenvalue problem, other non-linear fits can be solved using a variant of the Gauss-Newton algorithm (1). Linear fits use slightly different definitions of residuals. Linear least squares algorithms exist for fitting lines, planes, circles and cylinders, and are commonly used in CMM practice, linearised solutions exist for ellipses and cones, but they usually rely on the assumption that the residuals have a Gaussian amplitude distribution (3).

If the data are accurate, and the speed of calculation is important, as is the case in coordinate measurement, then the linear model is appropriate (1), as this solution has far

greater computational efficiency.

3. Geometric Fitting Routines

In respect to the placing of data used in the experiments outlined in this chapter, the guidelines set out in B.S. 7172 are not adhered followed for the following reason. B.S. 7172 makes reference to the ways in which data points should be distributed upon a geometric entity in order to best represent the surface and therefore minimise error. These recommendations are merely based upon general heuristic and qualitative considerations, and it is the purpose of this thesis is to develop a quantitative analysis of these errors.

3.1. The Linear Least Squares Line

The calculation of the parameters of the least squares line are particularly important in coordinate metrology as they are used in the process of aligning instrument and component coordinate frames at the begining of the inspection process , indeed the expression of the residuals in component coordinates is directly dependent upon the value of slope (4), the validity of the inspection process would therefore be adversely affected by errors in this estimate.

Where two dimensional surface data is represented by a series of coordinates, (x_1,y_1), (x_2,y_2),..., (x_n,y_n), and these data loosely describe a straight line, the value of b may be predicted thus (5):

$$y_i = \beta_o + \beta x_i + e_i \tag{3.1}$$

Where e_i is the quantity by which the prediction of b_i differs from the raw data and is termed the residual. For the least squares crtierion states that the sum of the squares of these residuals is minimised,

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_o - \beta x_i)^2 \tag{3.2}$$

β and β_o will be replaced by values b and b_o respectively that will minimise S. In

order to obtain these values, equation (3.2) must be differentiated with respect to β and β_o , thus:

$$\frac{\partial S}{\partial \beta_o} = -2\sum_{i=1}^{n}(y_i - \beta_o - \beta x_i) \tag{3.3}$$

$$\frac{\partial S}{\partial \beta} = -2\sum_{i=1}^{n} x_i\,(y_i - \beta_o - \beta x_i)$$

The estimates b_o and b are given by equating to zero:

$$\sum_{i=1}^{n}(y_i - b_o - b x_i) = 0 \tag{3.4}$$

Which can also be expressed in the following way:

$$\sum_{i=1}^{n} y_i - n b_o - b\sum_{i=1}^{n} x_i = 0 \tag{3.5}$$

$$\sum_{i=1}^{n} x_i y_i - b_o \sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0$$

$$b_o n + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{3.6}$$

$$b_o \sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Solving equation (3.6) for b gives:

$$b = \frac{\sum x_i y_i - ((\sum x_i)(\sum y_i))/n}{\sum x_i^2 - (\sum x_i^2)/n} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{3.7}$$

The solution of equation (3.6) for b_o is:

$$b_o = \bar{y} + b(x - \bar{x}) \tag{3.8}$$

3.1.1. Investigation into the Sampling Error Incurred in the Least-Squares Estimates of Slope and Intercept.

Gaussian techniques for the analysis of errors in the estimates of b and b_o are well documented, but hitherto, no formal attempt has been made to quantify the sampling error incurred in these estimates. Sampling error is not only a function of the number and distribution of data points used in the calculations, but also of the surface profile characteristics. Therefore a series of experiments were performed in order to investigate this phenomenon with respect to systematic changes in the distribution of data and the nature of the surface. In order to achieve this, it is neccessary to perform the calculations

using deterministic simulated surface profile. The use of deterministic surface data is doubly appropriate as many engineering surfaces bear predominantly periodic characteristics resulting from manufacturing processes, (Chapter 4).

3.1.1.1. Experimental Procedure

The purpose of the following section is to investigate the sensitivity of the estimates of b and b_o with respect to changes in the surface parameters and the distribution of data points used in the calculations. It was decided to begin the investigation using a simulated surface consisting of a sinusoid, and to alter the values of amplitude, frequency, phase, sampling interval and number of data points with respect to each other in order to observe the relative sensitivity of b and bo.

The sampling errors were obtained by calculating the value of each parameter using a particular number and distribution of data points upon the surface, calculating the parameter again, using an unrealistically high number of data points (in the order of 500) distributed over the same length of profile. The latter estimate was subtracted from the former, and this error was regarded as the sampling error. This procedure was caried out using sets of data consisting of between 2 and 50 equispaced data points and varying each input parameter in turn, (all other input parameters remaining constant) so that a series of three dimensional plots representing sampling error against the number of data points used and one other parameter were formed.

All surface simulations and calculations were performed on a SUN 3/60 computer using the MATLAB software package. Table 3.1 shows the values of the parameters in the surface data and number and distribution of data used in the experiments, where

f = frequency of sinusoid

w = amplitude of sinusoid

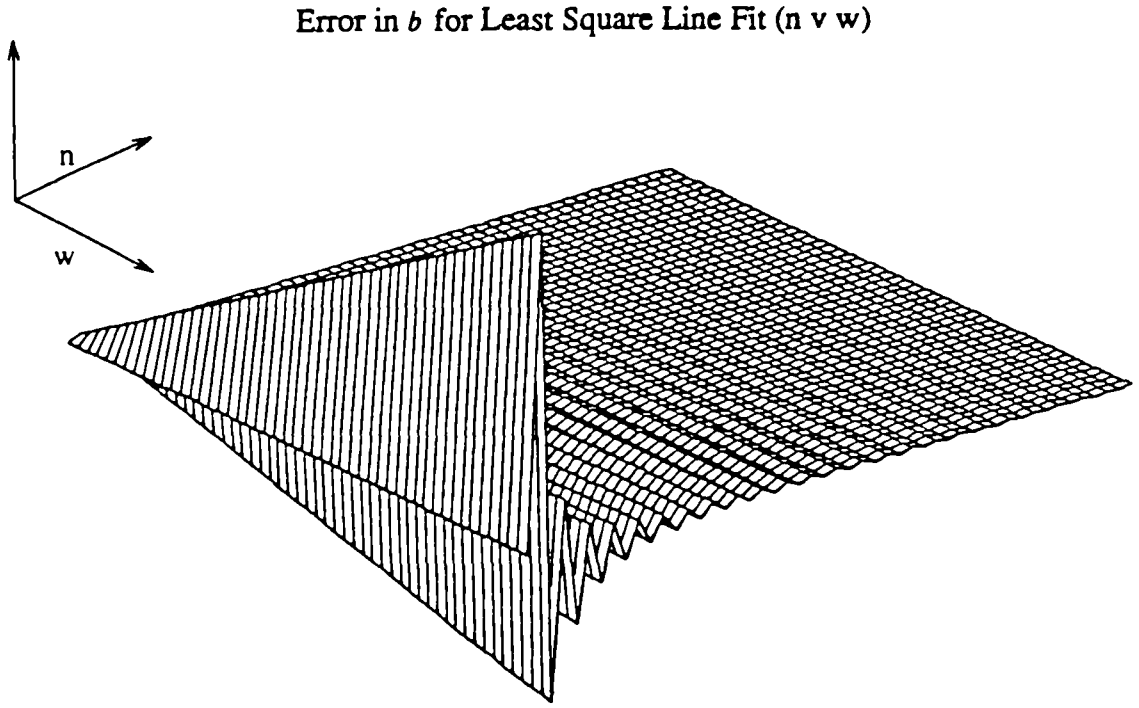
phi = phase of sinusoid

dx = sampling interval

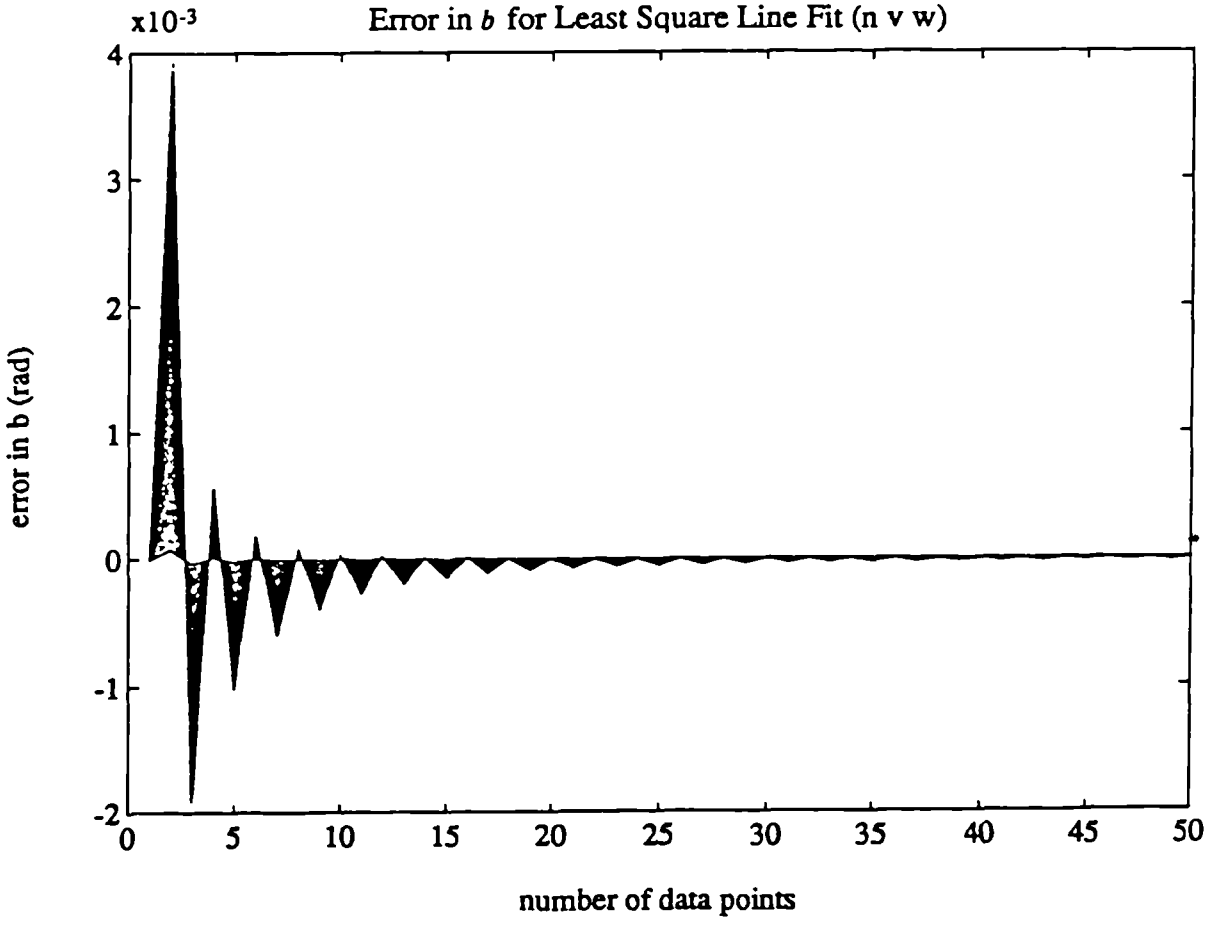
n = number of data points

Simulated Data Parameters		
Parameter	Constant	Variable
w	0.005(mm)	0.001:0.500(mm)
phi	0.000(rad)	0.000:2 π(rad)
f	0.500(Hz)	0.020:1.000(Hz)
dx	10.000(mm)	0.500:25.000(mm)

Table 3.1

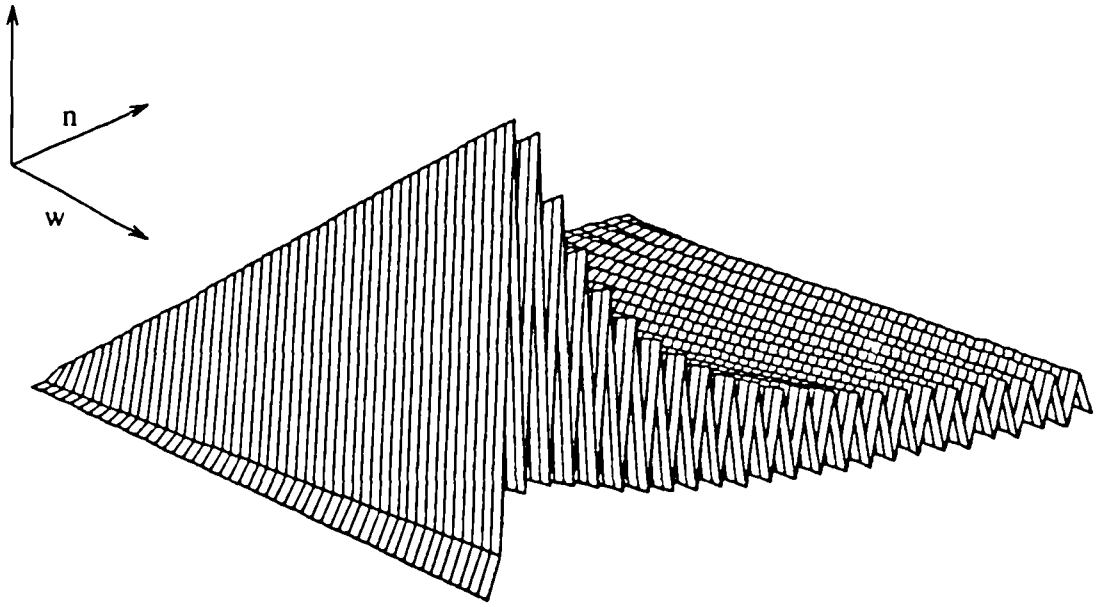


Graph 3.1 (a)

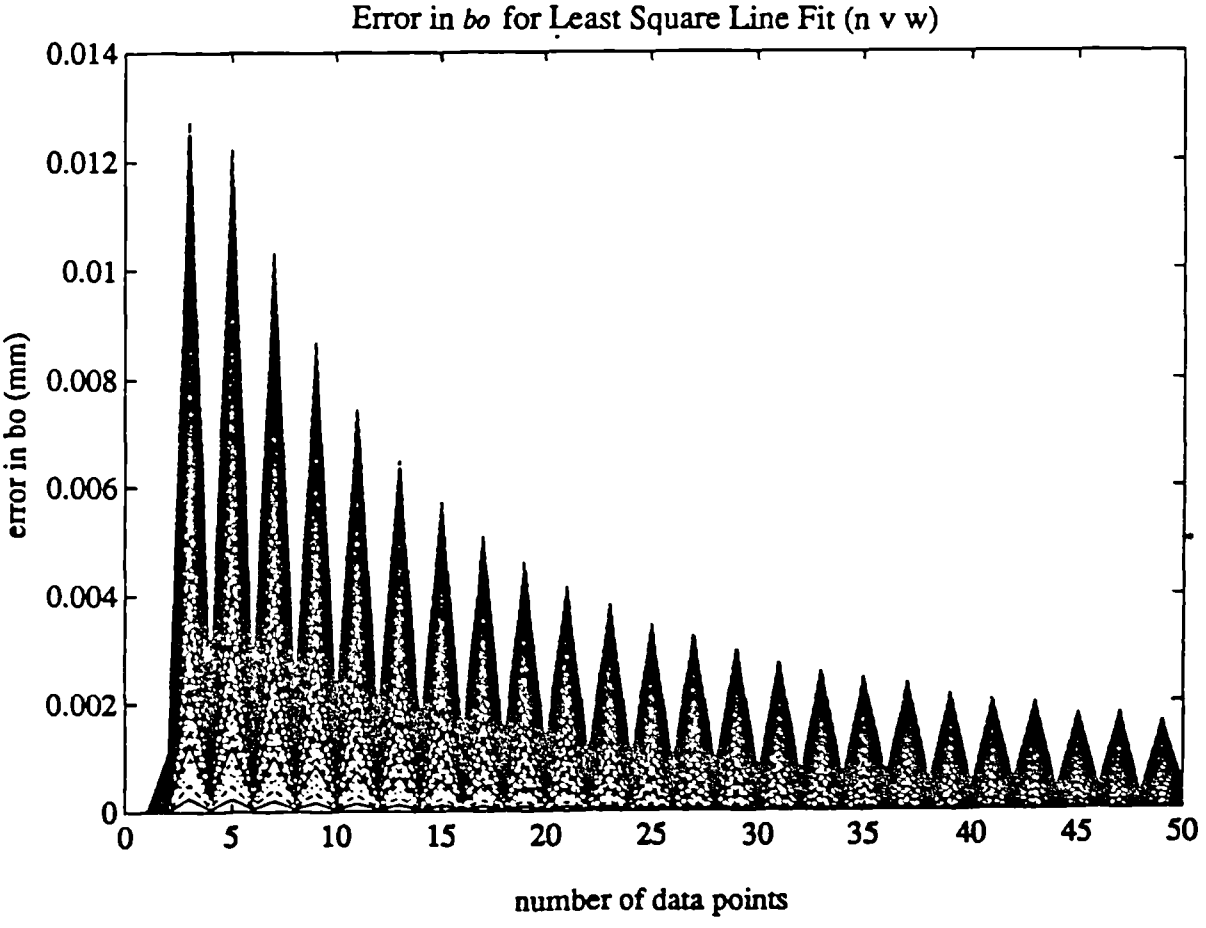


Graph 3.1 (b)

Error in *bo* for Least Square Line Fit (n v w)

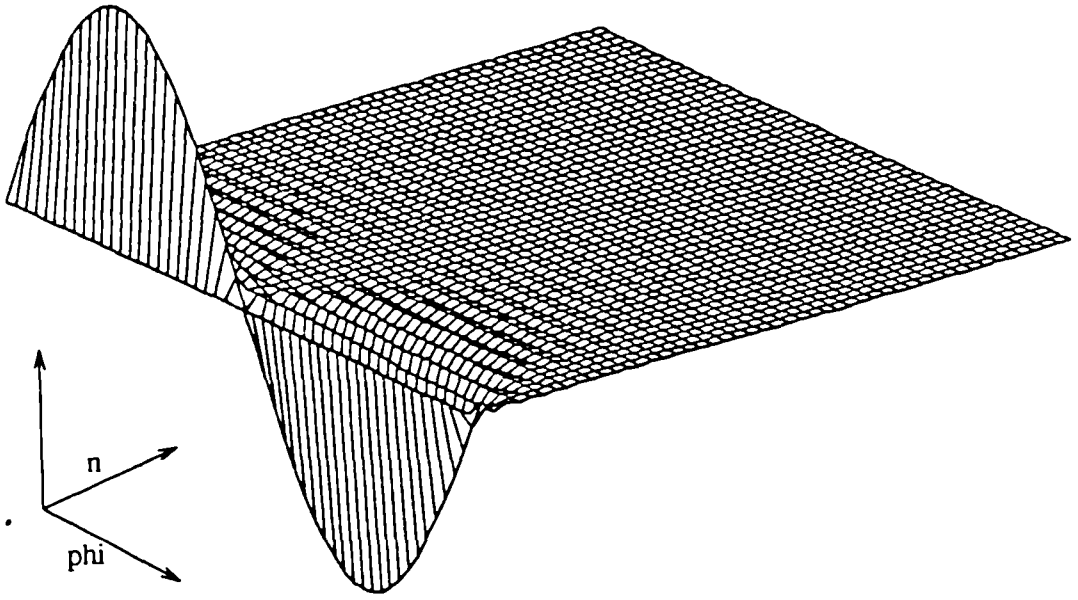


Graph 3.2 (a)

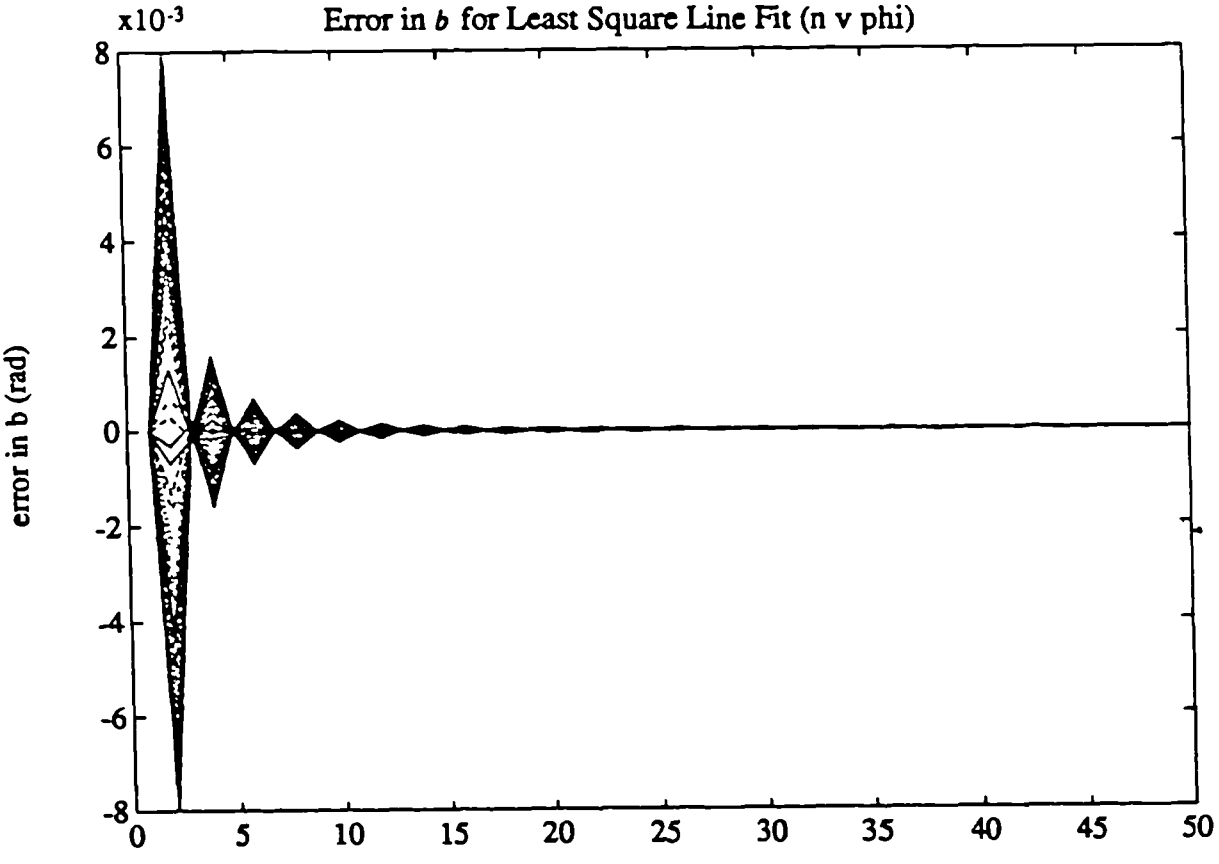


Graph 3.2 (b)

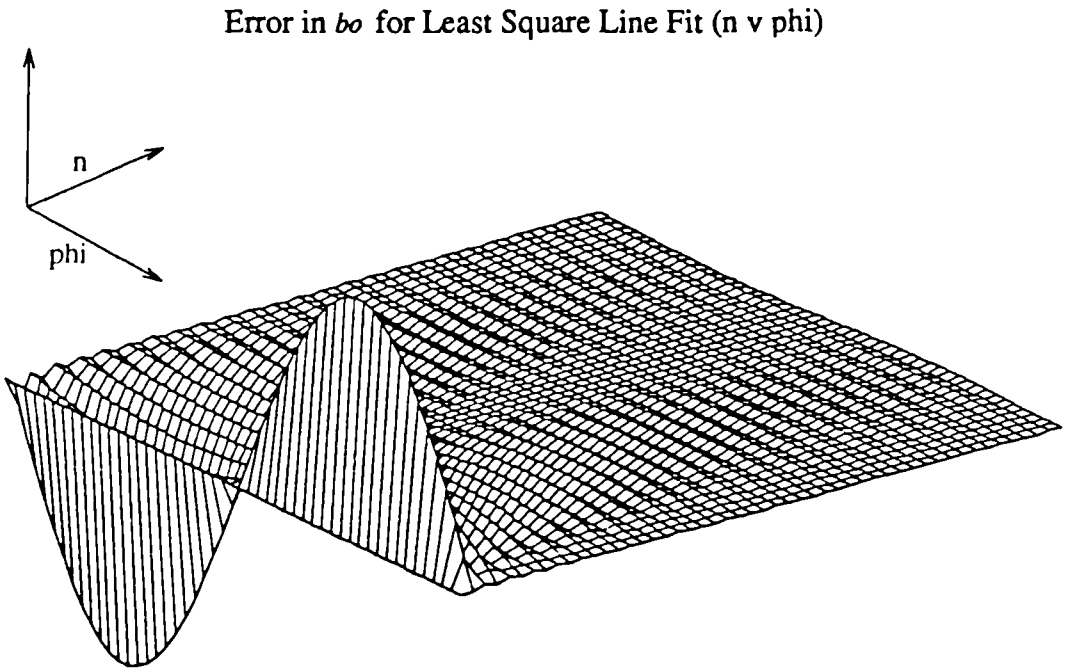
Error in *b* for Least Square Line Fit (n v phi)



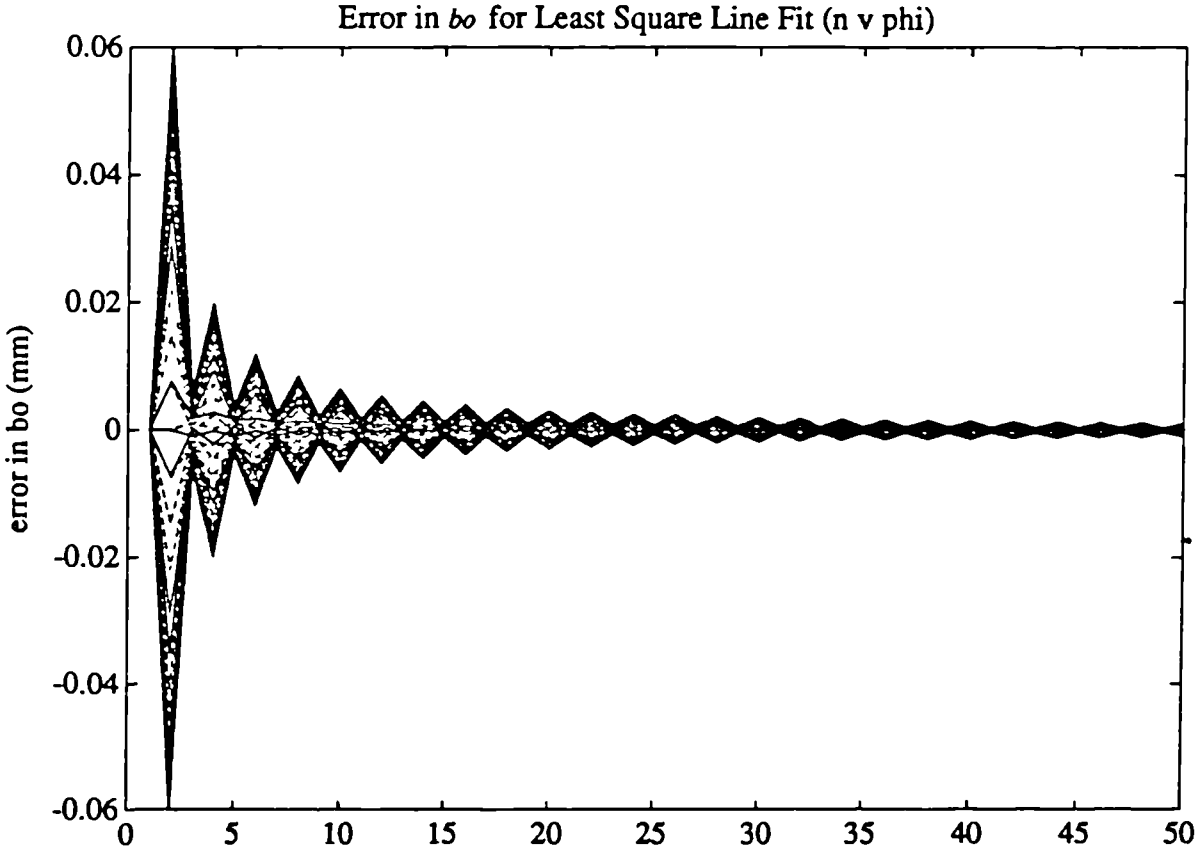
Graph 3.3 (a)



Graph 3.3 (b)

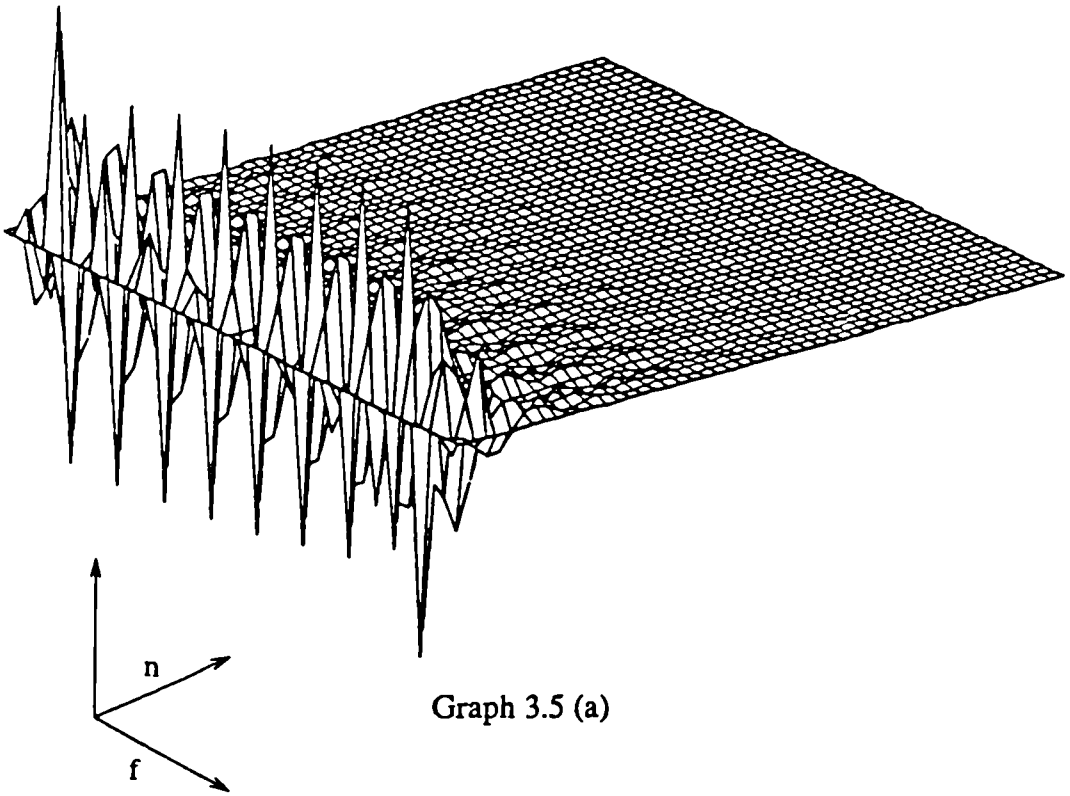


Graph 3.4 (a)

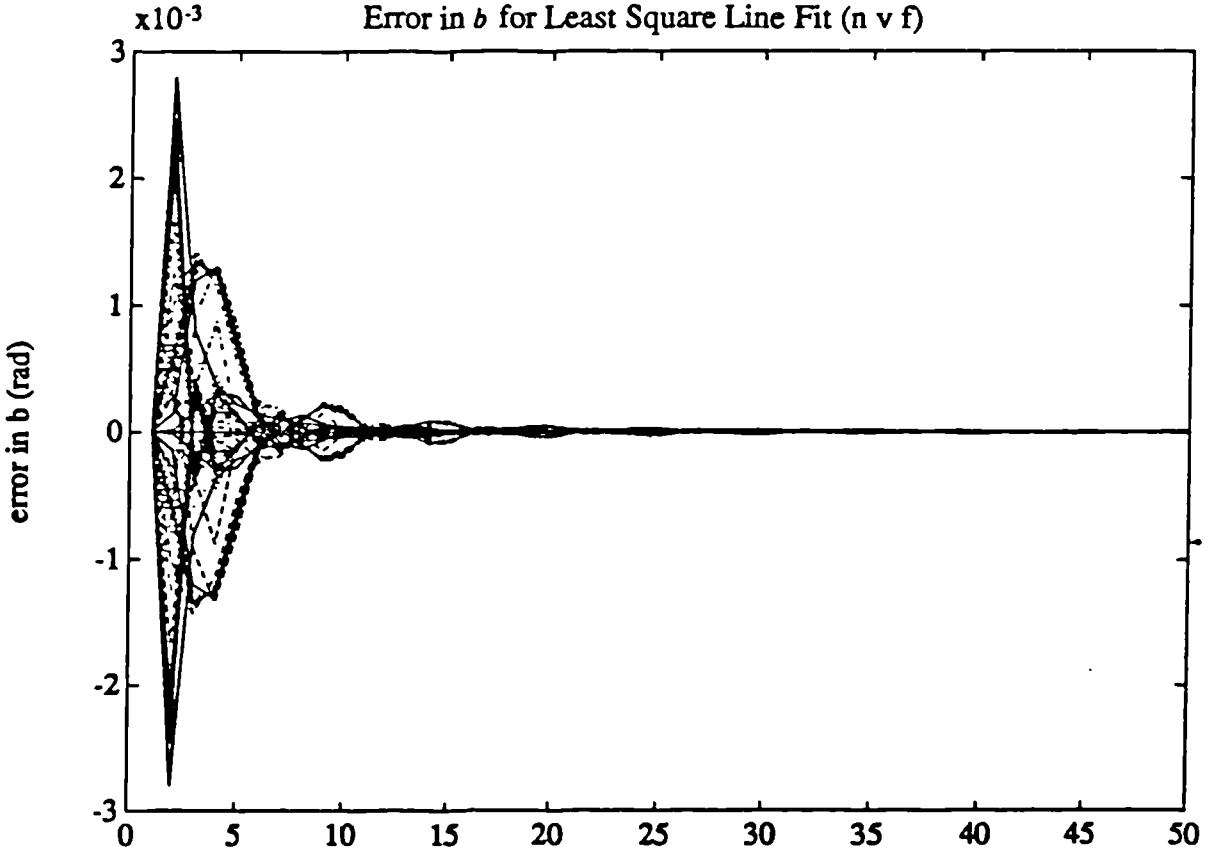


Graph 3.4 (b)

Error in *b* for Least Square Line Fit (n v f)

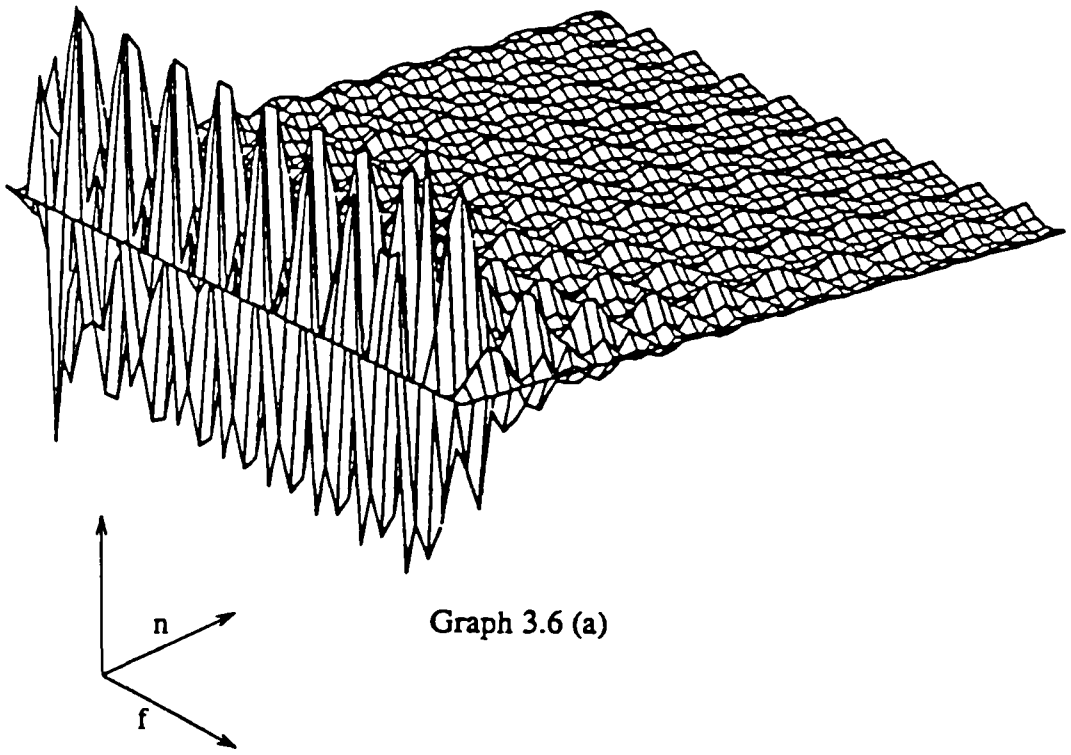


Graph 3.5 (a)

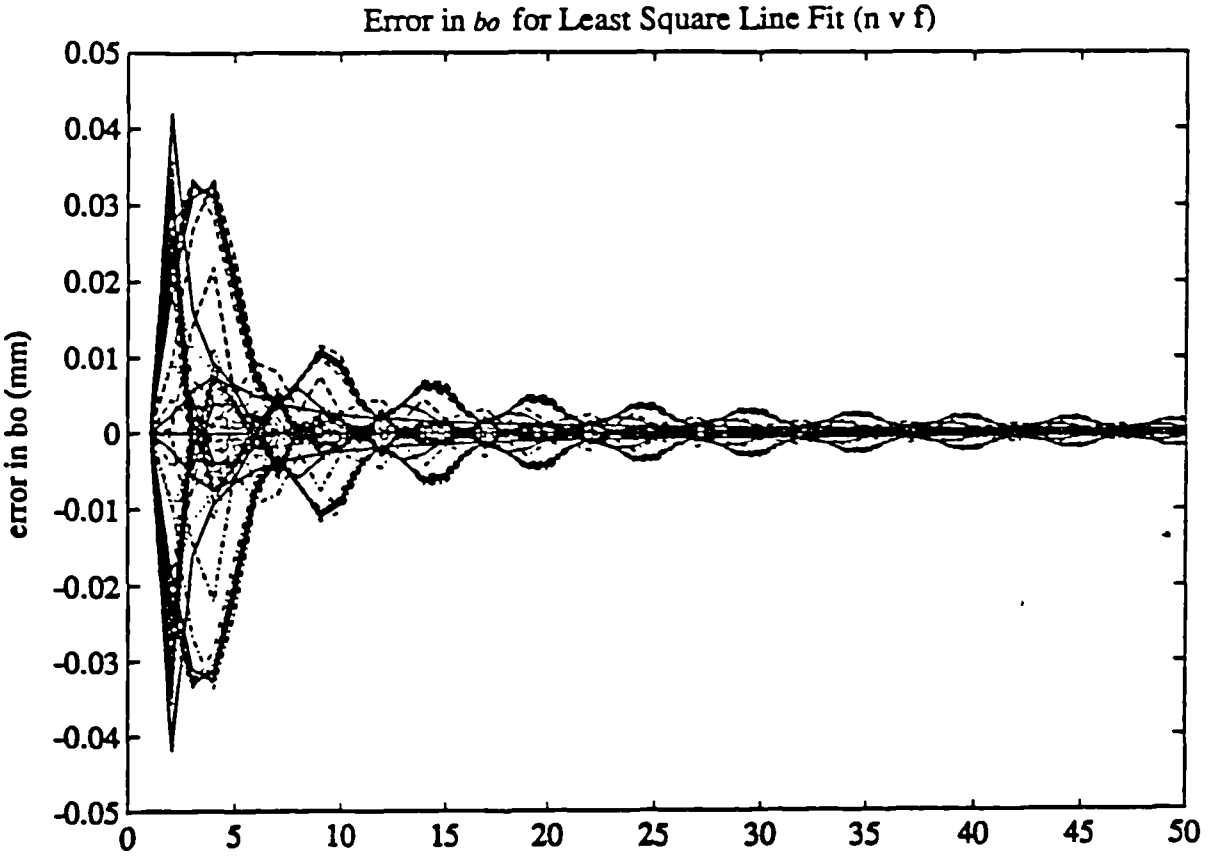


Graph 3.5 (b)

Error in *bo* for Least Square Line Fit (n v f)

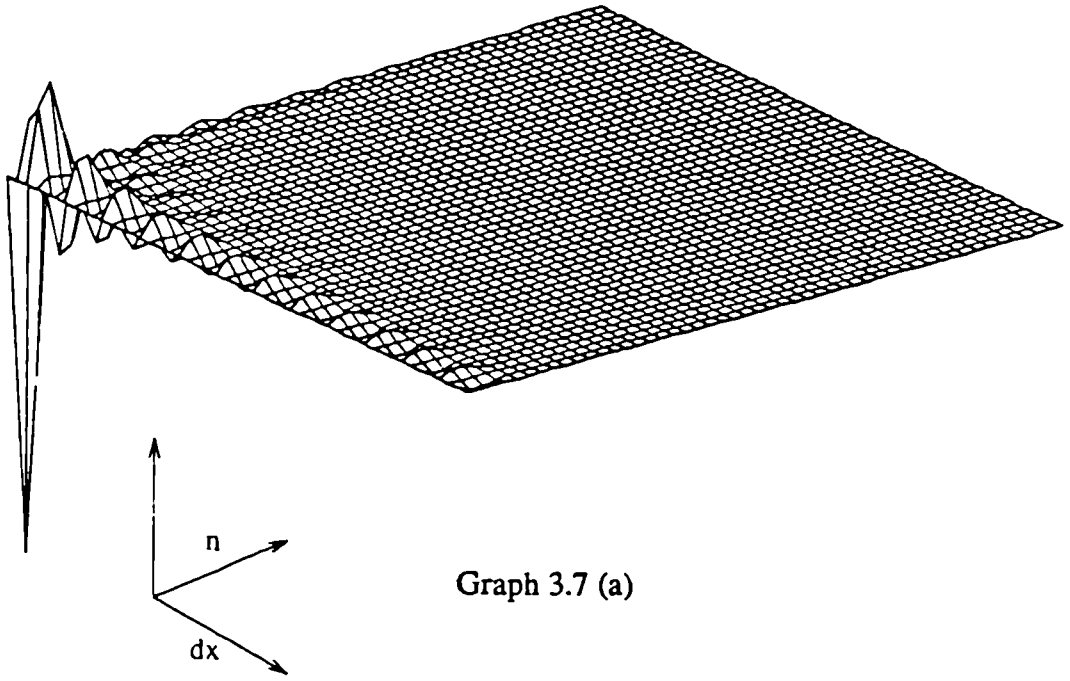


Graph 3.6 (a)

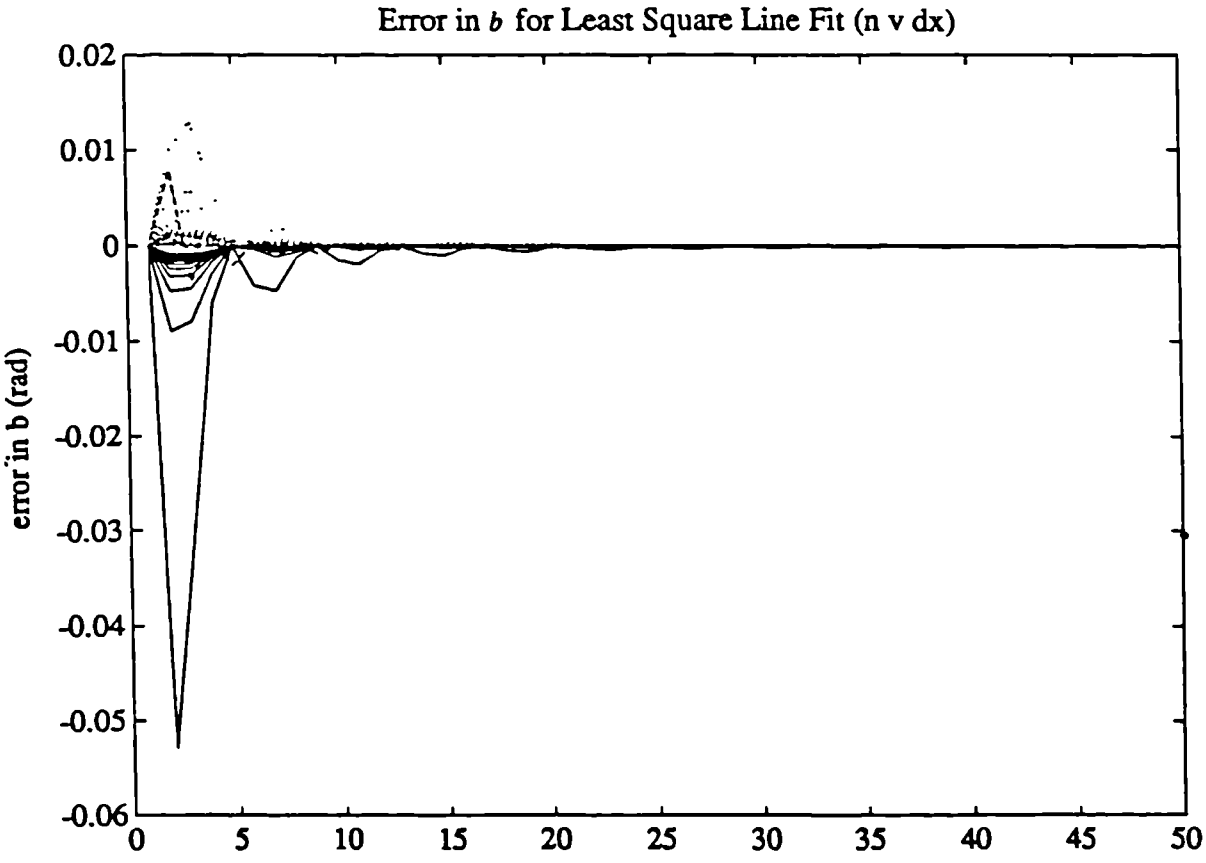


Graph 3.6 (b)

Error in *b* for Least Square Line Fit (n v dx)

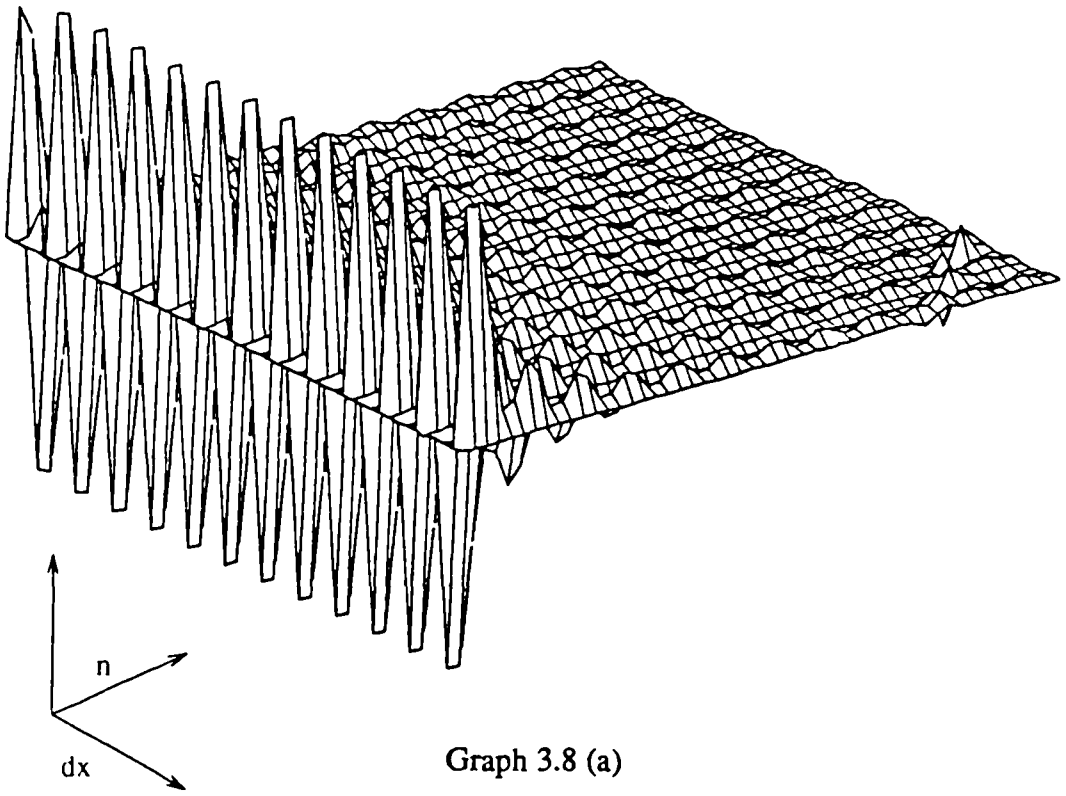


Graph 3.7 (a)



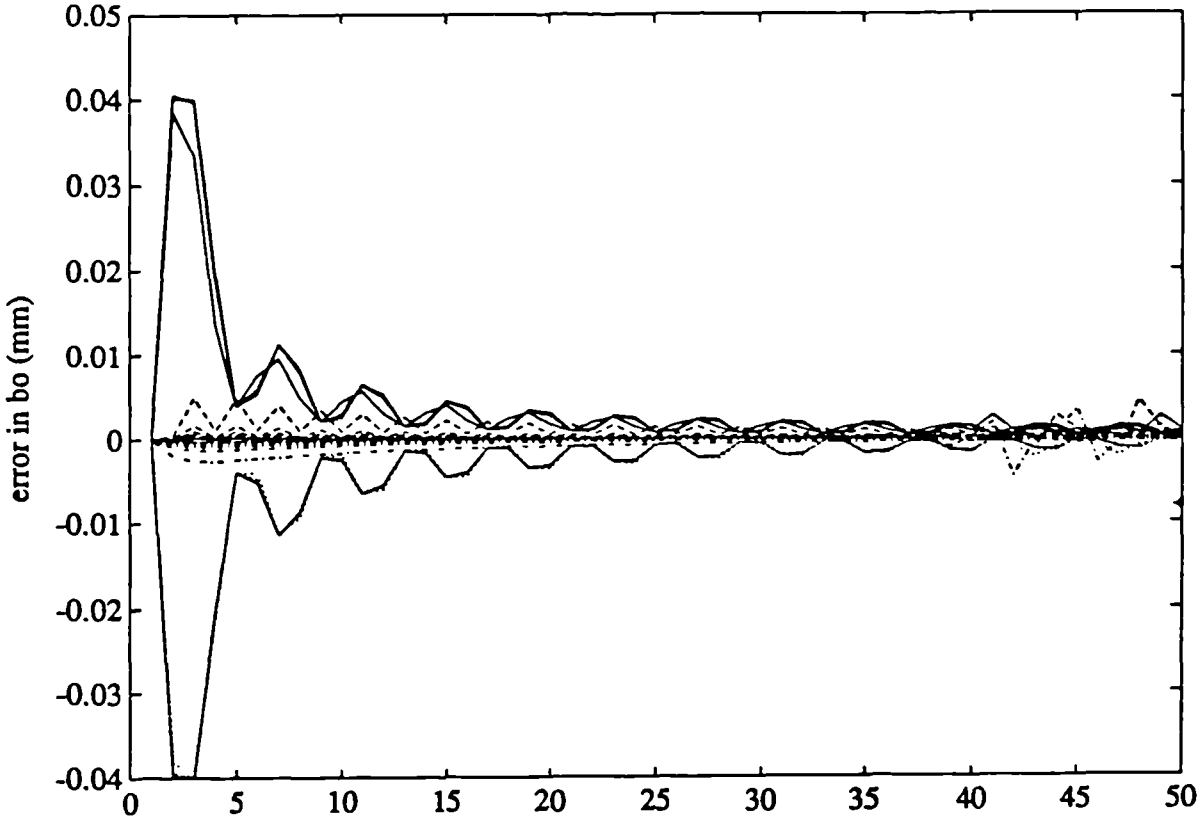
Graph 3.7 (b)

Error in *bo* for Least Square Line Fit (n v dx)



Graph 3.8 (a)

Error in *bo* for Least Square Line Fit (n v dx)



Graph 3.8 (b)

3.1.1.2. Results

On engineering component surfaces, surface characteristics in the order 0.001 to 0.01mm are common (chapter 5 includes a large range of typical engineering surface profiles.) The results illustrate that the sampling errors incurred by surfaces with characteristics within this range of amplitude are in the same order of magnitude as typical machine errors (see summarisation of results in chapter 1.) It must also be borne in mind that the machine errors refered to were measured from a fairly old, uncorrected machine. Existing error compensation techniques have the capacity to reduce these errors substantially, and the significance of sampling error in the overall measuring error would increase accordingly.

With reference to graphs 3.1 to 3.4 it can be seen that the sampling errors decrease steadily with respect to the number of data points used, although there is a sucession of local maxima. Similar behaviour occurs with respect to increasing sampling interval, this effect indicates that in general coordinate measurement practice it would be advisable to distribute the data points over the widest possible expanse of surface in order to minimise error.

Graphs 3.2a to 3.2b illustrate that there is a direct linear relationship between sampling error and the amplitude of surface characterisitics. In graphs 3.3a and 3.3b the cyclic behaviour seen is an excellent illustration of the phenomenon known as aliasing (6). At certain frequencies the data lies upon the waveform in such a way as to effectively represent a waveform of a different frequency. The frequencies at which this phenomenon occurs is dependent upon the sampling interval. This implies that if the samping interval is greater than the Nyquist criterion i.e.

$$dt \leq \frac{1}{2\delta f}$$

(where δf is the highest frequency component present on the surface) then the parameters being calculated will have the same value as those for some other surface of a different frequency. This phenomenon will occur cyclically, the cycle being dependent

upon the sampling interval. The plots of *b* and *bo* against (n v phi) show that the errors are in direct correlation to the location of the distribution of the data on the sinusoid.

Some of the above plots illustrate the effects of quantisation errors, where the peaks of the sampling errors appear to be foreshortened, i.e. the true peak would have appeared at a variable value either side of this truncated peak. The actual peak was not detected due to the fact that the sampling errors were calculated at too infrequent variable values.

3.1.1.3. Conclusions and Recommendations

From these initial investigations it is possible to make some general recommendations for coordinate measurement practice:

(i) Using less than six data points in the calculation of slope and intercept invariably leads to substantial sampling errors.

(ii) Distributing the data over the longest possible distance will serve to minimise errors.

(iii) Surface having characteristics of relatively large amplitude incurr correspondingly high sampling errors

(iv) Information about the harmonic content of the surface profile would be extremely useful in order to avoid sampling at rates that would cause instability in the estimates of the parameters.

3.2. Plane Fit Algorithms

The following section is concerned with the accuracy and reliability of plane fit algorithms, which may be used to assess flatness, but more importantly as a reference figure. In order to take coordinate measurements of a component with any computer assisted measuring machine, a reference plane must first be established. This is usually achieved by taking a number of measurements on the uppermost plane surface of the component. However, the reference plane is a machined surface, and consequently not

perfectly flat. Errors incurred in the determination of a reference plane subsequently create abberations in the inspection accuracy of the geometrical features of the component. Therefore, for accurate part inspection the choice of a suitable algorithm for the determination of a reference plane is critical.

Several types of plane fit algorithms exist; some of the most frequently implemented in coordinate measurement are the following:

(i) The exact fit algorithm; this algorithm consists of taking three points and doing an exact fit of a plane through these points.

(ii) The best of exact fits algorithm; this algorithm considers all possible combinations of the given data points, three at a time. The combination of the three data points which produces the smallest flatness value is selected as being the most suitable reference plane, the exact fit algorithm being used for all combinations of data points. For accurate analysis, it is advised to use of 20 reference points and 1140 appliations of the best fit algorithm (7).

(iii) The average of best fits algorithm;this algorithm considers the best possible combination of three points, but in such a way as the normal vector and the distance to the reference plane is taken as the average of the normal vectors and the distance to the plane of all possible combinations of the data points taken three at a time.

(iv) The root mean square plane (least-squares plane);this algorithm considers all data points and fits a least squares plane through them.

J. Hurt et. al. (7) performed a series of experiments upon the above algorithms in order to determine the most accurate and reliable procedure to use. Hurt's experiments were all performed on a digital computer using data points generated by using exact data points and adding random errors. These errors were a uniform random variable over a cube of side 2e (the quantity e being regarded as machine inaccuracy or surface deviation) centred at an exact data point . The experiment was repeated for values of e from 1 μm to 15 μm in increments of 1μm. Each algorithm was applied to these points and the

comparison made between the errors in :

(i) Each component of the normal vector to the plane

(ii) The Euclidean norm of the error in the distance from the origin of the inspection machine to the reference plane

(iii) The root mean square plane, (the error being defined as the discrepancy between the parameters of the exact and calculated planes.

(iv) The flatness

Of the four algorithms that were tested, the least squares fit was deemed to be by far the best with respect to accuracy, and it is upon this algorithm that attention will now be focussed.

3.2.1. The Least Squares Plane

It is proposed to investigate the properties of plane descriptors (L,M,N,D) by using Eigenvalue, Eigenvector methods (8). The plane may be defined thus:

$$Lx + My + Nz - D = 0 \tag{3.9}$$

Equation (3.9) may be re-written in matrix algebra form:

$$\mathbf{a}^T\mathbf{v} - D = 0 \tag{3.10}$$

Where:

$$\mathbf{v}=(x, y, z)^T$$

and

$$\mathbf{a} = (L, M, N)$$

The residuals are expressed normal to the plane such that the distance of the i th coordinate, defined by $\mathbf{v}_i = (x_i, y_i, z_i)$ from the plane is

$$\begin{aligned} e_i &= Lx_i + My_i + Nz_i - D \\ &= \mathbf{a}^T\mathbf{v}_i - D \end{aligned}$$

Application of the least squares criterion now requires that the sum of the squares of these distances (residuals) be minimised.

$$\begin{aligned} S = \sum e_i^2 &= \sum (\mathbf{a}^T \mathbf{v}_i - D)^2 \\ &= \sum [(\mathbf{a}^T \mathbf{v}_i)^2 - 2D\,\mathbf{a}^T \sum_i \mathbf{v}_i + nD^2 \\ &= \mathbf{a}^T (\sum_i \mathbf{v}_i \mathbf{v}_i^T)\mathbf{a} - 2D\,\mathbf{a}^T \sum_i \mathbf{v}_i + nD^2 \end{aligned} \tag{3.11}$$

To find the optimal solution, differentiating partially with respect to D and equating to zero.

$$\frac{\partial S}{\partial D} = -2\mathbf{a}^T \sum_i \mathbf{v}\mathbf{z}_i + 2nD = 0 \tag{3.12}$$

Therefore

$$D = \frac{\mathbf{a}^T \sum_i \mathbf{v}_i}{n} \tag{3.13}$$

Now, in order to simplify the problem, and to aid in the numerical stability of the calculations, it is desirable to move the origin of the coordinate system to the centroid of the data; i.e. the elements of $\mathbf{v}_i$ under go the following transformation to $\mathbf{v}'_i = (x'_i, y'_i, z'_i)$.

$$\begin{aligned} x'_i &= x_i - \bar{x} \\ y'_i &= y_i - \bar{y} \\ z'_i &= z_i - \bar{z} \end{aligned}$$

therefore

$$\sum_i \mathbf{v}_i = 0$$

so that

$$S = \sum e_i^2 = \mathbf{a}^T (\sum \mathbf{v}_i \mathbf{v}_i^T)\mathbf{a} = \mathbf{a}^T \mathbf{V}\mathbf{a}$$

Where

$$\mathbf{V} = \sum_i \mathbf{v}'_i \mathbf{v}'^T_i = \begin{bmatrix} \sum x'^2_i & \sum x'_i y'_i & \sum x'_i z'_i \\ \sum x'_i y'_i & \sum y'^2_i & \sum y'_i z'_i \\ \sum x'_i z'_i & \sum y'_i z'_i & \sum z'^2_i \end{bmatrix} \tag{3.14}$$

However, the minimisation of S must take into account the constraint

$$\mathbf{a}^T \mathbf{a} = 1$$

or

$$\mathbf{a}^T\mathbf{a} - 1 = 0$$

i.e. the solutions are orthogonal.

An appropriate way of incorporating constraints is by the use of Lagrange multipliers, thus:

$$S' = \mathbf{a}^T\mathbf{a} - \lambda(\mathbf{a}^T\mathbf{a} - 1)$$

Where $-\lambda$ is a Lagrange multiplier

$$S' = \mathbf{a}^T(\mathbf{V} - \lambda I)\mathbf{a} + \lambda \tag{3.15}$$

To find optimal values; partially differentiating with respect to L,M and N.

$$\Delta S' = 2(\mathbf{V} - \lambda I)\mathbf{a} \tag{3.16}$$

$$\Delta = \left[\frac{\partial}{\partial L}, \frac{\partial}{\partial M}, \frac{\partial}{\partial N}\right] \tag{3.17}$$

In order for the above equation to be valid $\mathbf{a}$ must be an eigenvector of $\mathbf{V}$ and λ the corresponding eigenvalue.

Therefore to find $\mathbf{a}$ it is necessary to find the minimum eigenvalue of $\mathbf{V}$ and its corresponding eigenvector. This was achieved using MATLAB.

The parameter D was calculated in the following way; the values of L, M, and N calculated as per the above method the transformation to the centroid coordinate system was reversed and using equation (3.13).

The resultant value was taken as being the mean of the vector D.

To ensure significant results, the validity of this algorithm must be checked. This was done by inputing particularly simple permutations of coordinate data, for which the values of L, M, and N are known. For the purposes of this test the following data were used:

$$x = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \quad y = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \quad z = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The algorithm was run with the following result :

$$\begin{bmatrix} L \\ M \\ N \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

This result is as expected, therefore the algorithm is valid.

3.2.2. Investigation into the Sampling Error Incurred in the Least-Squares Estimates of Plane Parameters.

With the exception of the work described in section 3.3, the analysis of systematic and random errors in the estimation of plane parameters appears to have remained relatively neglected. The following is an investigation into the sampling error incurred in these estimates. The simulated surface is sinusoidal in both planes. The parameters used in the experiments are listed with the resulting error plots.

3.2.2.1. Experimental Procedure

The simulated surface data used in the following experiments consisted of a plane described by two sinusoidal elements oriented along the X and Y axes respectively. The sampling errors incurred in the estimates of plane parameters were assessed in the following way; the parameters were calculated using data arranged in a straight lines of equal numbers of equispaced data. A minimum number of two lines were used to avoid ill-conditioning. Another set of parameters was then calculated upon a surface of similar description to that used in the above, using a large quantity of coordinate data ($2/mm^2$) on a plate of 100mm × 100mm, the latter estimate is regarded as the 'actual' parameter value, and is subtracted from the former, the result is deemed to be the sampling error. This process was performed using between two and fifty data sets, each consisting of 5 data points and varying each of the input parameters in turn, against varying number of data sets, all other parameters remaining constant. The constant values and variable range of all input parameters used in the experiments are given in table 3.2, where

wy = amplitude of sinusoid oriented along the Y axis

wx = amplitude of sinusoid oriented along the X axis

phiy = phase of sinusoid oriented along the Y axis

phix = phase of sinusoid oriented along the X axis

fy = frequency of sinusoid oriented along the Y axis

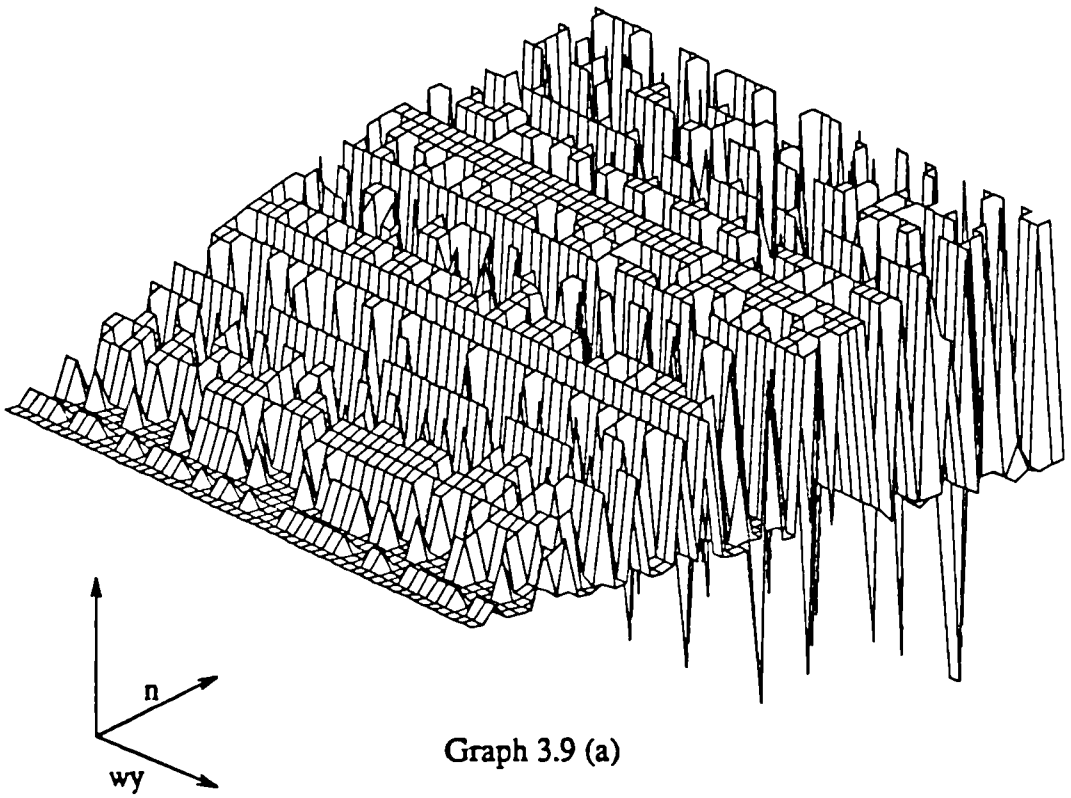
fx = frequency of sinusoid oriented along the X axis

dx= spacing of data points

Simulated Data Parameters		
Parameter	Constant	Variable
wy	0.020(mm)	0.005:0.25(mm)
wx	0.001(mm)	0.005:0.25(mm)
phiy	0.000(rad)	0:2π(rad)
phix	0.000(rad)	0:2π(rad)
fx	0.060(Hz)	0.020:1(Hz)
fy	0.300(Hz)	0.020:1(Hz)
dx	20 (mm)	0.500:25.000 (mm)

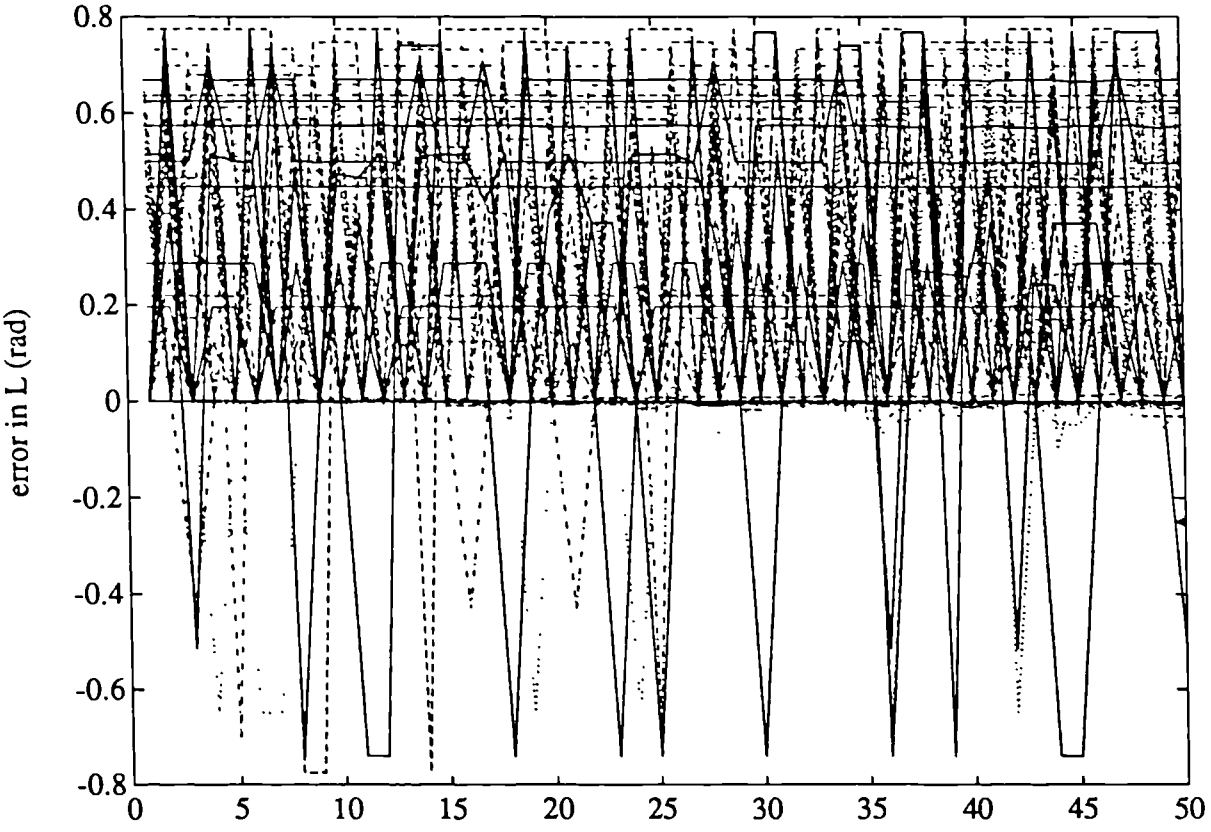
Table 3.2

Sampling Error in L (wy v n)

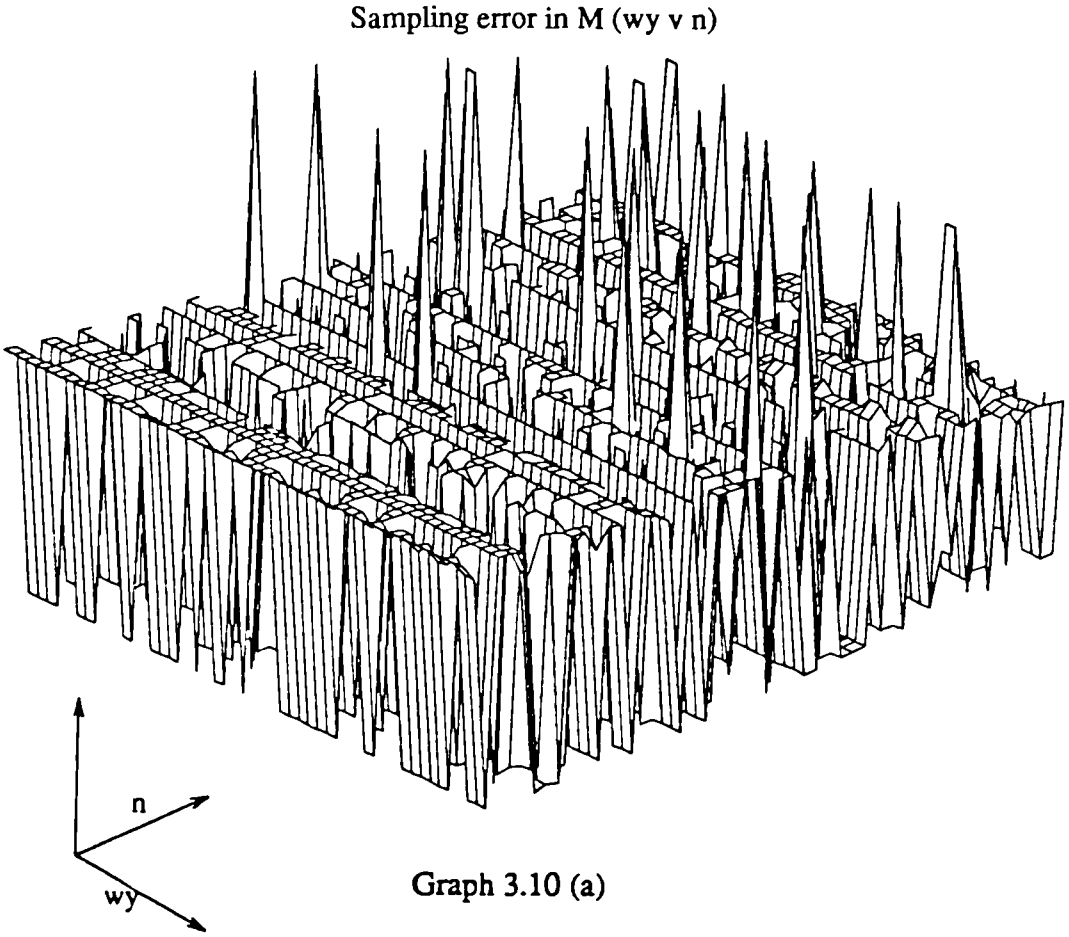


Graph 3.9 (a)

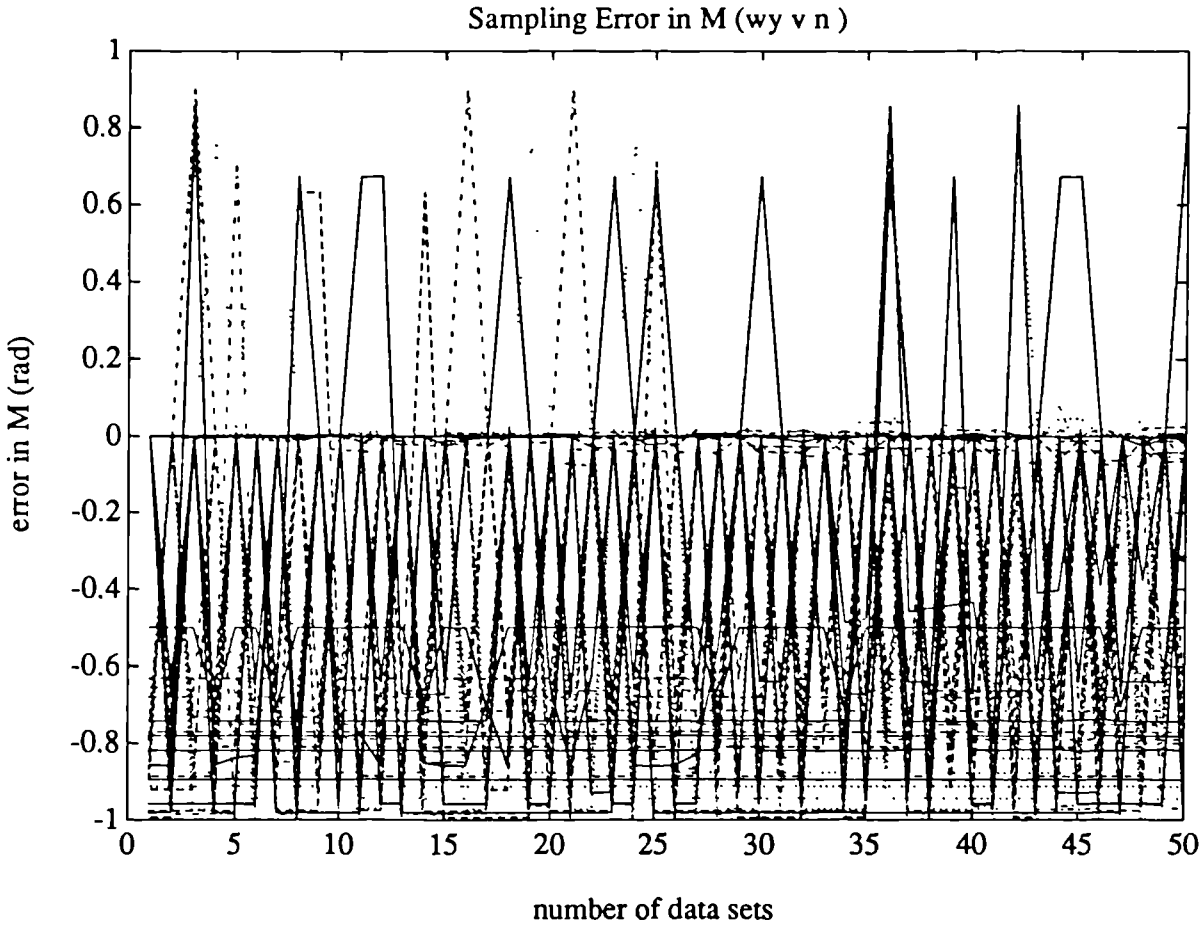
Sampling Error in L (wy v n)



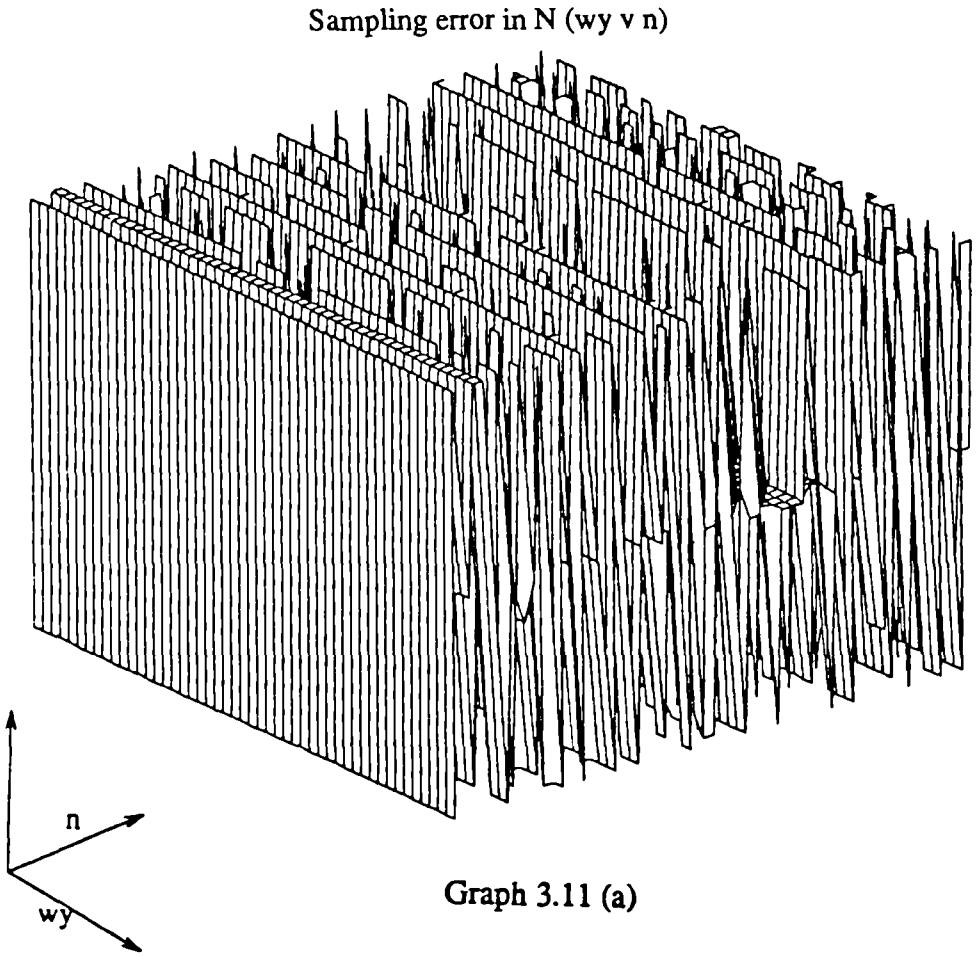
Graph 3.9 (b)



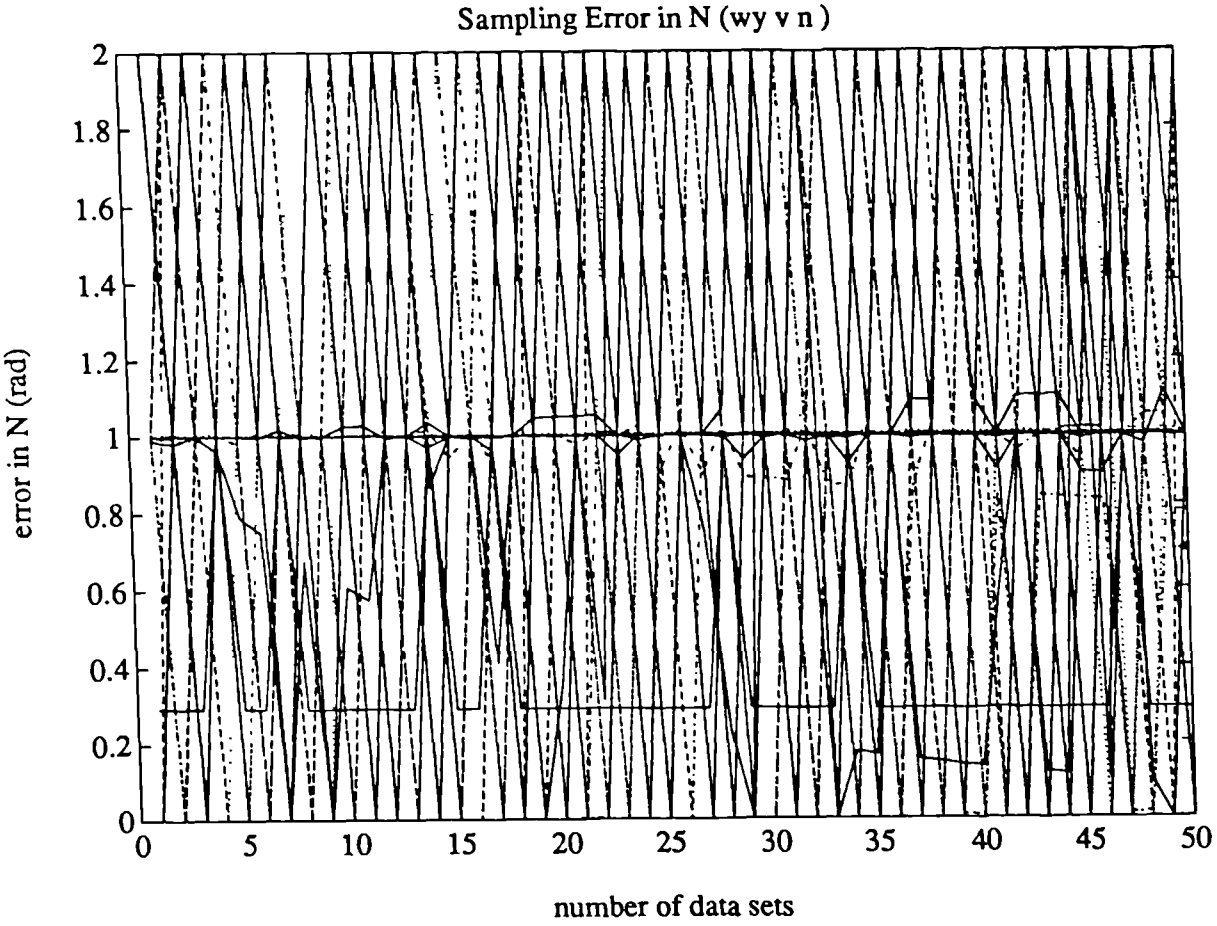
Graph 3.10 (a)



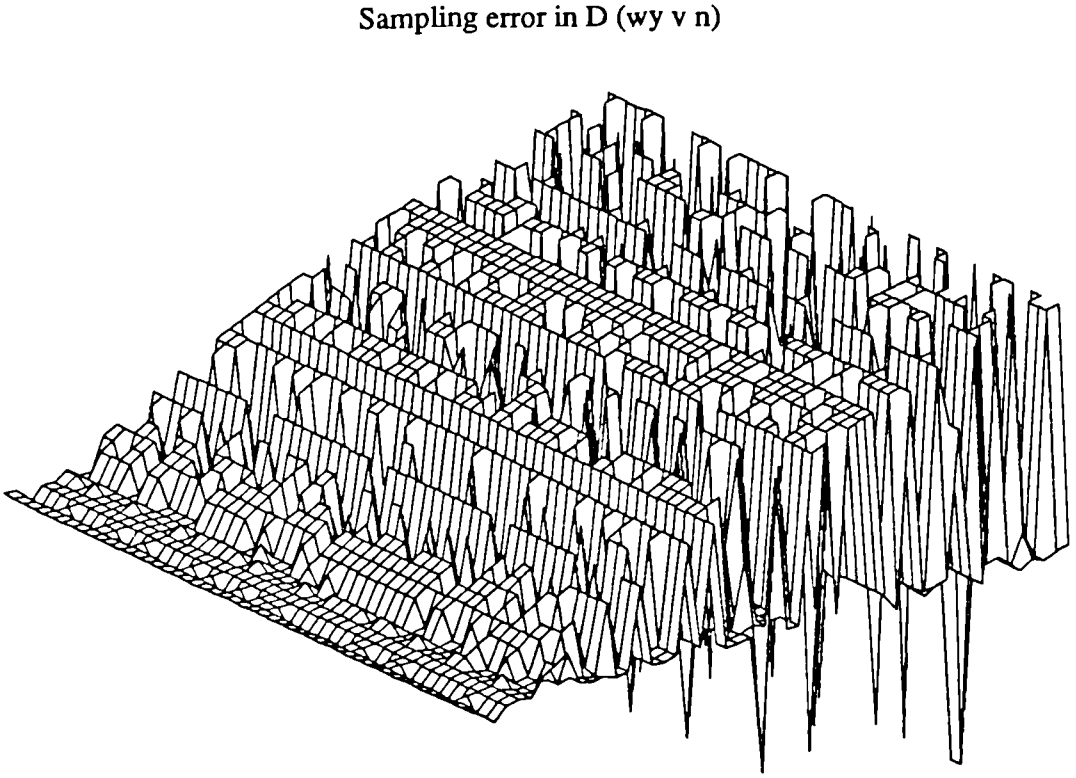
Graph 3.10 (b)



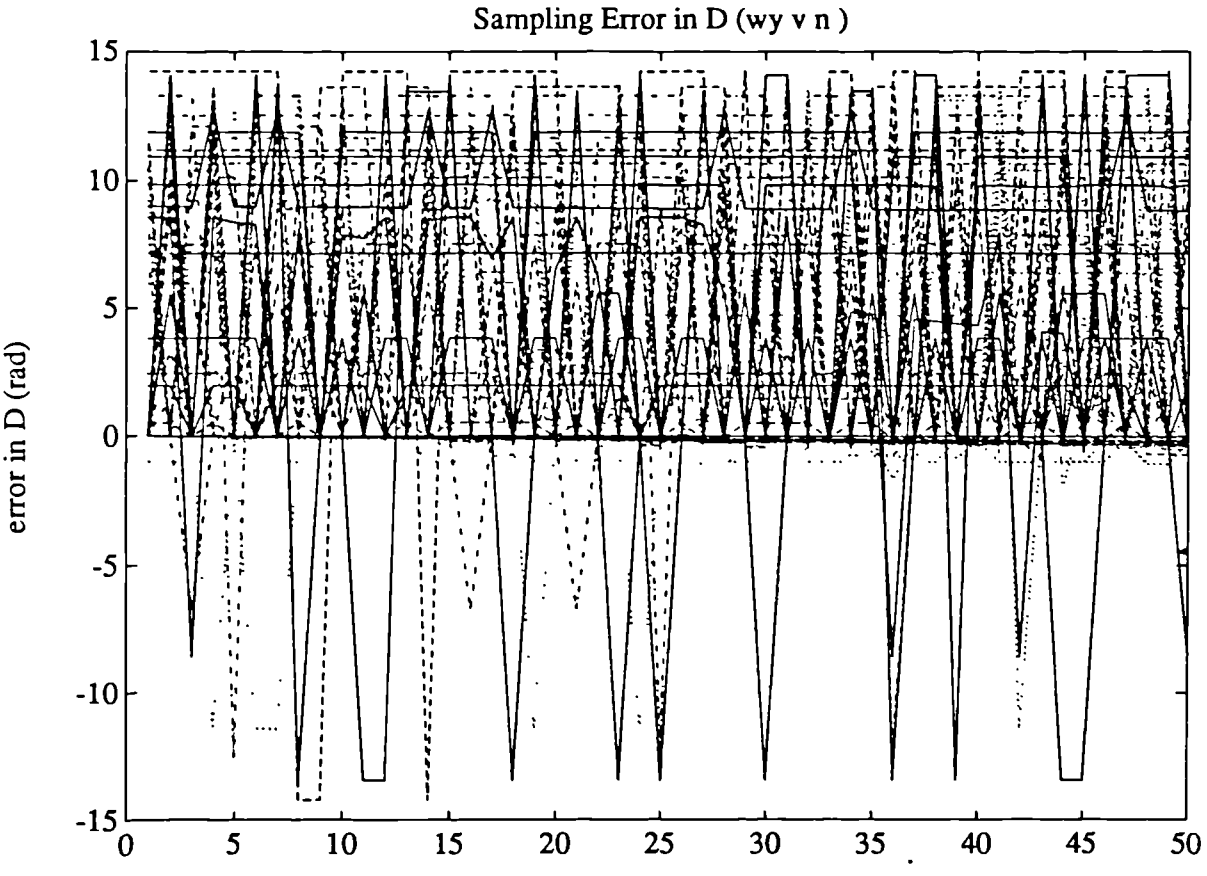
Graph 3.11 (a)



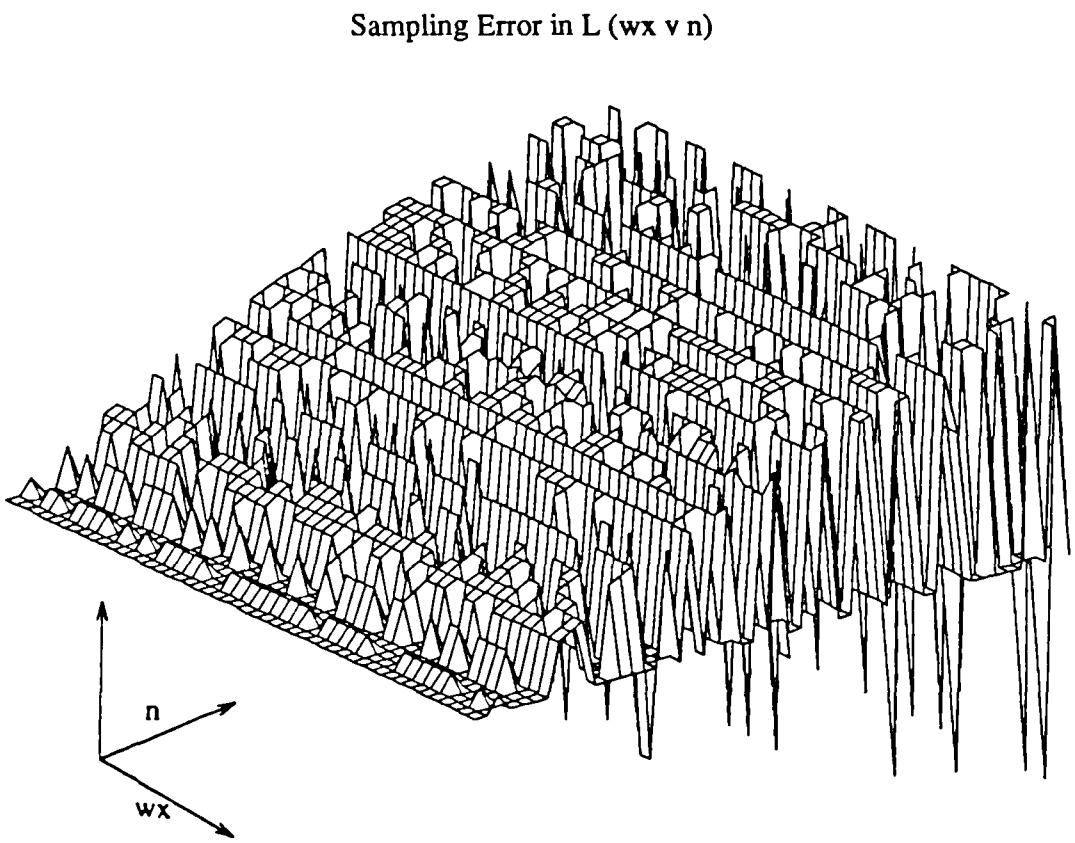
Graph 3.11 (b)



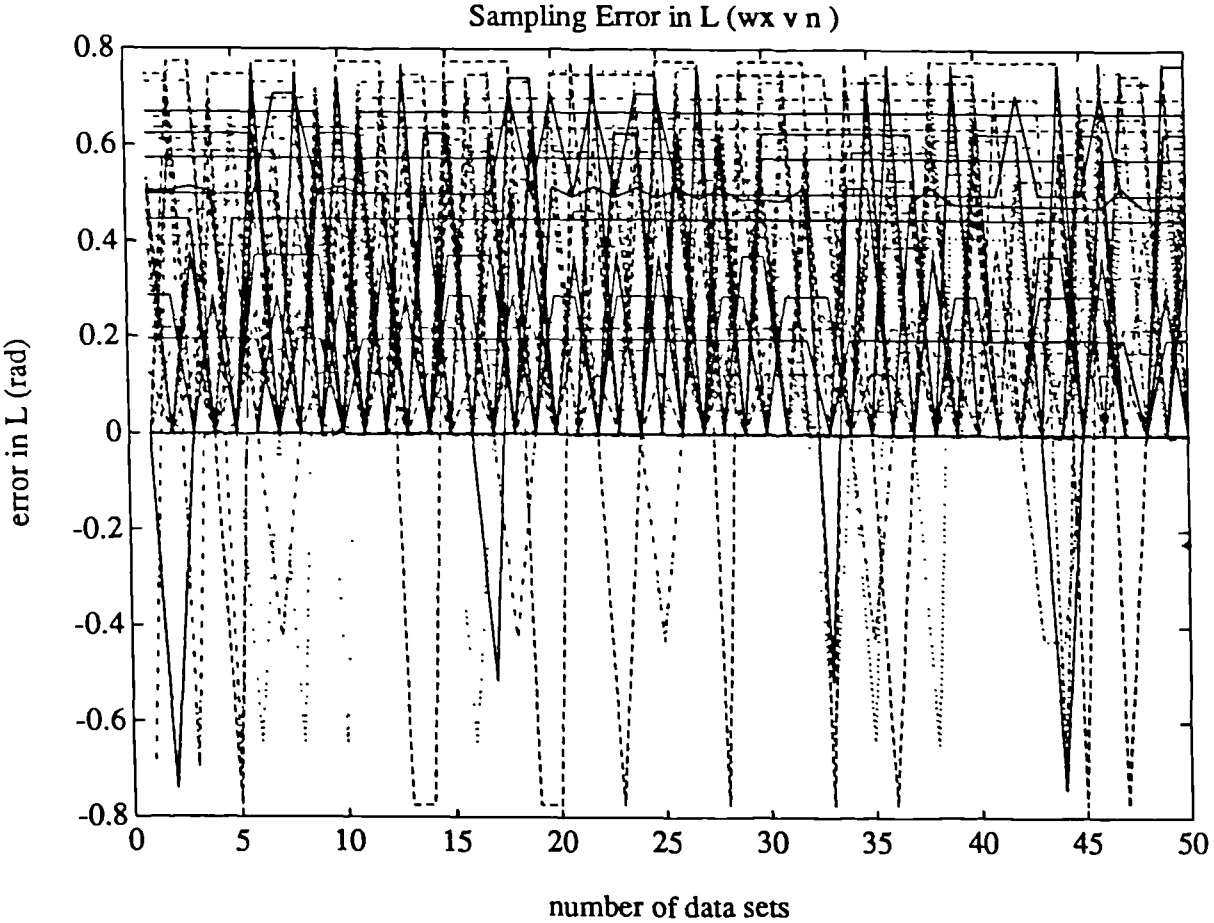
Graph 3.12 (a)



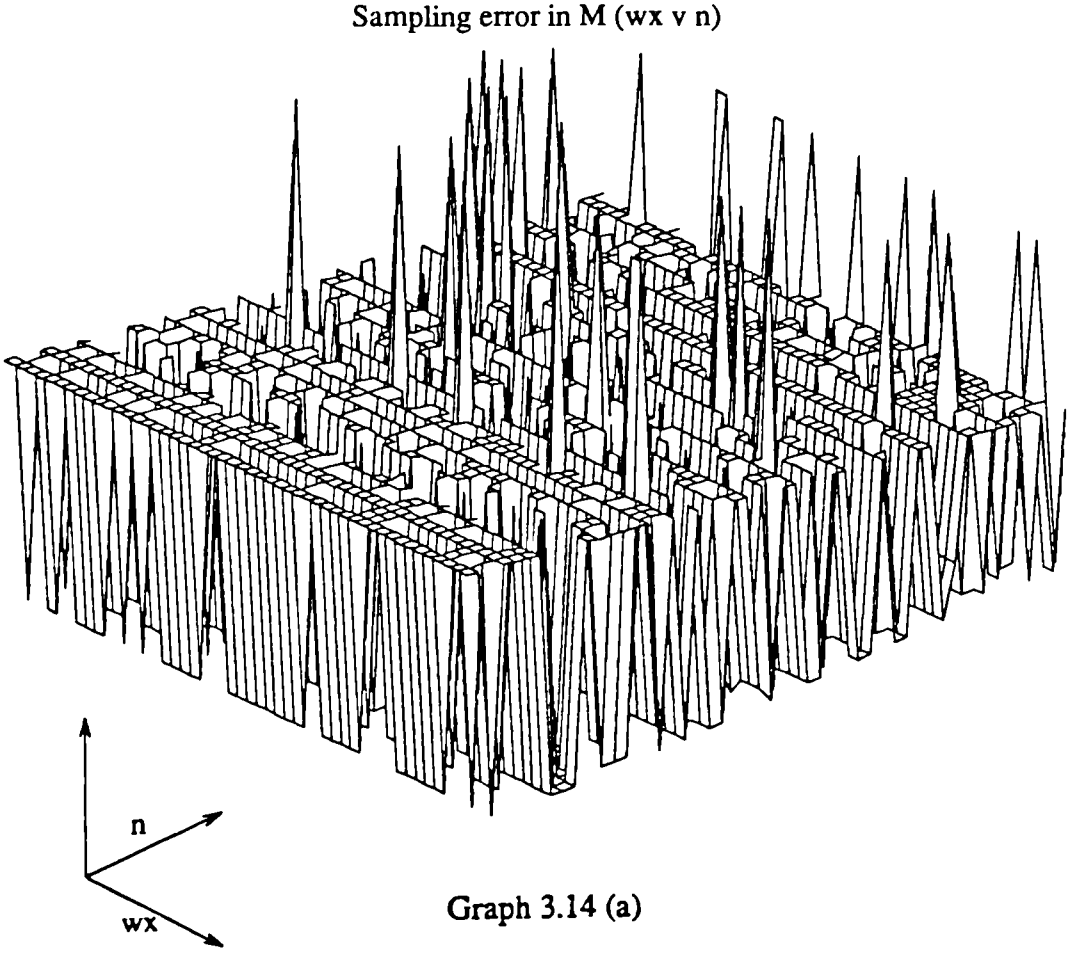
Graph 3.12 (b)



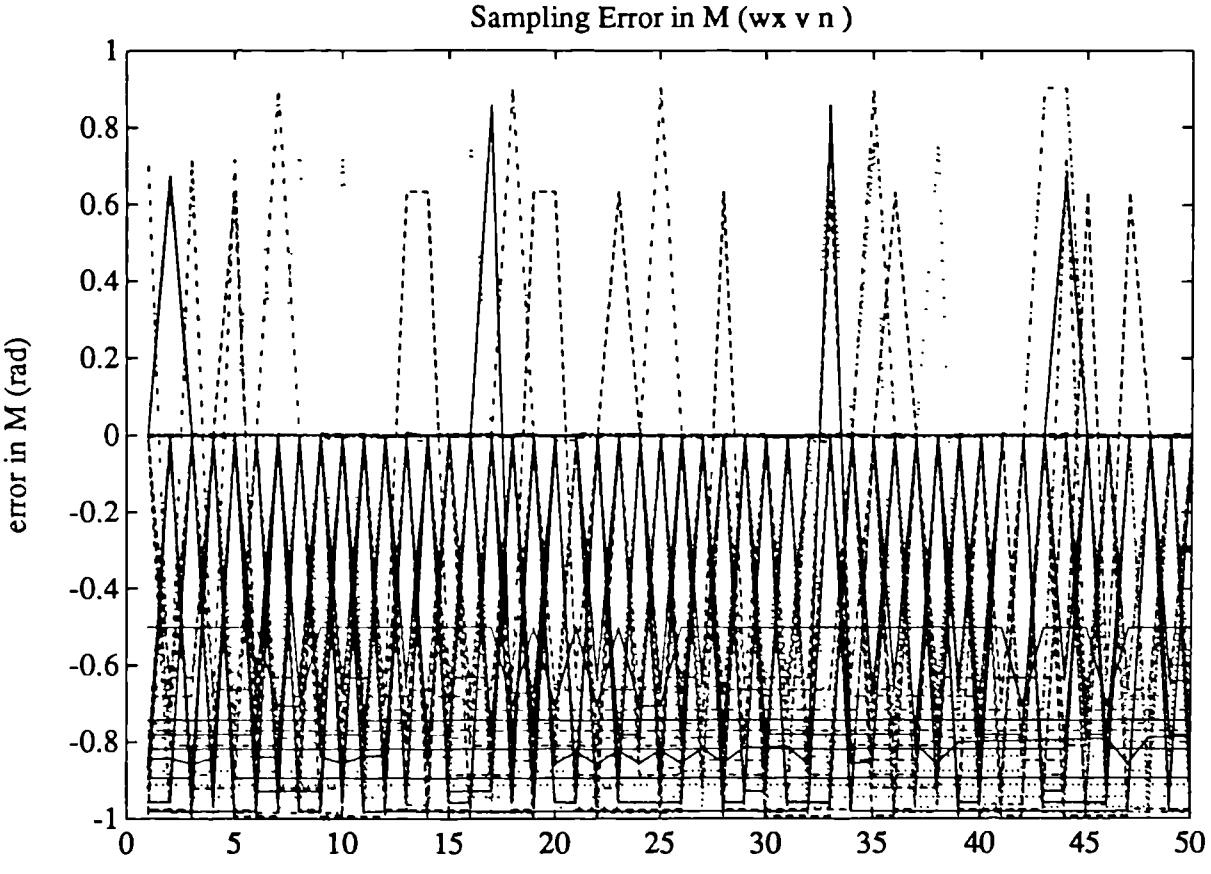
Graph 3.13 (a)



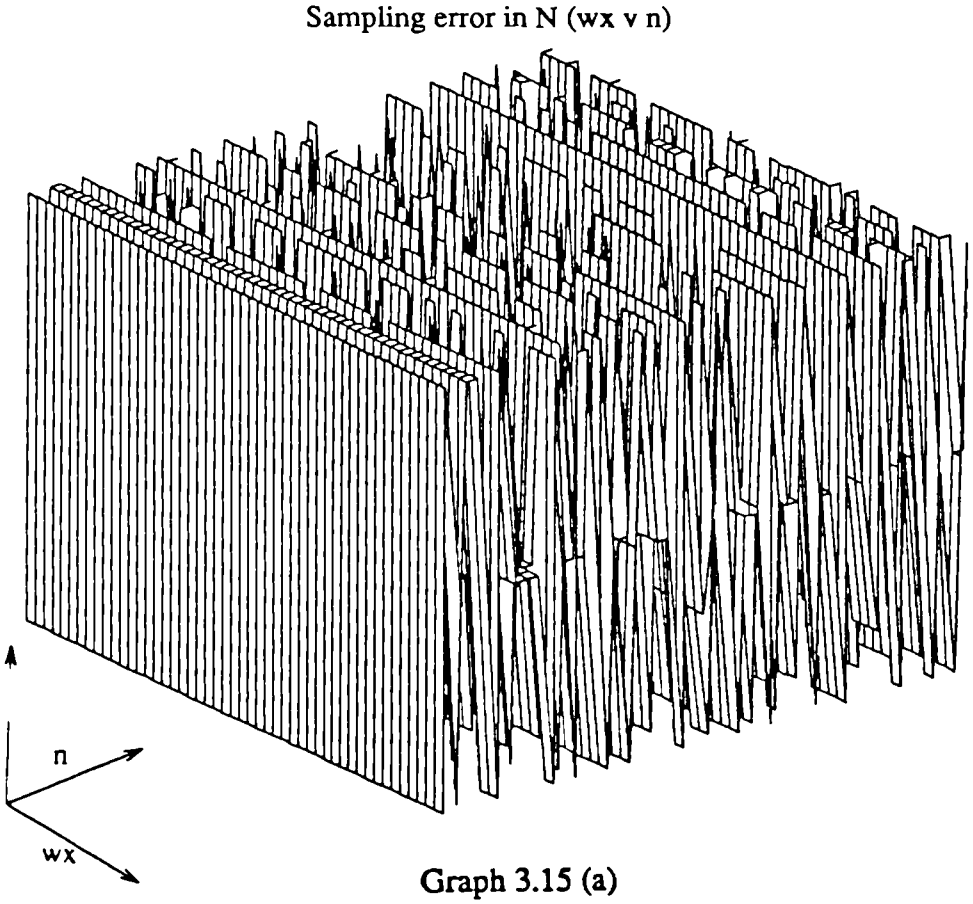
Graph 3.13 (b)



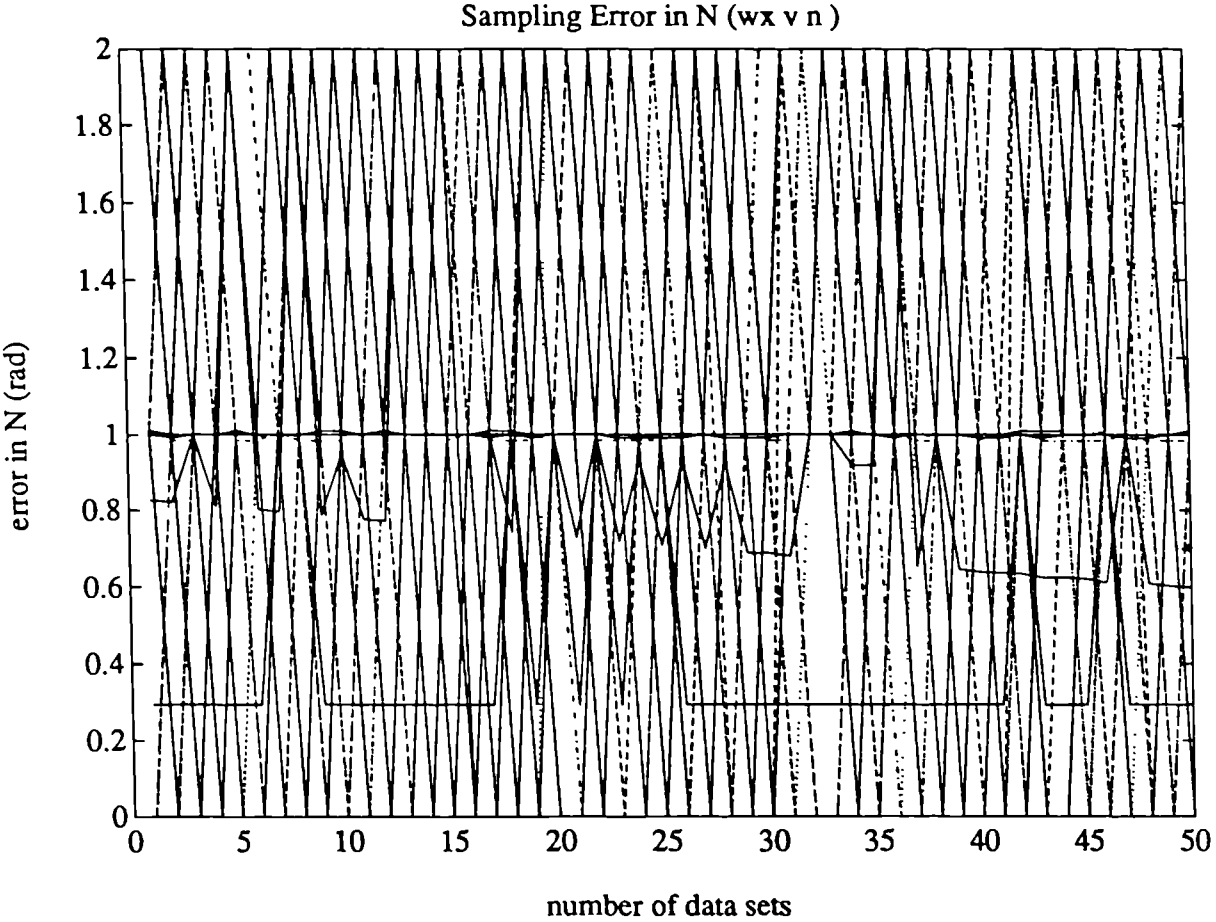
Graph 3.14 (a)



Graph 3.14 (b)

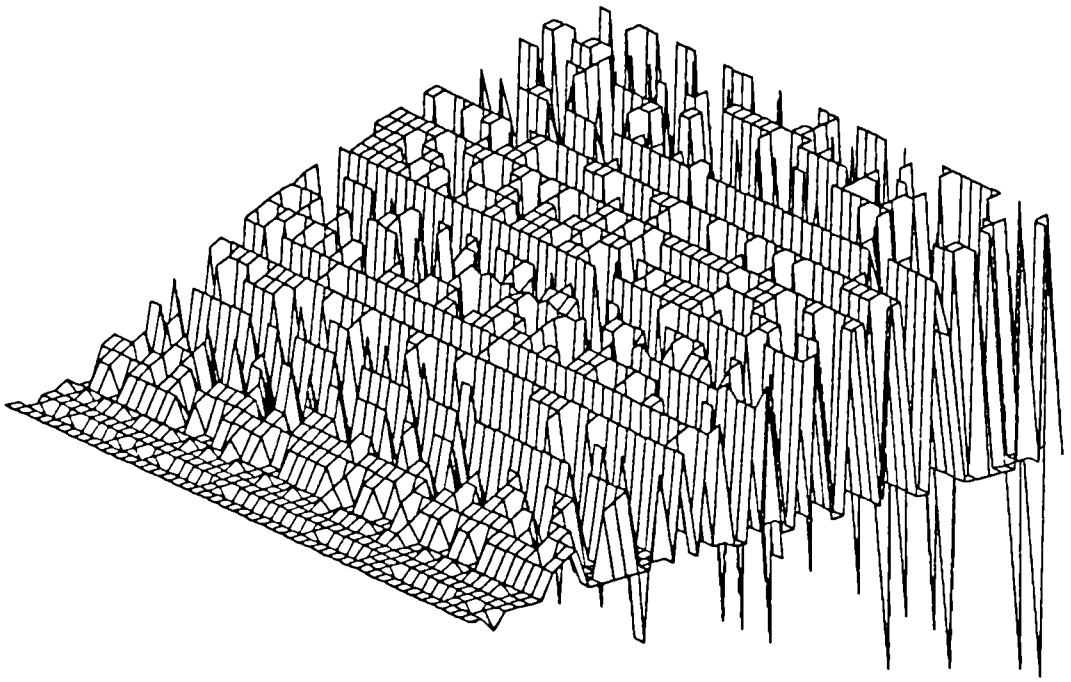


Graph 3.15 (a)

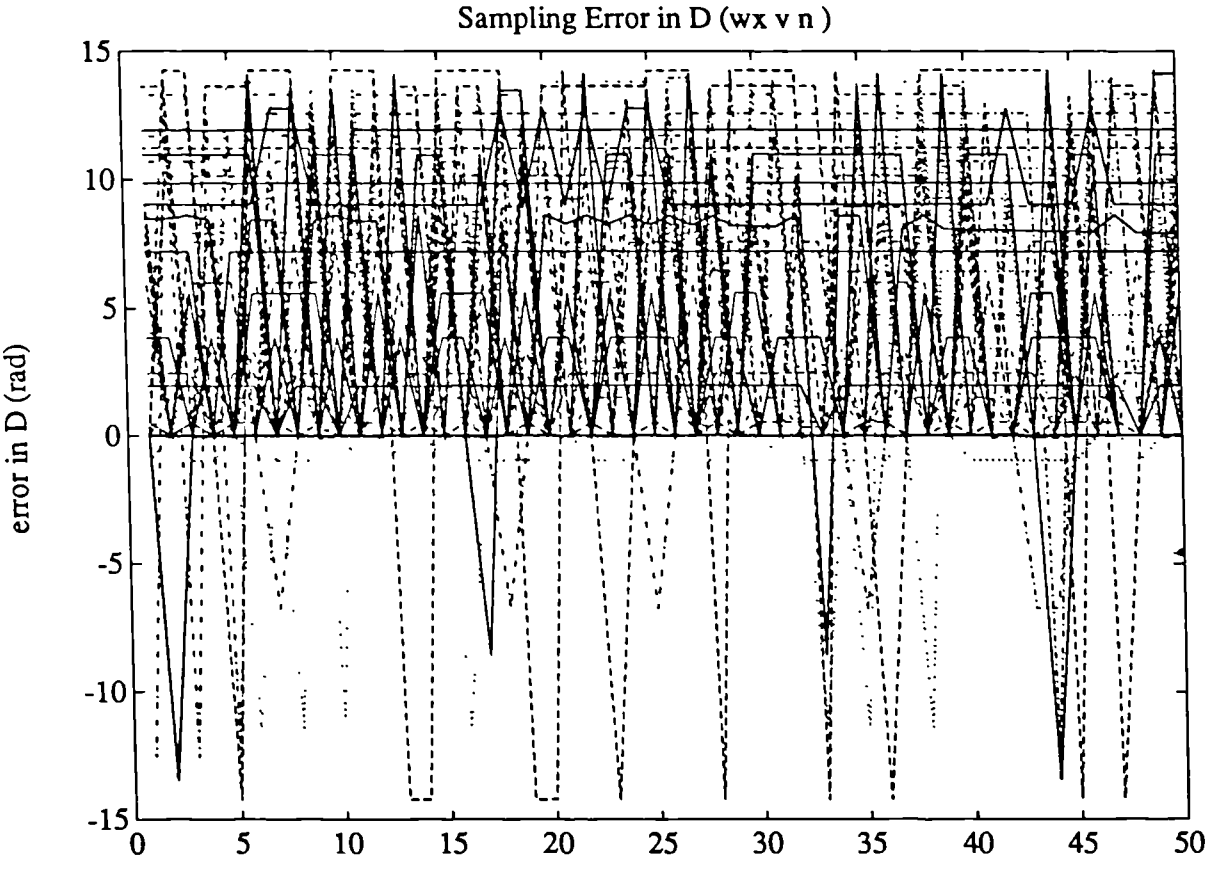


Graph 3.15 (b)

Sampling error in D (wx v n)

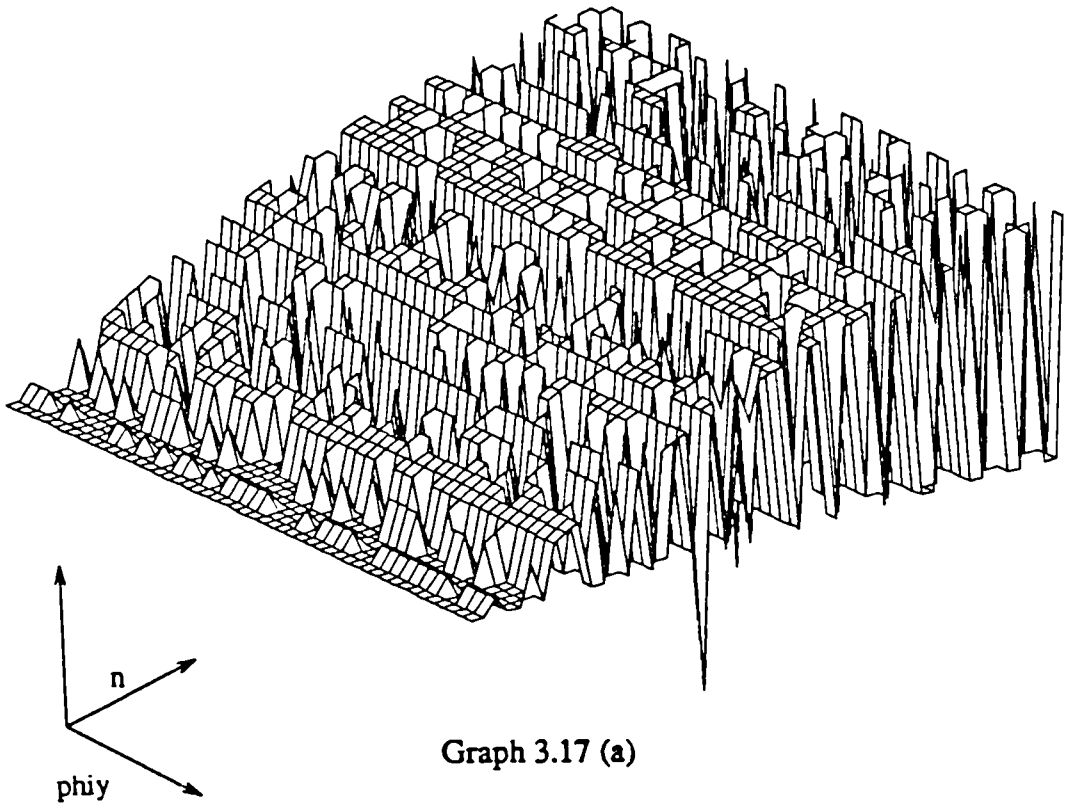


Graph 3.16 (a)

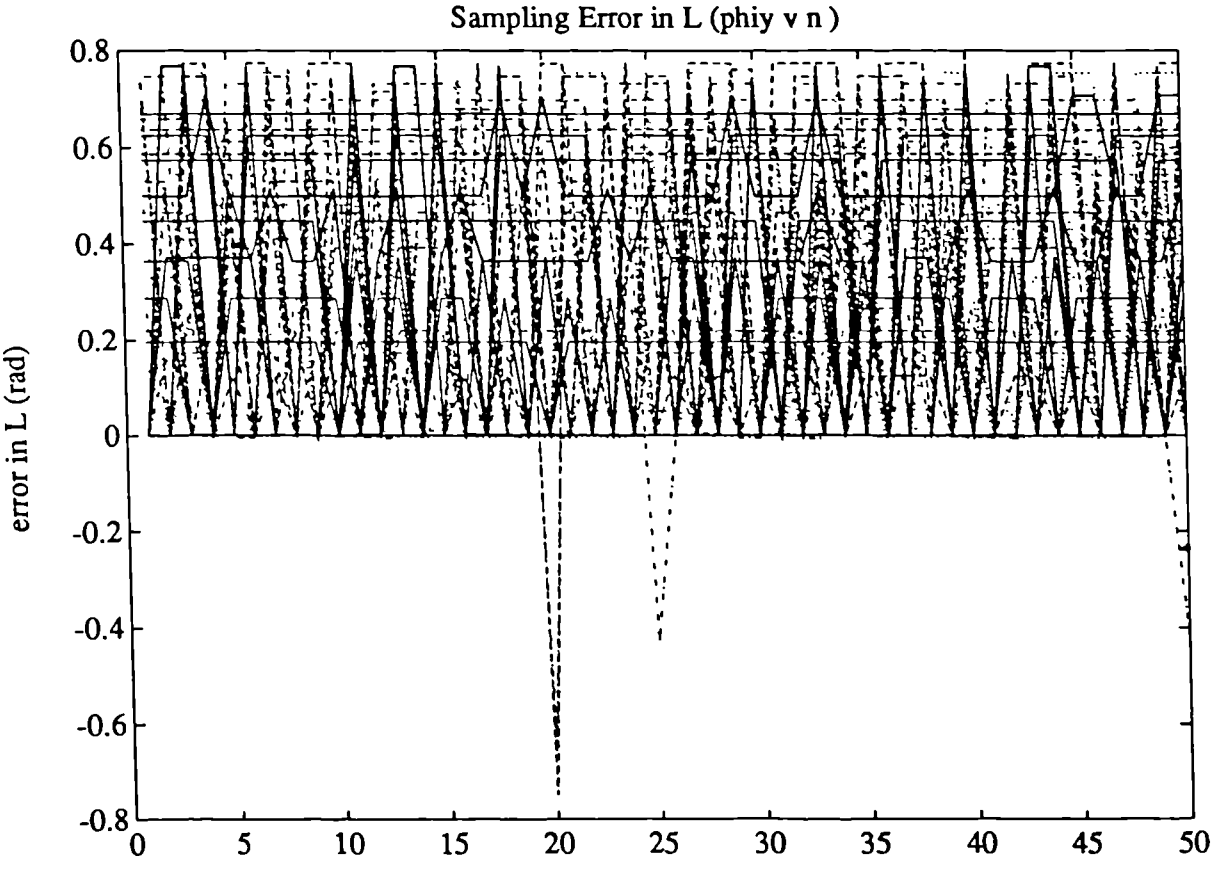


Graph 3.16 (b)

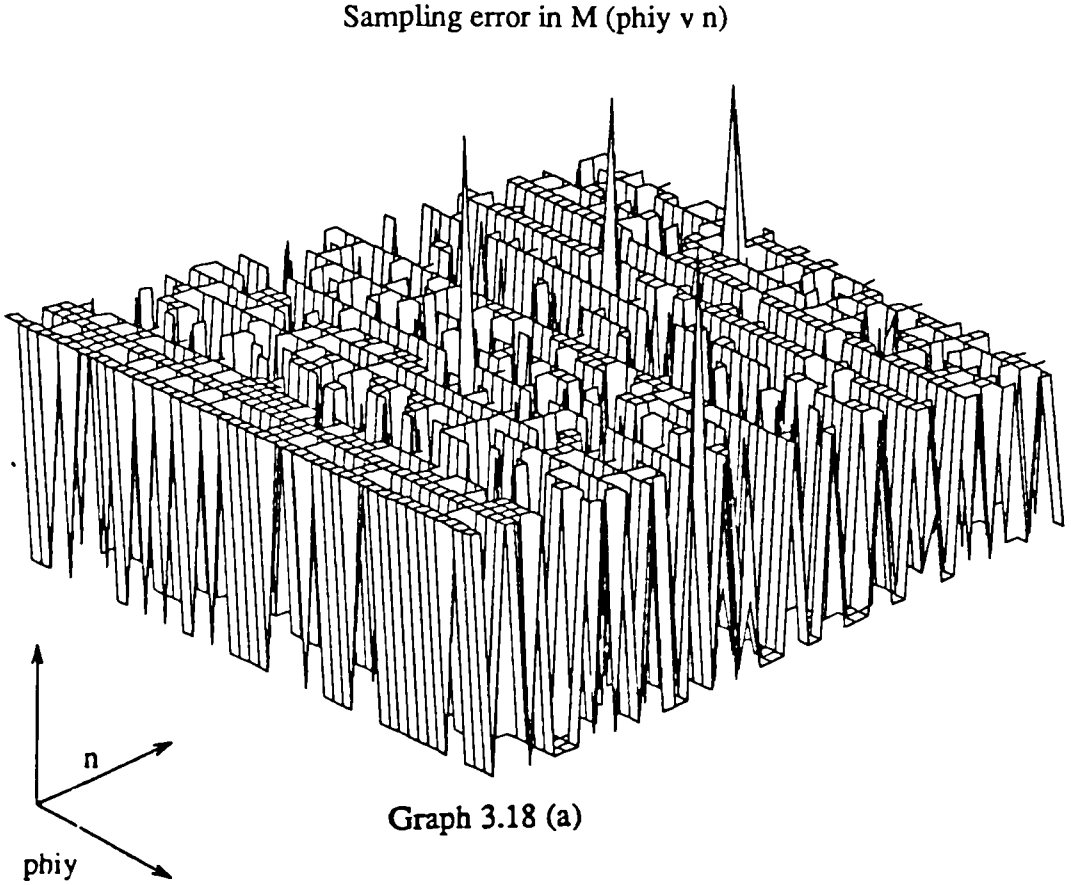
Sampling Error in L (phiy v n)



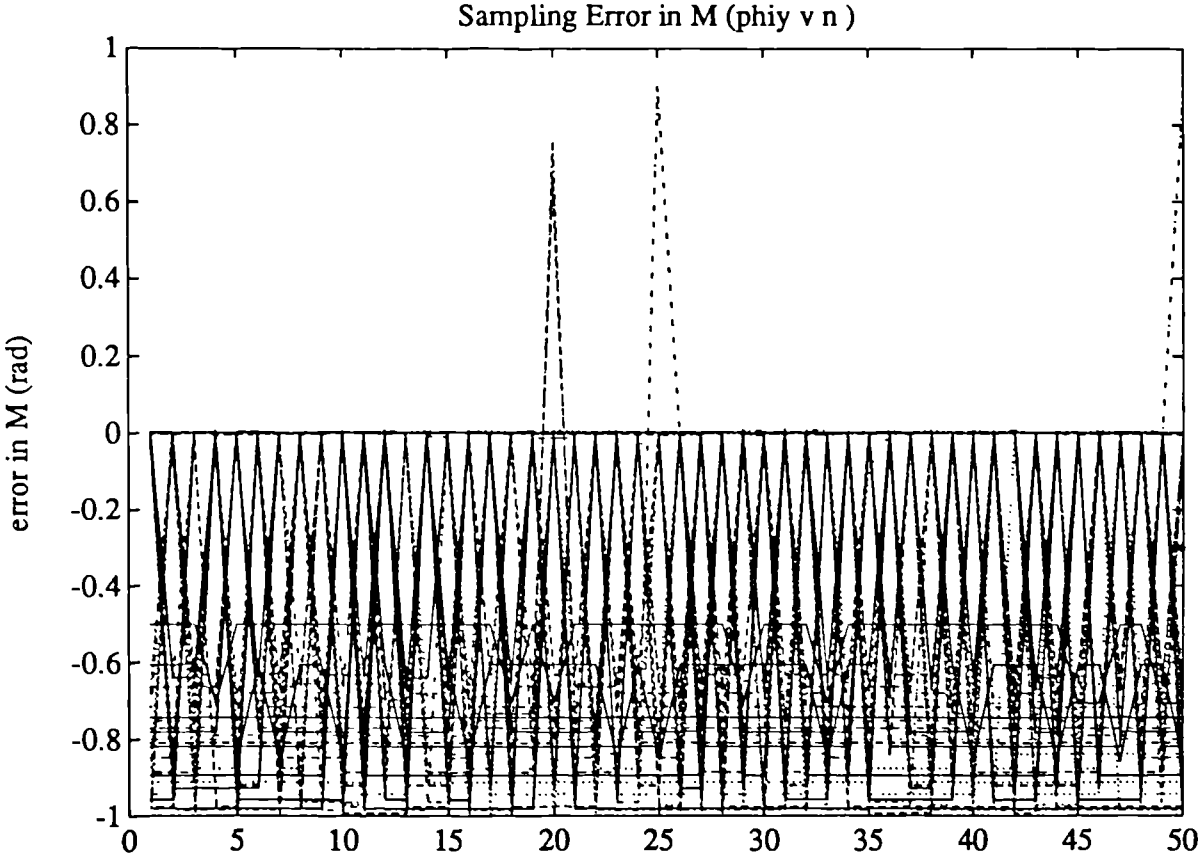
Graph 3.17 (a)



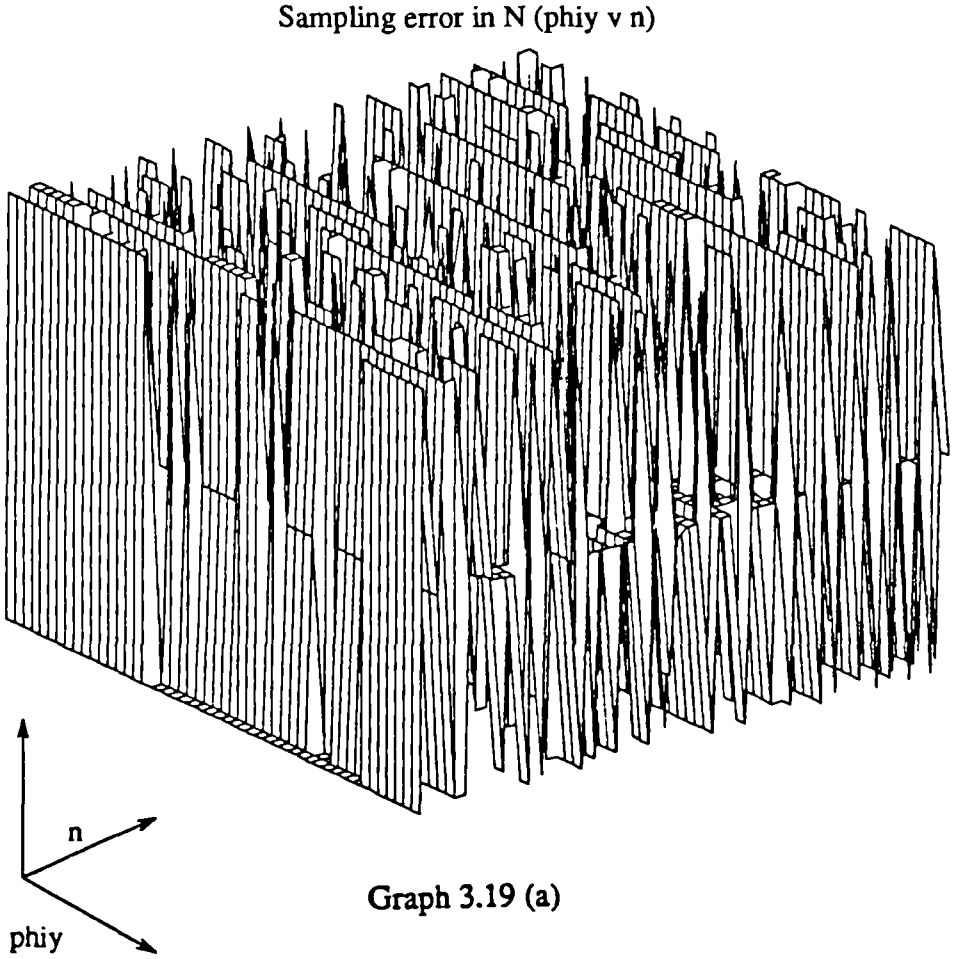
Graph 3.17 (b)



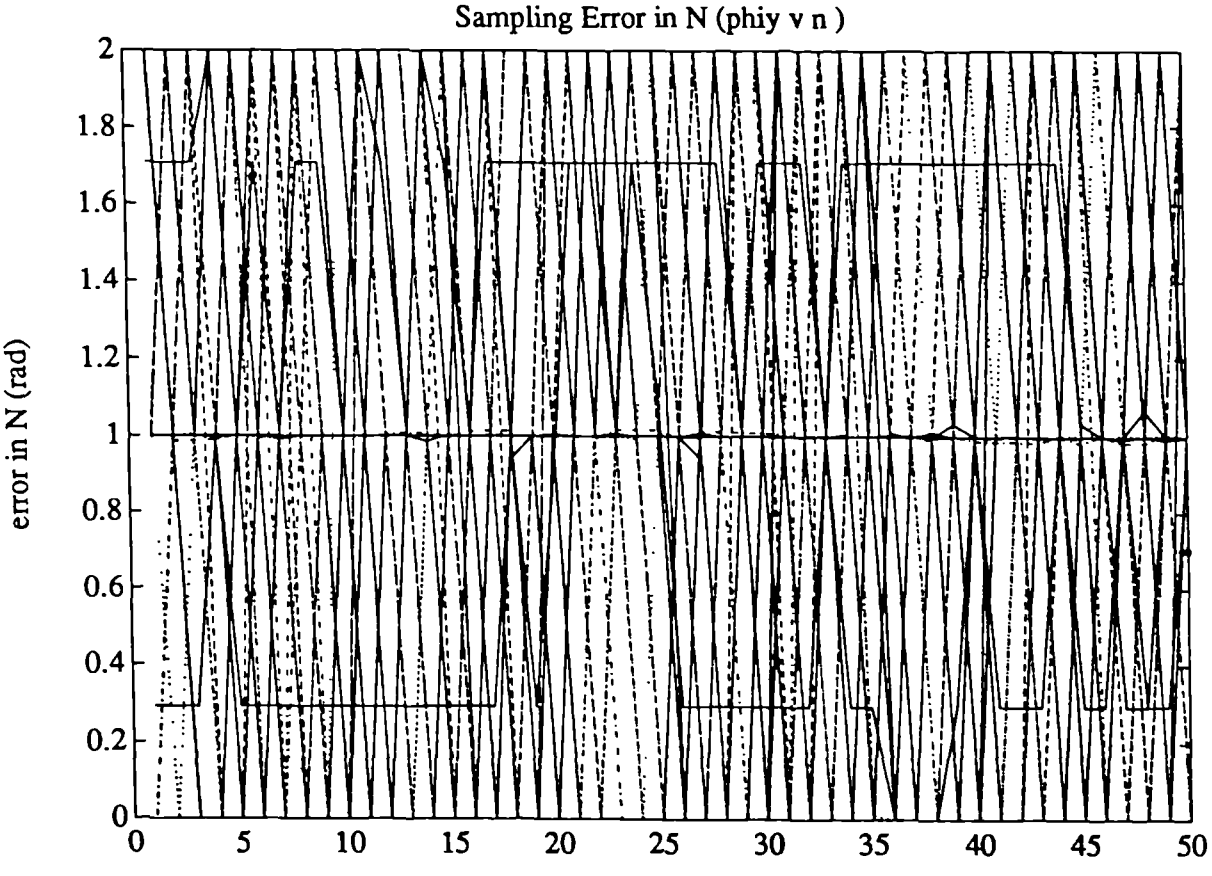
Graph 3.18 (a)



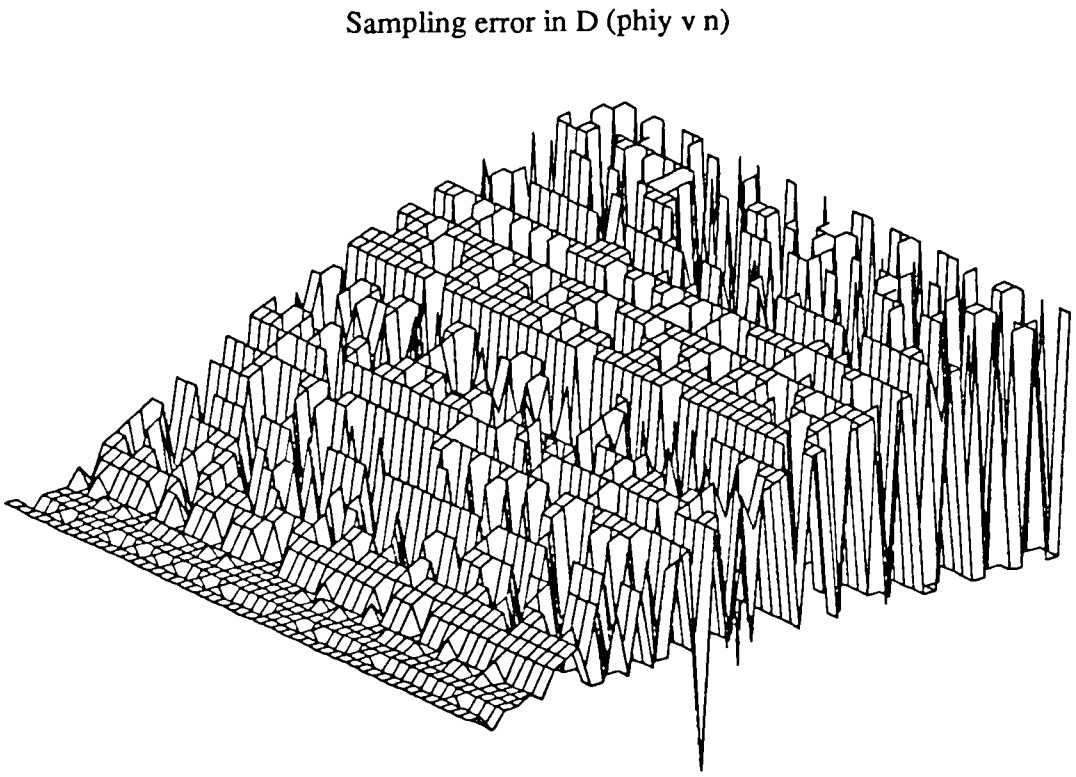
Graph 3.18 (b)



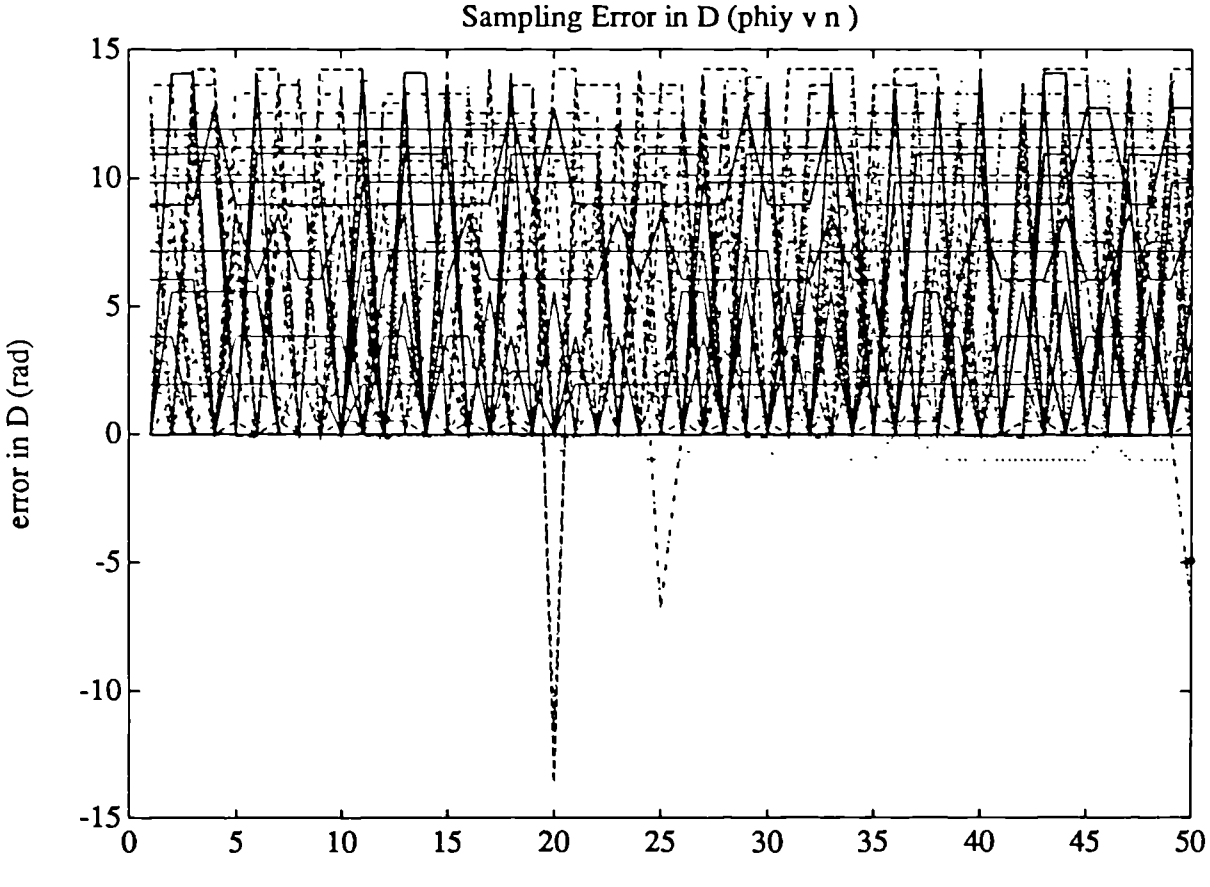
Graph 3.19 (a)



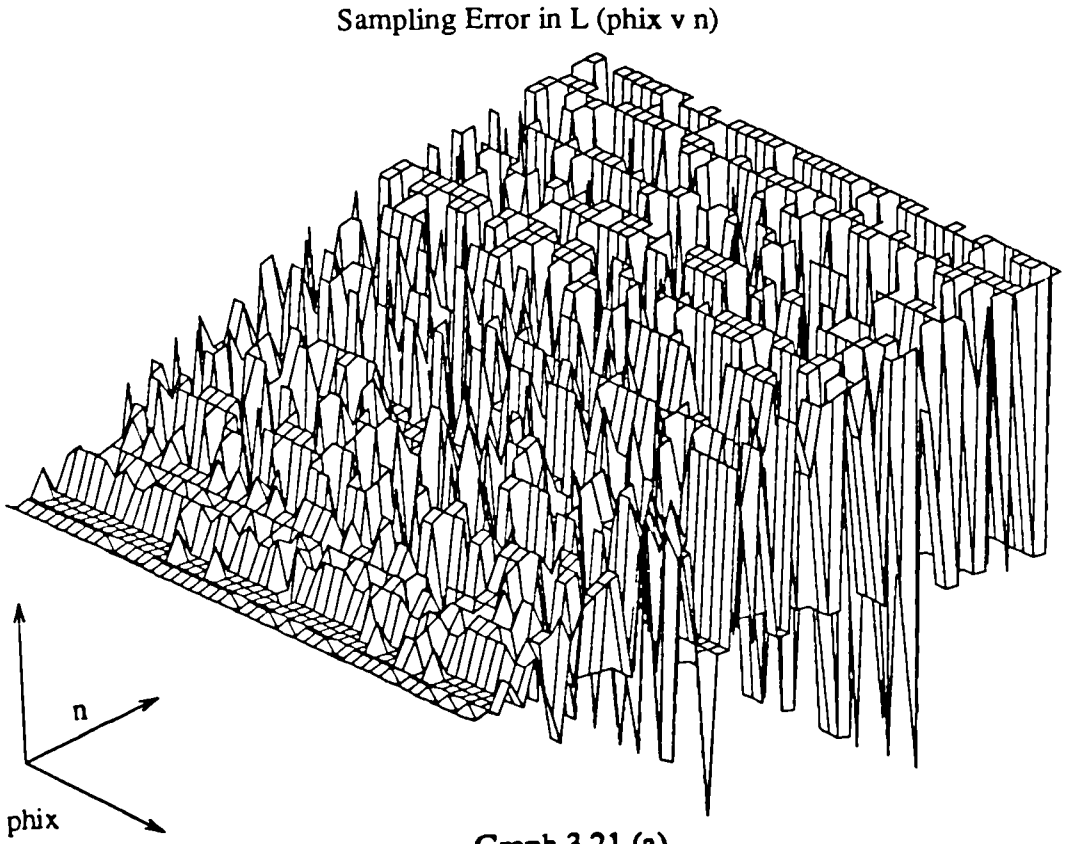
Graph 3.19 (b)



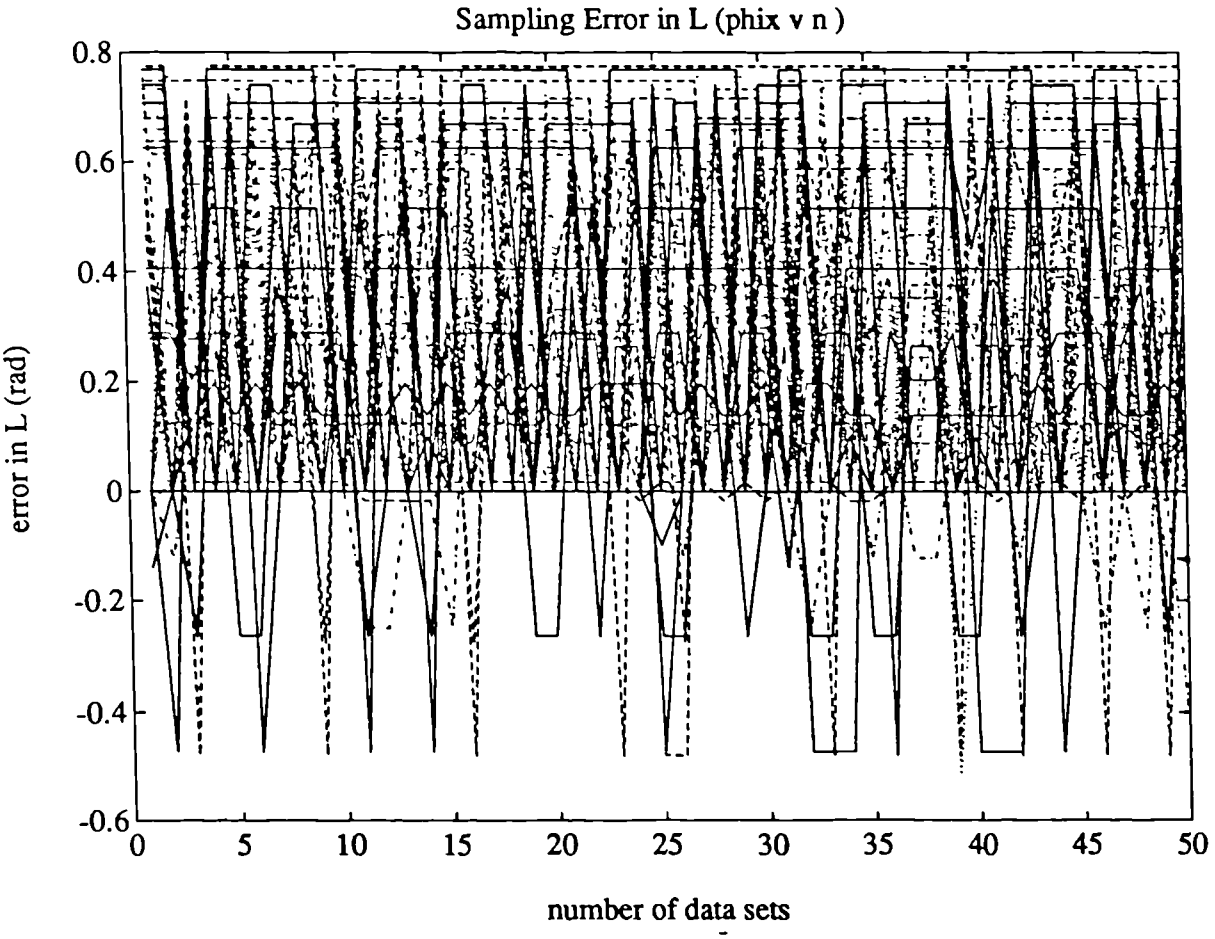
Graph 3.20 (a)



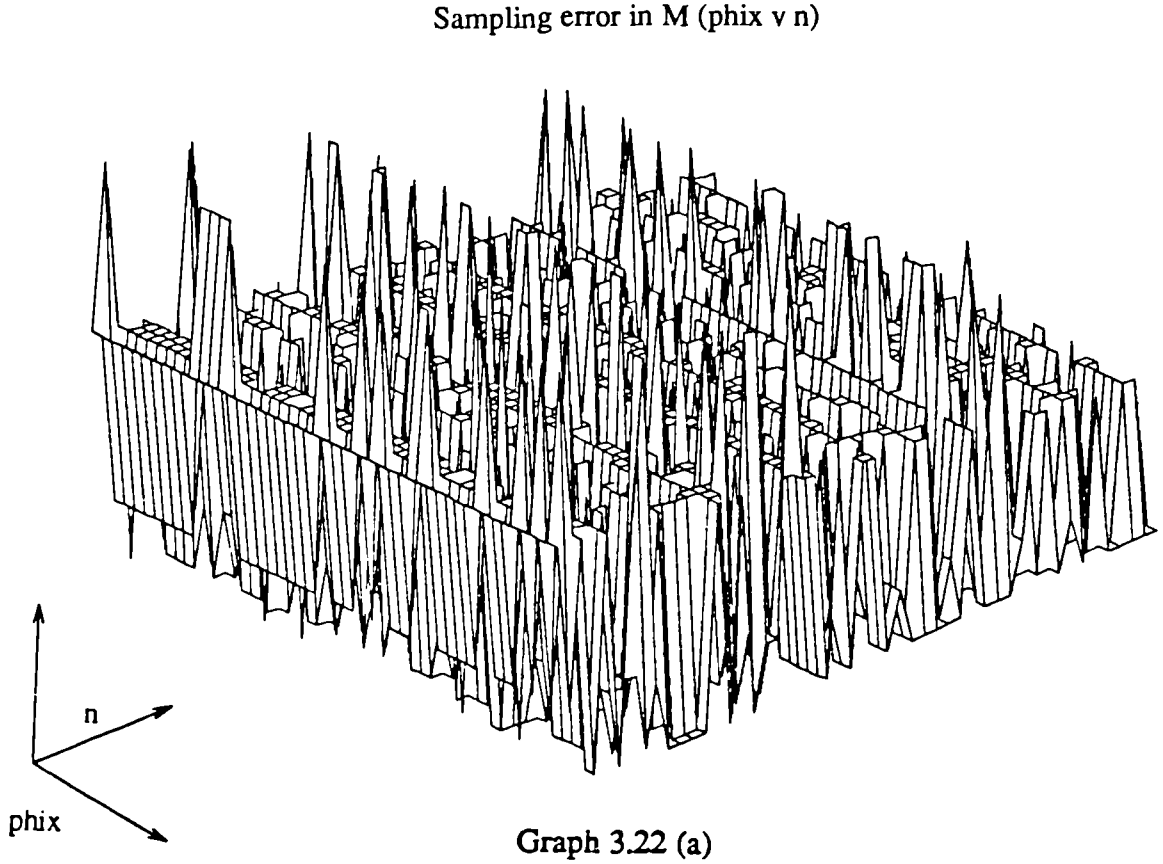
Graph 3.20 (b)



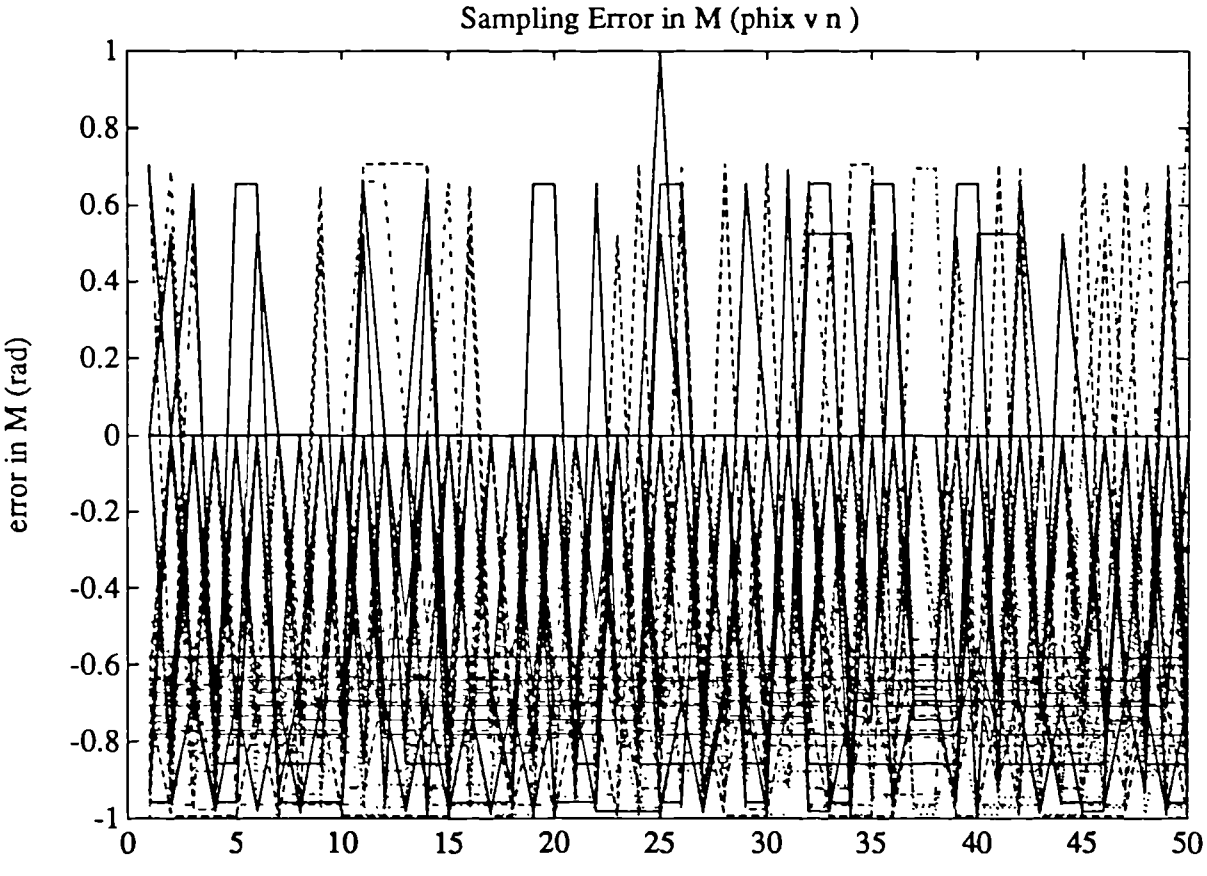
Graph 3.21 (a)



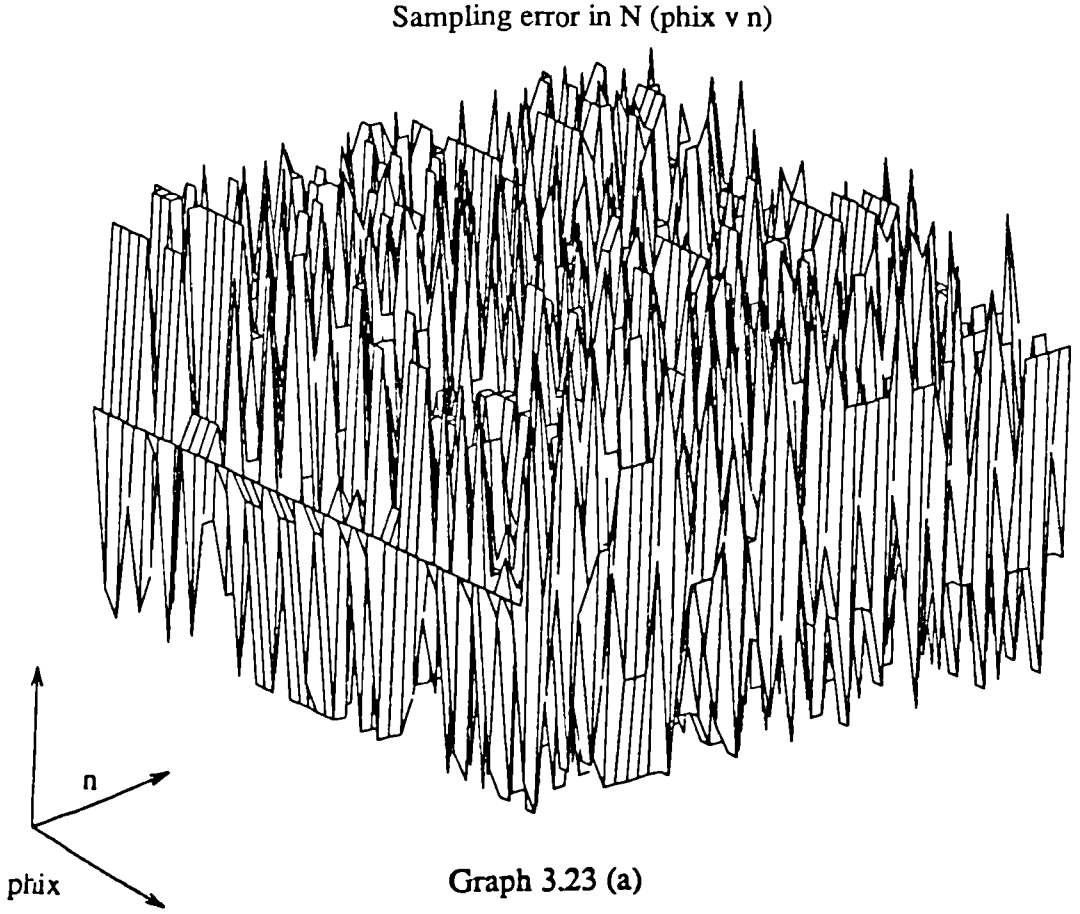
Graph 3.21 (b)



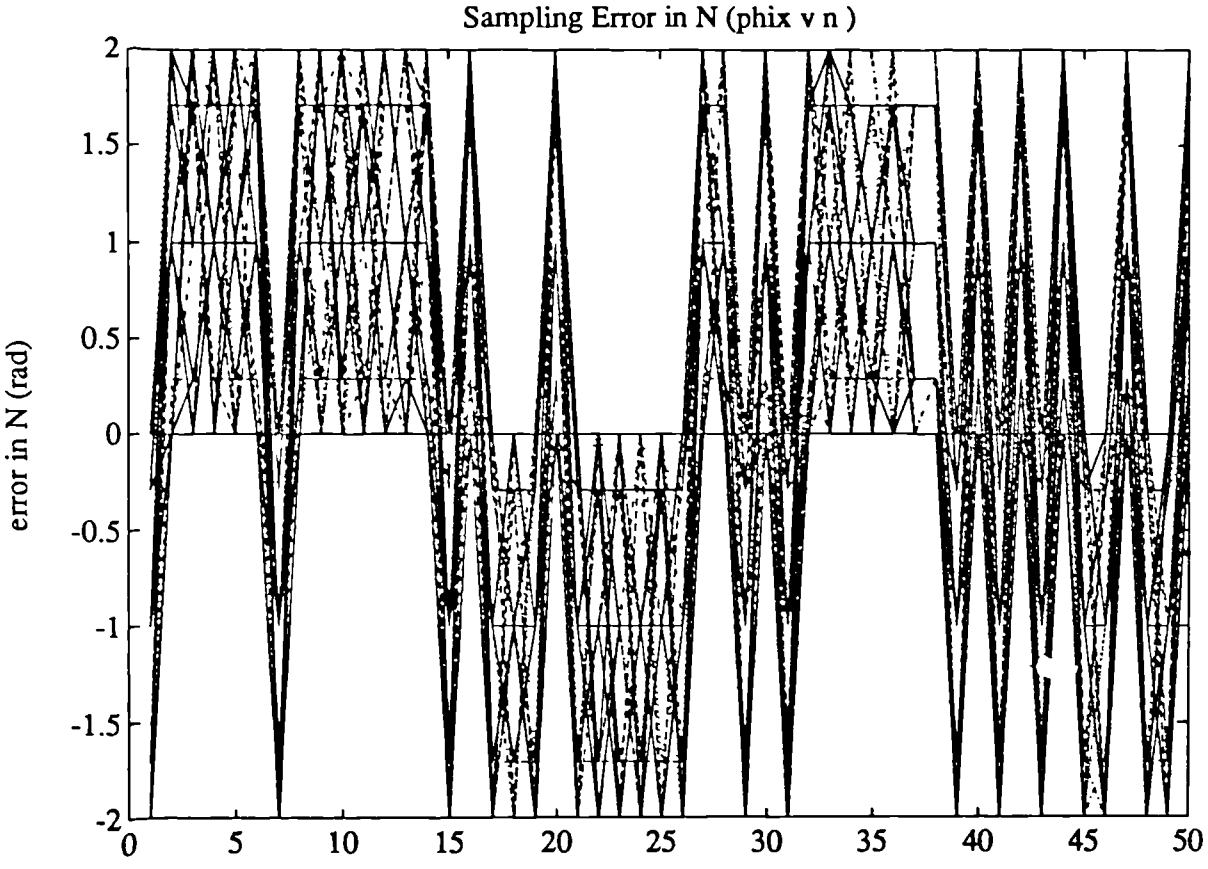
Graph 3.22 (a)



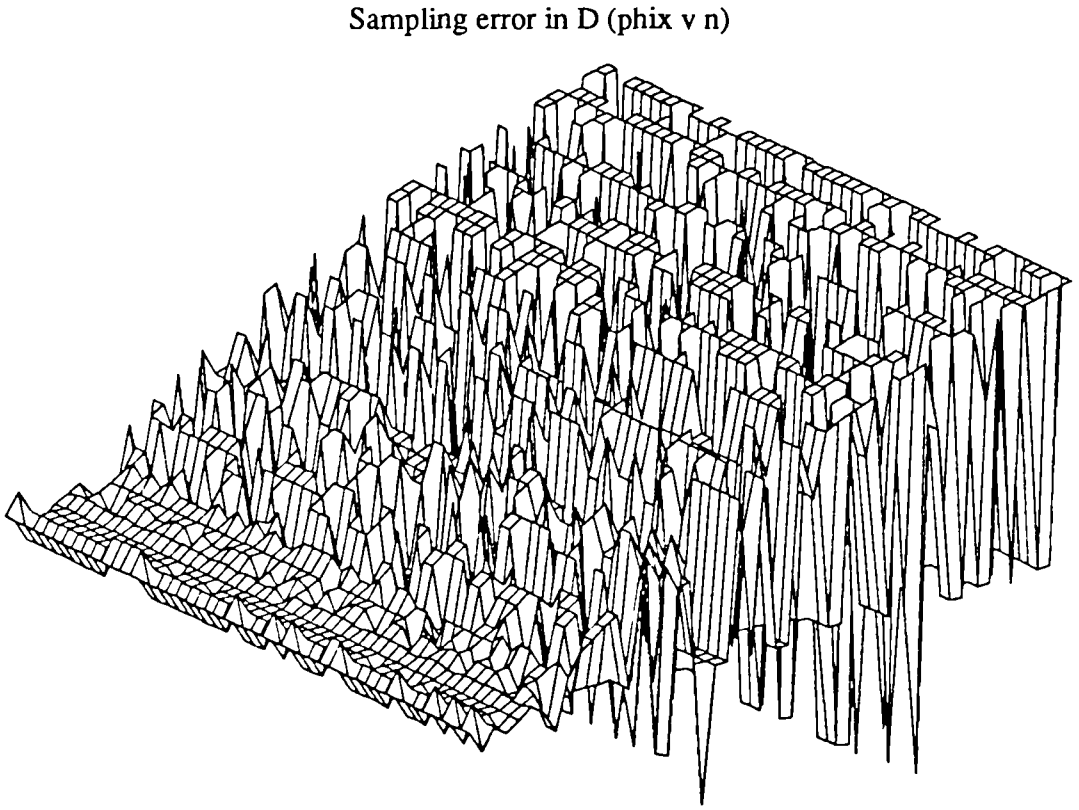
Graph 3.22 (b)



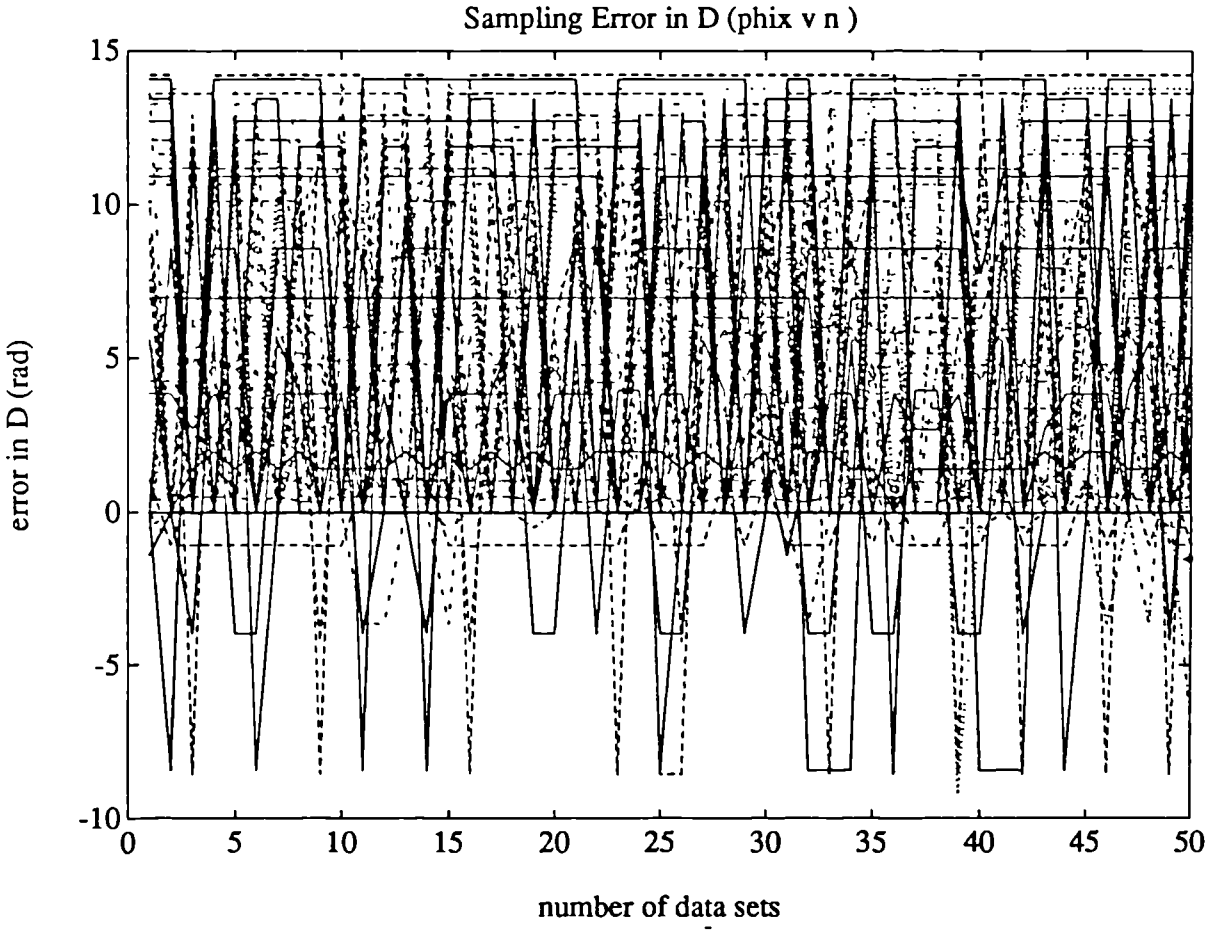
Graph 3.23 (a)



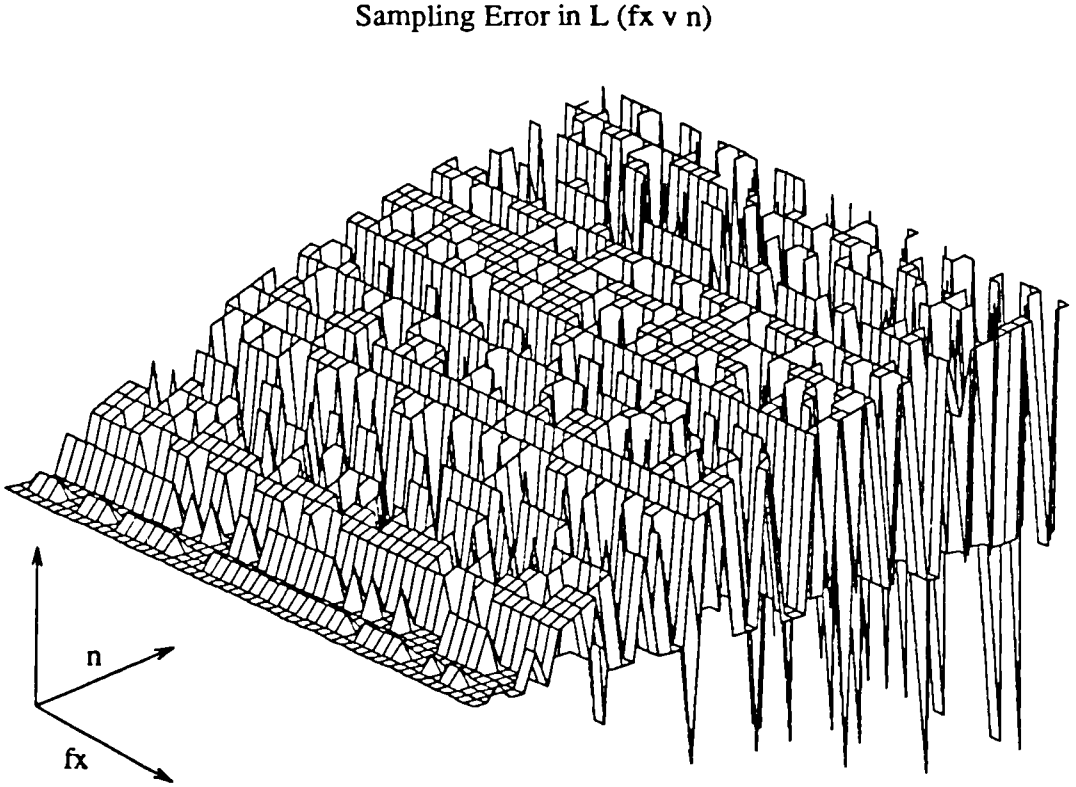
Graph 3.23 (b)



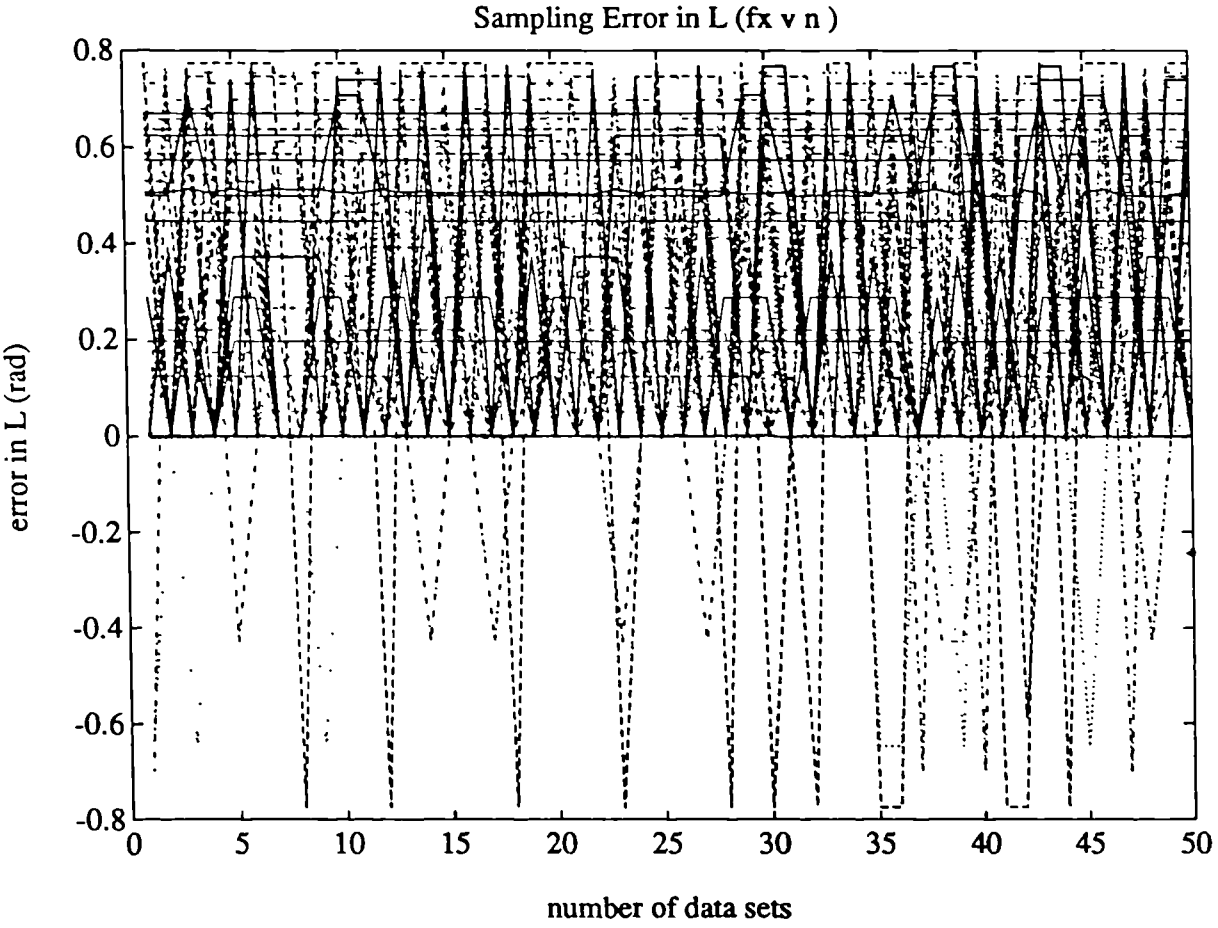
Graph 3.24 (a)



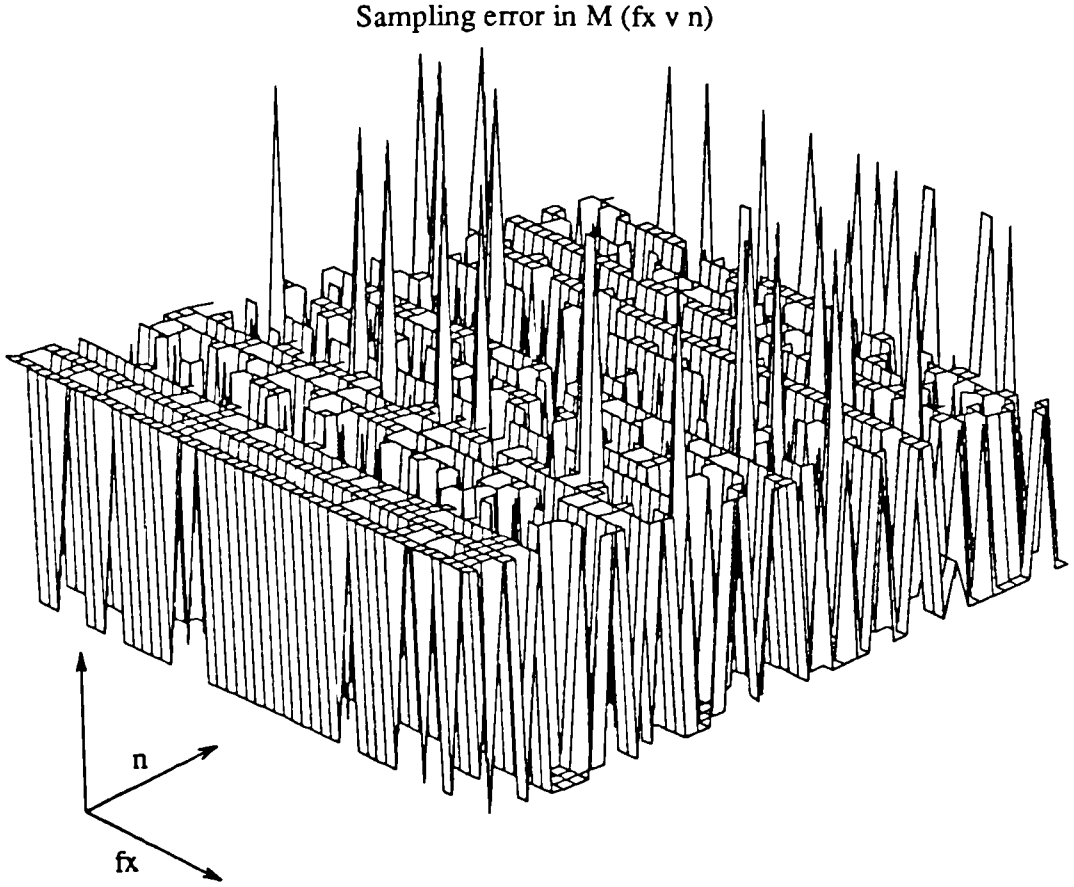
Graph 3.24 (b)



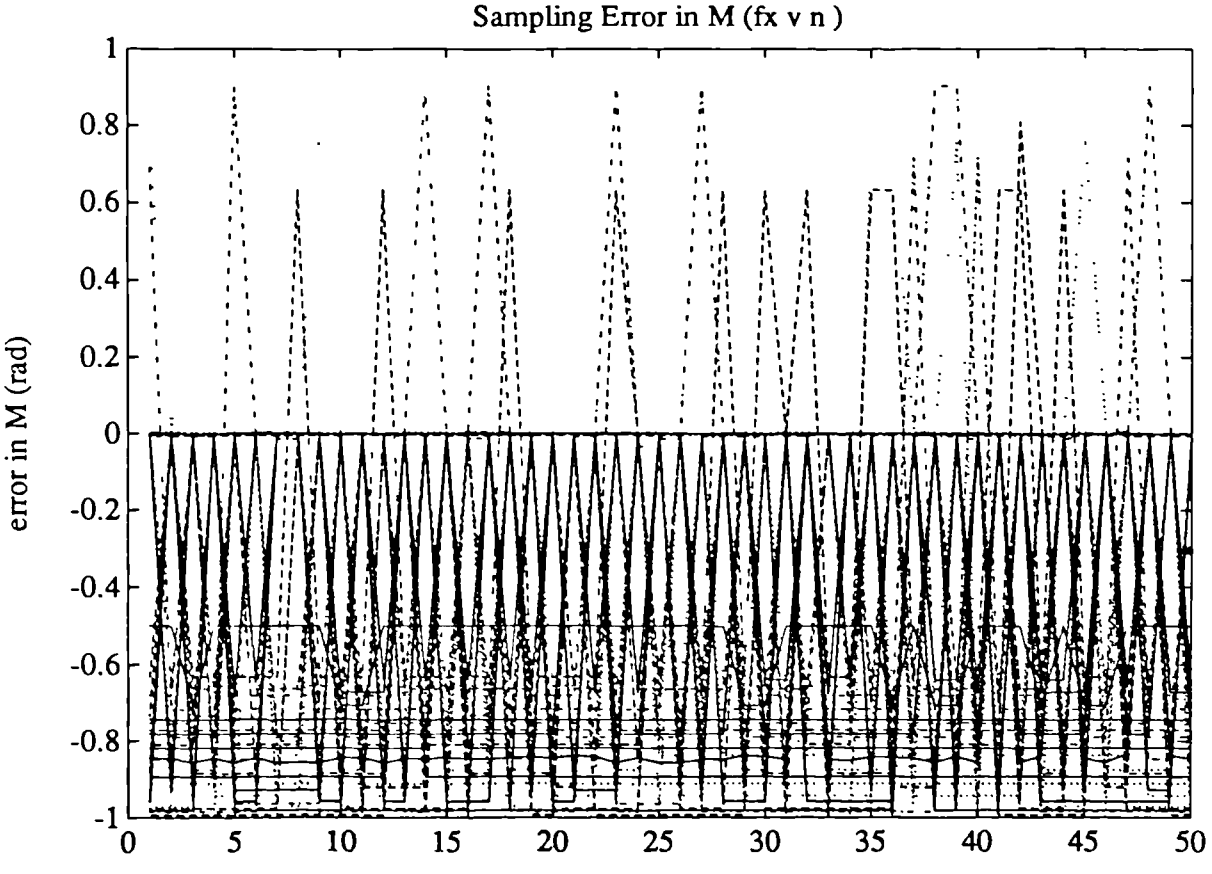
Graph 3.25 (a)



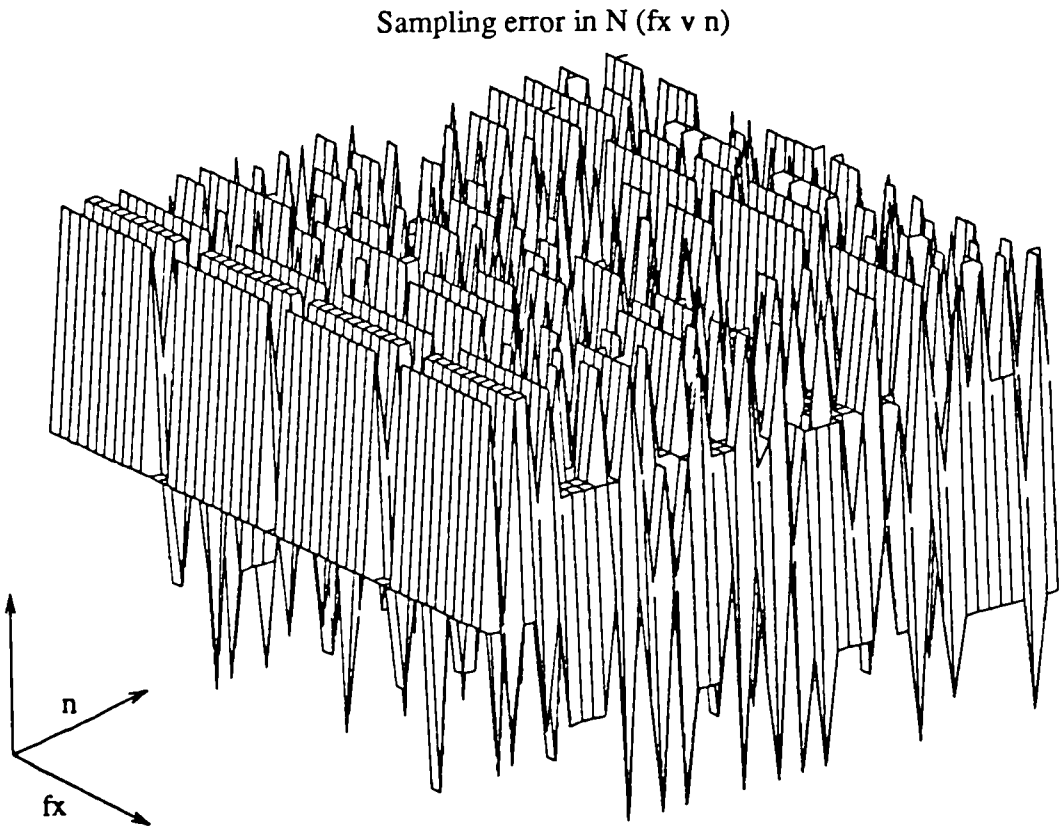
Graph 3.25 (b)



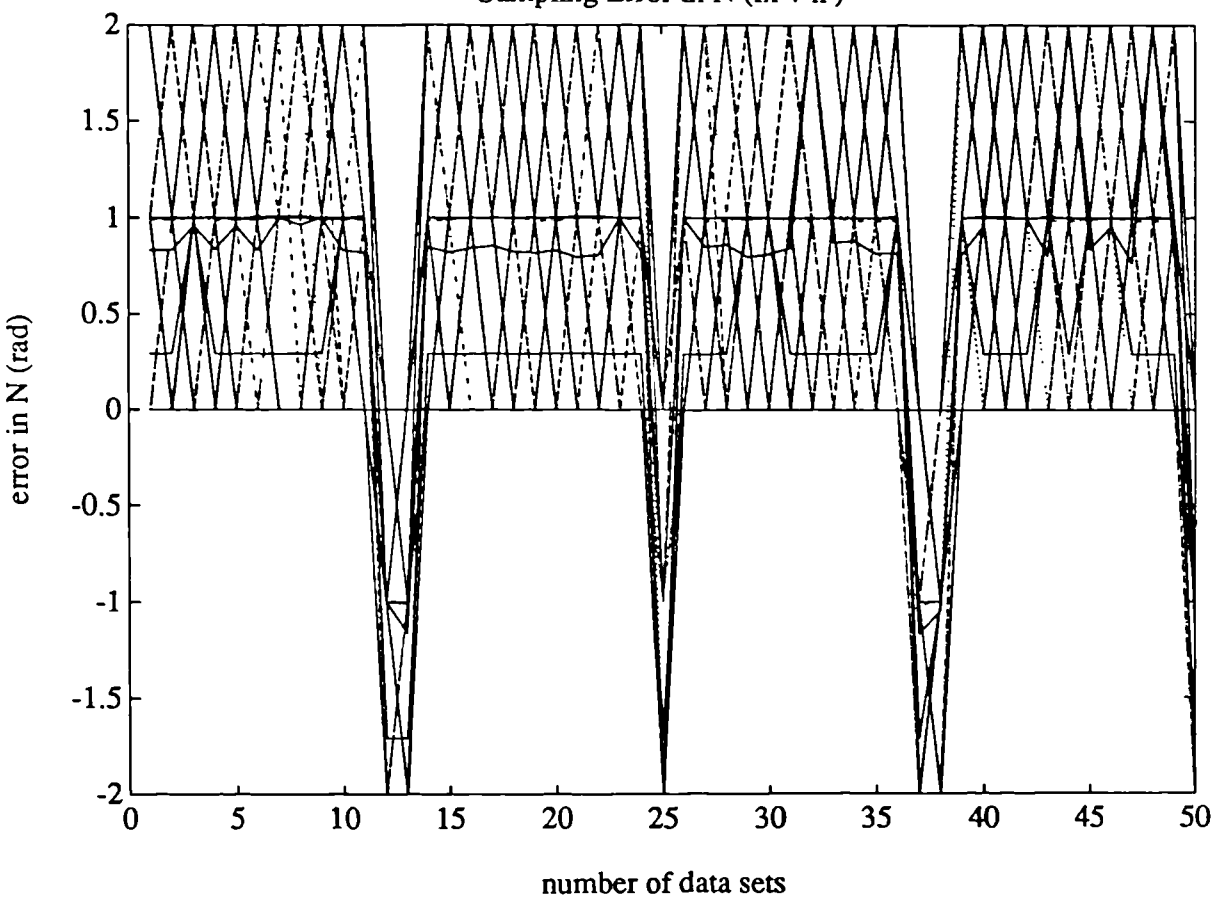
Graph 3.26 (a)



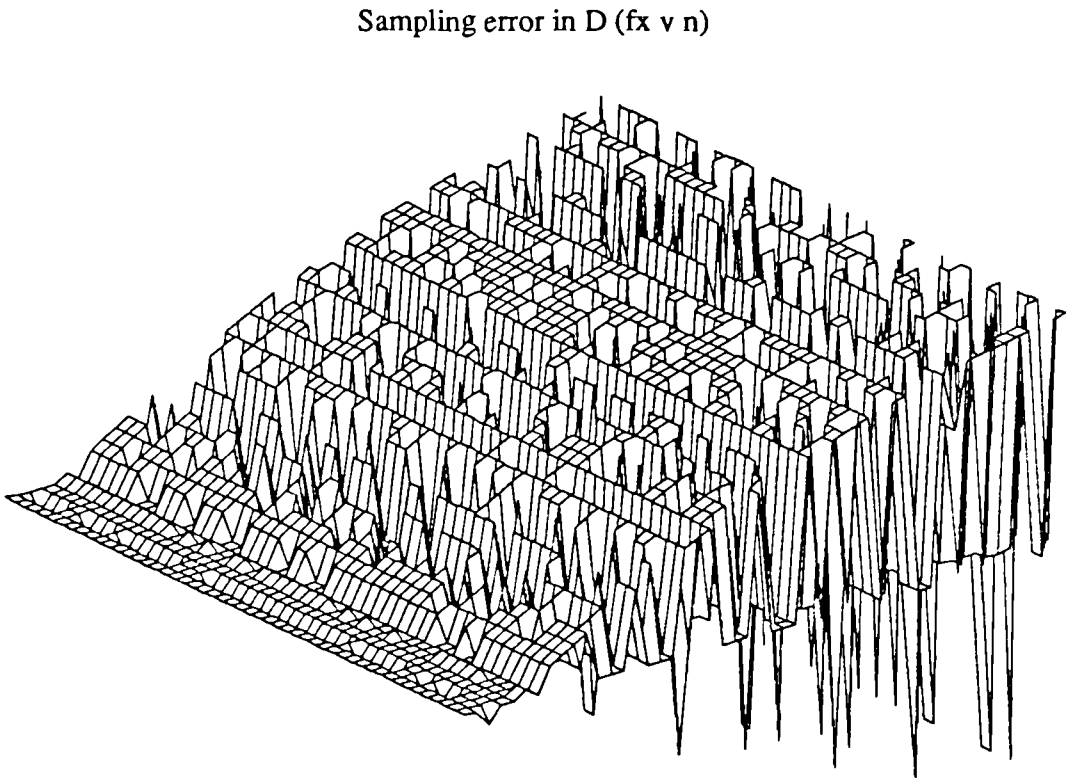
Graph 3.26 (b)



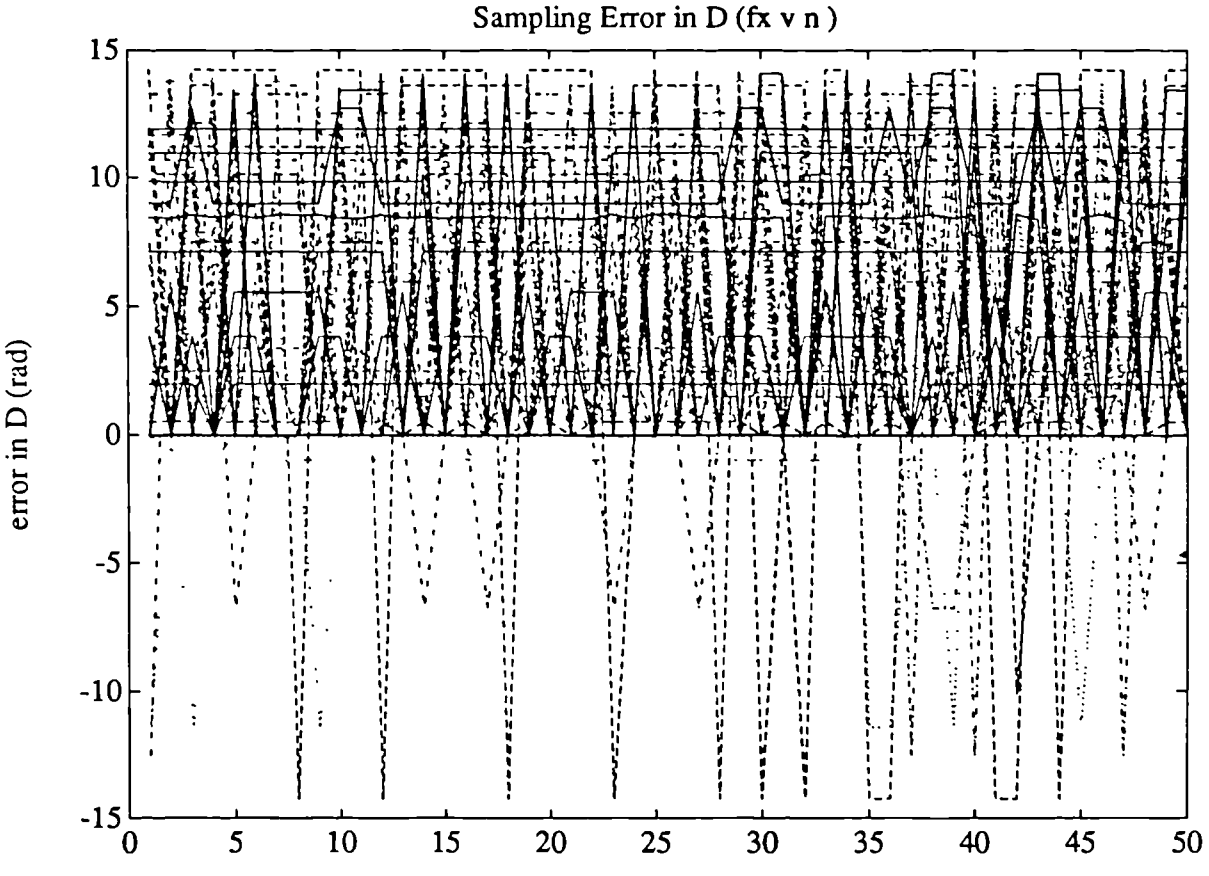
Graph 3.27 (a)



Graph 3.27 (b)

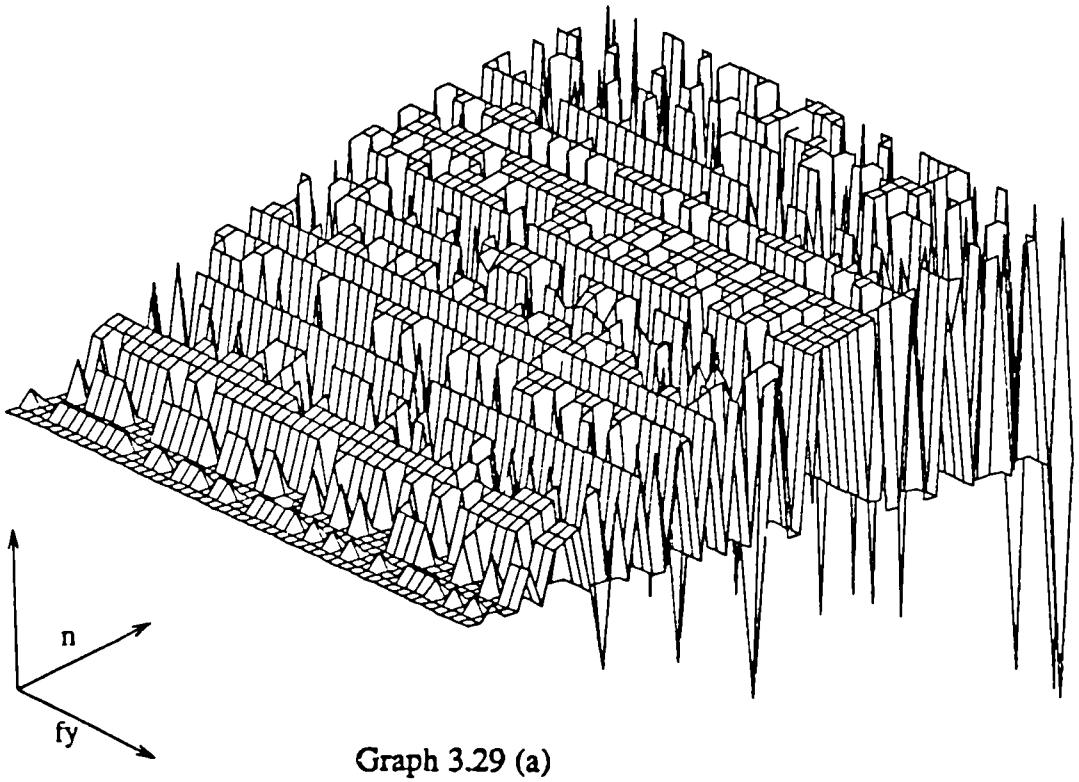


Graph 3.28 (a)



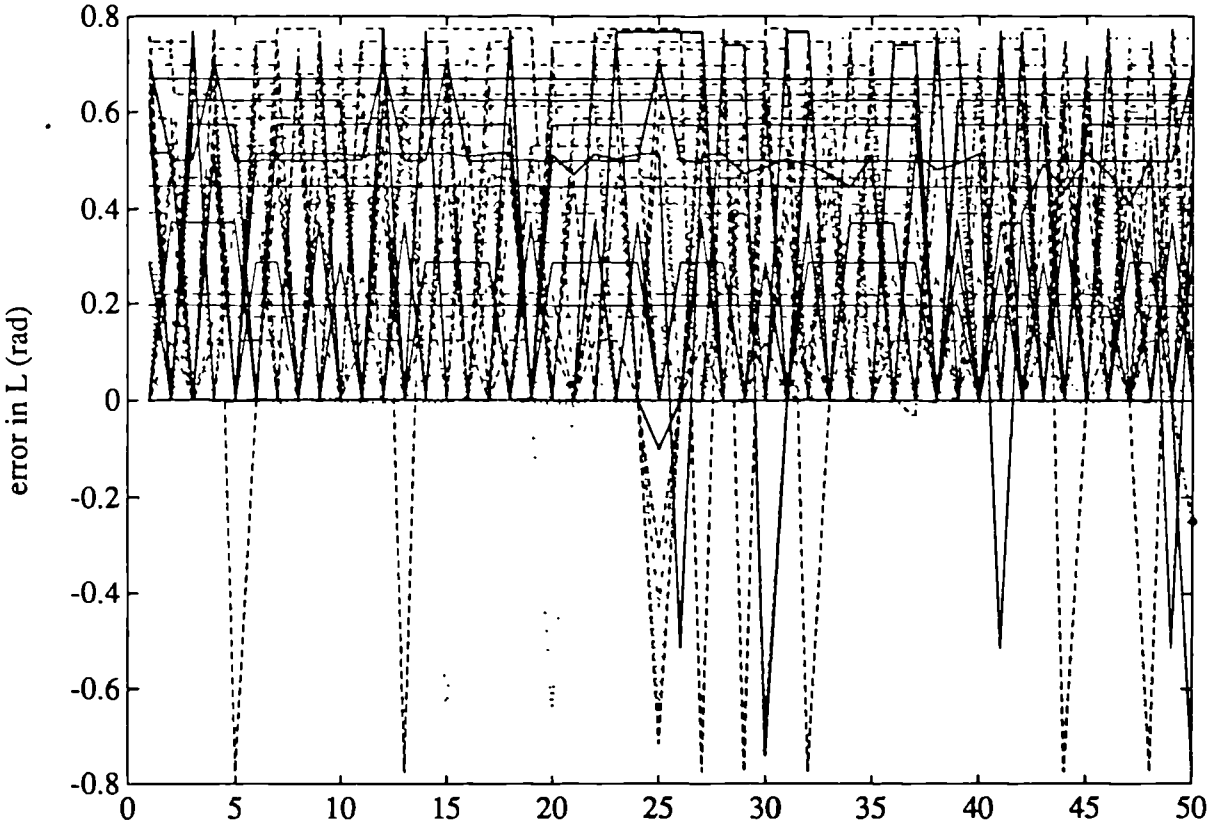
Graph 3.28 (b)

Sampling Error in L (fy v n)

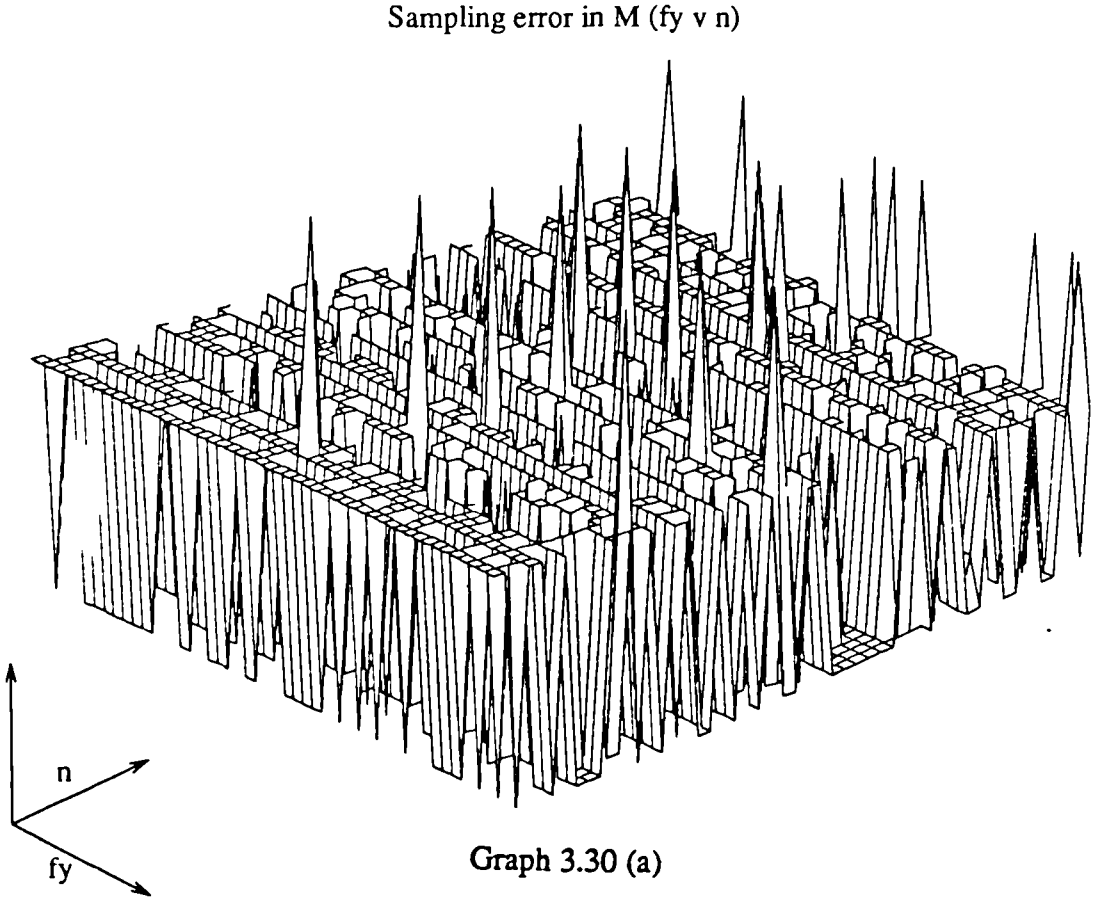


Graph 3.29 (a)

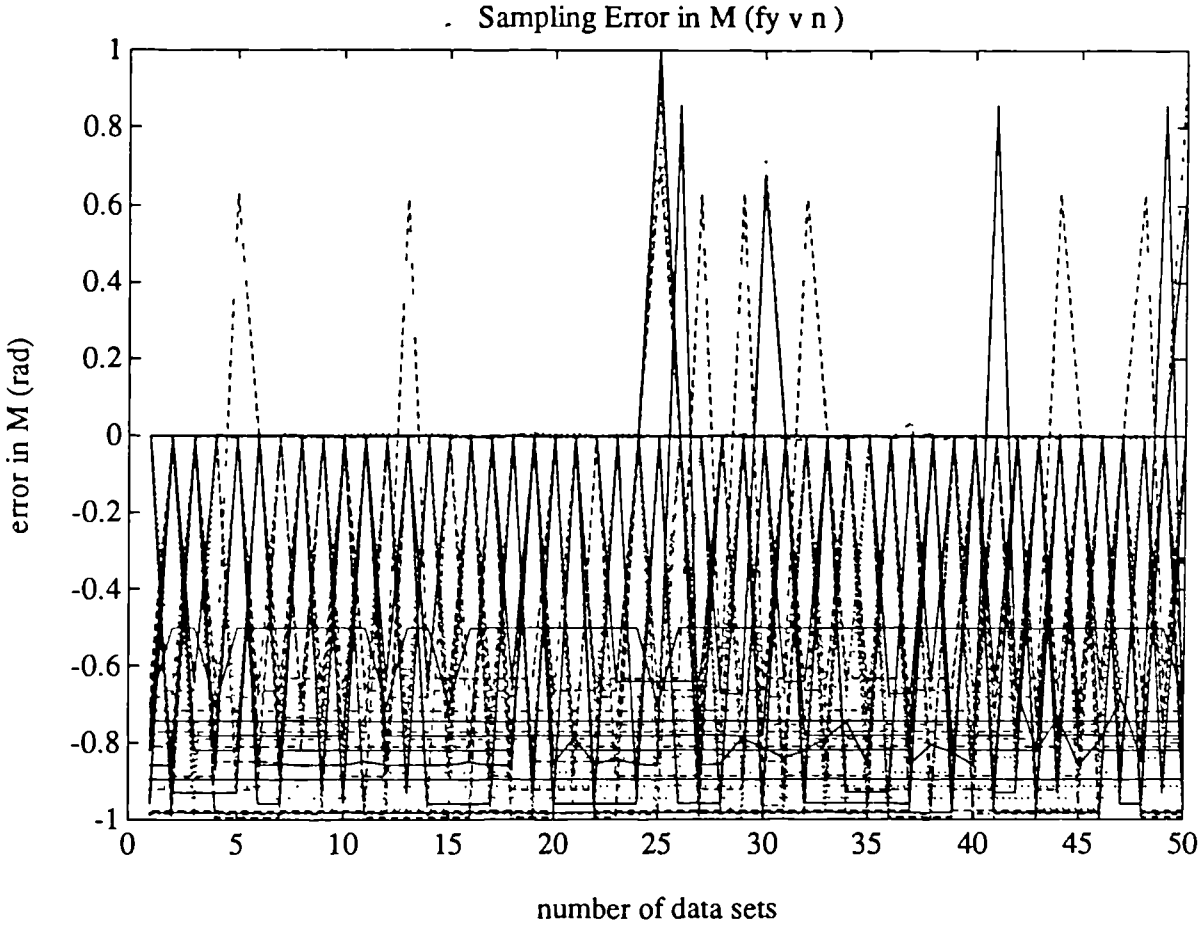
Sampling Error in L (fy v n)



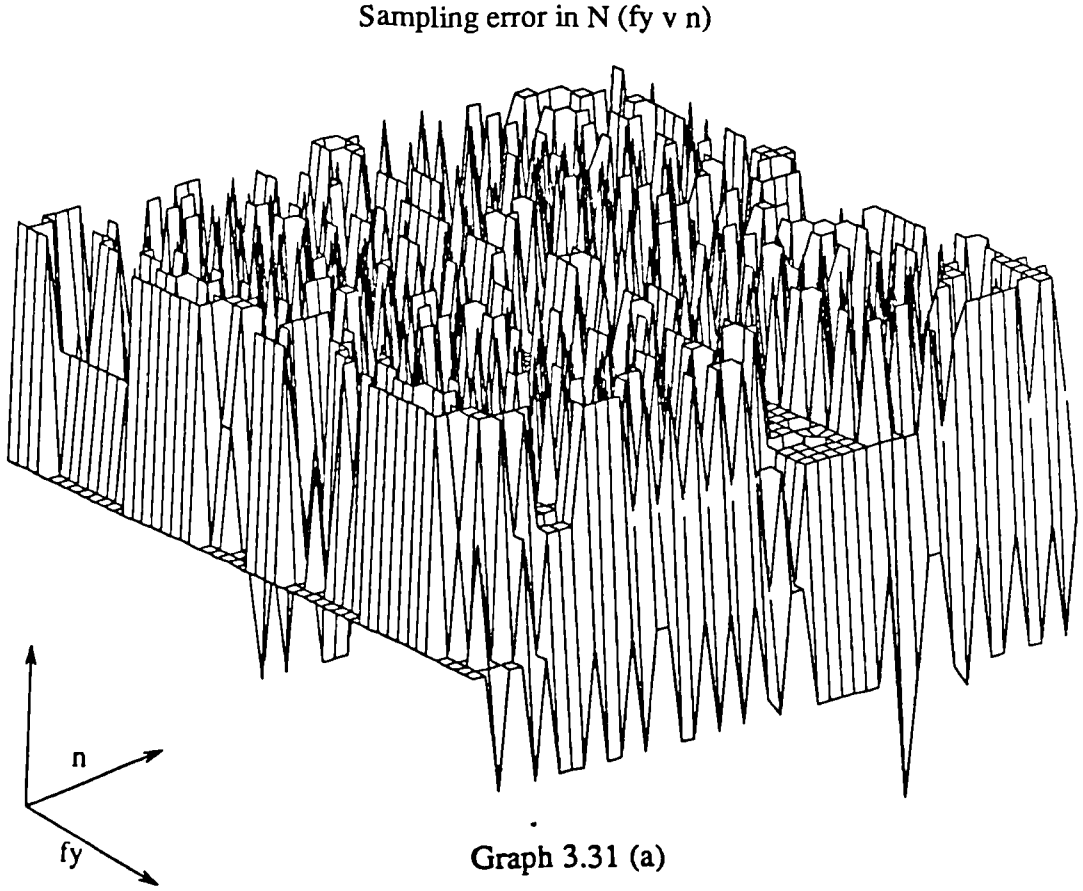
Graph 3.29 (b)



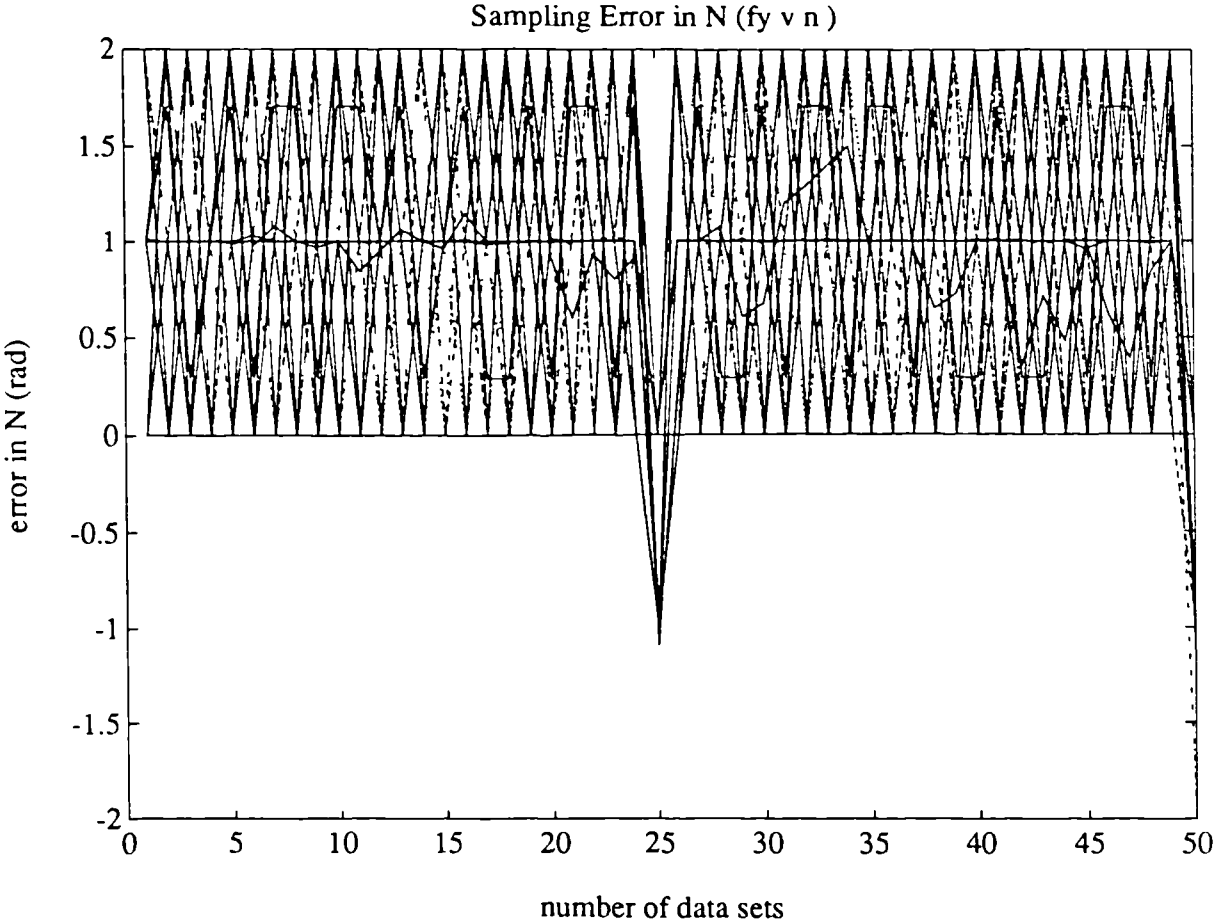
Graph 3.30 (a)



Graph 3.30 (b)

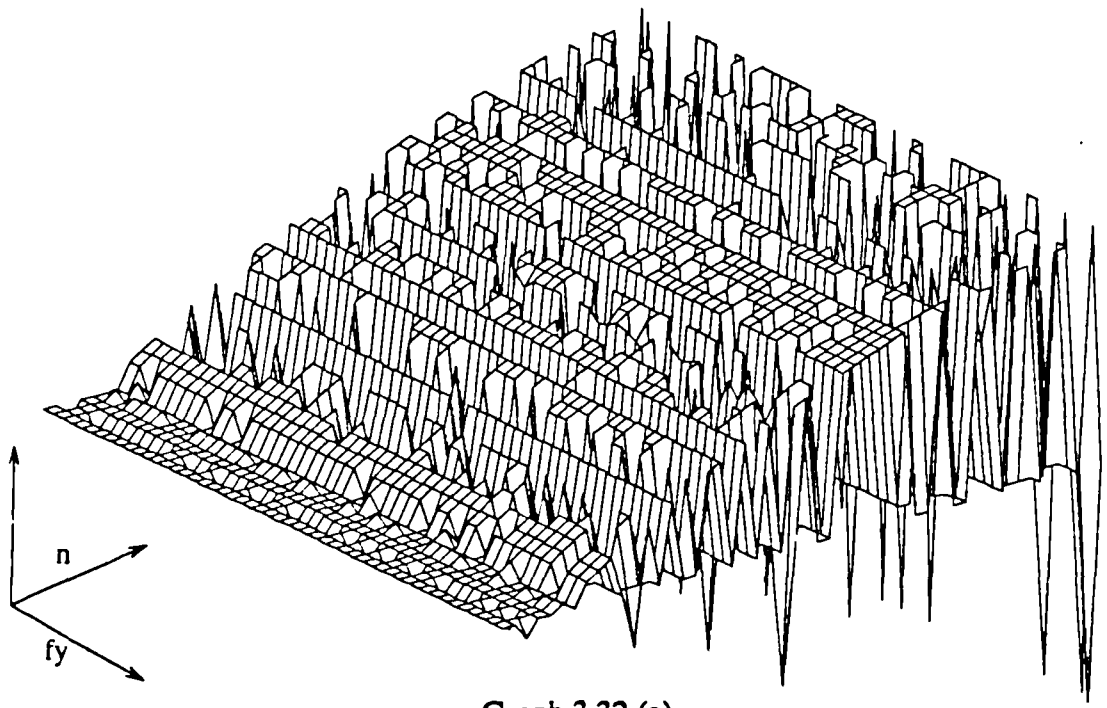


Graph 3.31 (a)



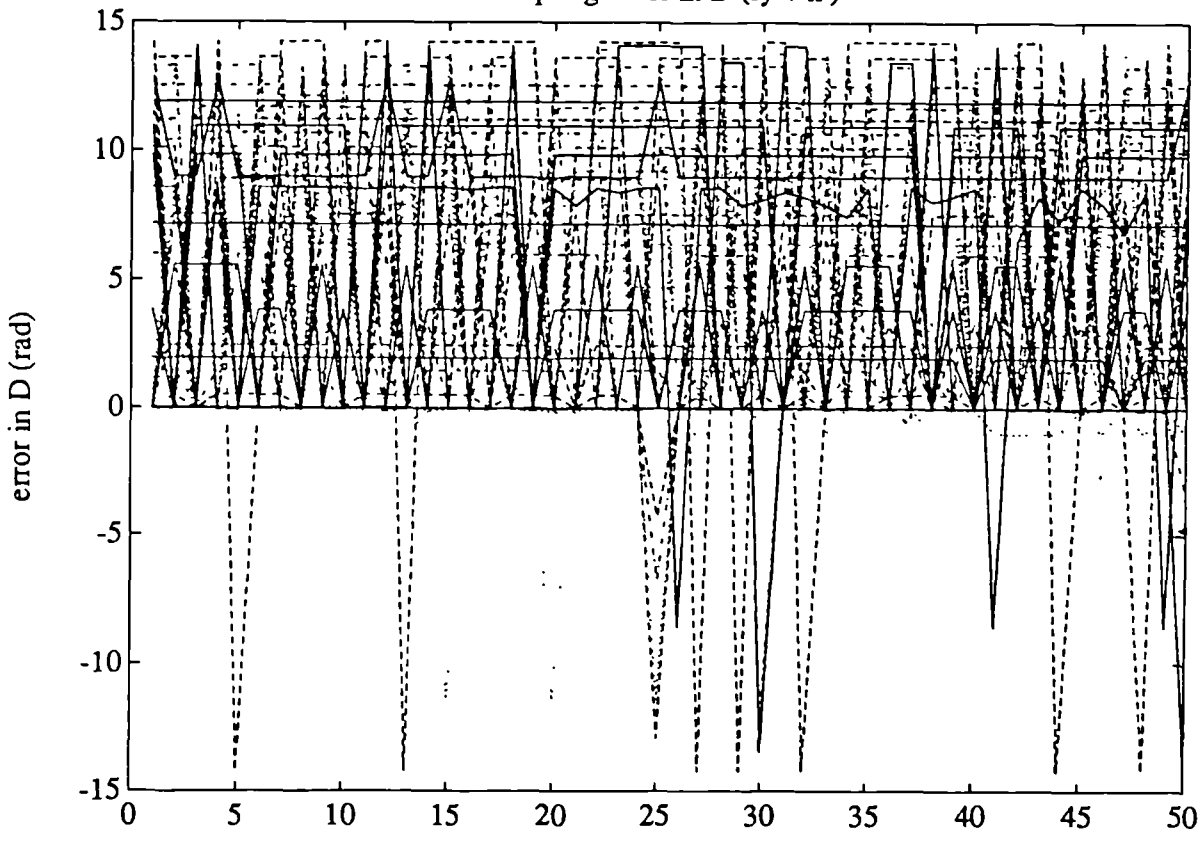
Graph 3.31 (b)

Sampling error in D (fy v n)

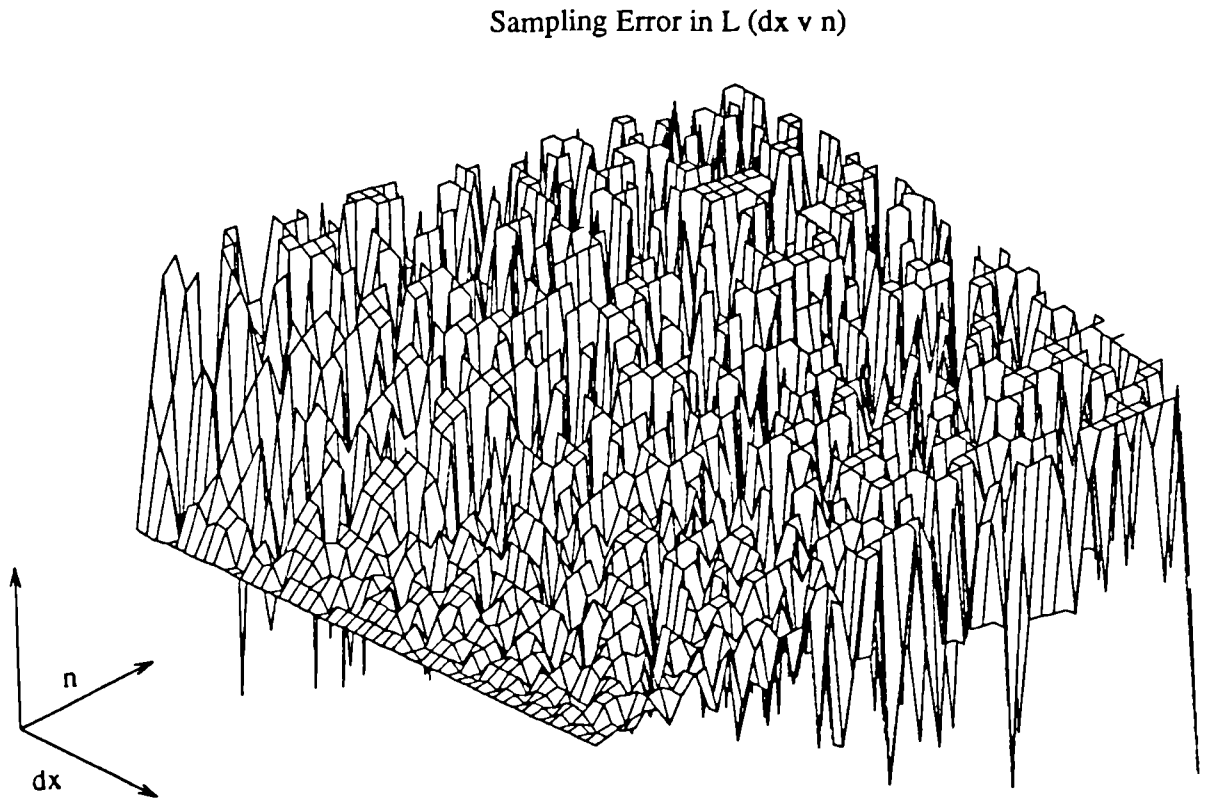


Graph 3.32 (a),

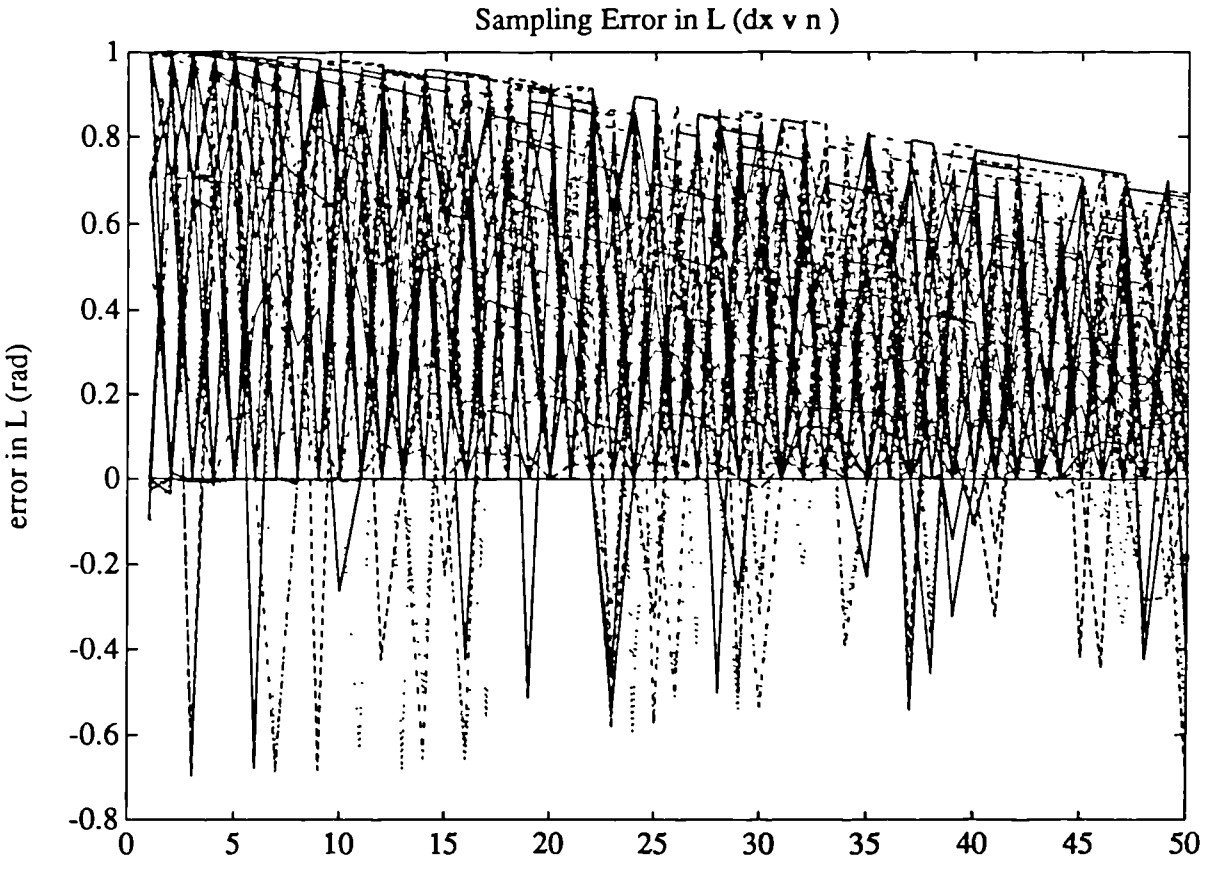
Sampling Error in D (fy v n)



Graph 3.32 (b)

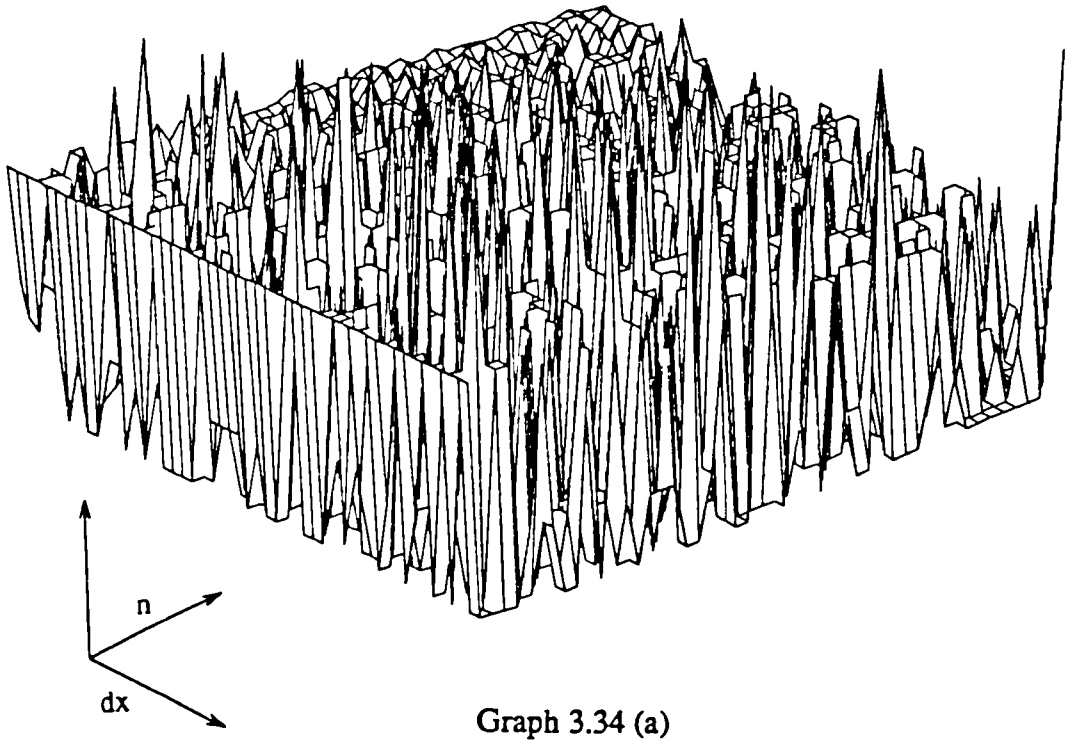


Graph 3.33 (a)



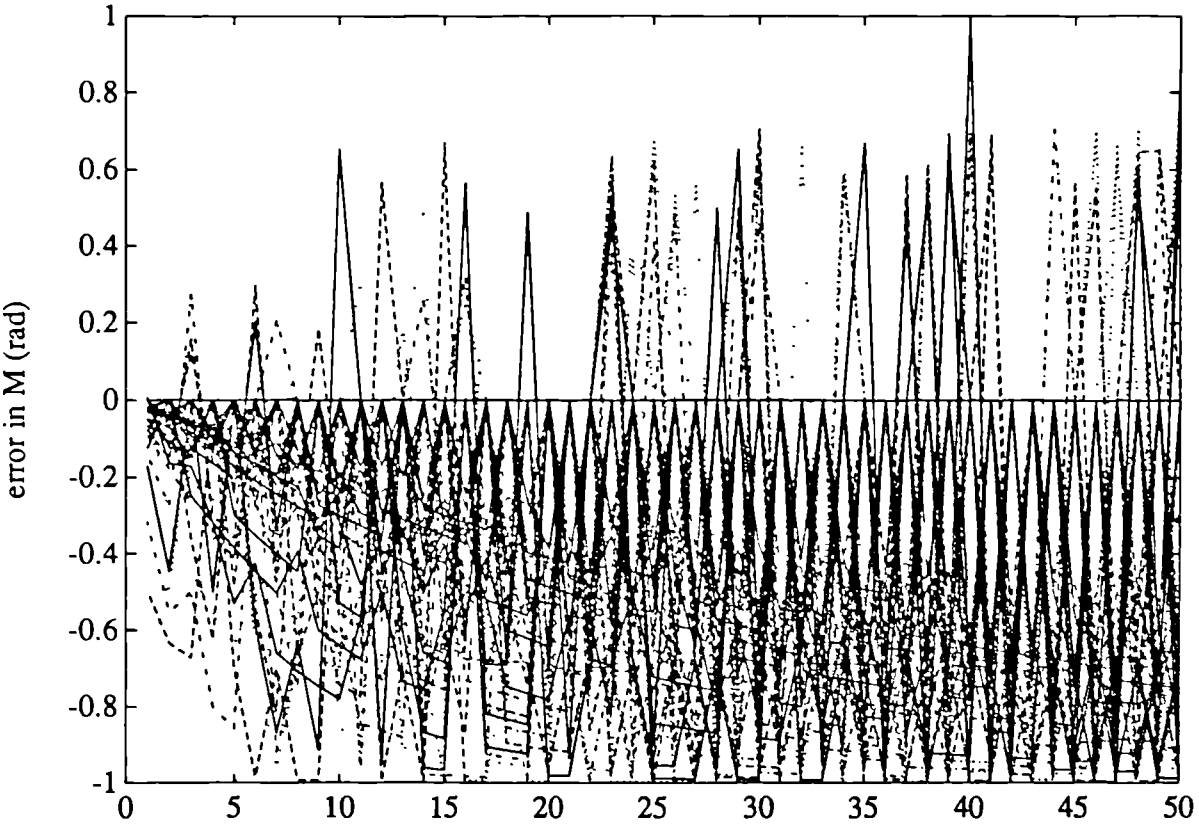
Graph 3.33 (b)

Sampling error in M (dx v n)

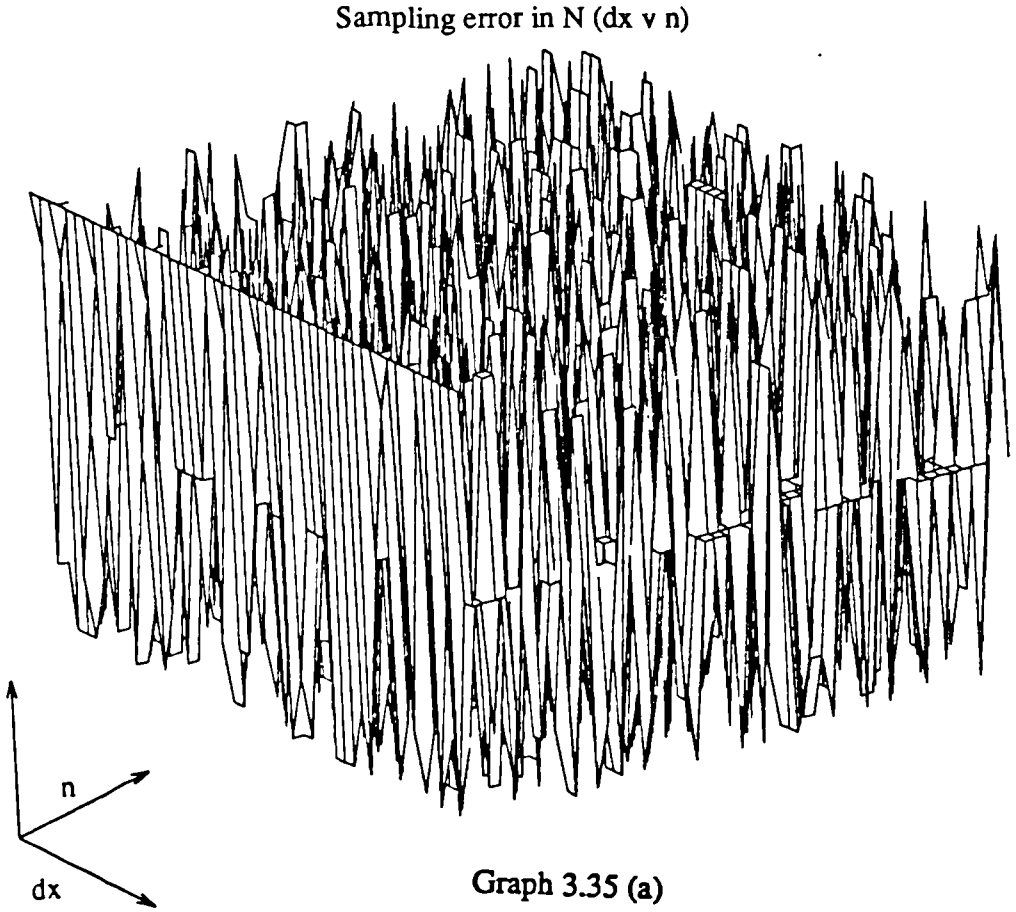


Graph 3.34 (a)

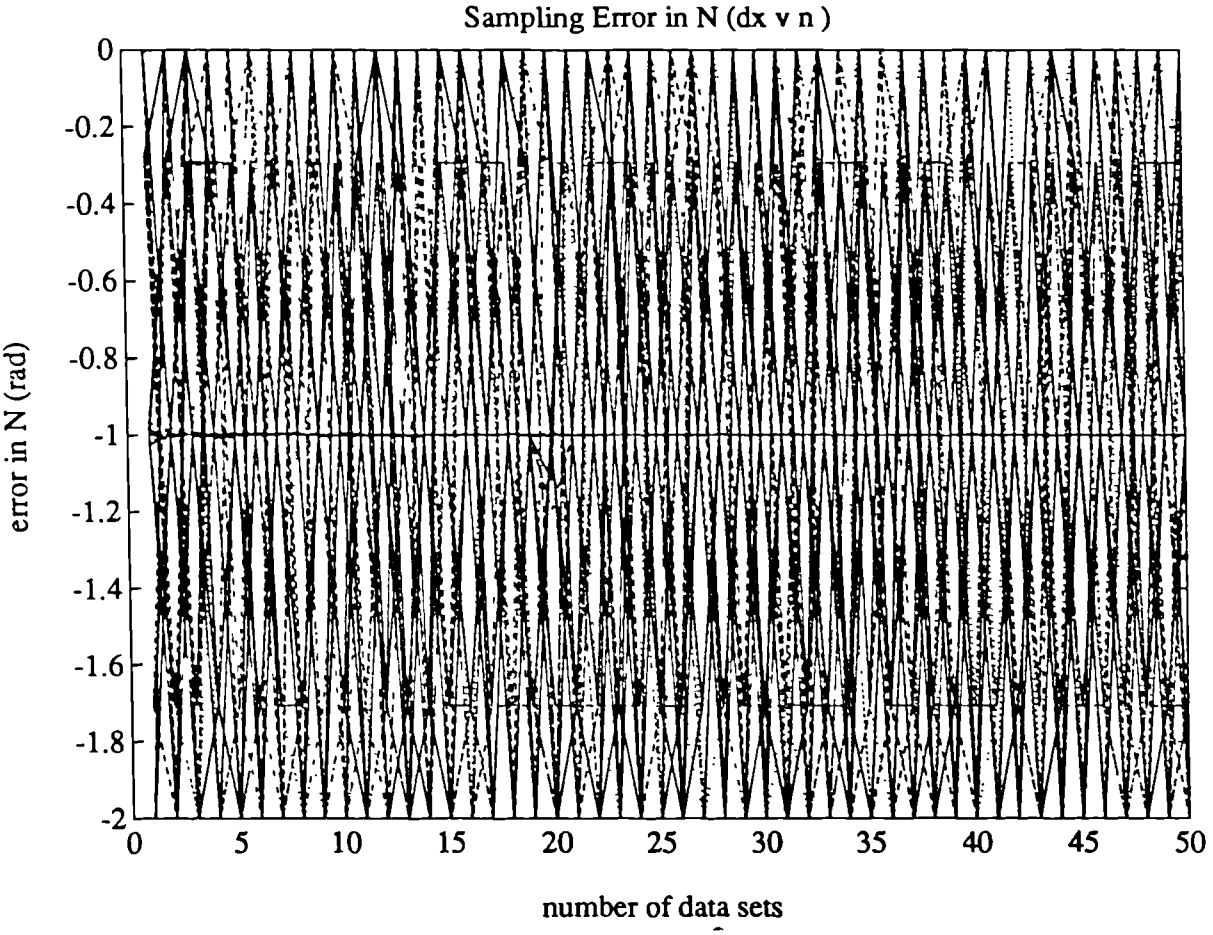
Sampling Error in M (dx v n)



Graph 3.34 (b)

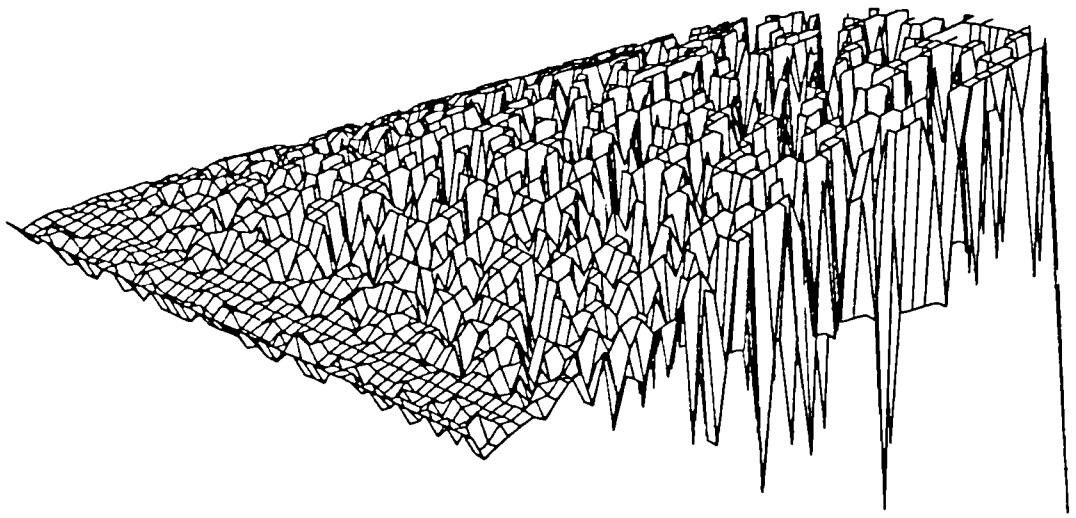


Graph 3.35 (a)

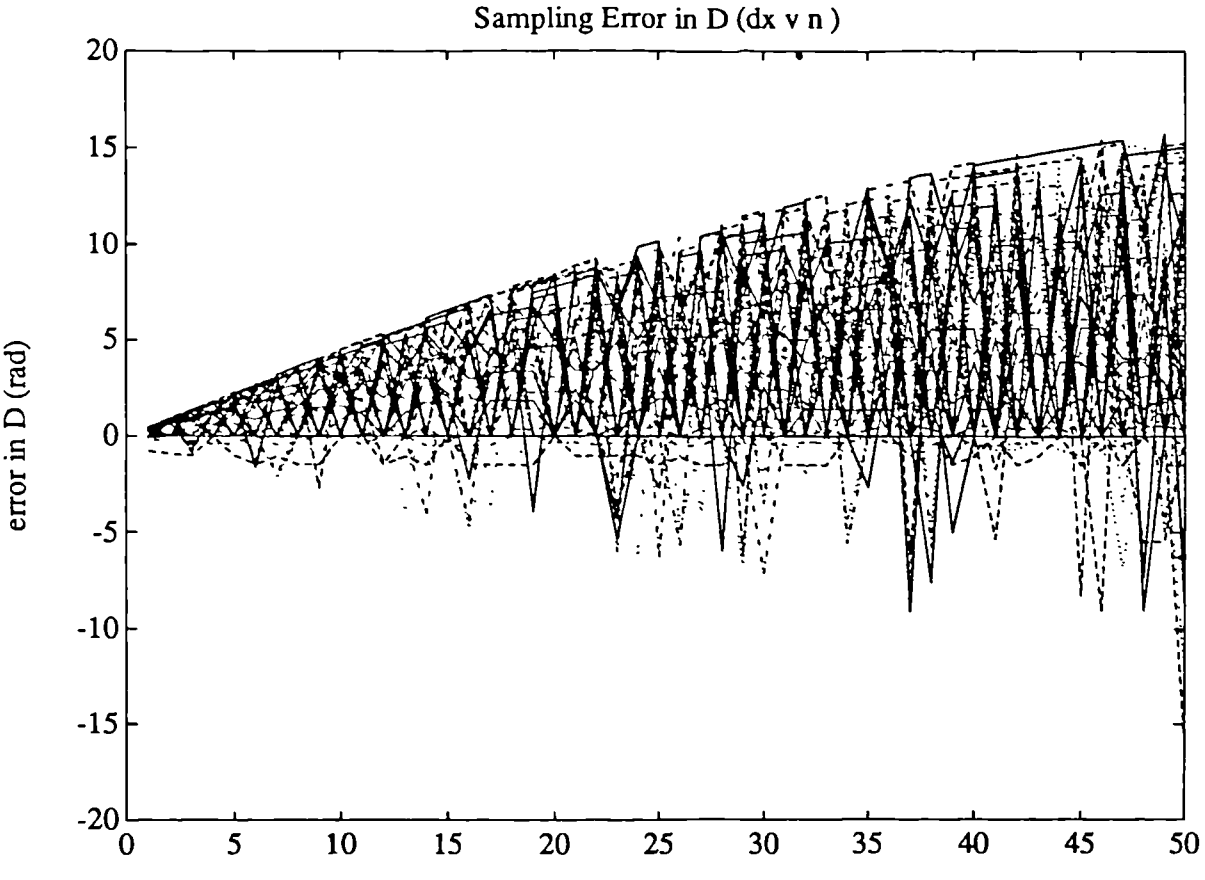


Graph 3.35 (b)

Sampling error in D (dx v n)



Graph 3.36 (a)



Graph 3.36 (b)

3.2.2.2. Results

The results obtained by the use of the Eigenvalue method are somewhat surprising. It is seen that the accuracy of the estimates of L, M, N and D are not necessarily improved by the use of more data (rows of data). Errors will be incurred by the data used in the experiment, i.e. evenly placed rows of equidistributed data points. This effectively means that the plane is not neccessarily more fully represented by the use of increased data placed in this way; but as explained earlier, there is little choice for the purposes of this investigation, as the experiments must remain systematic.

Additional errors in this experiment may result from compromises made to prevent prohibitively long computation times, such that, it was necessary to limit the number of data used to fully represent the plane, and hence give the estimate of the actual plane parameters. For these experiments, the data was limited to one point per $2mm^2$. Even using this amount of data, the programmes took in excess of 9 hours to run! The effect of this limitation is that where surfaces of relatively high frequency characteristics (> 0.25Hz) are concerned, this data distribution will not provide adequate estimates of true geometric parameters.

3.3. Circle Fitting Algorithms

Many algorithms exist for the estimation of circular parameters. Most of these routines are based on least squares criteria or a second degree equation (9)and predict best-estimates of radius and centre coordinates, but a series of alternative reference figures exist called limacons which are formulated to predict the required parameters according to ring-gauge, plug-gauge, minimum zone and least-squares criteria (2,10). However the very nature of the limacon renders it unsuitable for coordinate measurement applications, i.e. it is a circle linearised in its parameters about its own origin. The derivation of the limacon relies upon the assumption that the eccentricity is approximately zero. In coordinate measurement terms this effectively means that the initial estimate of the location

of the circle centre would have to be virtually perfect.

Algorithms based on least squares criteria are many and varied. Some rely upon the assumption that the residuals from the reference figure are independent i.e. Gaussian or normally distributed (11). Hitherto investigations into the random and systematic errors incurred in the least squares estimation of circular parameters has been confined to the specific case of Gaussian distributed residuals. Two least-squares derivations for the estimations of circular parameters and subsequent error analyses are outlined, each formulated so that explicit expressions can be given for both the random and systematic errors in these estimates. The fundamental difference between these two algorithms is that in the first, the assumption of Gaussian data distribution is made in the actual formulation of the parameter estimation procedure, the second does not make this assumption until the derivation of the error analyses.

An original investigation into the sensitivity of the least-squares estimates of all circular parameters is performed, using simulated surface data in which the residuals from the reference figure are periodic.

3.3.1. The Least Squares Circle(1)

Consider the circle represented by figure 3.2, of nominal radius R, having centre coordinates (a,b).

The following analysis has been formulated in order to ultimately derive explicit expressions for the systematic and random errors incurred in least-squares estimation of circular contours with respect to changes in the angle of arc upon which the data points are distributed and the closeness of the reference and contour centres (11).

Consider a circular contour of radius $\hat{R}$ that has upon its profile a random fluctuation having an expected value of zero . Also consider a reference circle of radius R whose centre is displaced from the centre of the original circle by a distance having orthogonal components a and b with respect to a set of cartesian coordinates established at the refer-

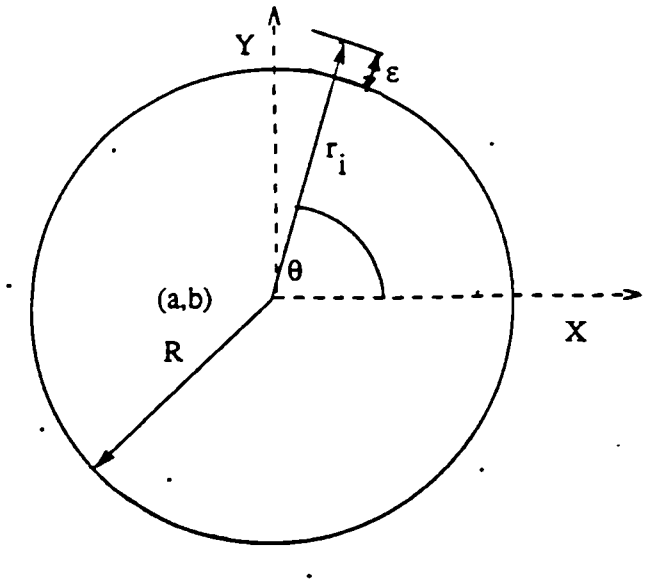


Figure 3.1

ence circle centre. Finally consider a radial measurement r_i made from the reference centre to the circular contour at an angle θ_i with respect to the X axis.

This geometry gives:

$$[r_i \cos\theta_i - b]^2 + [r_i \sin\theta_i - a]^2 = R^2 \tag{3.18}$$

Squaring, combining the terms and solving for r_i gives:

$$r_i = b\cos\theta_i = a\sin\theta_i + [(b\cos\theta_k + a\sin\theta_i)^2 - (a^2 + b^2 - R^2)]^{1/2} \tag{3.19}$$

If a and b are small relative to R equation (3.19) simplifies to the following:

$$r_i \simeq a\cos\theta_i + b\sin\theta_i + R \tag{3.20}$$

Random deviations due to either surface characteristics of the component under inspection and measurement error 'contaminate' the measurement r_i by an additive error term ε_i, having a mean value of zero. i.e.

$$r_k \equiv r_i + \varepsilon_i + b\cos\theta_i + a\sin\theta_i + R + \varepsilon_i \tag{3.21}$$
$$k = 1,..,n$$

Providing that the angles $\theta_1 \rightarrow \theta_i$ are sufficiently large for the errors ε_i to be considered independent, equation (3.21) may be re-written in matrix form as:

$$\begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ r_n \end{bmatrix} = \begin{bmatrix} 1 & \cos\theta_1 & \sin\theta_1 \\ 1 & \cos\theta_2 & \sin\theta_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \cos\theta_n & \sin\theta_n \end{bmatrix} \begin{bmatrix} R \\ b \\ a \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \tag{3.22}$$

or

$$\underline{r} = \underline{aB} + \underline{\varepsilon} \tag{3.23}$$

The estimates of the unknown parameters are given by the matrix equation (3.24)

$$\underline{\hat{B}} = [\underline{a}'\,\underline{a}]^{-1}\,\underline{a}'\,\underline{r} \tag{3.24}$$

Where $\underline{\hat{B}}$ is the vector of the estimates [R, b, a]. The symmetric product matrix is calculated to be :

$$\underline{a}'\,\underline{a} = \begin{bmatrix} n & \sum\cos\theta_i & \sum\sin\theta_i \\ \sum\cos\theta_i & \sum\cos^2\theta_i & \sum\sin\theta_i\cos\theta_i \\ \sum\sin\theta_i & \sum\sin\theta_i\cos\theta_i & \sum\sin^2\theta_i \end{bmatrix} \tag{3.25}$$

This matrix may be greatly simplified by placing the n data points equiangularly and symmetrically about the y and x axes, i.e.

$$\underline{a}'\,\underline{a} = \begin{bmatrix} n & \sum\cos i\,\delta\theta & 0 \\ \sum\cos i\,\delta\theta & \sum\cos^2 i\,\delta\theta & 0 \\ 0 & 0 & \sum\sin^2 i\,\delta\theta \end{bmatrix} \tag{3.26}$$

Getting the inverse of this matrix may be greatly simplfied by partitioning the matrix into the form

$$\underline{a}'\underline{a} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{3.27}$$

Where A is a 2×2 matrix:

$$A = \begin{bmatrix} n & \sum \cos i\,\delta\theta \\ \sum \cos i\,\delta\theta & \sum \cos^2 i\,\delta\theta \end{bmatrix} \tag{3.28}$$

The inverse of $\underline{a}'\underline{a}$ is

$$[\underline{a}'\underline{a}]^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \tag{3.29}$$

Which can be easily evaluated to be:

$$[\underline{a}'\underline{a}]^{-1} = \frac{1}{D}\begin{bmatrix} \cos^2 i\,\delta\theta & -\cos i\,\delta\theta & 0 \\ -\cos i\,\delta\theta & n & 0 \\ 0 & 0 & \frac{D}{\sum \sin^2 i\,\delta\theta} \end{bmatrix} \tag{3.30}$$

where

$$D \equiv n\sum \cos^2 i\,\delta\theta - (\sum \cos i\,\delta\theta)^2 \tag{3.31}$$

Postmultiplying $[\underline{a}'\underline{a}]^{-1}$ by $\underline{a}'$ gives a $3 \times n$ matrix product

$$[a'a]^{-1}a' = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \\ c_1 & c_2 & \cdots & c_n \end{bmatrix} \tag{3.32}$$

Where the ith element of each type is given by :

$$a_i = \frac{\sum \cos^2 i\,\delta\theta - \cos i\,\delta\theta \sum \cos k\,\delta\theta}{n\sum \cos^2 i\,\delta\theta - (\sum \cos i\,\delta\theta)^2} \tag{3.33}$$

$$b_i = \frac{-\sum \cos i\,\delta\theta + n\,\cos\delta\theta}{n\sum \cos^2 i\,\delta\theta - (\sum \cos\delta\theta)^2} \tag{3.34}$$

$$c_i = \frac{\sin i\,\delta\theta}{\sum \sin^2 i\,\delta\theta} \tag{3.35}$$

Expressing the least squares estimates as linear combinations of the measurements a_i, b_i, c_i and r_i:

$$R = \sum a_i r_i \tag{3.36}$$

$$b = \sum b_i r_i \tag{3.37}$$

$$a = \sum c_i r_i \tag{3.38}$$

3.3.1.1. Systematic Errors

From equation (3.19) it is evident that equation (3.21) is true only for $a = b = 0$, in which case

$$r_i = R + \varepsilon_i \tag{3.39}$$

Substituting equation (3.39) into (3.36) and taking expectations gives

$$E(\hat{R}) = \sum a_i\ (R + \varepsilon_i) = R \sum a_i + \sum a_i\ E\ (\varepsilon_i) \tag{3.40}$$
$$E(\varepsilon_i) = 0$$

In order for $\hat{R}$ to be an unbiased estimate, i.e. for $E(\hat{R}) = R$ the condition

$$\sum_{-m}^{m} a_i = 1.0 \tag{3.41}$$

must be satisfied, so summing equation (3.31) between -m and m equation (3.41) holds. Similarly, for an unbiased estimation

$$\sum b_i = \sum c_i = 0 \tag{3.42}$$

Summing these between -m and m equation (3.41) still holds. It follows that when a and/or b differ from zero, the estimates of R, a and b are biased. The degree of bias or systematic error is calculated as follows.

For prescribed values of R, a, and b, the exact values of r_i are calculated from equation (3.19), the expected value of $\hat{R}$ is then found by taking expectations in equation (3.36) i.e.

$$E(\hat{R}) = \sum a_i r_i \qquad as\ E(\varepsilon_i) = 0 \tag{3.43}$$

3.3.1.2. Random Errors

A variance-covariance matrix can be written for the least squares estimates:

$$V(\hat{B}) = (a'\ a)^{-1}\sigma^2 = \begin{bmatrix} \sigma^2(\hat{R}) & cov(\hat{b}\hat{R}) & cov(\hat{a}\hat{R}) \\ cov(\hat{b}\hat{R}) & \sigma^2(\hat{b}) & cov(\hat{b}\hat{a}) \\ cov(\hat{R}\hat{a}) & cov(\hat{a}\hat{b}) & \sigma^2(\hat{a}) \end{bmatrix} \tag{3.44}$$

where σ^2 is the variance of the error terms ε_i. From equation (3.30), the elements of this matrix can be expressed in terms of σ^2, i.e.

$$\frac{\sigma^2(\hat{R})}{\sigma^2} = \frac{\sum \cos^2 i\,\delta\theta}{n \sum \cos^2 i\,\delta\theta - (\sum \cos i\,\delta\theta)^2} \tag{3.45}$$

$$\frac{\sigma^2(\hat{b})}{\sigma^2} = \frac{n}{n\sum\cos^2 k\,\delta\theta - (\sum\cos i\,\delta\theta)^2} \tag{3.46}$$

$$\frac{\sigma^2(\hat{a})}{\sigma^2} = \frac{1}{n\sum\sin^2 i\,\delta\theta} \tag{3.47}$$

$$cov(\hat{R}\hat{b}) = \frac{-\sum\cos i\,\delta\theta}{n\sum\cos^2 k\,\delta\theta - (\sum\cos i\,\delta\theta)^2} \tag{3.48}$$

$$cov(\hat{a}\hat{R}) = 0 \tag{3.49}$$

$$cov(\hat{a}\hat{b}) = 0 \tag{3.50}$$

The correlation coefficient between R and b can be expressed as:

$$r(\hat{R}\hat{b}) = \frac{cov(\hat{R}\hat{b})}{\sigma(\hat{R})\sigma(\hat{b})} = \frac{-\sum\cos i\,\delta\theta}{(n\sum\cos^2 i\,\delta\theta)^{1/2}} \tag{3.51}$$

Surface roughness and measurement error information can then be used in equations (3.45) to (3.49) in order to calculate the random errors in the estimates of the parameters.

3.3.2. Least Squares Circle (2)

The least squares criterion for circle fitting is :

$$E_s = \sum_{i=1}^{i=N}(r_i - R)^2 = \min \tag{3.52}$$

where

$$r_i = \sqrt{(x_i - a)^2 + (y_i - b)^2} \tag{3.53}$$

This expression is computationally difficult to handle, but according to Delonge (12), a modified criterion may be used:

$$\sum_{i=1}^{i=N}(r_i^{\,2} - R^2)^2 = \min \tag{3.54}$$

If (3.53) is substituted into (3.52), then E_s may be written in the approximate form:

$$E_s = \sum_{i=1}^{i=N}(r_i - R - a\,\cos\theta_i - b\,\sin\theta_i)^2 \tag{3.55}$$

where

$$x_i = R\,\cos\theta_i$$

$$y_i = R\,\sin\theta_i$$

$$a = e\,\cos\alpha$$

$$b = e\,\sin\alpha$$

$$e^2 = a^2 + b^2$$

Performing partial differentiation with respect to a, b and R on E_s gives the matrix equation for the solution of the parameters a,b and R as follows:

$$\begin{bmatrix} \sum\cos^2\theta_i & \sum\cos\theta_i\sin\theta_i & \sum\cos\theta_i \\ \sum\cos\theta_i\sin\theta_i & \sum\sin^2\theta_i & \sum\sin\theta_i \\ \sum\cos\theta_i & \sum\sin\theta_i & \sum 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ R \end{bmatrix} = \begin{bmatrix} \sum x_i \\ \sum y_i \\ \sum z_i \end{bmatrix} \tag{3.56}$$

If the spacing of the data is equiangular i.e. $\theta_1 = \theta_2 ,...,= \theta_N$, then this matrix equation reduces to :

$$a = \frac{2\sum x_i}{N} \tag{3.57}$$

$$b = \frac{2\sum y_i}{N} \tag{3.58}$$

$$R = \frac{r_i}{N} \tag{3.59}$$

This formulation can be represented in cartesian coordinates:

Introducing the new variable C i.e.

$$C = R^2 - A^2 - B^2$$

the solution may be rewritten

$$A\ 2\sum x_i + B\ 2\sum y_i + C\ N = \sum(x_i^2 + y_i^2) \tag{3.60}$$

$$A\ 2\sum x_i^2 + B\ 2\sum x_i y_i + C\sum x_i = \sum(x_i^3 + x_i y_i^2) \tag{3.61}$$

$$A\ 2\sum x_i y_i + B\ 2\sum y_i^2 + \sum y_i = \sum(x_i^2 y_i + y_i^3) \tag{3.62}$$

Choosing the variables

$$D = \begin{bmatrix} 2\sum x_i & 2\sum y_i & N \\ 2\sum x_i^2 & 2\sum x_i y_i & \sum x_i \\ 2\sum x_i y_i & 2\sum y_i^2 & \sum y_i \end{bmatrix} \tag{3.63}$$

$$E = \begin{bmatrix} \sum(x_i^2 + y_i^2) \\ \sum(x_i^3 + x_i y_i^2) \\ \sum(x_i^2 y_i + y_i^3) \end{bmatrix} \tag{3.64}$$

and

$$Q = \begin{bmatrix} C \\ B \\ A \end{bmatrix} \tag{3.65}$$

The solution can be written in matrix form:

$$E = D\ Q \tag{3.66}$$

and the solution to the parameters as:

$$Q = D^{-1} E \tag{3.67}$$

This analysis provides a readily implementable algorithm that is reliable and computationally efficient provided that the centre from which the coordinates are related is relatively close to the true centre of the circle, and the entire circumference is adequately represented by data. However, if these criterion are not satisfied, then systematic errors will occur.

3.3.2.1. Random Errors

Another method of calculating the random errors in the calculation of circular parameters was developed by Kasa (12):

Data is sampled from points on a circle :

$$(a - A_o)^2 + (b - B_o)^2 = R_o^2 \tag{3.68}$$

N coordinates (x_i , y_i), i=1,...,N, are used and are such that:

$$x_i = x_{i0} + \xi_i \tag{3.69}$$

$$y_i = y_{i0} + \eta_i \tag{3.70}$$

where (x_{i0}, y_{i0}) satisfy equation (3.68) and (ξ_i, η_i) represents a pure random component i.e $E(\xi_i) = E(\eta_i) = 0$, and the aim of this section is to quantify the effect of this random component upon the estimates of A, B and R. Errors which are relatively small compared to the values of the estimates can be calculated by using the Taylor series expansion of the error variances, thus:

$$\alpha = A - A_o = \sum_{i=1}^{N}\left[\frac{\partial A}{\partial x_{i0}}\xi_i + \frac{\partial A}{\partial y_{i0}}\eta_i\right] + \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\partial^2 A}{\partial x_{i0}^2}\xi_i^2 + 2\frac{\partial^2 A}{\partial x_{i0}y_{i0}}\xi_i\eta_i + \frac{\partial^2 A}{\partial y_{i0}^2}\eta_i^2\right] + \ldots\ldots \tag{3.71}$$

The second order and higher terms are deemed to be negligably small. Expansion for the error terms in A, B, C and R results in similar series, therefore the first order terms for the parameters are, respectively

$$\alpha = \sum_{i=1}^{N}\left[\frac{\partial A}{\partial x_{i0}}\xi_i + \frac{\delta A}{\partial y_{i0}}\eta_i\right] \tag{3.72}$$

$$\beta = \sum_{i=1}^{N} \left\{ \frac{\partial B}{\partial x_{i0}} \xi_i + \frac{\partial B}{\partial y_{i0}} \eta_i \right\} \tag{3.73}$$

$$\gamma = \sum_{i=1}^{N} \left\{ \frac{\partial C}{\partial x_{i0}} \xi_i + \frac{\partial C}{\partial y_{i0}} \eta_i \right\} \tag{3.74}$$

$$r = \sum_{i=1}^{N} \left\{ \frac{\partial R}{\partial x_{i0}} \xi_i + \frac{\partial C}{\partial y_{i0}} \eta_i \right\} \tag{3.75}$$

where

$$\frac{\partial R}{\partial Z} = \frac{1}{2R_o} \left[2A_o \frac{\partial A}{\partial Z} + 2B_o \frac{\partial B}{\partial Z} + \frac{\partial C}{\partial Z} \right] \tag{3.76}$$

For this first order approximation :

$$E(\alpha) = E(\beta) = E(\gamma) = E(r) = 0$$

and

$$\sigma_A^2 = \sigma^2 \sum_{i=1}^{N} \left\{ \left\{ \frac{\partial A}{\partial x_{i0}} \right\}^2 + \left\{ \frac{\partial A}{\partial y_{i0}} \right\}^2 \right\} \tag{3.77}$$

$$\sigma_B^2 = \sigma^2 \sum_{i=1}^{N} \left\{ \left\{ \frac{\partial B}{\partial x_{i0}} \right\}^2 + \left\{ \frac{\partial B}{\partial y_{i0}} \right\}^2 \right\} \tag{3.78}$$

$$\sigma_{AB} = \sigma^2 \sum_{i=1}^{N} \left\{ \left\{ \frac{\partial A . \partial B}{\partial x_{i0}} \right\} + \left\{ \frac{\partial A . \partial B}{\partial y_{i0}} \right\} \right\} \tag{3.79}$$

$$\sigma_C = \sigma^2 \sum_{i=1}^{N} \left\{ \left\{ \frac{\partial C}{\partial x_{i0}} \right\}^2 + \left\{ \frac{\partial C}{\partial y_{i0}} \right\}^2 \right\} \tag{3.80}$$

$$\sigma_R = \sigma^2 \sum_{i=1}^{N} \left\{ \left\{ \frac{\partial R}{\partial x_{i0}} \right\}^2 + \left\{ \frac{\partial R}{\partial y_{i0}} \right\}^2 \right\} \tag{3.81}$$

where

$$\sigma^2 = E(\xi_i^2) = E(\eta_i^2)$$

and the covariance

$$\sigma_{\xi\eta} = E(\xi_i, \eta_i) = 0$$

As all first order terms are proportional to the quantity σ, it is convenient to normalise with respect to it , the error sensitivities of the estimated parameters can be written:

$$S_A = \frac{\sigma_A}{\sigma} \tag{3.82}$$

$$S_B = \frac{\sigma_B}{\sigma} \tag{3.83}$$

$$S_{AB} = \frac{\sigma_{AB}}{\sigma} \tag{3.84}$$

$$S_C = \frac{\sigma_C}{\sigma} \tag{3.85}$$

$$S_R = \frac{\sigma_R}{\sigma} \tag{3.86}$$

In order to calculate these sensitivities , eqn 3.62 must be differentiated partially i.e.

$$\dot{D}Q + D\dot{Q} = \dot{E} \tag{3.87}$$

$$\dot{Q} = \begin{bmatrix} \dot{A} \\ \dot{B} \\ \dot{C} \end{bmatrix} \tag{3.88}$$

Therefore:

$$\dot{Q} = D^{-1}(\dot{E} - \dot{D}Q) \tag{3.89}$$

The explicit derivatives with respect to x_i and y_i are :

$$\frac{\partial E}{\partial x_i} = \begin{bmatrix} 2x_i \\ 3x_i^2 + y_i^2 \\ 2x_i y_i \end{bmatrix} \tag{3.90}$$

$$\frac{\partial E}{\partial y_i} = \begin{bmatrix} 2y_i \\ 2x_i y_i \\ x_i^2 + 3y_i^2 \end{bmatrix} \tag{3.91}$$

$$\frac{\partial D}{\partial x_i} = \begin{bmatrix} 2 & 0 & 0 \\ 4x_i & 2y_i & 1 \\ 2y_i & 0 & 0 \end{bmatrix} \tag{3.92}$$

$$\frac{\partial D}{\partial y_i} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 2x_i & 0 \\ 2x_i & 4y_i & 1 \end{bmatrix} \tag{3.93}$$

3.3.2.2. The Sensitivities of the Parameters with Respect to the Number of Data Points Used

If the centre coordinates of the circle are A_o and B_o and

$$A_o = B_o = 0$$

then

$$x_{i0} = R_o \cos\phi_i$$

$$y_{i0} = R_o \sin\phi_i$$

where

R = radius of the circle

ϕ_i = angle subtended at the centre of the circle by the ith point and the a axis

similarly

$$x_i = R_o \cos\phi_i + \xi_i$$
$$y_i = R_o \sin\phi_i + \eta_i$$

If the data are equally spaced

$$\phi_i = \frac{i2\pi}{N} \qquad i=1,...,N$$

For a first order approximation

$$D = N\begin{bmatrix} 0 & 0 & 1 \\ R_o^2 & 0 & 0 \\ 0 & R_o^2 & 0 \end{bmatrix} \tag{3.94}$$

Therefore the square of the sensitivity of the parameter A with respect to the number of points used is

$$D^{-1} = \frac{1}{N}\begin{bmatrix} 0 & R_o^{-2} & 0 \\ 0 & 0 & R_o^{-2} \\ 1 & 0 & 0 \end{bmatrix} \tag{3.95}$$

and

$$\frac{\partial Q}{\partial x_i} = \frac{2}{N}\begin{bmatrix} \cos^2\phi_i \\ \cos\phi_i \sin\phi_i \\ R_o \cos\phi_i \end{bmatrix} \tag{3.96}$$

$$\frac{\partial Q}{\partial y_i} = \begin{bmatrix} \cos\phi_i \sin\phi_i \\ \frac{2}{N}\sin^2\phi_i \\ R_o \sin\phi_i \end{bmatrix} \tag{3.97}$$

Therefore

$$S_A{}^2 = \left[\frac{2}{N}\right]\sum_{i=1}^{N}(\cos^4\phi_i - \cos^2\phi_i \sin^2\phi_i) \simeq \frac{2}{N} \tag{3.98}$$

Therefore

$$S_A = \sqrt{\frac{2}{N}} \tag{3.99}$$

Similarly the sensitivities of the other parameters may be written

$$S_B = \sqrt{\frac{2}{N}} \quad S_{AB} = 0 \quad S_C = \frac{2R_o}{\sqrt{N}} \quad S_R = \frac{1}{\sqrt{N}} \tag{3.100}$$

It is clear from these equations that the stability of the estimates continues to improve with an increasing number of data points used in the calculations. Although this

is useful information for a broad guideline, no indication can be given as to how many data points would be required to achieve acceptable limits of stability, nor of the point at which the use of more data would not render any significant improvement in accuracy. The following investigation gives a far more detailed set of guidelines for coordinate measurement practice.

3.3.3. Investigation into the Sampling Error Incurred in the Least Squares Estimates of Circular Parameters.

The above analyses of errors in the estimation of circular parameters all make the assumption that errors due to measurement and surface profile effects are random and a nominally Gaussian or normal distribution. In circular measurement this assumption is not neccessarily true. Circular components are usually manufactured by turning, which involves the workpiece being constrained in a chuck and therefore compressed at several equispaced points around the circumference, when the component is removed from the chuck, the compression is released and 'lobing' appears as shown in the 'Talyrond' profile traces in fig. 3.3 Additionally, there is a degree of probe pre-travel variation due to varying triggering force with respect to the direction of deflection, the triggering forces , and therefore probe errors, vary sinusoidally with respect to the angular position of the deflection. The measuring errors incurred by CMMs in the inspection of circular components cannot therefore be realistically considered random Both phenomena occur frequently, so for circular measurement on CMMs, any error analysis including the assumption that the errors due to surface profile and measurement are near Gaussian, would in many cases result in substantial inaccuracies. In this section attention is now focussed upon instabilities in the estimated centre coordinates and radius due to sinusoidal elements in the circular profile.

Nawara et. al.(13) investigated the effects of the two most common types of periodic error of form i.e. ovality and triangularity upon the estimates of the circular parameters calculated using both the three and four point methods. The experiments

were designed to ascertain which of these algorithms was the most suitable to use when inspecting components with one or other of these errors of form. The findings of these experiments pre-suppose a knowledge of the form of the component before measurement commences and therefore have little general applicability.

The intention of the following section is to investigate these instabilities, and then to discover the extent to which these effects are due to sampling error.

3.3.3.1. Experimental Procedure

All experiments were carried out by simulation methods on a SUN 3/60 computer. Due to the arguments set out in section 3.3.3 the radius of the simulated circular profile was modeled as being sinusoidal, thus:

$$r = r_o + w \sin(fln + phi) \tag{3.101}$$

where

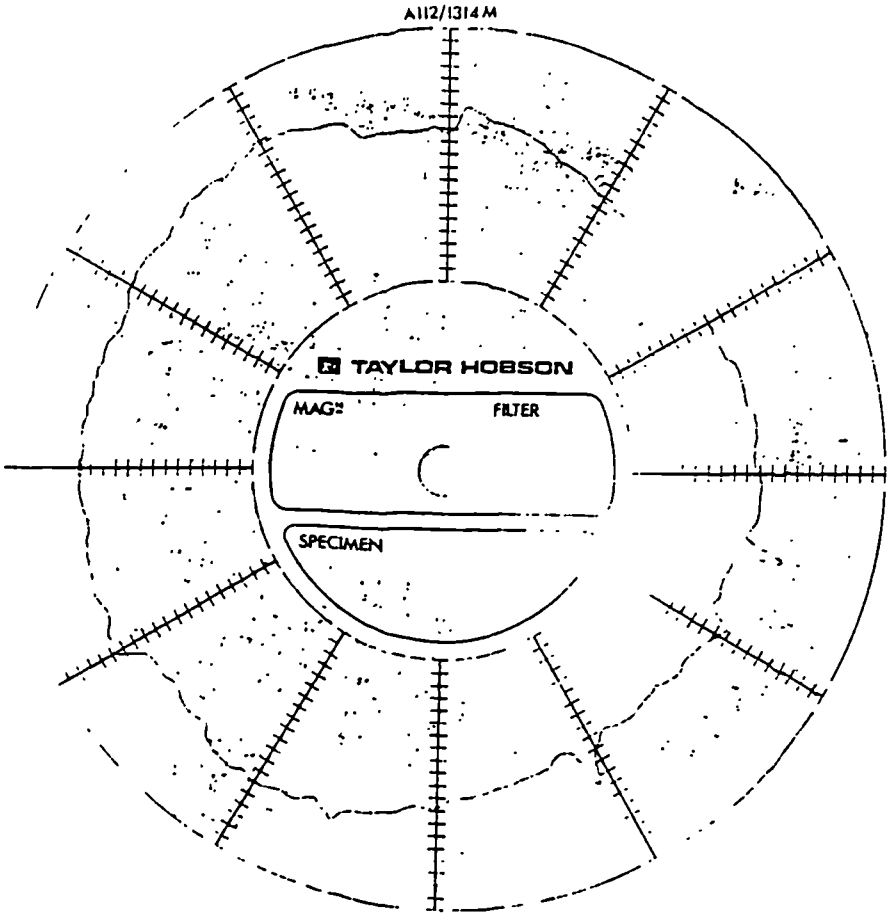
$$l = \frac{2\pi r_o \theta}{N}$$

N = number of data points

n = [1:N]

In order to ascertain the instability of the estimated values of centre coordinates and radius with respect to the input parameters of N, f, ϕ w and r, a series of experiments were performed using sets of data having between 3 and 50 data points estimating the radial and axial parameters using data having three of the input values remaining constant and the other changing at a constant rate so that a series of three dimensional plots of sampling error against number of data points used and each other parameter in turn were constructed.

The initial values of a, b, and r were then subtracted from these estimates. Care was


A112/1314M
TAYLOR HOBSON
MAG
FILTER
SPECIMEN

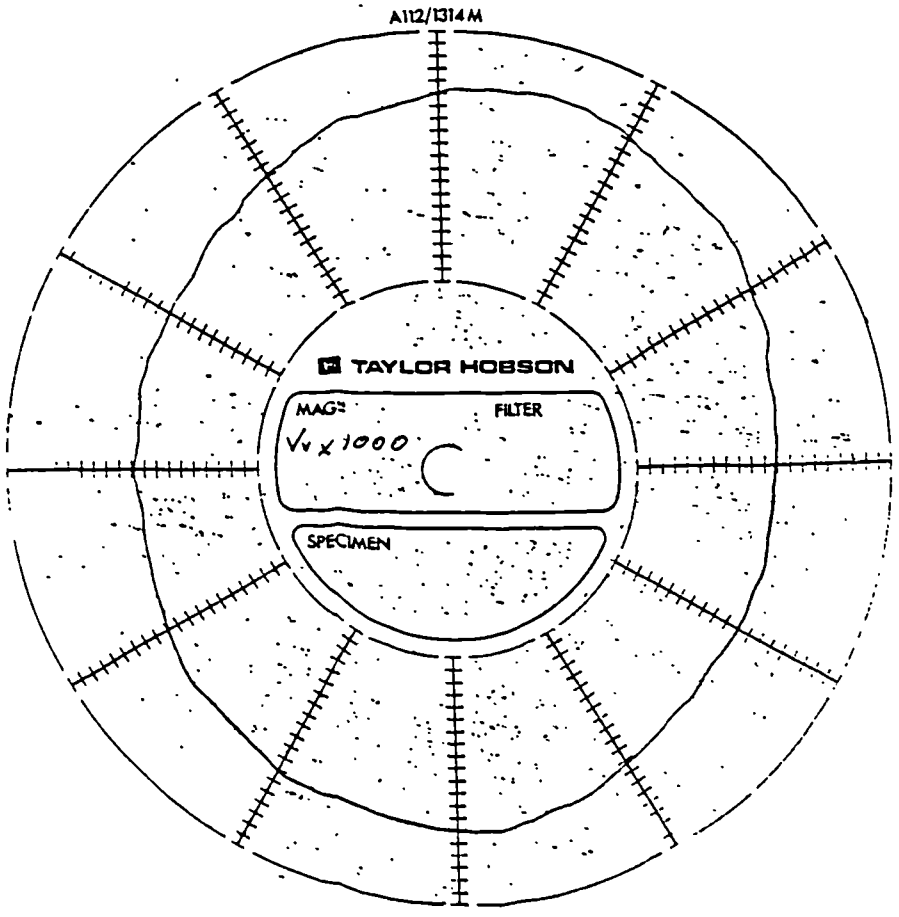

A112/1314M
TAYLOR HOBSON
MAG
FILTER
√v x 1000
SPECIMEN

Figure 3. 2

taken so that the entire circumference of the profile was represented by data, i.e. no partial arcs were used.

To find the effects of sampling error the circular parameters were calculated in a similar manner to that above, but additional estimates were made using 500 data points (in order to adequately represent the true surface profile), and these estimates were taken as being the true values, and were subtracted from the former set of estimates. The result was taken as the sampling error.

The input parameters used in the experiments described above are given in table 3.3, where

w = amplitude of sinusoidal circumferencial distortion

phi = phase of sinusoidal circumferencial distortion

f = frequency of sinusoidal circumferencial distortion

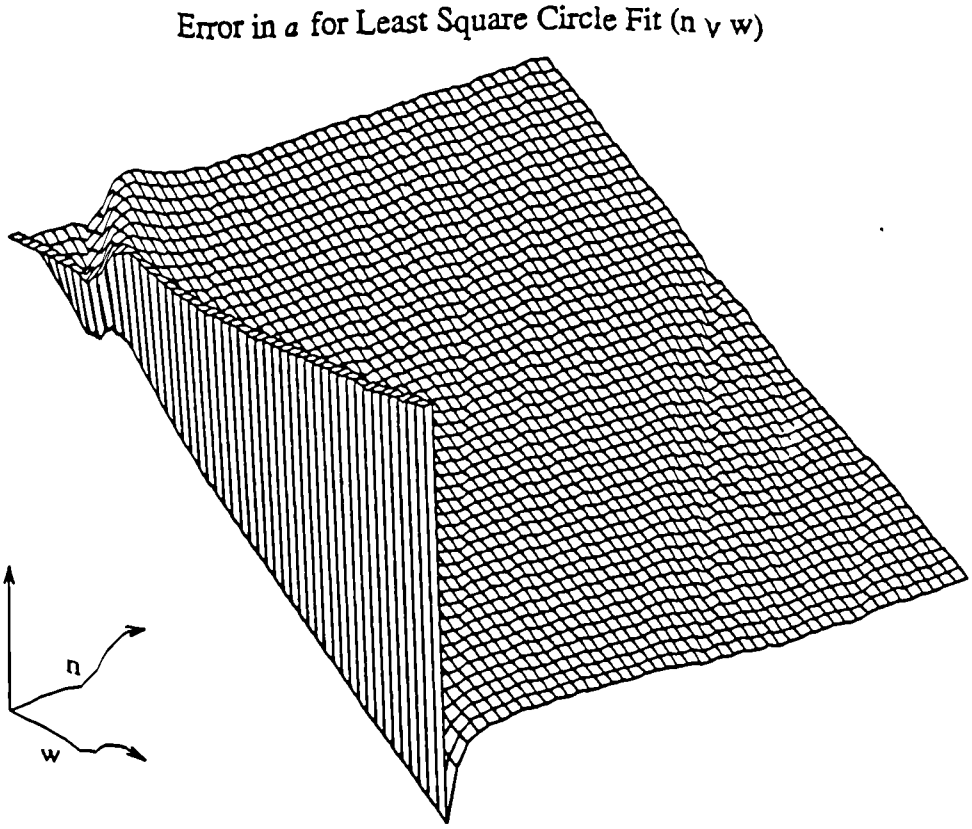
r = nominal radius of cicular profile

a = horizontal coordinate of circle centre

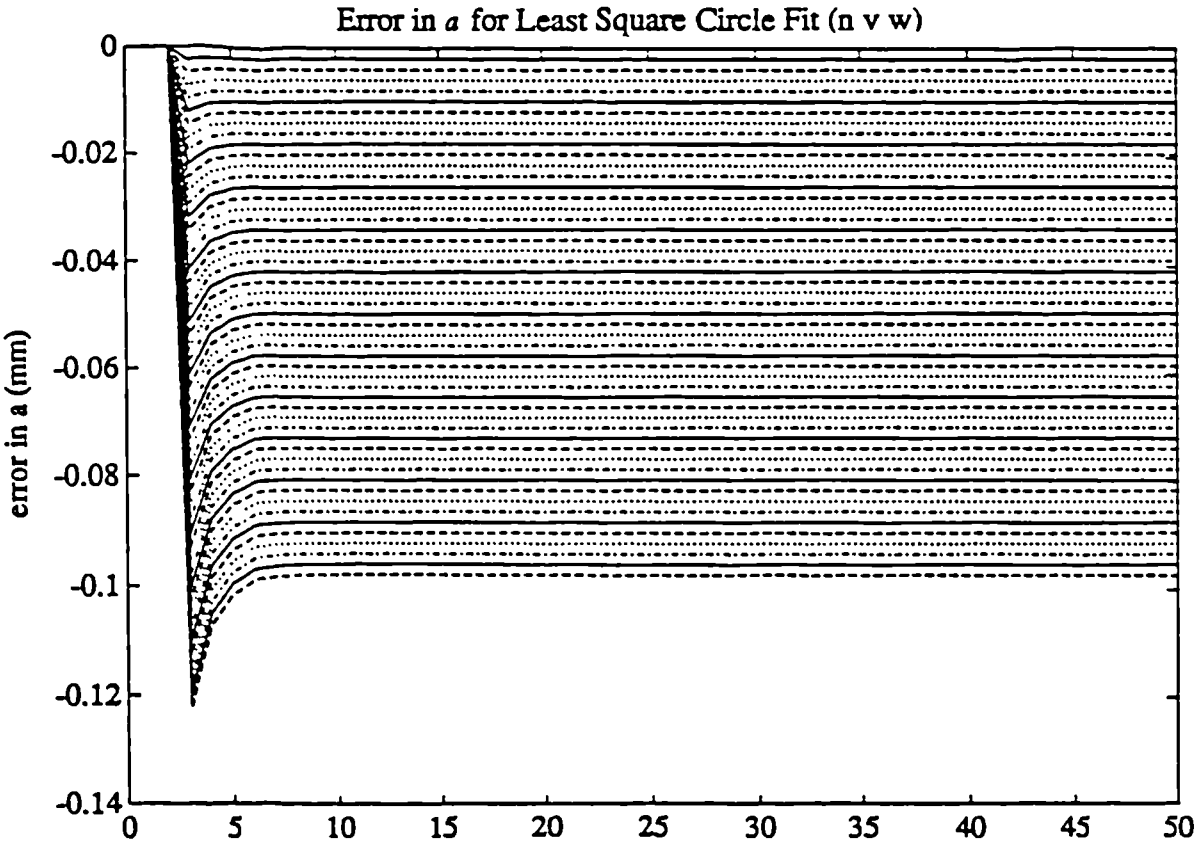
b = vertical coordinate of circle centre

Simulated Data Parameters		
Parameter	Constant	Variable
w	0.010(mm)	0.005:0.250(mm)
phi	0.000(rad)	0:2π(rad)
f	0.024(Hz)	0.008:0.398(Hz)
r	20.000(mm)	-
a	0.000(mm)	-
b	0.000(mm)	-

Table 3.3

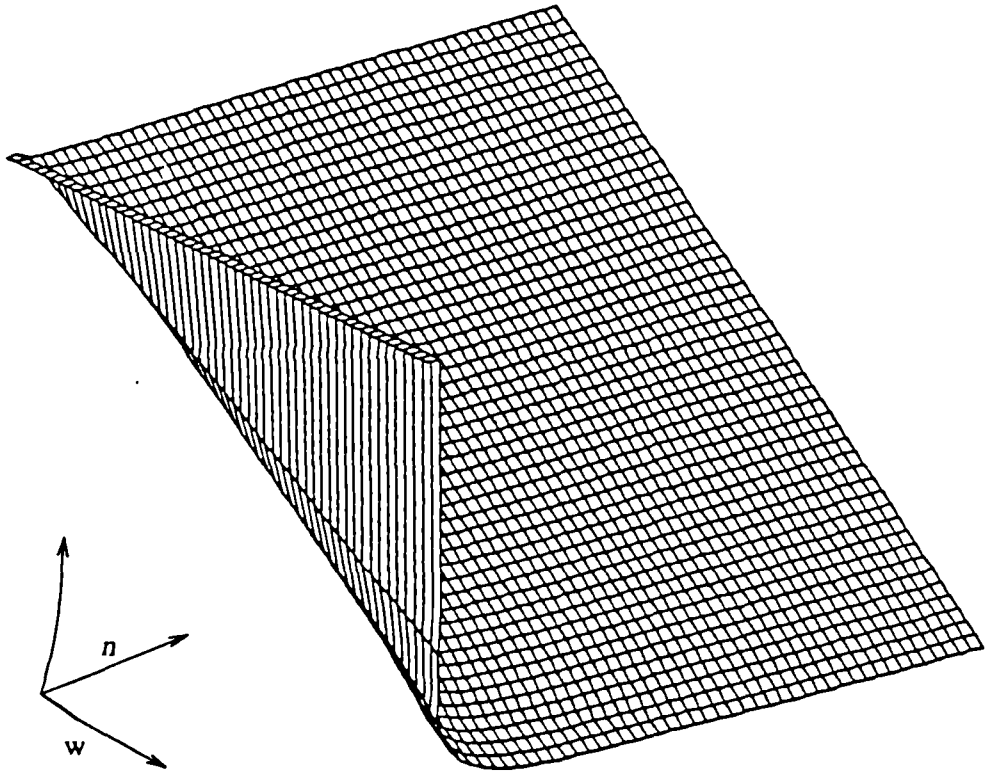


Graph 3.37 (b)



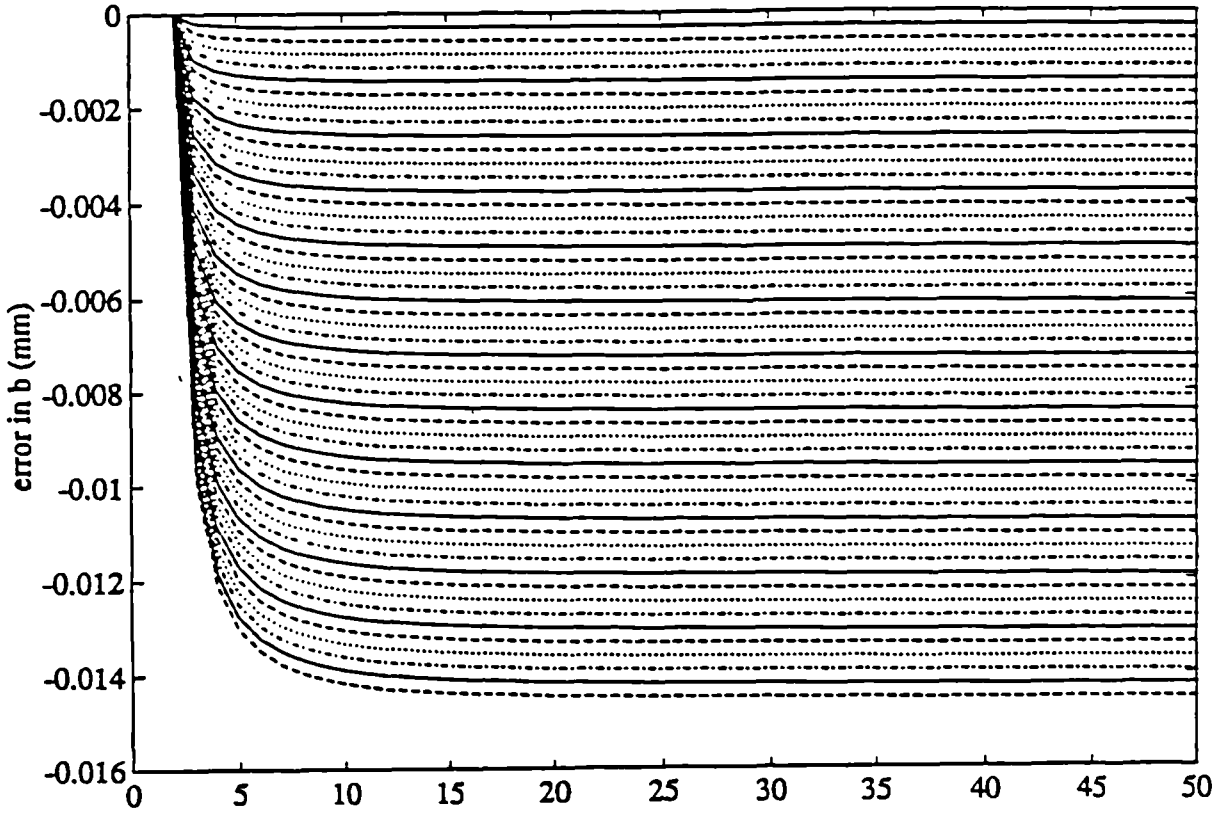
Graph 3.37 (a)

Error in *b* for Least Square Circle Fit (n v w)



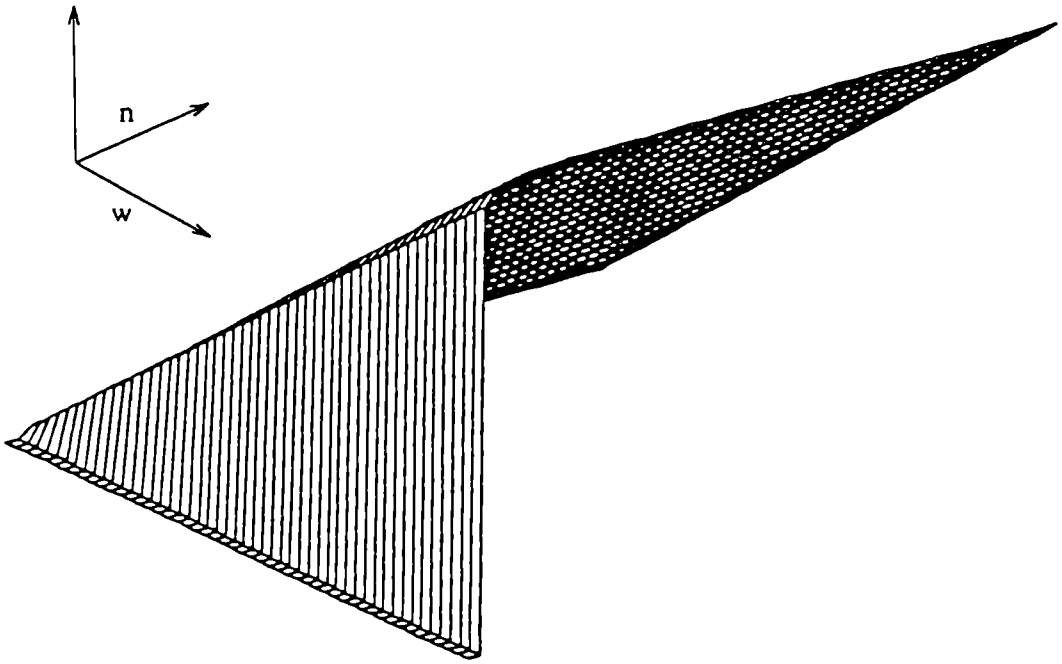
Graph 3.38 (a)

Error in *b* for Least Square Circle Fit (n v w)

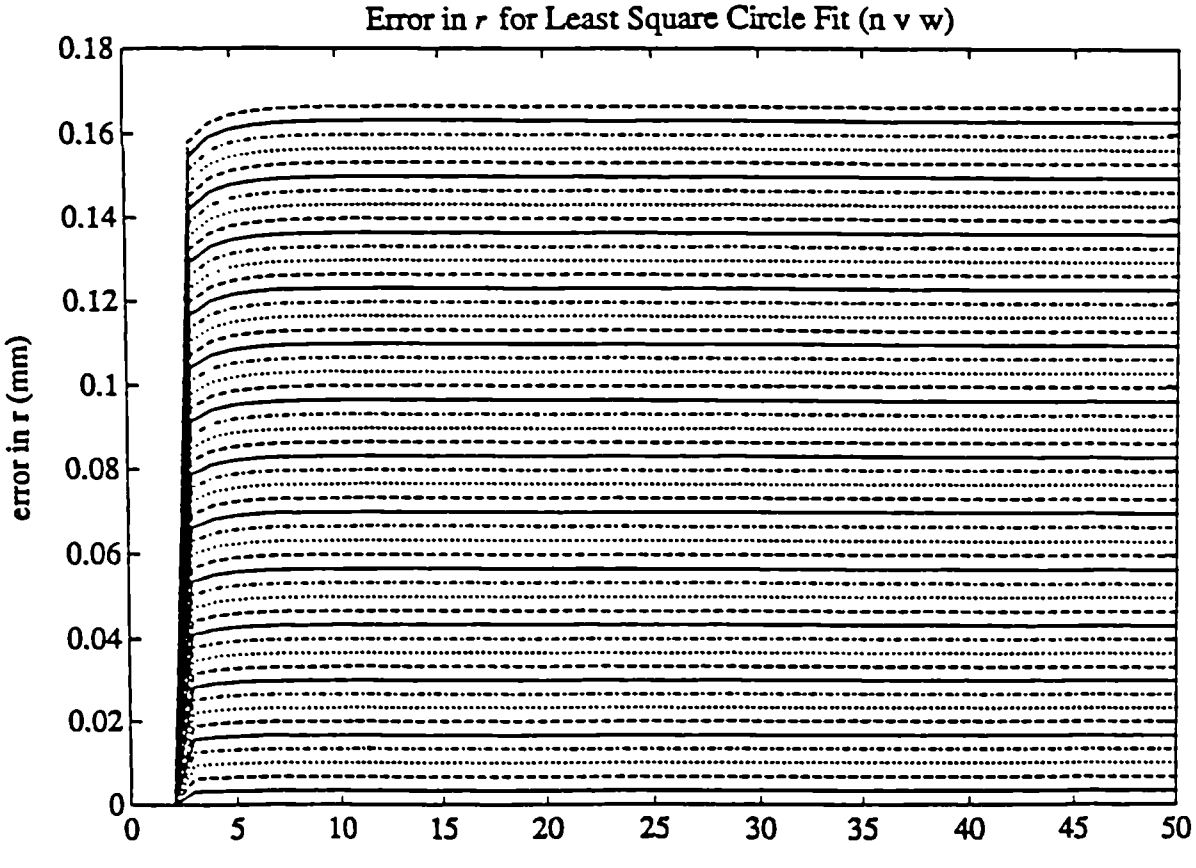


Graph 3.34 (b)

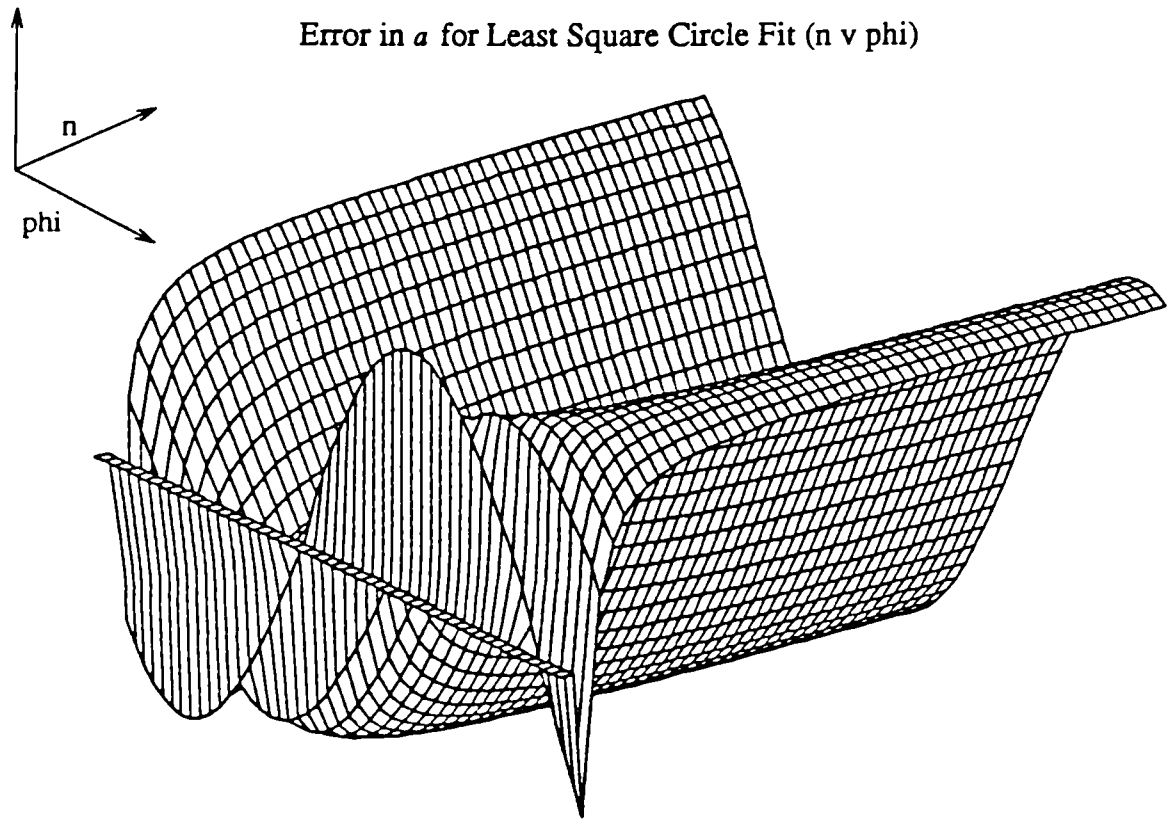
Error in r for Least Square Circle Fit (n v w)



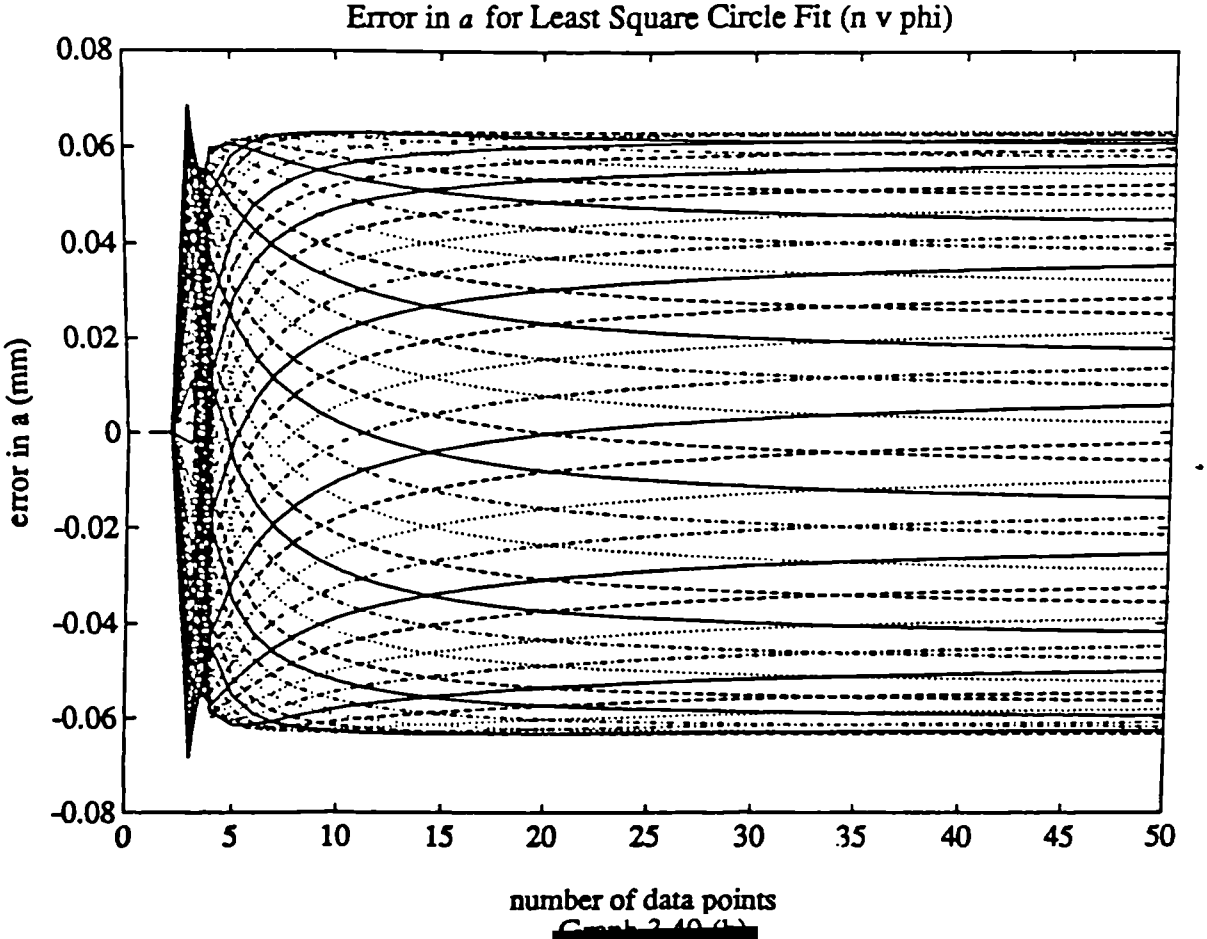
Graph 3.39 (a)



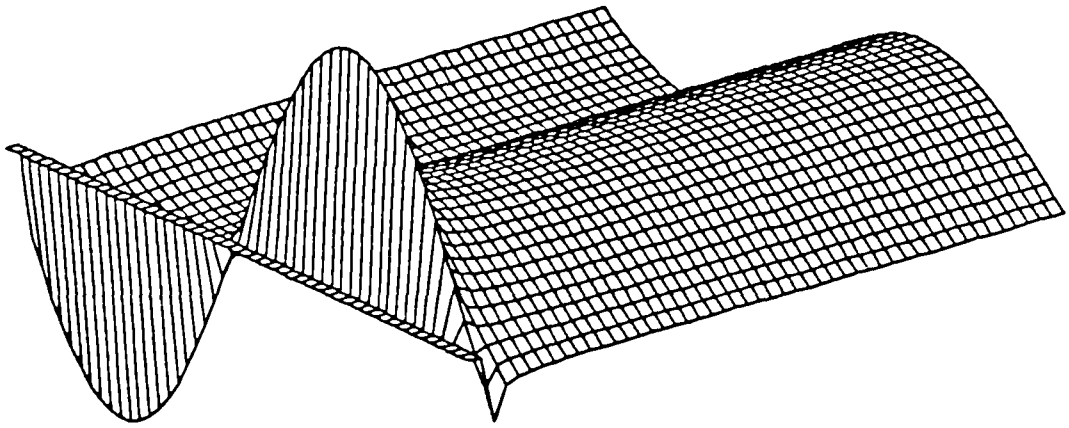
Graph 3.39 (b)

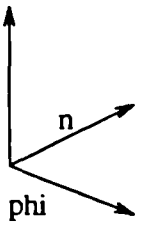

Error in a for Least Square Circle Fit (n v phi)
n
phi

Graph 3.40 (a)

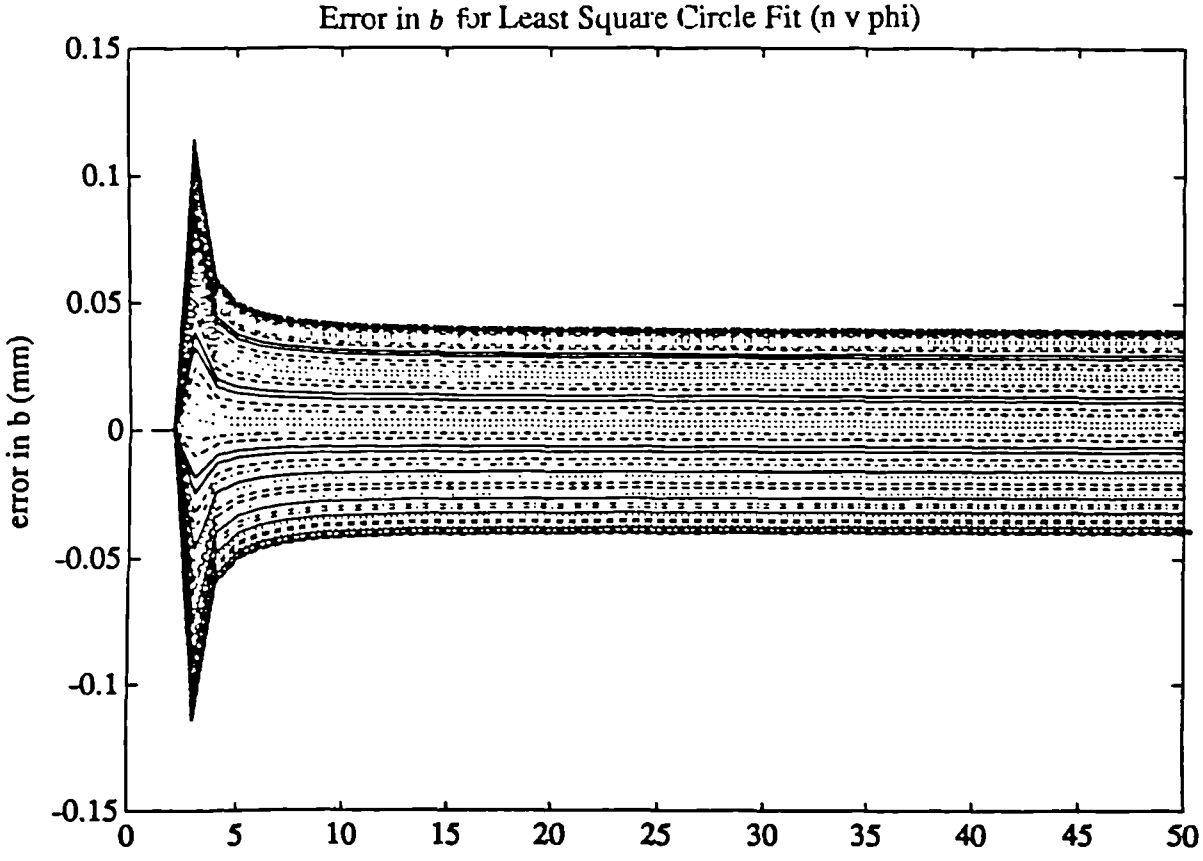

Error in a for Least Square Circle Fit (n v phi)
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
error in a (mm)
0
5
10
15
20
25
30
35
40
45
50
number of data points

Error in *b* for Least Square Circle Fit (n v phi)



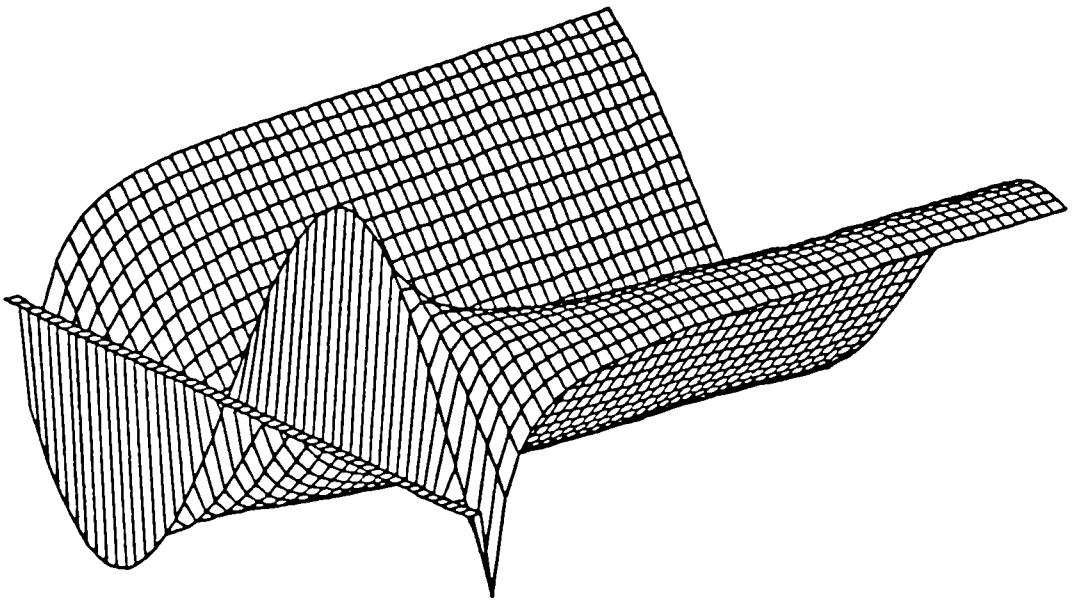


Graph 3.41 (a)

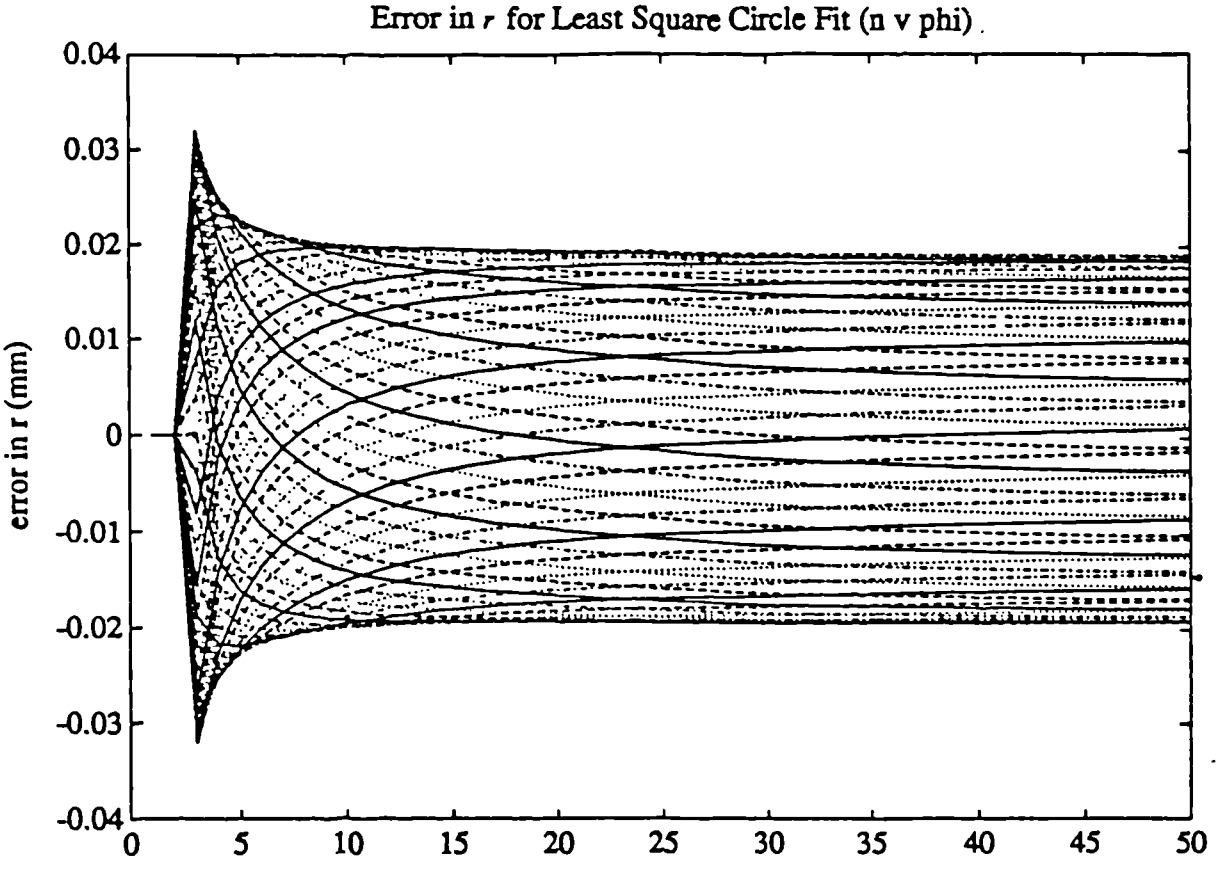


Graph 3.41 (b)

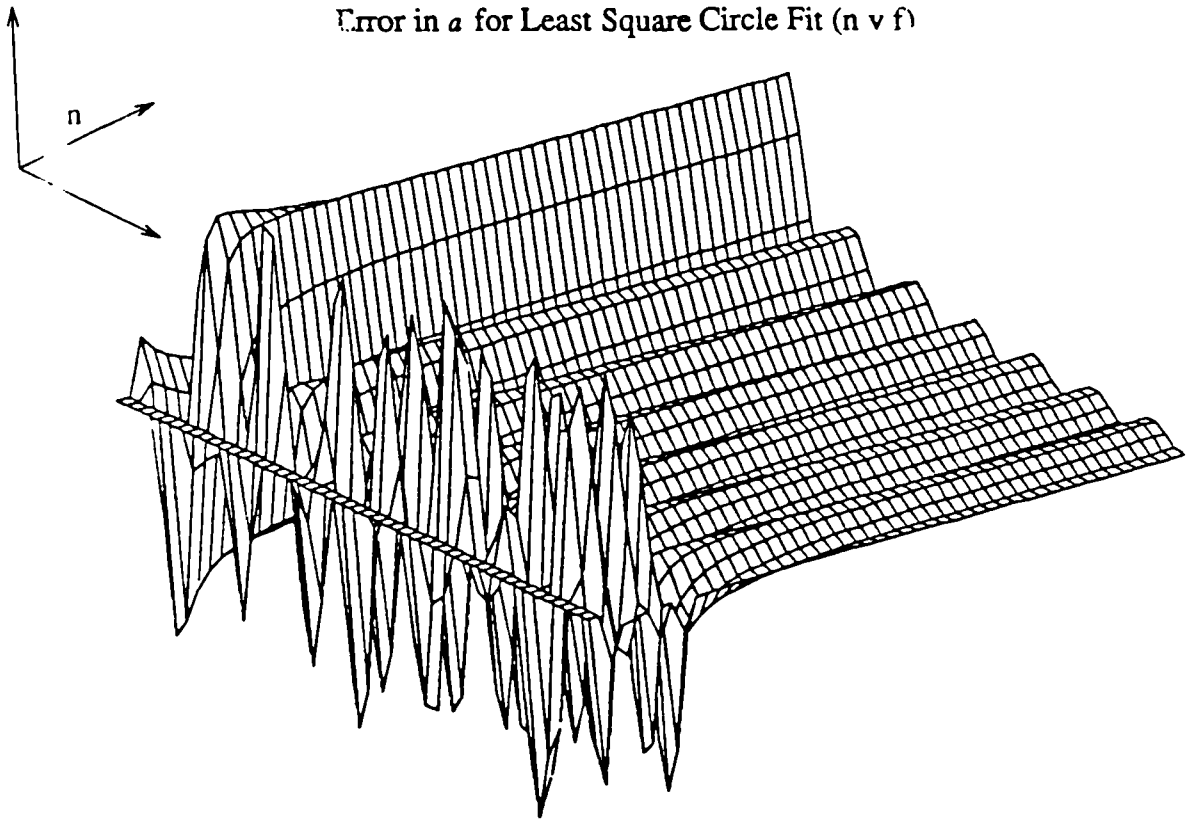
Error in *r* for Least Square Circle Fit (n v phi)



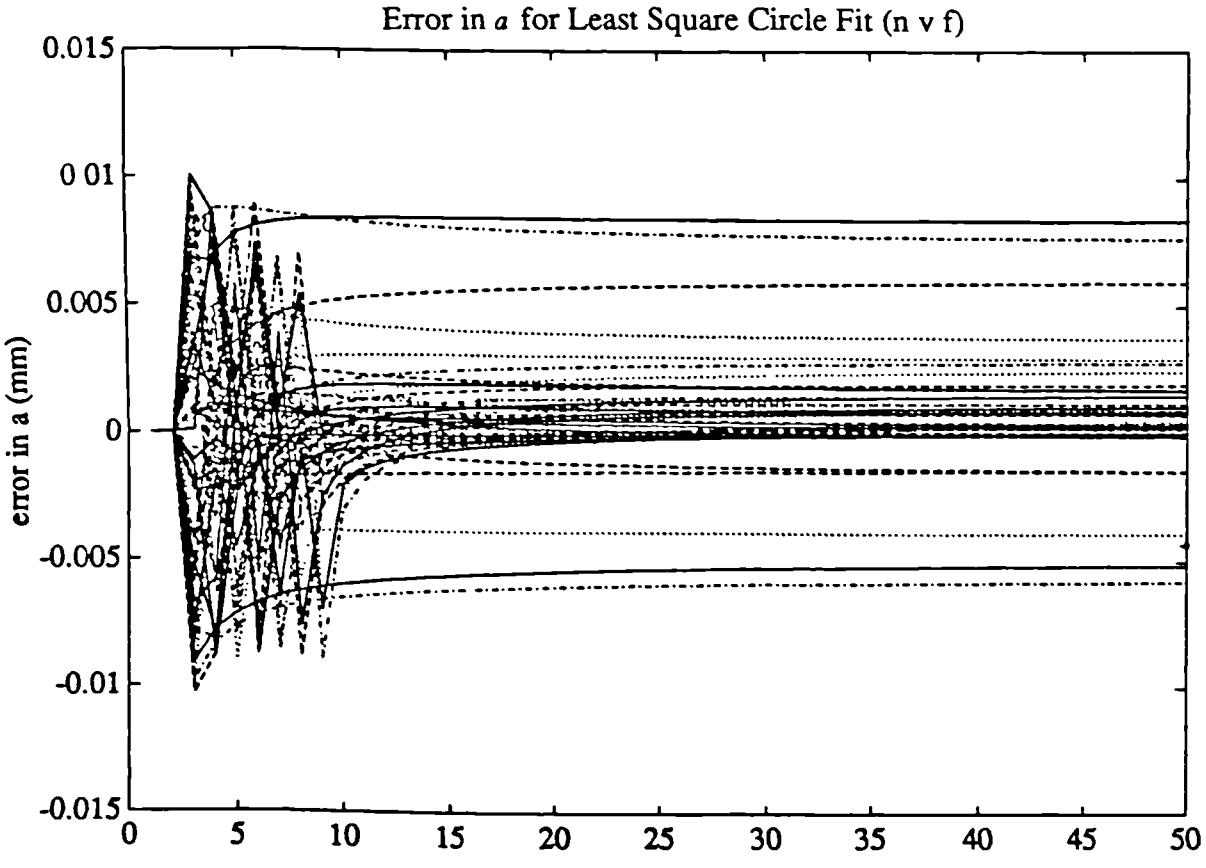
Graph 3.42 (a)



Graph 3.42 (b)

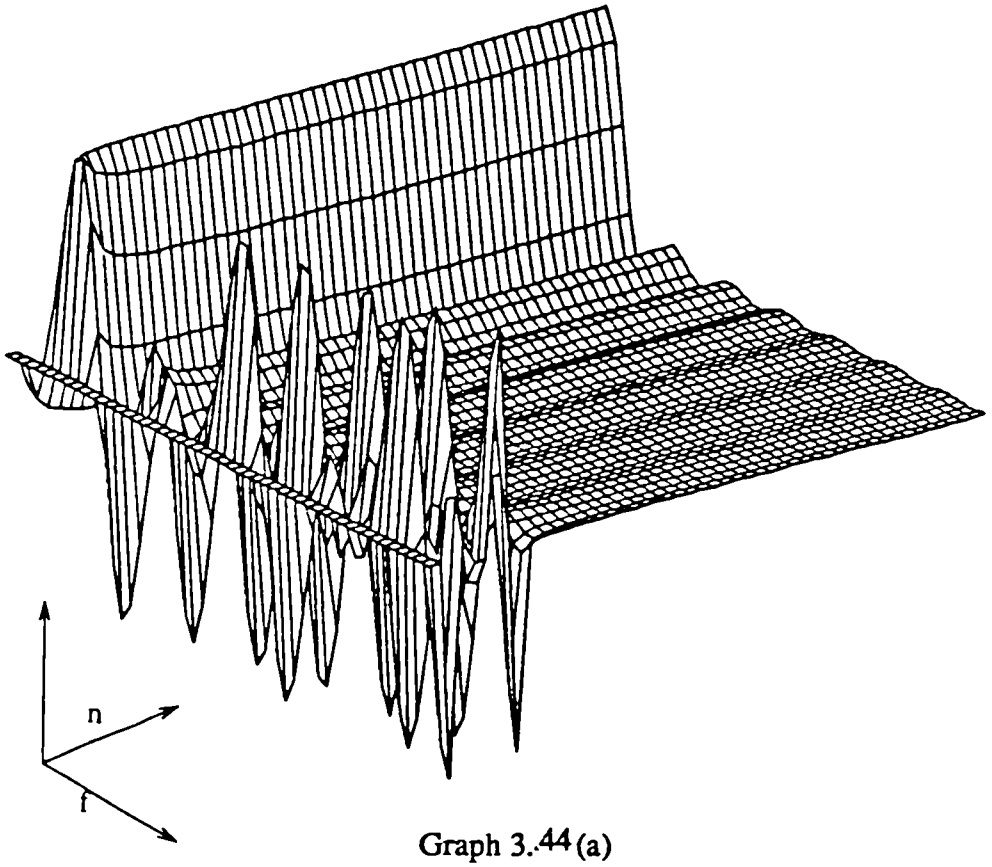


Graph 3- 43(a)



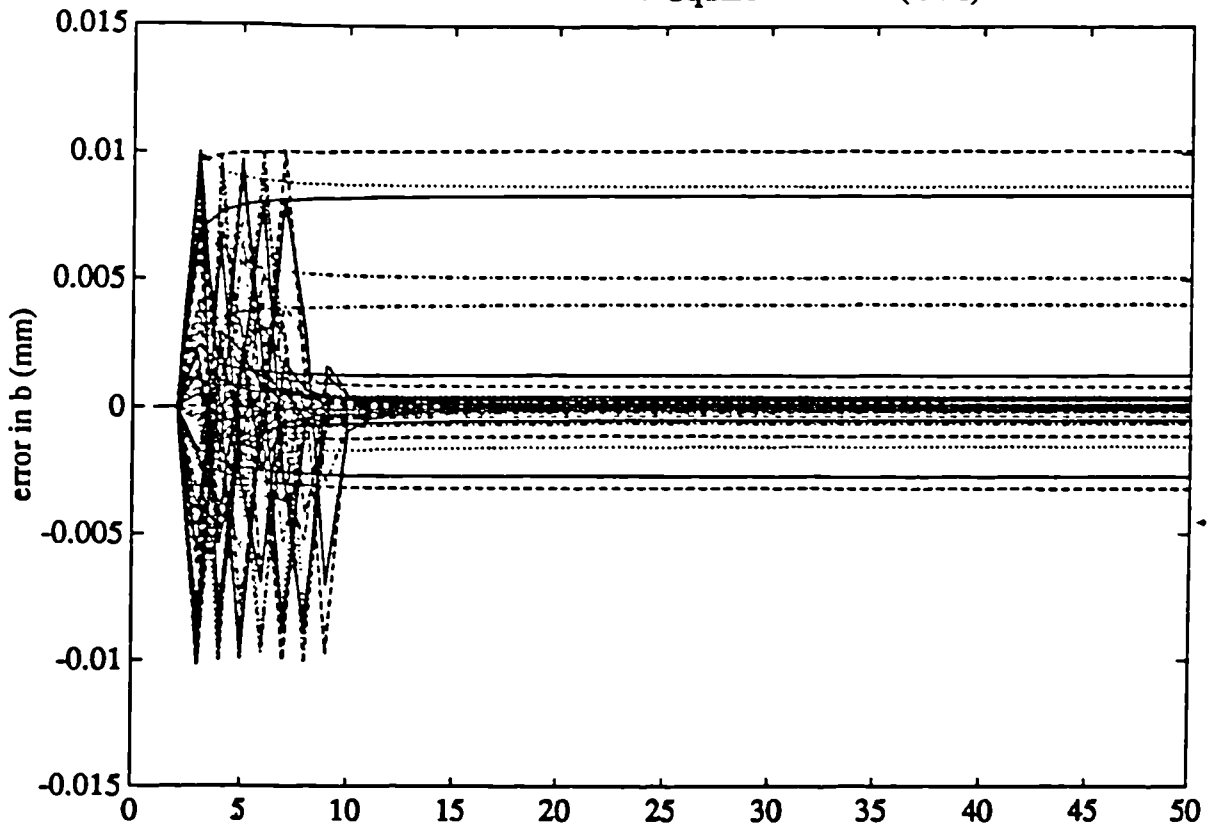
Graph 3. 4: (b)

Error in *b* for Least Square Circle Fit (n v f)

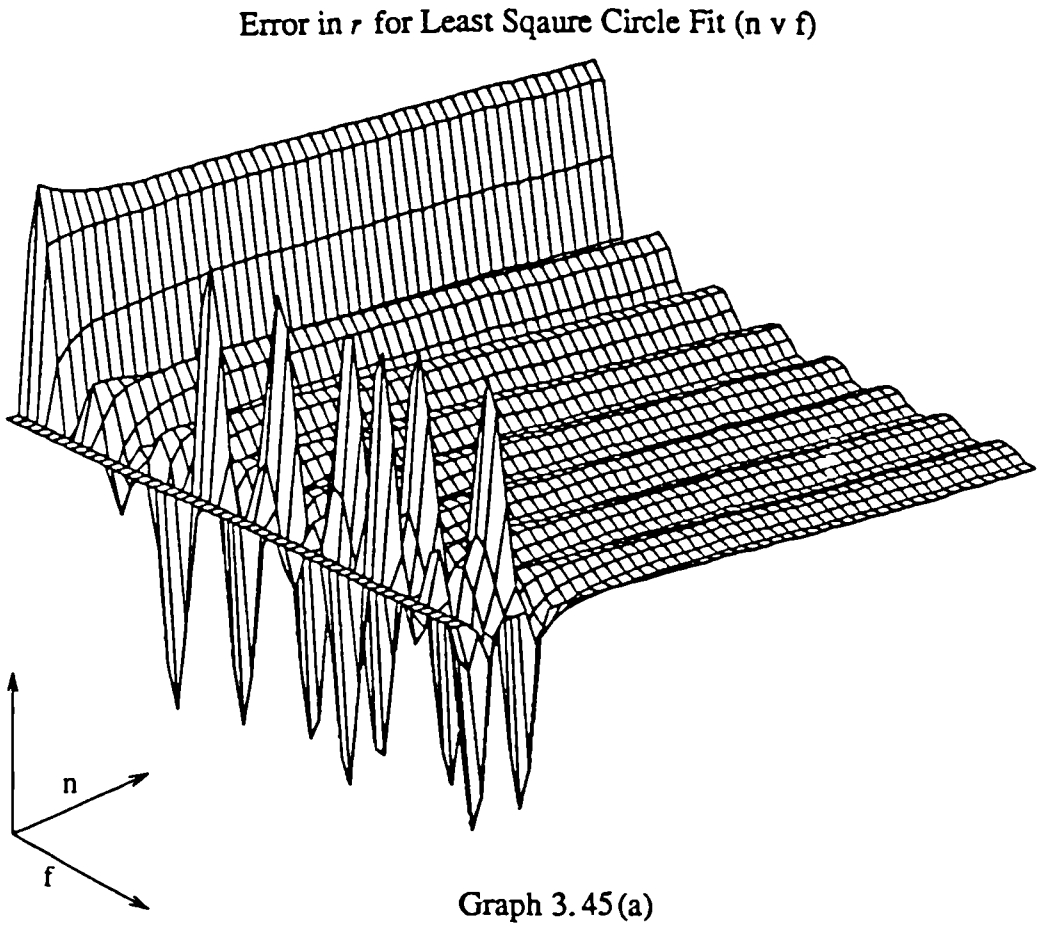


Graph 3.44(a)

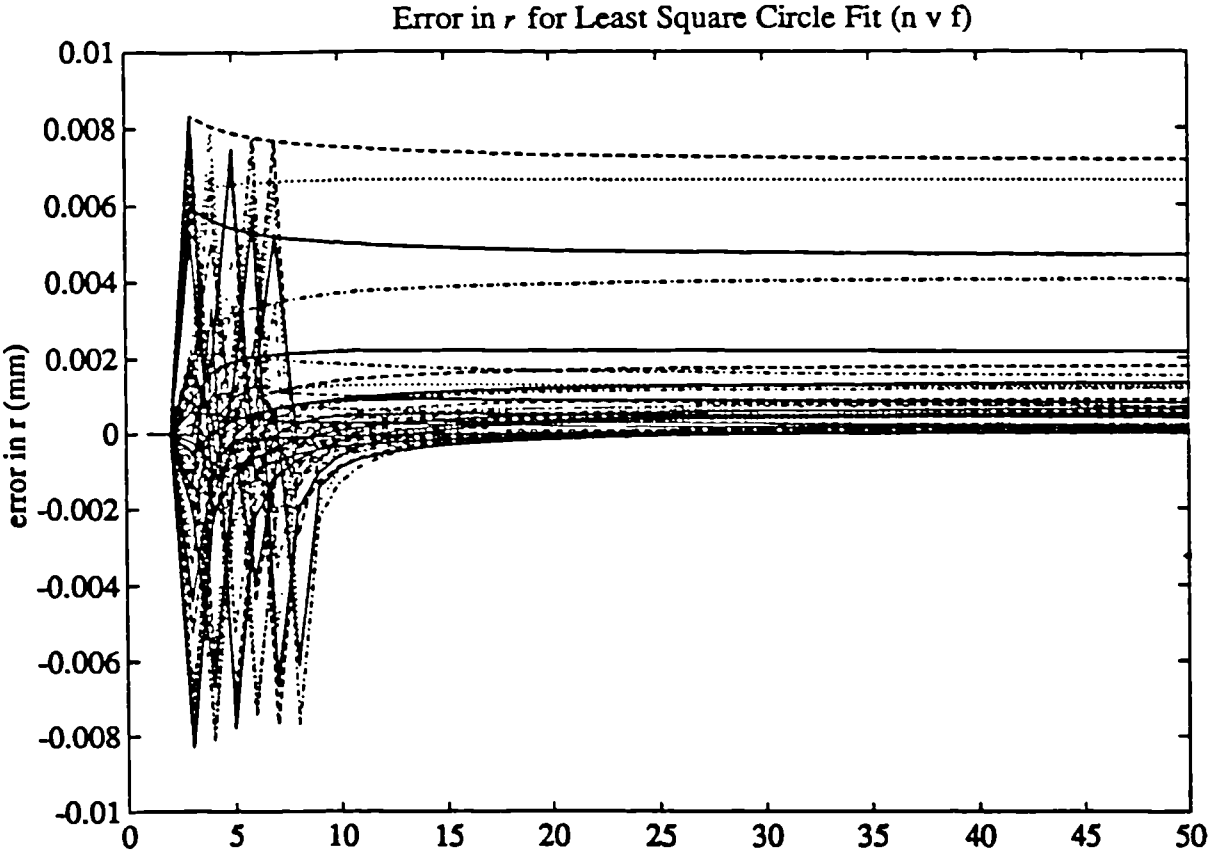
Error in *b* for Least Square Circle Fit (n v f)



Graph 3.44(b)



Graph 3.45(a)



Graph 3_45(b)

3.3.3.2. Results

Graphs 3.10 to 3.12 show that the sampling errors again diminish exponentially with respect to the quantity of data used, but to a lesser extent than that observed in line and plane fits. The diminished effect is due to the nature of the placement of the data, i.e. the data are equispaced around the full circumference of the circular profile, so that the less data that is used, the greater the sampling interval. Previous results indicate that these effects would in this situation tend to cancel each other. The strategy of using data of varying quantities at a fixed interval would not be valid as ill-conditioning would colour the results.

Changes in amplitude and phase of surface features produce similar effects upon the parameter estimates to those described in previous sections. Frequency changes appear to produce a different effect, however, this is merely due to the fact that aliasing has not occurred within the frequncy range used.

From these results it is possible to make the following conclusions and recommendations.

(i) Provided that the data are equispaced around the full circumference of the circular profile, there is no advantage to be gained, with regard to accuracy of the estimates, by using more than 12 data points.

(ii) Within the range of frequency specified in table 6.3, it can be said that the greater the frequency, the less the sampling error accrued, i.e. circular profiles having errors of triangularity will produce less sampling error than those having errors of ovality (fig 3.3).

4. Conclusion

The investigation into sampling errors in the estimation of geometric parameters of lines planes and circles clearly illustrate that samping error behaves very similarly with respect to changes in all of the possible variables. It is therefore possible to conclude that

the formulation of a formal solution for the quantification of the sampling error for all least-squares element fitting routines is viable, and that such a solution for one element would indicate a way in which similar solutions for the remaining elements could be found.

References

[1]:P. FORBES: NPL report unpublished.

[2] D.G. CHETWYND: Ph.D. Thesis, University of Leicester 1980.

[3] .T.S.R. MURTHY (1983): " A Method For the Evaluation of Elliptical Profiles"; Precision Engineering, Vol 5, Pt. 2, pp77-81.

[4] D.G. CHETWYND (1985): "On the Definition of Geometric Residuals"; NPL. Proc. Conf. Software for Coordinate Measuring Machines, 25/9/1985.

[5] N.R. DRAPER & H. SMITH (1966): "Applied Regression Analysis"; John Wiley & Sons, Inc., New York.

[6] D.E. NEWLAND (1975): "An Introduction to Random Vibration and Spectral Analysis"; Longman .

[7] J.J. HURT (1980): "A Comparison of Several Plane Fit Algorithms"; Annals of the CIRP, Vol 29.

[8] E.J. MURDEN (1989): " Algorithms for the Assessment of Geometric Form"; M. Phil. Thesis, University of Leicester.

[9] T.S.R. MURTHY (1986): "A Comparison of Different Algorithms for Circularity Evaluation"; Precision Engineering, Vol 8, No. 1.

[10] D.G. CHETWYND (1985): "Linearized Exchange Algorithms in Metrology" ; Proc. NPL Conf. on Software for CMMs.

[11] J.C. MC. COOL (1979): "Systematic and Random Errors in Least Squares Estimation of Circular Contours"; Precision Engineering, Vol 1, Pt. 4, pp 215-20.

[12] I. KASA (1976): "A Circle Fitting Procedure and its Error Analysis"; IEEE Trans. Inst. & Meas.

[13] L. NAWARA & M. KOWALSKI (1981): "Investigations on the Variability on the Results of Measurements Realised by Multicoordinate Measuring Machines on Circular Sections with Determined Types of Roundness Errors"; Annals of the CIRP, Vol 30.

The Development of an Algorithm for the Quantification of Sampling Error

1. Introduction

At present, an unacceptably high proportion of engineering components inspected by coordinate measurement techniques are not assembling with adequate precision. A significant contribution to these errors can be attributed to inadequacies in the geometric evaluation software algorithms. The use of least squares analysis, which is common in coordinate measurement practice, can produce estimates of geometric parameters which are in excess of five times the true value (1). These discrepancies arise from several possible sources of numerical instability, most notably ill-conditioning and sampling error, i.e. the error incurred due to not considering all possible data. The latter problem has hitherto remained uninvestigated by the CMM community despite the widely recognised need for the quantification of these errors to enable CMM techniques to compete with more traditional inspection methods. Practitioners have instead relied upon heuristic considerations, which practice has shown to be inadequate.

The advantages offered by a quantitative sampling error evaluation are considerably enhanced by the prospect of also solving the inverse problem; being able to specify the optimum number and distribution of coordinate points required in order to satisfy predetermined tolerances and accuracy specifications. In addition to the obvious advantages accrued by such an approach, inspection accuracy is further improved by the subsequent minimisation of the effects of machine and temperature drift, as the time taken to gather the coordinate data will be minimised.

It is the intention of this chapter to consider the nature of the phenomenon of sampling error, to formulate an original expression for the variability of sampling error and to perform a series of experiments to systematically quantify the behaviour of this expression, developing an algorithm that has the capability of quantifying sampling error.

2. Analysis of the problem

2.1. Assessment of Sampling Error

In order to formulate an effective algorithm for the quantification of sampling error in the estimation of geomteric parameters, the nature of the problem must be carefully considered. If a parameter is calculated first from a set of data, and then from a relatively small subset of the original data, the difference between the two parameter values would depend upon the characteristics of the data, or more specifically in the case of coordinate measurement, its variation with respect to horizontal displacement of the data used. Obviously if all of the data lies upon the reference figure, the difference between the two values of the parameter will be zero. If the data is distributed about the reference figure, there will be a difference in the two parameter values that is dependent upon the nature of the data, i.e. the surface profile characteristics, the horizontal placing of the data upon this profile, and the quantity of data points used in the calculations.

In the light of the above discussion, it is logical to base the formulation of an algorithm for the quantification of sampling error upon an expression whose terms are a descriptor of surface variation, and a function of both the placing of this data upon this variation and the number of data points used.

2.2. Representation of Spatial Variation

In order to select the most appropriate measure of spatial variation it is appropriate to consider the characteristics of the surfaces encountered in coordinate measurement. It must also be realised that coordinate measurement techniques involve the inspection of relatively large surfaces (≈ 5mm-5m), and consequently the effects of deterministic or

periodic surface characteristics resulting from machining or slight errors of form cannot be disregarded as they often are in the measurement of surface texture parameters. The problem cannot therfore be considered in terms of simple Gaussian statisitcs.

Manufactured surfaces can be broadly catagorised into three main types:

i) Surfaces in which there is an almost entirely random waveform which is commonly produced by lapping, grinding or polishing etc.

ii) Surfaces in which there is a random term that is modulated by a periodic waveform. The short wavelength deterministic characteristics can be produced by the cutting action of a machine tool, the frequency and shape of the waveform are dictated by feed rates, cutting forces, cutting speeds and the shape of the machine tool (2). Higher order harmonics are usually present and are often due to machine chatter vibration and tool damage. Low order deterministic elements can also occur as a result of machine tool wear (particularly in main bearings) or misalignment. The random characteristics can be attributed to either vibrational sources within or external to the machine tool, a grit based abrasive material removal process or the nature of the material.

iii) Surfaces containing both random and deterministic elements, but that are added together in contrast to being modulated, this can happen in type i) where the machine may be subject to a degree of chatter, or in type ii), where the slideway may have a systematic error of movement (3).

This considerable variation in surface character presents a problem of how to describe the spatial variation so that all types of surface are adequately represented.

The most common way of representing spatial variation is by using the variance of an amplitude probability density function, $p(x)$. Surfaces of type 1 can be adequately described in terms of the classical Gaussian distribution (4) as follows:

$$p(z) = (\sigma_z \sqrt{2\pi})^{-1} \exp^{\frac{-z^2}{2\sigma_z^2}} \tag{4.1}$$

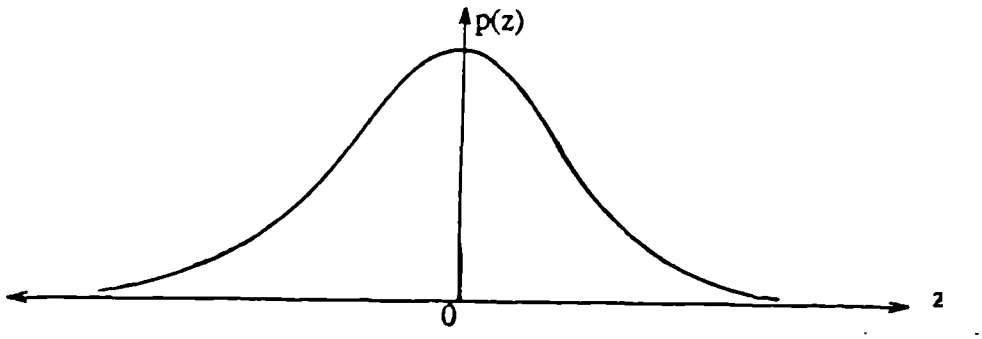


Figure 4.1

However, this distribution will not represent the deterministic elements of a profile. Periodic elements can be described by alternative probability distributions, for example, the most basic periodic element, the sine wave, can be represented as follows:

$$p(z) = \pi\sqrt{(X^2 - z^2)^{-1}} \qquad |z| < X \tag{4.2}$$

$$p(z) = 0 \qquad |z| \geq X$$

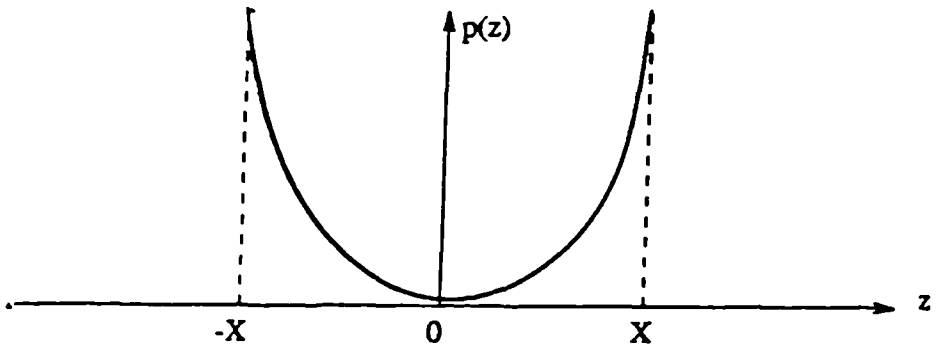


Figure 4.2

Therefore surface types ii) and iii) may be reperesented by a probability distribution having the predomoinant characteristics of both the above (4) i.e.:

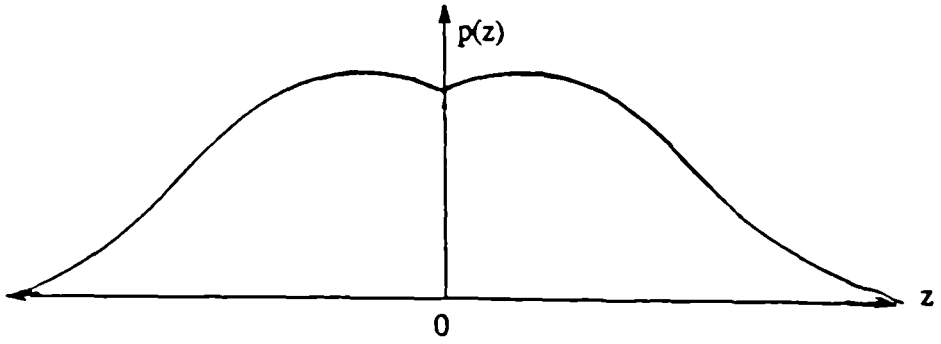


Figure 4.3

However, surfaces of types ii) and iii) which may be described by this general form would still vary considerably in either their degrees of randomness and periodicity, the probability distribution would alter accordingly, so that the use of a fixed model would be of somewhat limited value.

Probability distributions could be calculated for each set of data, thus:

$p(z)dz$ = Fraction of total length of profile for which $z(x)$ lies in the z to $z+dz$ band

$$p(z)dz = \frac{dz_1 + dz_2 + dz_3 + \ldots.}{L} = \sum \frac{dz}{L} \tag{4.3}$$

For the above equation to be mathematically correct, L must be infinite and the profile must not change its characteristics with respect to horizontal displacement (5). Surface types ii) and iii) have deterministic characteristics and therfore do not satisfy this criterion, i.e they are non-stationary. Obtaining a value for the variance of the surface data requires further computation (6):

$$\sigma^2 = \sum_{all\,x} (z - \mu)^2 z(x) \tag{4.4}$$

Where μ is the mean of the distribution.

The Gram-Charlier series has also been used as a means to express surface amplitude distribution (7). The method involves comparing the experimental distribution with a normal or Gaussian distribution by a numerical method. In its normalised form the Gram-Charlier series is

$$f(x) = \frac{1}{\sqrt{(2\pi)}} e^{\frac{-t^2}{2}} \left[1 + \frac{\mu_1 - 3}{8} - \frac{\mu_3 t^2}{2} - \frac{\mu_4 - 3}{4} t^2 + \frac{\mu_3 t^3}{6} + \frac{\mu_4 - 3}{2} t^4 \right] \tag{4.5}$$

where

$$t = \frac{x - m_x}{\sigma_x^2}$$

and

$$\mu_n = \int_{-\infty}^{\infty} (x - m_x)^n f(x) dx$$

and

$$m_x = \int_{-\infty}^{\infty} x f(x) dx$$

Although this method goes some way to achieving a comprehensive description of the surface texture, it relies upon the assumption of predominantly Gaussian behaviour. Also its relative complexity may result in an avalanche of numerical problems in the formation of the algorithm

In the light of the difficulties and potential mathematical inaccuracies in the computation of variance via a probability distribution for each surface, an alternative and more computationally efficient function to represent spatial variation and therefore to calculate data variance is proposed, namely the autocorrelation function.

The autocorrelation function has excited more interest amongst metrologists in recent years than perhaps any other surface descriptor. It is a function that is recognised as having great potential use in industry (3). Using this function and its close relatives such as the structure function (8), it is possible to obtain information about signals and surfaces having both random and periodic structure.

Much work has already been done in using the autocorrelation function to express surface parameters. Whitehouse (9) describes a situation concerning the inspection of miniture of circular components in which it was realised that the reduced size of the workpiece leads to considerable difficulty in distinguishing between errors of form and deviations from the nominal figure due to surface irregularities; the result being that the data is no longer independent i.e. it cannot be considered as uncorrelated and Gaussian statistics are no longer applicable. Instead the autocorrelation function of has been used as a measure of spatial variation for the radius of the component. Indeed, in any situation involving surface data that cannot truly be regarded as Gaussian, or as having any other definable amplitude distribution, the autocorrelation function is an appropritate way to represent the spatial variation. Other work has been based on the pioneering two dimensional research of Longuet-Higgins (9). However, although the function is theoretically useful, it is not completely understood (10), and the use of the above techniques only

gives part of the story, in that only the theoretical relationships between the surface parameters and the autocorrelation function are given, it does not show the researcher how the measured values relate to the measured autocorrelation function (3). It may therefore by necessary to perform a series of experiments in the analysis of sampling error in order to provide an empirical relationship between the derived and actual sampling error.

3. Calculation of the Autocorrelation Function

Various methods already exist for the calculation of the autocorrelation function, most relying upon approximations to simplify and speed up the processing.

There are four basic methods of calculating the autocorrelation function: step by step, parallel, by the use of the Discrete Fourier Transform and by the Fast Fourier Transform.

3.1. Step by Step Method

Until recently the step by step method was the most commonly used, it involves the direct use of the formal definition

$$R(\tau) = \frac{1}{L-\tau} \int_0^{L-\tau} z(x)\, z(x+\tau) dx \tag{4.6}$$

where $z(x)$ and $z(x+\tau)$ are the height of the profile at the horizontal locations x and $(x+\tau)$ respectively and L is the length of the profile. This equation can be expressed in discrete form:

$$R(\tau) = \frac{1}{N-\tau} \sum_{i=0}^{N-1} z(i)\, z(i+\tau) \tag{4.7}$$

Where N is the total number of ordinates. Although the calculations involved are simple, they require a relatively large amount of computation time and cannot be performed until all the data are stored (8). However, this restriction is of decreasing importance with rapid and large data processing capabilities becoming increasingly available on modern computers.

3.2. The Parallel Method

In using the parallel method, m+1 ordinates have to be stored where m is the maximum lag desired. Subsequently, the contribution of any new ordinate Z_j to all m+1 lag positions is worked out as it is read; the z_{j-m+1} ordinate being dispensed with at the same time. Thus the contributions worked out at the jth ordinate will comprise Z_j^2 to go in array position 1, Z_j , Z_{j+1} in position 2,..., $Z_j Z_{j+m}$ in position m. The operation is carried out for j=1 to N-m. Each array slot is then normalised to the first. If the operation is carried out from j=1 to N, then array positions have to be scaled by N for array position 1, N-1 for position 2 to N-m for position m+1. The advantage of this method lies in the fact that only m+1 ordinates are stored at any one time rather than N. (8)

3.3. Fourier Methods

The procedure for calculating the autocorrelation function by the Discrete Fourier Transform method is as follows: a sample function z(x) is taken in the form of a continuous record of length x=0 to x=L. In the following analysis it is assumed that the sample function z(x) constitutes a whole cycle of the entire function. (This assumption is not necessarily true, the consequences of such a premise are discussed later in the section.) These records are then sampled at an interval of $\delta = \frac{L}{N}$; the DFTs of this series are then calculated, using

$$z_k = \frac{1}{N}\sum_{r=0}^{N-1} z_r e^{-i\left[\frac{2\pi kr}{N}\right]} \tag{4.8}$$

To calculate the autocorrelation function between records on the surface $z(x)$ and $z(x+\tau)$, remembering that it can only be estimated at 'lags' that are multiples of the sampling interval δ, an estimate of the autocorrelation function can now be written :

$$R(\tau) = \frac{1}{N}\sum_{s=0}^{N-1} z_s z_{s+r} \quad \ldots\ldots \quad r = 0,1,2,\ldots,(N-1) \tag{4.9}$$

The Fast Fourier Transform method is one which takes advantage of the Wiener-Kinchne relationship (8). In this procedure, the autocorrelation function is

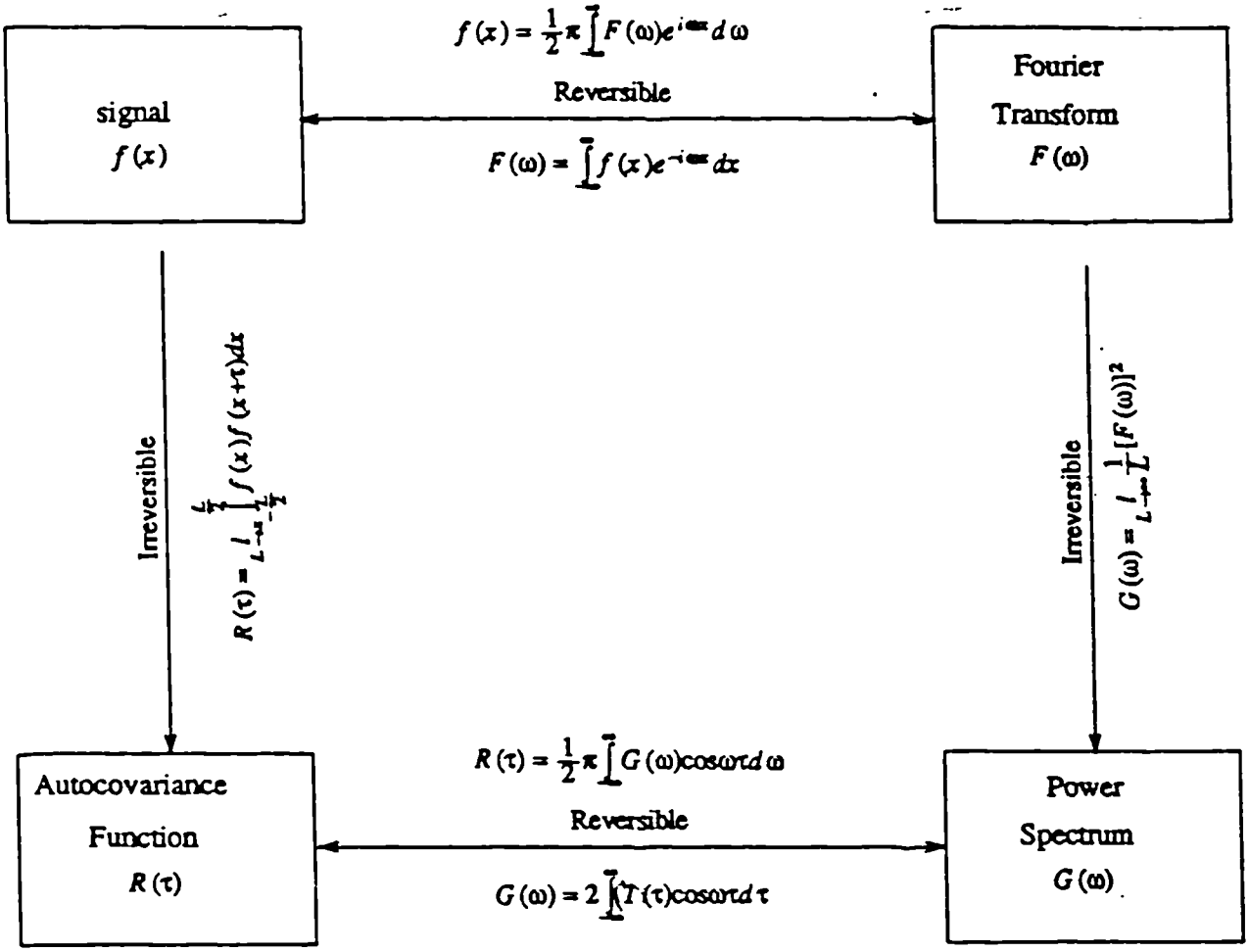

signal
$f(x)$
$f(x) = \frac{1}{2\pi}\int F(\omega)e^{i\omega x}d\omega$
Reversible
$F(\omega) = \int f(x)e^{-i\omega x}dx$
Fourier
Transform
$F(\omega)$
Irreversible
$R(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x)f(x+\tau)dx$
Irreversible
$G(\omega) = \lim_{T\to\infty}\frac{1}{T}[F(\omega)]^2$
Autocovariance
Function
$R(\tau)$
$R(\tau) = \frac{1}{2\pi}\int G(\omega)\cos\omega\tau d\omega$
Reversible
$G(\omega) = 2\int R(\tau)\cos\omega\tau d\tau$
Power
Spectrum
$G(\omega)$

Figure 4.4

derived form the power spectral density function by the use of the Fast Fourier Transformation, thus:

$$R(\tau) = \frac{1}{\pi}\int_0^{\infty} G(\omega)\cos\omega\tau d\,\omega \tag{4.10}$$

or alternatively, in digital form, where k is the wave number:

$$R(m) = \frac{1}{N}\sum_{i=0}^{N-1} G(k)\cos\left[\frac{-2\pi imk}{N}\right] \tag{4.11}$$

The FFT is essentially an algorithm which utilises possible symmetries in the exponential function and the properties of sparse matrices. The FFT can reduce the computation involved in a Fourier transformation by at least $\frac{2\log_{2N}}{N}$ relative to the direct method. Rounding-off errors are minimised due to the reduced number of calculations.

3.4. The 'Wrap Around Effect'

The approximations used in the above methods can result in a phenomenon know as the 'wrap around effect ' when the autocorrelation functions of heavily periodic signals are calculated (5). The 'wrap around effect ' may be illustrated by considering the definition of the autocorrelation function used in equation (4.9) periodicity implies:

$$z_s = z_{s+N} \tag{4.12}$$

This periodicity may result in R_r differing from the true correlation function. The reason for this possible discrepancy may be illustrated as follows:

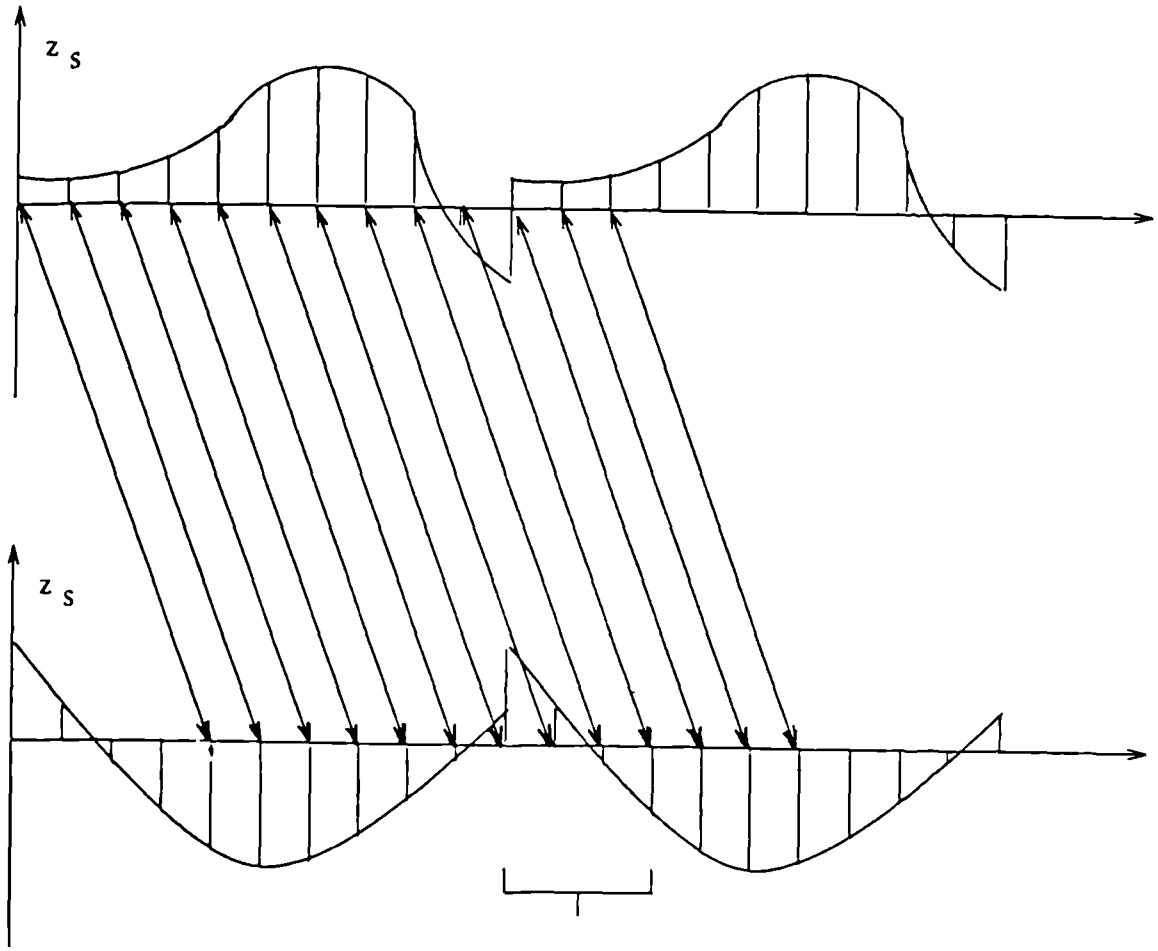


Figure 4.5

With reference to fig 4.5; for the case N=10 , r=4, in calculating the summation in equation (4.9, the first before the y_s process.

$$R_4 = \frac{1}{10}[z_0z_4 + z_1z_5 + z_2z_6 + z_3z_7 + z_4z_8 + z_5z_9] + \frac{1}{10}[z_6z_0 + z_7z_1 + z_8z_2 + z_9z_3] \qquad (4.13)$$

The second term has nothing to do with R($\tau = 4\delta$) and is instead an estimate of R($\tau = -6\delta$). Additionally, both terms are biased estimates for R(τ) since in the first case the denominator should be 6 and in the second case 4 (5). It may be concluded that the approximations in deriving the autocorrelation functions are corrupting the calculations, it is therefore advisable to process the calculated autocorrelation through an algorithm to rectify the effects, as described in appendix A4 prior to performing further analysis. The autocorrelation functions used in the following experiments were calculated using MATLAB software package on a SUN 3/50 computer.

4. Formulation of an Initial Expression for Sampling Error

It is proposed to use the autocorrelation function as means of expressing the variability characteristics of the surface and to consequently evaluate the sampling error incurred in the estimation of geometric parameters. Ideally, the solution should be formulated in such a way as to be expressible as a straightforward combination of readily calculable terms and a representation of the surface spatial variation.

It is intended to develop initial expressions for the variability of the sampling error with respect to the number of data points used in the estimation of the geometric parameters, and then to use experimental methods to further develop these expressions so that it gives an acceptable quantitaive analysis of sampling error.

In order to test the validity and potential implementability of such a proposal, its application to a couple of specific but commonly used coordinate measurement evaluation routines must first be investigated.

The means by which it is intended to develop expressions for sampling errors incurred in the estimation of geometric parameters is as follows. It is logical to begin

with the expressions for the calculation of the relevant parameters, and then to derive expressions for their variance and consequently standard error (standard error being the square root of variance). Using these equations as a starting point, it is intended to introduce the autocorrelation function in order to predict the nature of the independent variable and reflect the nature of the problem of sampling error. At this point it is not strictly permissable to call the resulting expressions 'variance' and 'standard error', therefore they will thenceforth be termed 'variability ' (V) and 'variability index' (V.I.) respectively. It is the intended combine these expressions with the arguments put forward in the qualitative discussion in section 2.1 in order to identify and further develop those expressions so that sampling error will be more accurately reflected.

The first part of this task is to develop an expression which varies at an approximately similar rate (with respect to the number of data points used) as the actual sampling error. This is a logical starting point as, if the sampling error for one subset of data used on a particular surface is known, then the sampling error for a subset of a different size may be roughly estimated. Then if no more sucess is forthcoming, at least some progress has been made. This expression will subsequently be used in experiments to establish a final descriptive expression for sampling error.

As this analysis is being developed from fundemental ideas, it is logical to begin by assessing the simplest possible case; i.e. with fitting parameters to data that is:

(a) Two dimensional

(b) Nominally straight

(c) The definition of this data is aligned with perfect precision to the coordinate frame of the instrument that is gathering and analysing it.

The parameters that can be fitted to data of this type (expressed in an (X,Y) coordinate frame) are those that describe its slope (b), and its intercept with the Y axis (b_o). Using least squares analysis, the familiar expression for slope b is:

$$b = \frac{\sum((X_i - \bar{X})(Y_i - \bar{Y}))}{(\sum(X_i - \bar{X})^2)} \tag{4.14}$$

since:

$$\sum(X_i - \bar{X})\bar{Y} = \bar{Y}\sum(X_i - \bar{X}) = 0$$

$$b = \frac{\sum(X_i - \bar{X})Y_i}{\sum(X_i - \bar{X})^2} \tag{4.15}$$

To develop an expression for the variance of b:

$$var(b) = E(b^2) \tag{4.16}$$

Now, b can be written

$$b = \sum(a_{b\,i}\ Y_i)$$

where

$$a_{b\,i} = \frac{(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}$$

As Y_i are correlated, it is not possible to say that the sum of the variances is equal to the variance of the sum, therefore,

$$var(b) = E(\underline{e}^T \underline{a}\underline{Y}.\underline{Y}^T \underline{a}^T \underline{e}) \tag{4.17}$$

Where

$$\underline{e}^T = (1\ 1\ 1\ 1\ .....)$$

and

$$\underline{a}^T = (a_1\ a_2\ a_3.....)$$

Therefore

$$var(b) = E[(a_1\ a_2\ a_3.....)\ var(Y) \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \end{bmatrix} \tag{4.18}$$

$$var(b) = \sum a_i^2(\underline{e}^T\ var(Y)\underline{e}) \tag{4.19}$$

$$var(b) = \frac{\sum(X_i - \bar{X})^2}{(\sum(X_i - \bar{X})^2)^2}(\underline{e}^T\ var(Y)\underline{e}) \tag{4.20}$$

Therefore

$$var(b) = \frac{1}{\sum(X_i - \bar{X})^2}(\underline{e}^T\ var(Y)\underline{e}) \tag{4.21}$$

Here, the autocorrelation function is introduced in order to express ($Var(Y)$). Henceforth, $R(\tau)$ represents the autocorrelation function.

At this point, it is neccessary to depart from standard statistics based upon variance. It is proposed to estimate the variability of the surface data by using the autocorrelation function. Due to the nature of sampling error and the type of data being considered, it is suggested that standard variances used in random data analyses would model sampling error and large scale surface profiles badly. This is because sampling error is not only a function of spatial variation but also of the manner in which the data points are placed upon this spatial variation, and the data in large scale surfaces is such that the height of the profile need not be independent of its horizontal location. It was therefore decided to estimate the likely variability of the data by using points on the autocorrelation function corresponding to particular data points and some other point on the same function corresponding to a known displacement (τ). As the displacement (τ) of each successive point from the first point is always known, it was decided to use the point on the autocorrelation corresponding to the first data point from which to estimate/express the variability of each subsequent point. Thus, the initial expression for the variability of each point of the surface is expressed thus:

$$\delta Y_i = R_i(\tau) - R_1(\tau)$$

At this stage, it is necessary to regard the nature of the problem; it is desired to find a true descriptor of the surface variation. If the above expression is summed, the resulting figures may not be a true reflection of the nature of the surface as successive positive/ negative values may, overall, mask the effects of a highly variable surface. It is therefore proposed to square these values so that changes of sign may be avoided.

However, the value of this expression will obviously increse with increasing number of data points used in the calculations, but the nature of sampling error dictates that error will nominally decrease with increasing number of data. Therefore, this expression must be 'normalised' by the number of data used, and so that the weight car-

ried by the term representing the number of data used is equal, it too shall be squared, so:

$$\delta Y_i = \frac{(R_i(\tau) - R_1(\tau))^2}{n^2} \tag{4.22}$$

As this is not strictly statistical 'variance', it is decided to henceforth term the expression equivalent to variance as 'variability', (V), and that equivalent to standard error as 'variability index' (V.I.).

$R_i(\tau)$ = Point on autocorrelation function corresponding to point i by means of horizontal displacement.

Now,

$$V(b) = \frac{1}{\sum(X_i - X)^2}(\underline{e}^T \begin{bmatrix} \delta Y_0 & \delta Y_1 & \dots & \delta Y_{n-1} \\ \delta Y_1 & & & \\ \delta Y_2 & & & \\ \vdots & \delta Y_1 & & \\ \delta Y_{n-1} & & & \end{bmatrix} \underline{e}) \tag{4.23}$$

$$V(b) = \frac{1}{\sum(X_i - X)^2}.(1\ 1\ 1\ 1....) \begin{bmatrix} \delta Y_0 & \delta Y_1 & \dots & \delta Y_{n-1} \\ \delta Y_1 & & & \\ \delta Y_2 & & & \\ \vdots & & & \\ \delta Y_{n-1} & & & \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \tag{4.24}$$

$$= \frac{1}{\sum(X_i - X)^2}.n\delta Y_0 + 2(n-1)\delta Y_1 + 2(n-2)\delta Y_2 + ...2\delta Y_{n-1} \tag{4.25}$$

$$= \frac{1}{\sum(X_i - X)^2}\left[(R_1(\tau) - R_1(\tau))^2 + \frac{2\sum_{i=1}^{n-1}(R_i(\tau) - R_1(\tau))^2}{n^2}\right] \tag{4.26}$$

Therefore,

$$V(b) = \frac{1}{\sum(X_i - X)^2}\left\{2\sum_{i=1}^{n-1}\frac{(R_i(\tau) - R_1(\tau))^2}{n^2}\right. \tag{4.28}$$

and

$$V.I.(b) = \left[\frac{1}{\sum(X_i - X)^2}\left\{2\sum_{i=1}^{n-1}\frac{(R_i(\tau) - R_1(\tau))^2}{n^2}\right\}\right]^{1/2} \tag{4.29}$$

The variability and variability index of b_o are derived in a similar manner.

$$b_o = \bar{Y} - b\bar{X} \tag{4.30}$$

Therefore

$$var(b_o) = V(\bar{Y} - b\bar{X}) \tag{4.31}$$

$$b_o = \frac{1}{n}\left[\sum Y - b\sum X\right] \tag{4.32}$$

$$b = \frac{\sum(X_i - \bar{X})Y_i}{\sum(X_i - \bar{X})^2} \tag{4.33}$$

$$V(b_o) = V(\bar{Y}) - \bar{X}^2 V(b) \tag{4.34}$$

$$V(b_o) = \frac{1}{n^2}V(\sum(Y_i)) - \bar{X}^2 V(b) \tag{4.35}$$

Therefore

$$V(b_o) = \left[\frac{1}{n^2} - \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] \frac{2\sum(R_i(\tau) - R_1(\tau))^2}{n^2} \tag{4.36}$$

Consequently,

$$V\!I.(b_o) = \left[\left[\frac{1}{n^2} - \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] \frac{2\sum(R_i(\tau) - R_1(\tau))^2}{n^2}\right]^{1/2} \tag{4.37}$$

At this stage, it is proposed to remove terms that do not contribute to the description of sampling error, i.e. those terms that are not a function of the variation of the data Y_i and the placing of the data (X_i) upon the profile Y_i; i.e. the variability indeces for b and b_o now become:

$$V\!I.(b) = \left[\frac{1}{\sum(X_i - \bar{X})^2} \sum_{i=1}^{n-1} \frac{(R_i(\tau) - R_1(\tau))^2}{n^2}\right]^{1/2} \tag{4.38}$$

$$V\!I.(b_o) = \left[\left[\frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] \frac{\sum(R_i(\tau) - R_1(\tau))^2}{n^2}\right]^{1/2} \tag{4.39}$$

4.1. Experimental Procedure

The expressions for the V.I. s above are used to predict the sampling errors in the following experiments. Initial analysis was of the simplest possible case: the estimation of sampling errors incurred on a simulated surface profile represented by a single sinusoid, the intention then being to increase the complexity of the data used.

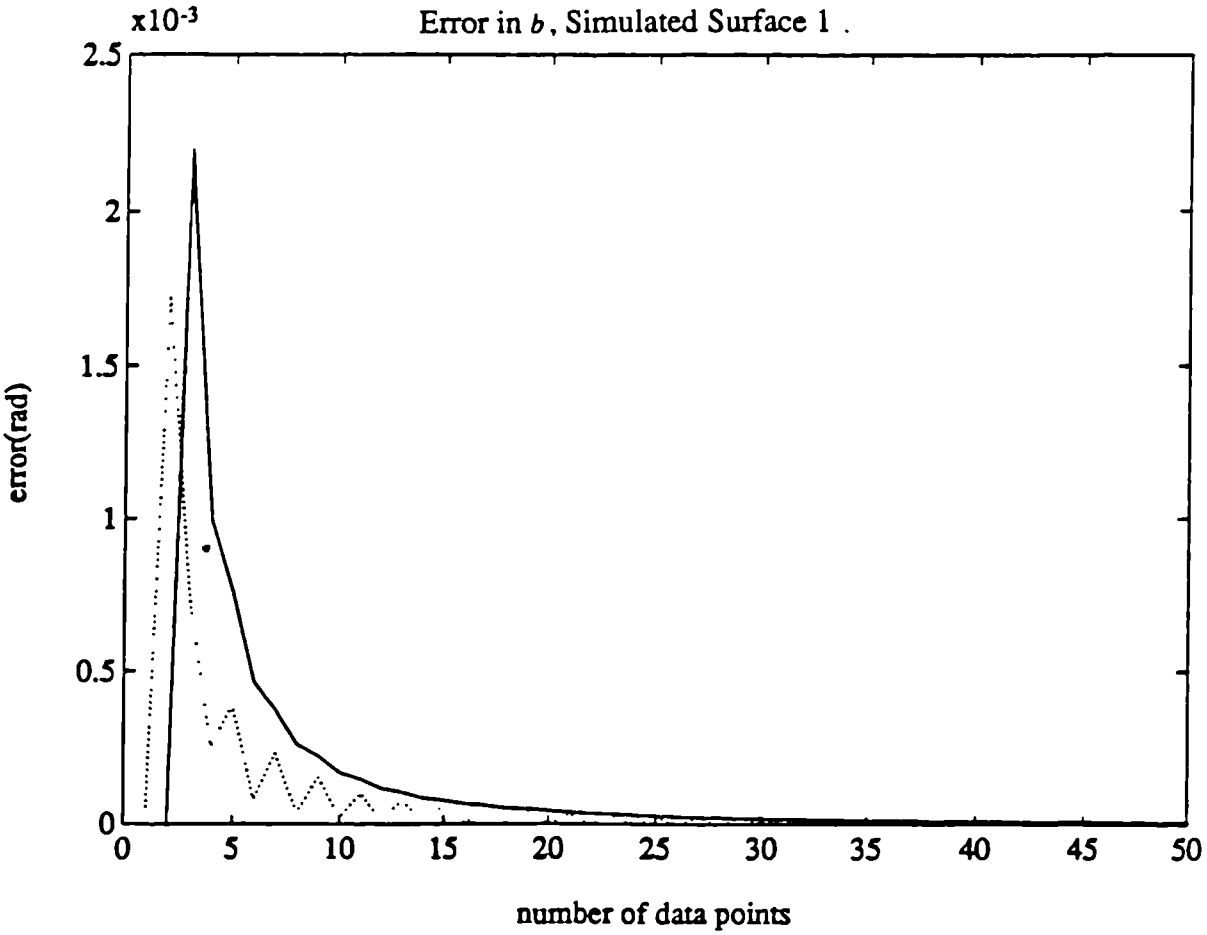
The derived expressions for the V.I.s were tested by comparing the values it produced to the actual sampling error for the same quantity and placing of data. This was achieved by calculating the theoretical values as per equations (4.38) and (4.39) using data set consisting of 2 to 50 equispaced data points, the results being plotted against the number of data points used in the calculations. The relevant parameters were then calculated using similar placing and quantity of data, and against using approximately 500 data points. The latter set of parameter estimates were taken as being the true value, and were subtracted from the former estimates in order to give the sampling error, the results of these calculations were then plotted in a similar manner to the theoretical values, and the two plots compared.

A number of experiments have been carried out upon sinusoids of varying amplitude, phase and frequency. Graphs 4.1. and 4.2 show the scaled (see later) sampling error distributions predicted using data from simulated surfaces 1 and 2 by both approaches. Throughout numerous experiments it was found that using less than six data points to predict geometric parameters of any profile leads to relatively large errors. The following analysis therefore assumes that more than five data points are to be used. It can be seen from these primary results that the autocorrelation method for estimation of slope error very closely envelopes the 'actual'sampling error. Discrepancies occurred between the derived and actual sampling error values in terms of a factor of multiplication, such that when this factor was applied to the theoretical values its plot matched or just enveloped the actual sampling error. Provided these discrepancies occur in a systematic manner in the input parameters of amplitude, phase (phi), frequency (f) and sampling interval (dx), then the sampling errors will be predictable.

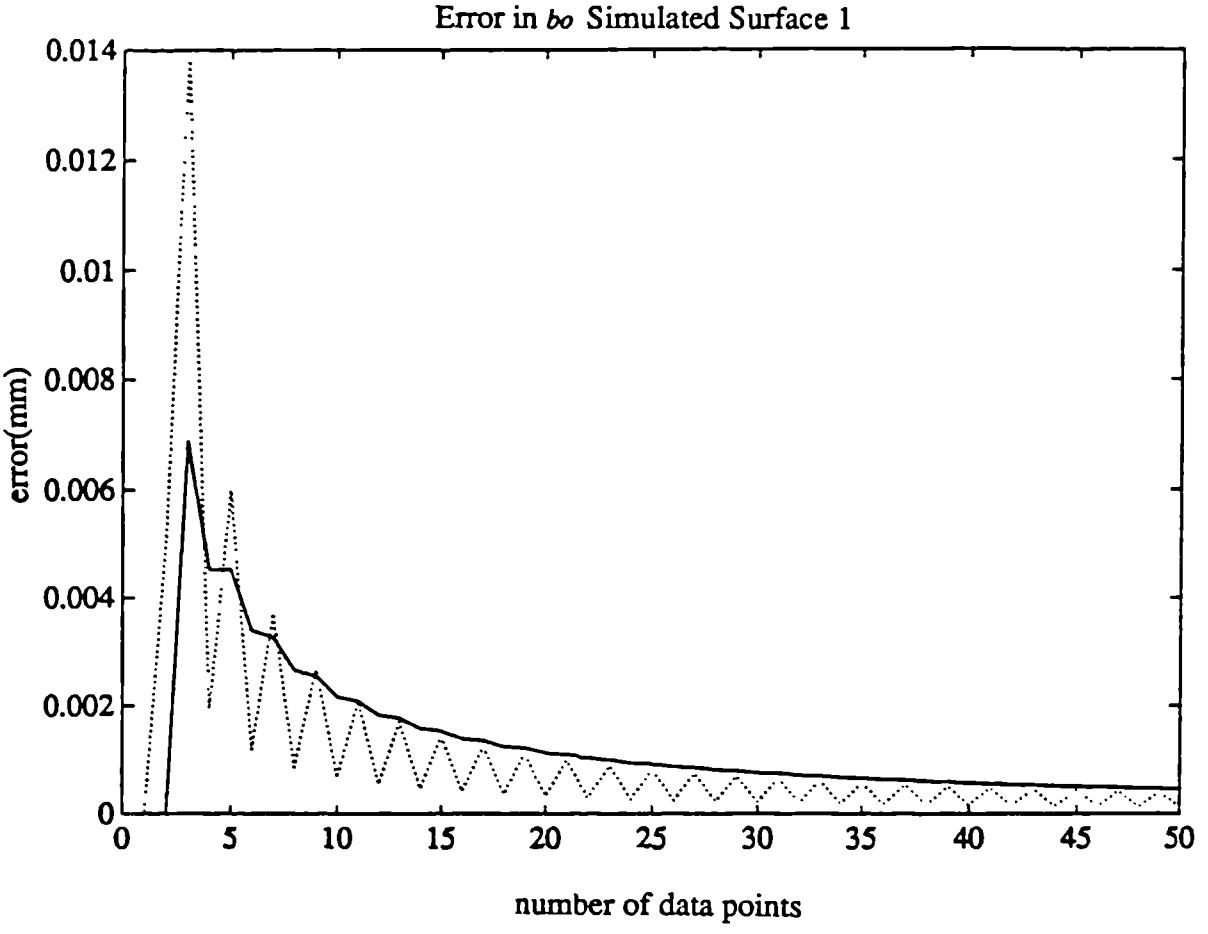
The key for all the following graphs is as follows:

_____: Theoretical Sampling Error

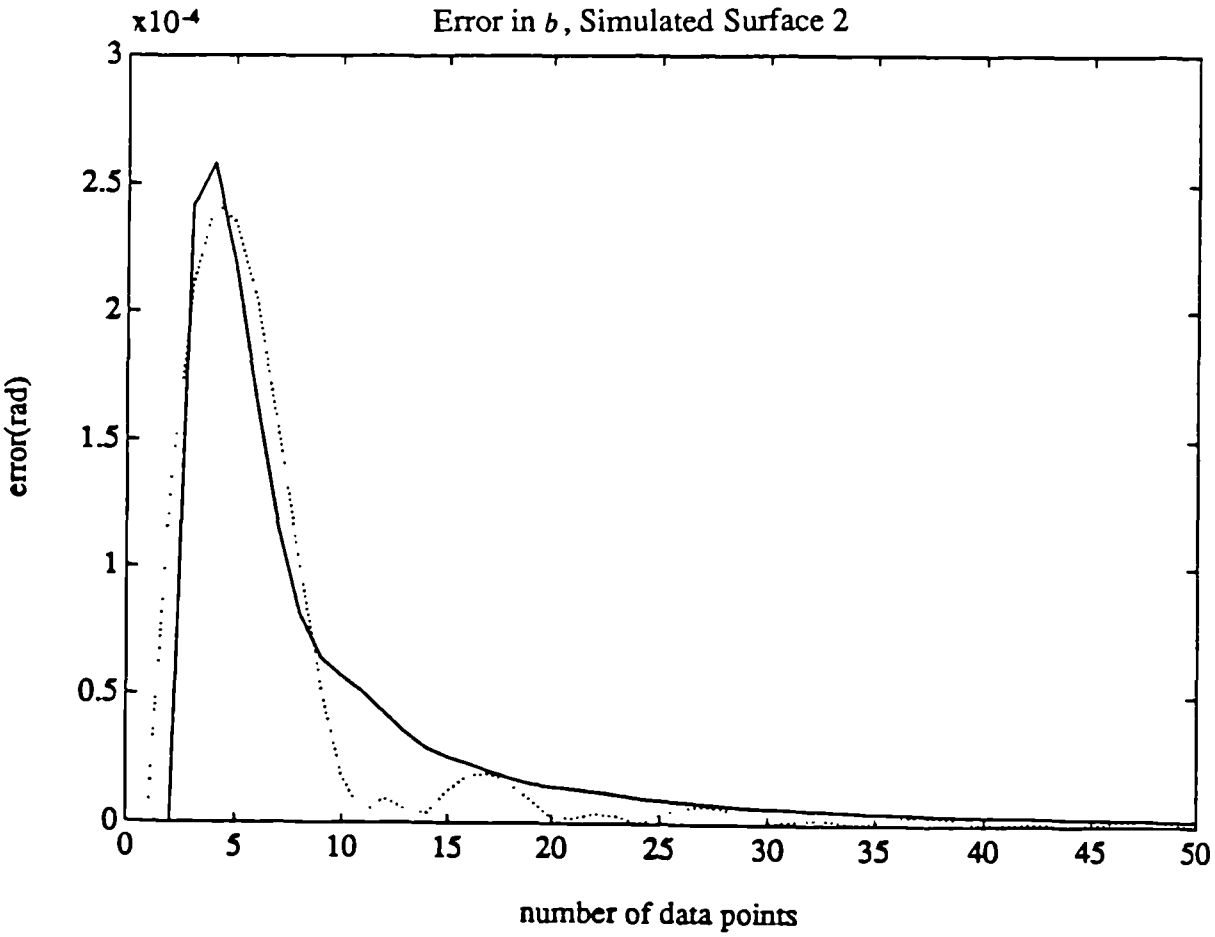
.....: Actual Sampling Error



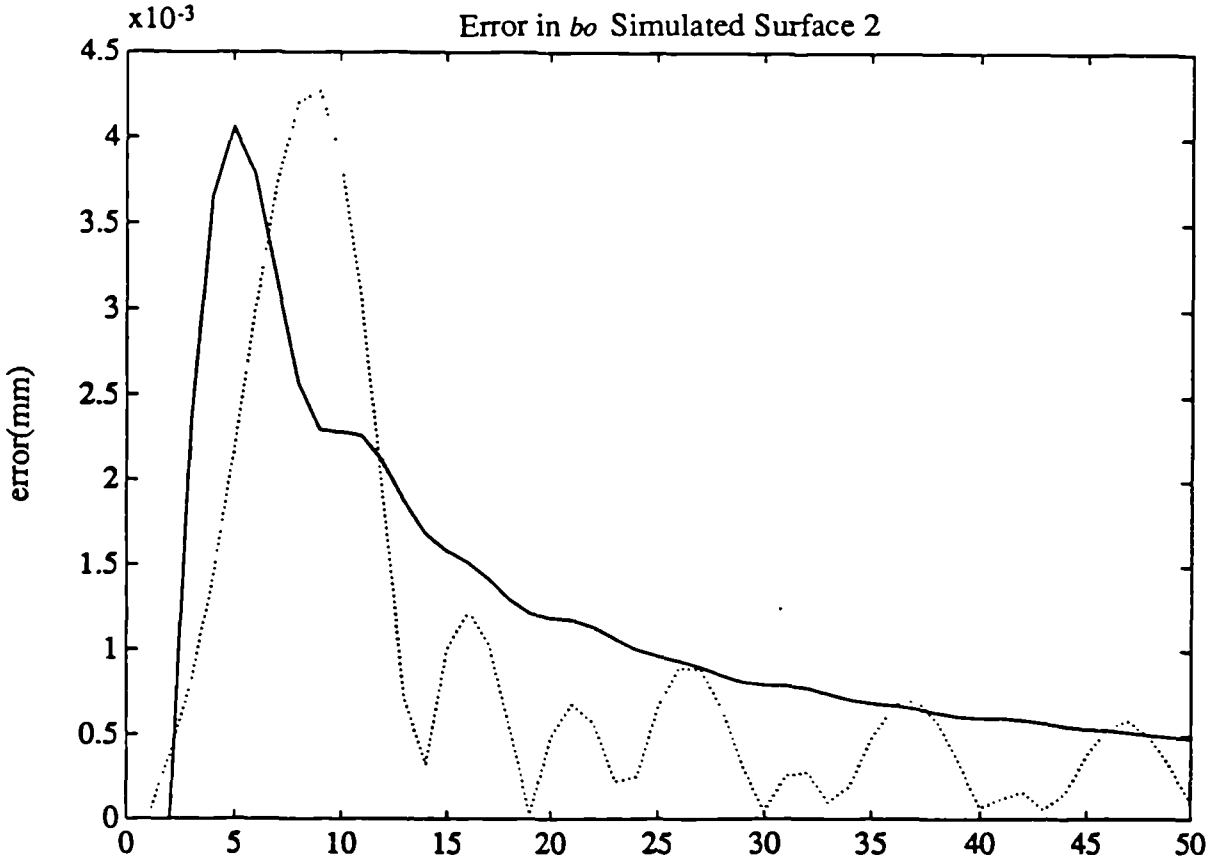
Graph 4.1 (a)



Graph 4.1 (b)



Graph 4.2 (a)



Graph 4.2 (b)

4.2. Behaviour of the Discrepancies Between Derived and Actual Sampling Errors

A series of experiments showed that when the discrepancies, expressed as a factor of multiplication were divided by the respective amplitude of the original signal and plotted against the product of frequency and sampling interval ($f \times dx$), the resulting plot remained the same for all permutations of frequency, sampling interval and amplitude.

Graph 4.3 shows this normalised (by the amplitude of the sinusoid) scaling factor plotted for values of frequency, sampling interval product ($f \times dx$) of between 0 and 1Hz mm and for phase (of sinusoid) varying between $0 \rightarrow 2\pi$. Graph 4.4 shows similar information when ($f \times dx$) lies between 1 and 2 Hz mm. The plot for values of ($f \times dx$) lying between any two succesive non zero integers is similar to Graph 4.4. Both of these plots can be represented by a series of fourteenth degree polynomials with a standard deviation of 0.04468 where $0 < (f \times dx) < 1$, and 0.02228 where $n < (f \times dx) < (n+1)$.

The sampling error on a sinusoidal surface of known form can now be predicted using the polynomials in graphs 4.3 and 4.4 , graphs 4.5, 4.6, and 4.7 show the results obtained by using the above procedure on simulated surfaces 3, 4 and 5 The parameters of which are given in table 4.1.

Surface		
1	0.010 sin (2π0.1000x)	220.000
2	0.005 sin (2π0.0200x)	900.000
3	0.020 sin (2π0.0900x) 0.010 sin (2π0.9000x)	110.000 220.000
4	0.005 sin (2π0.0600x + 4 π/10) 0.030 sin (2π0.1200 x + 9 π/10)	600.000 70.000
5	0.005 sin (2π0.0100x+ π/3) 0.010 sin (2π0.1600x + 19 π/20) 0.200 sin (2π0.0300x+ π) 0.020 sin (2π0.0200x+ π/5)	900.000 600.000 22.500 55.00
6	0.005 sin (2π0.0600x + 4 π /10) 0.030 sin (2π0.1200x + 9 π /10) rand, std=0.0644	600.000 70.000
7	0.010 sin (2π0.1000x + π /5) 0.220 sin (2π0.0600x + 7 π /10) 0.066 sin (2π0.1800x + 19 π /10) 0.007 sin (2π0.0300x + π) 0.030 sin (2π0.1400x + 13 π /10 rand, std=0.0177	130.000 13.640 48.488 514.230 166.667
8	0.003 sin (2π0.0900x + 7 π /10) 0.020 sin (2π0.0300x + π /5) 0.030 sin (2π0.0600x + 16 π /10) 0.069 sin (2π0.1400x + 19 π /10) 0.052 sin (2π0.1600 x) rand, std=0.0295	866.667 80.000 100.000 72.464 115.385

Table 4.1

The sampling error occuring as a result of random surface characterisitics was evaluated by two alternative approaches; using the autocorrelation function or the Gaussian distributed standard deviations the measure of spatial variation, the most appropriate choice would be that which is the most reliable and computationally efficient. The first approach gave results that were vaguely systematic in dx and $\sum y$, but were far too erratic to be used as a basis for prediction. However, the second approach was investigated using the following expressions:

$$s.e.(b) = \frac{\sigma}{\sqrt{\sum (X_i - X)^2}} \tag{4.30}$$

and

$$s.e.(bo) = \sqrt{\frac{\sum X_i^2}{n \sum (X_i - X)^2}}\sigma \tag{4.31}$$

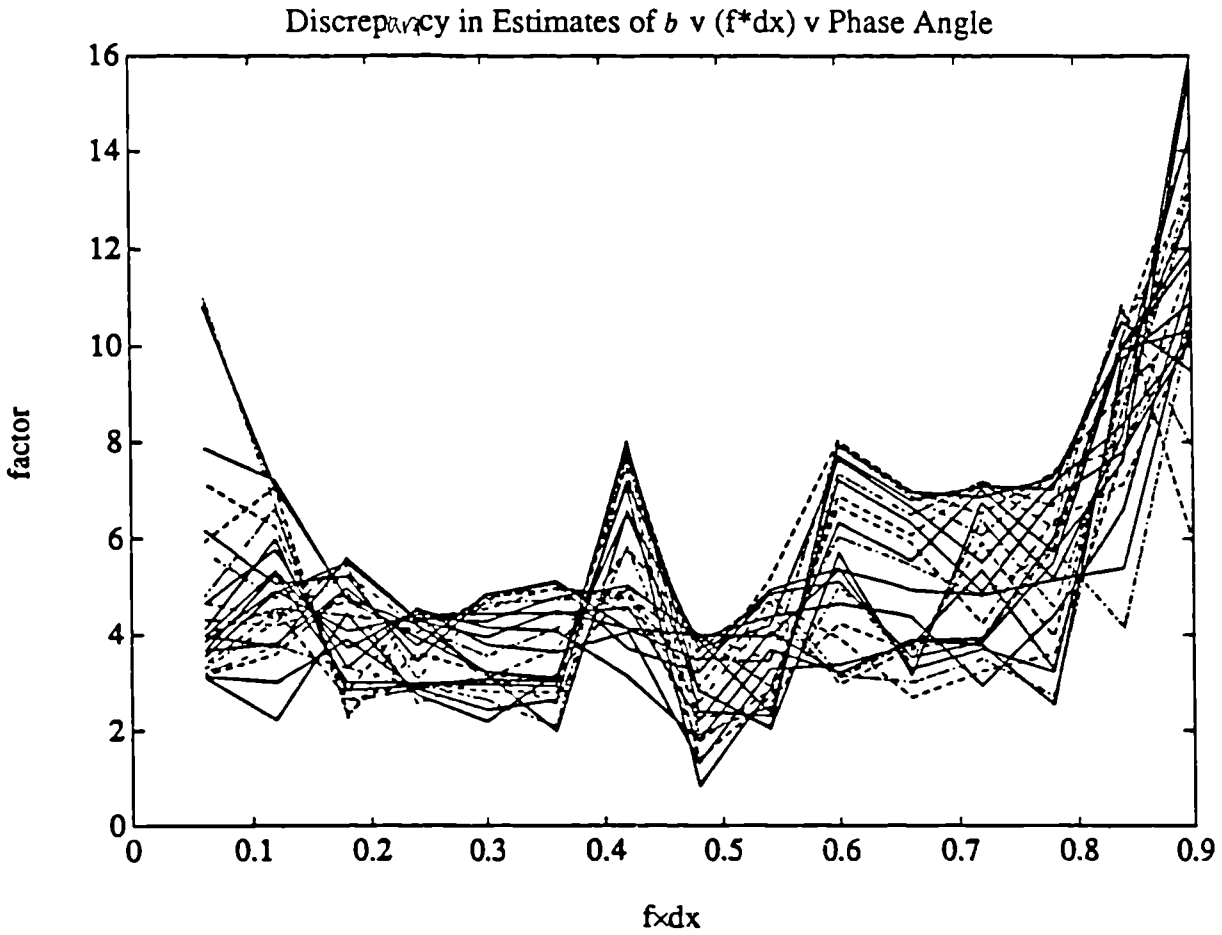
The data used was created using the random number generator on the MATLAB package. The data output form this package was in the form of 16 character digits between 0 and 1, (y_r) and was processed thus:

$$y = w(y_r - 0.5) \tag{4.32}$$

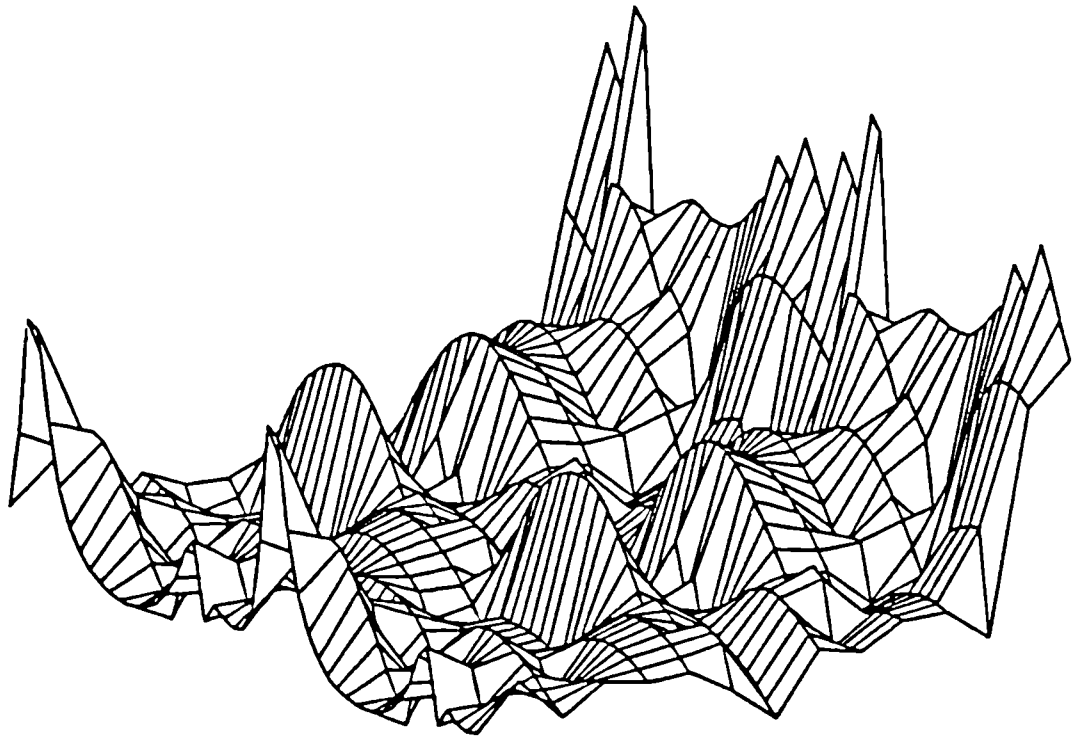
Where w denotes a real number less than one to scale the signal. It was found that using the equations (4.30) and (4.31) again resulted in discrepancies between the predicted and actual sampling error, these differences were systematic in sampling interval (dx), i.e. equations (4.30) and (4.31) should be postmultiplied by (dx/4). This method was tried on random data as described above. Graphs 4.8, 4.9 4.10 and 4.11 show the results obtained on surfaces having standard deviations of 0.003, 0.0295, 0.0148 and 0.0018 respectively.

The overall estimate of sampling error is to sum the predictions made for each of the surface elements. Graphs 4.12 to 4.15 show the predicted and actual sampling errors obtained on the simulated surface profiles 6, 7 and 8 described in table 4.1.

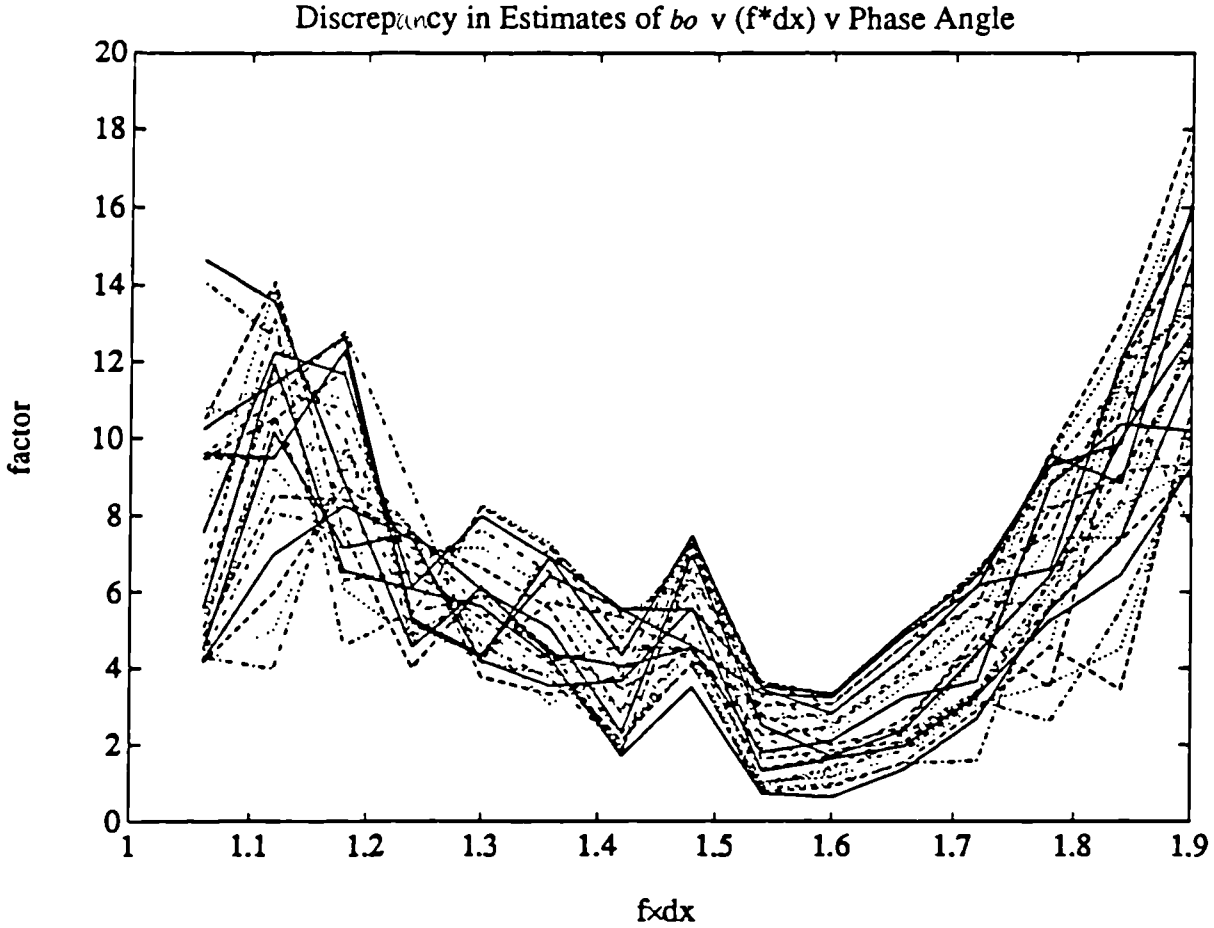
The above type of analysis is best suited to batch sample applications, particularly in ascertaining how many data points would be required to satisfy given tolerances. The



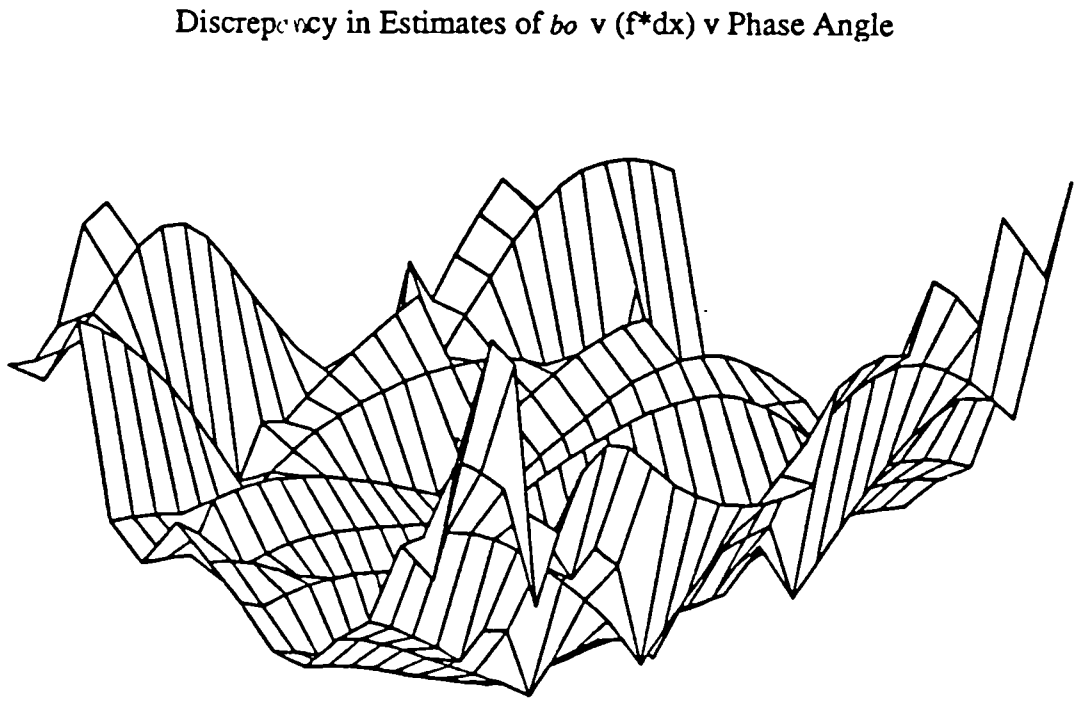
Graph 4.3 (a)



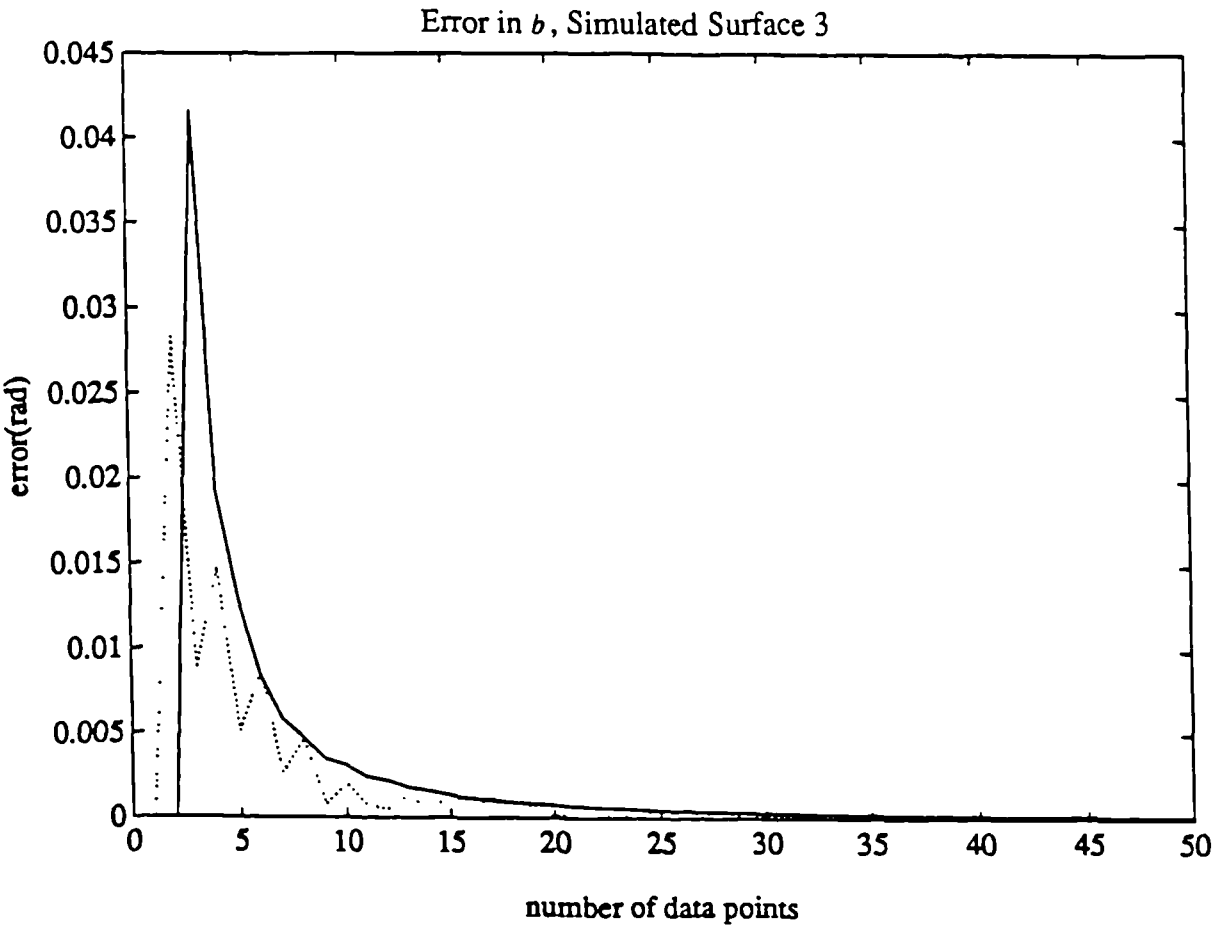
Graph 4.3 (b)



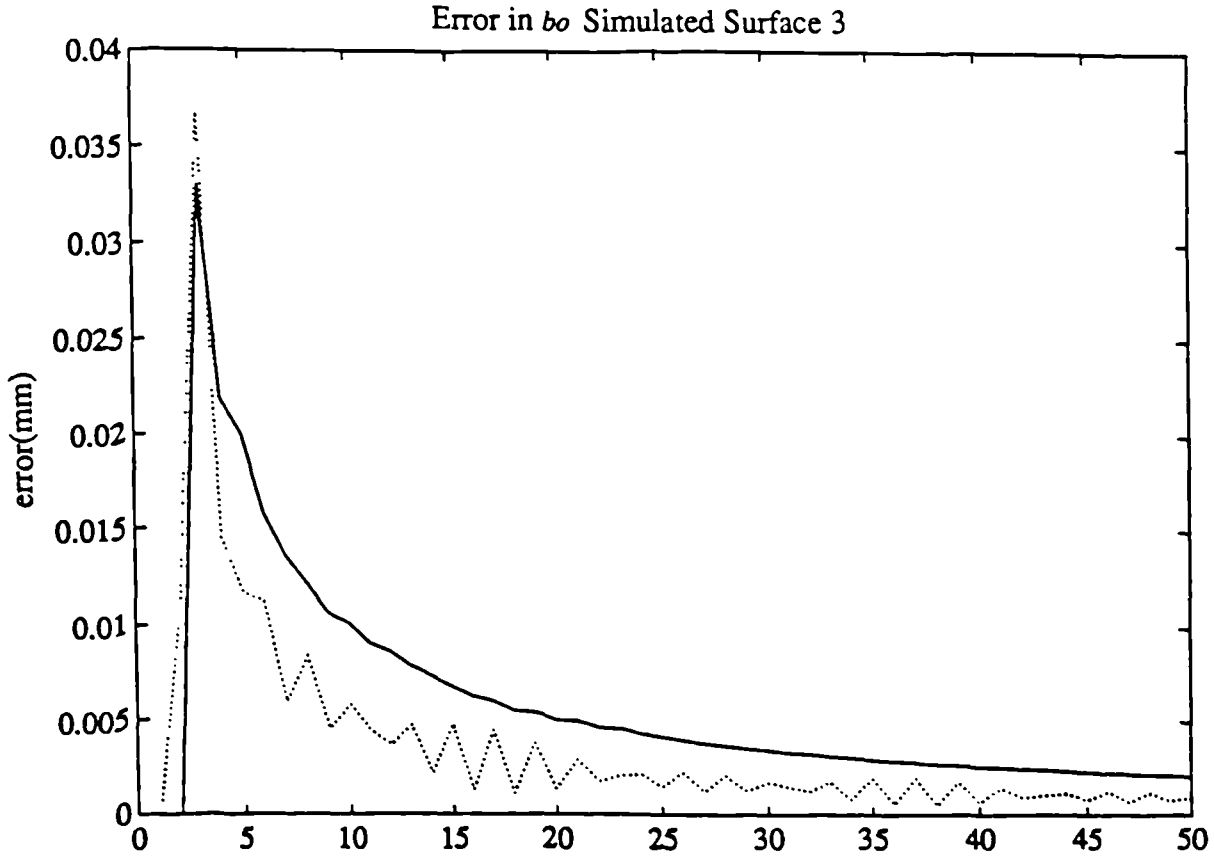
Graph 4.4 (a)



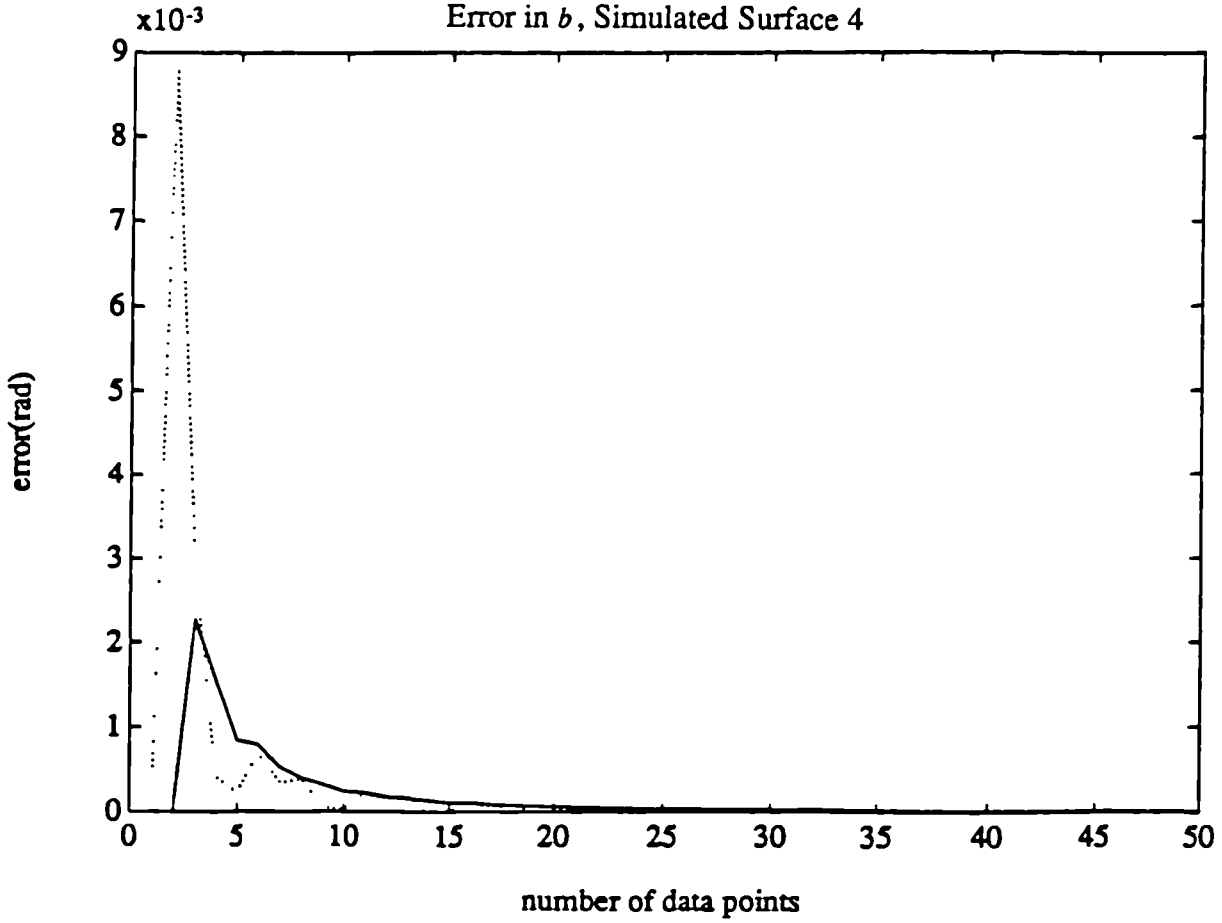
Graph 4.4 (b)



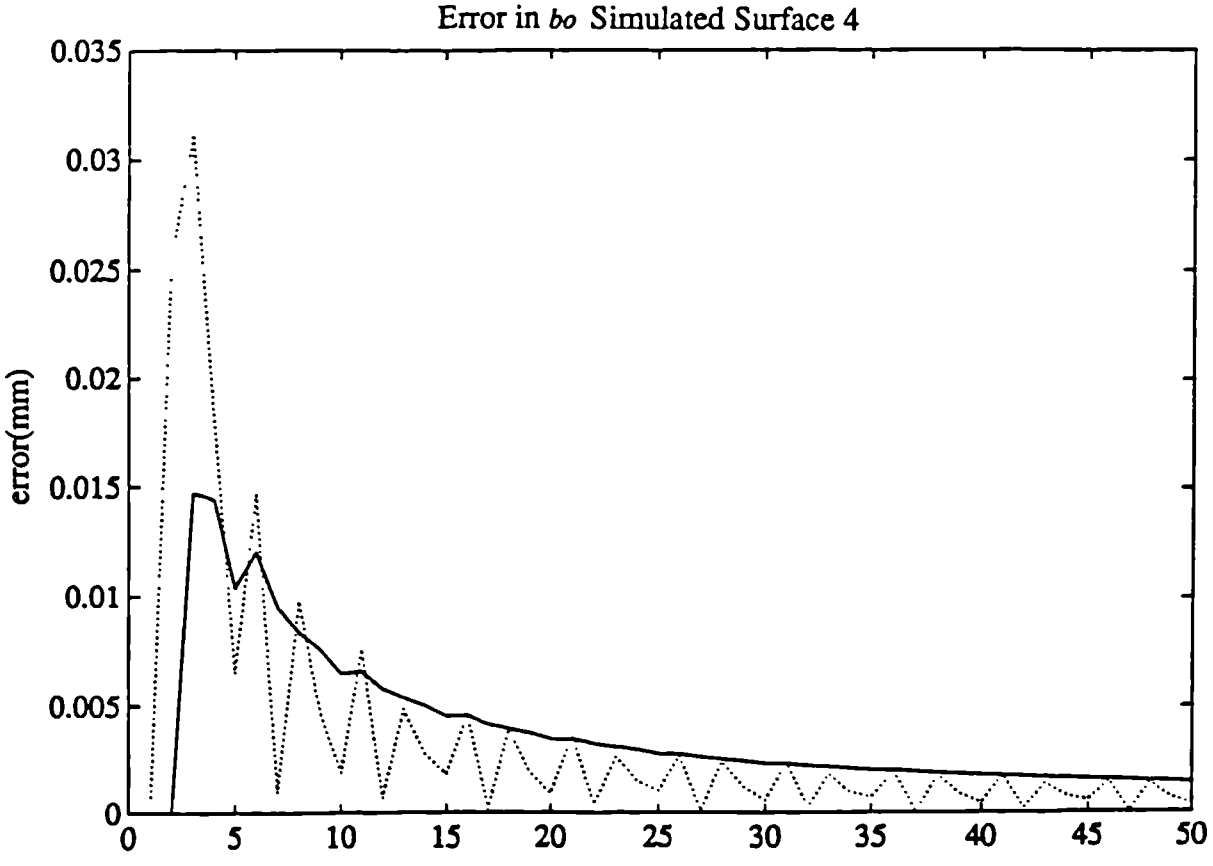
Graph 4.5 (a)



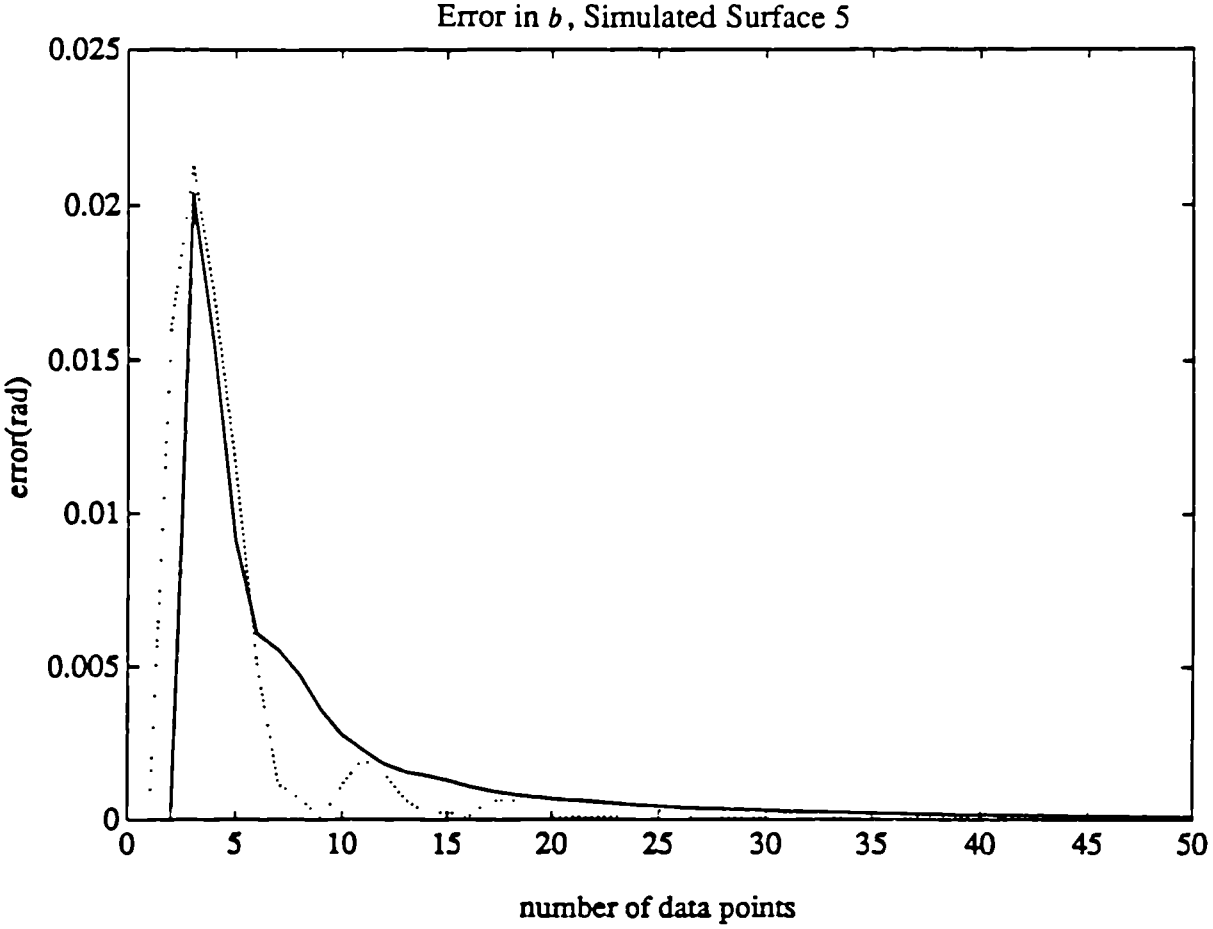
Graph 4.5 (b)



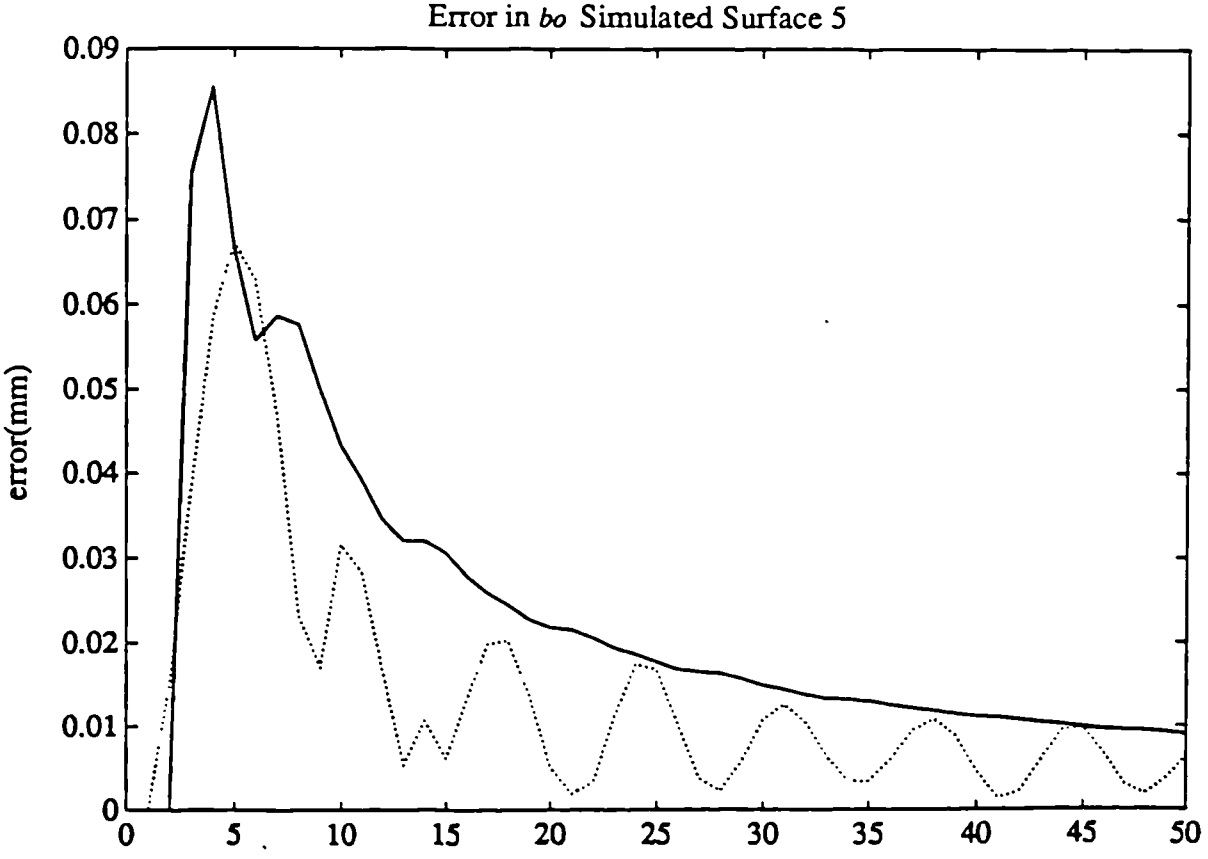
Graph 4.6 (a)



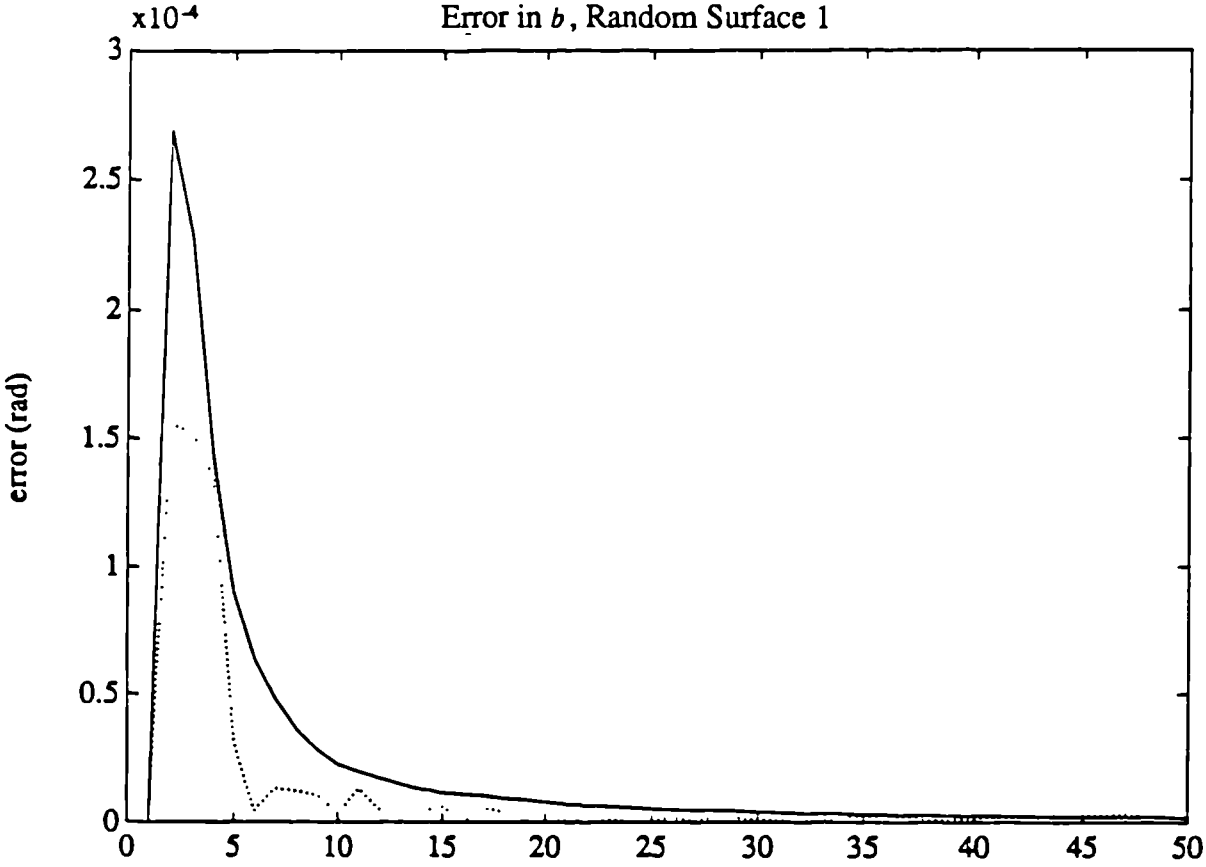
Graph 4.6 (b)



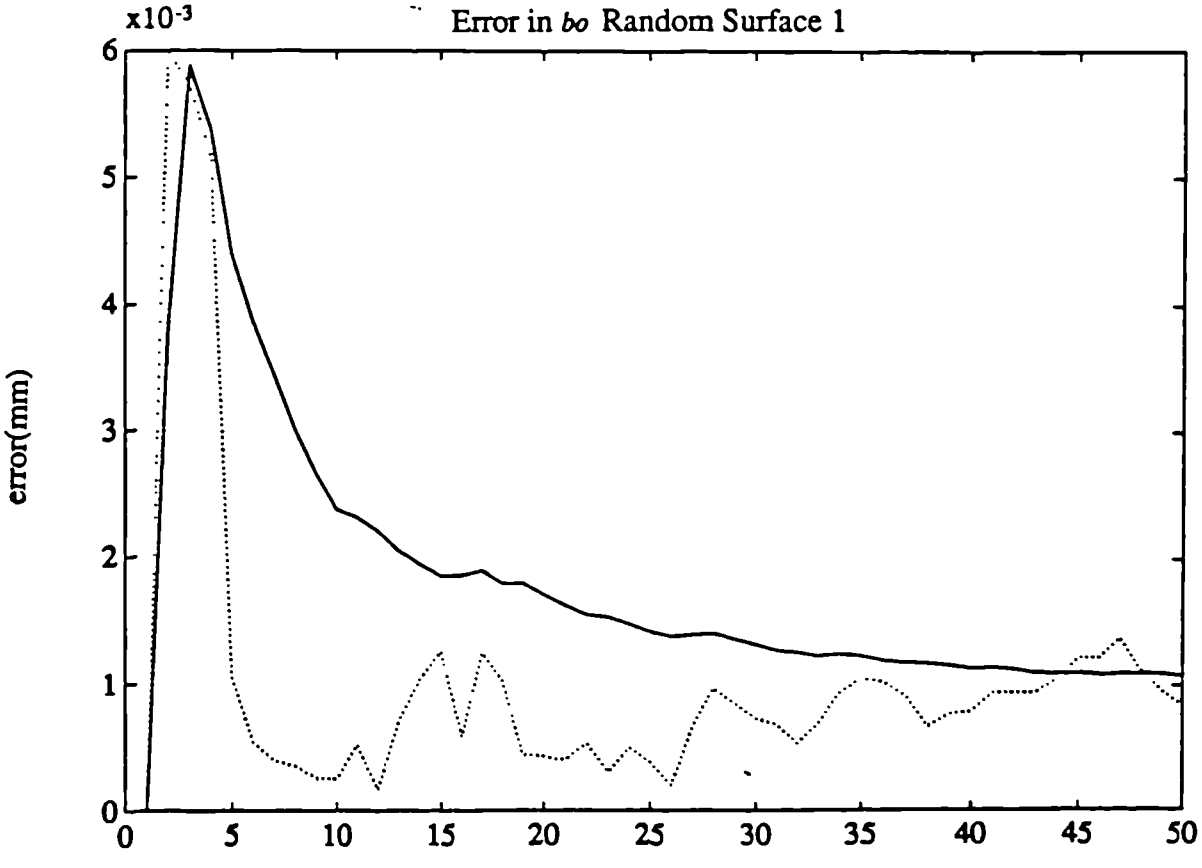
Graph 4.7 (a)



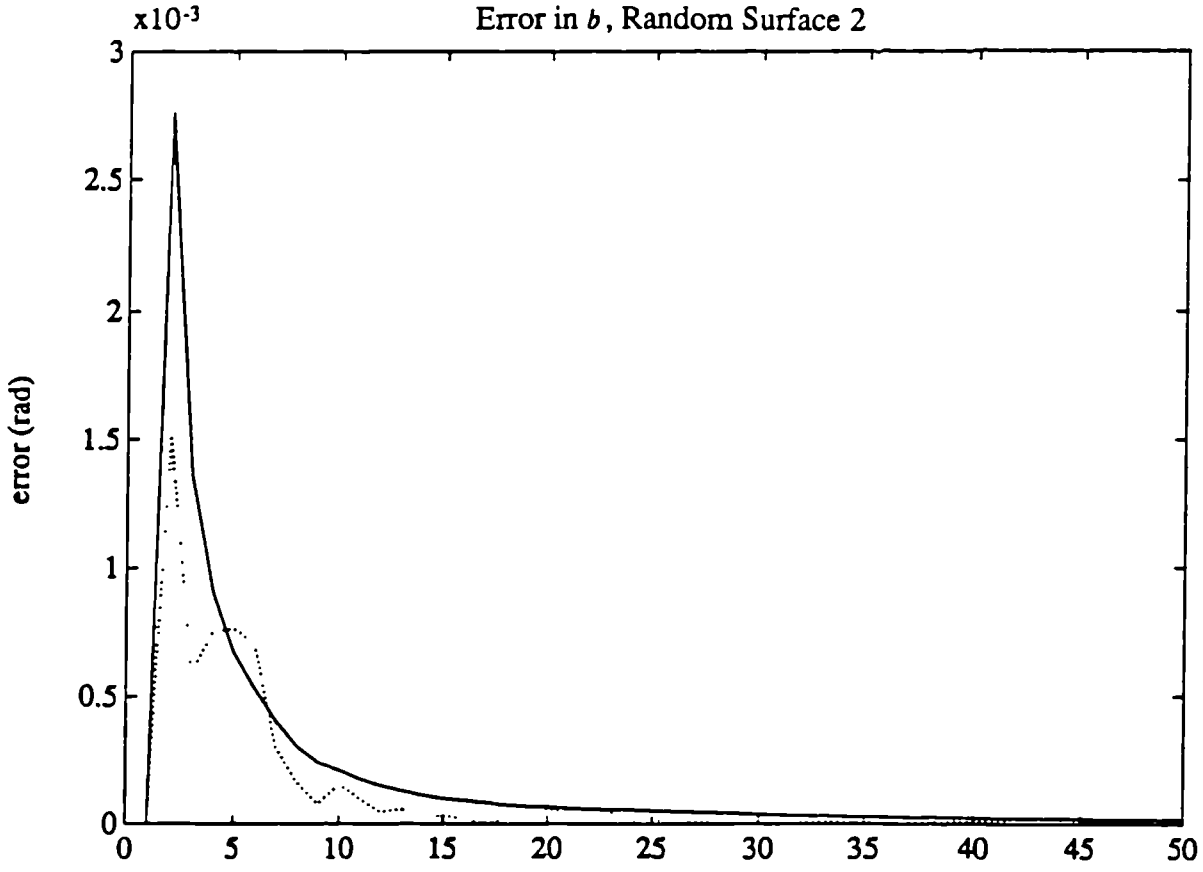
Graph 4.7 (b)



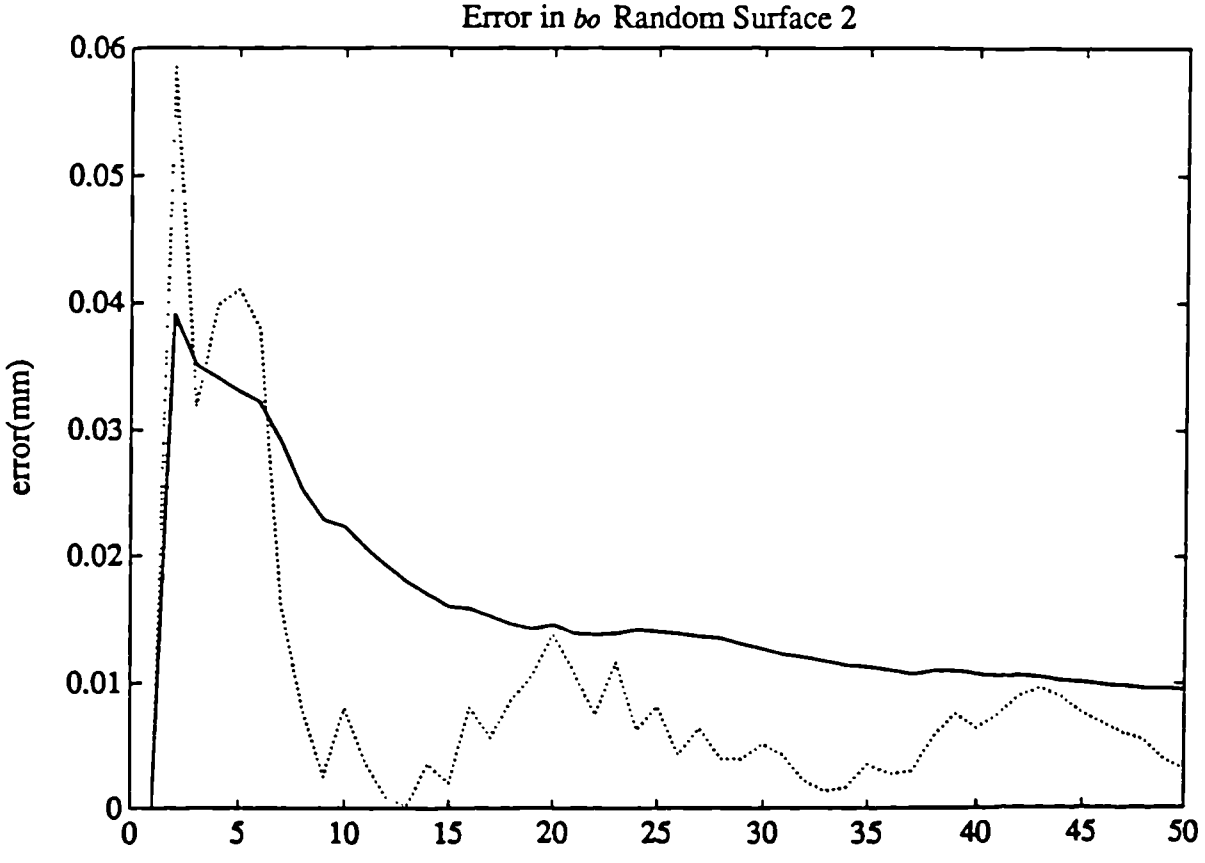
Graph 4.8 (a)



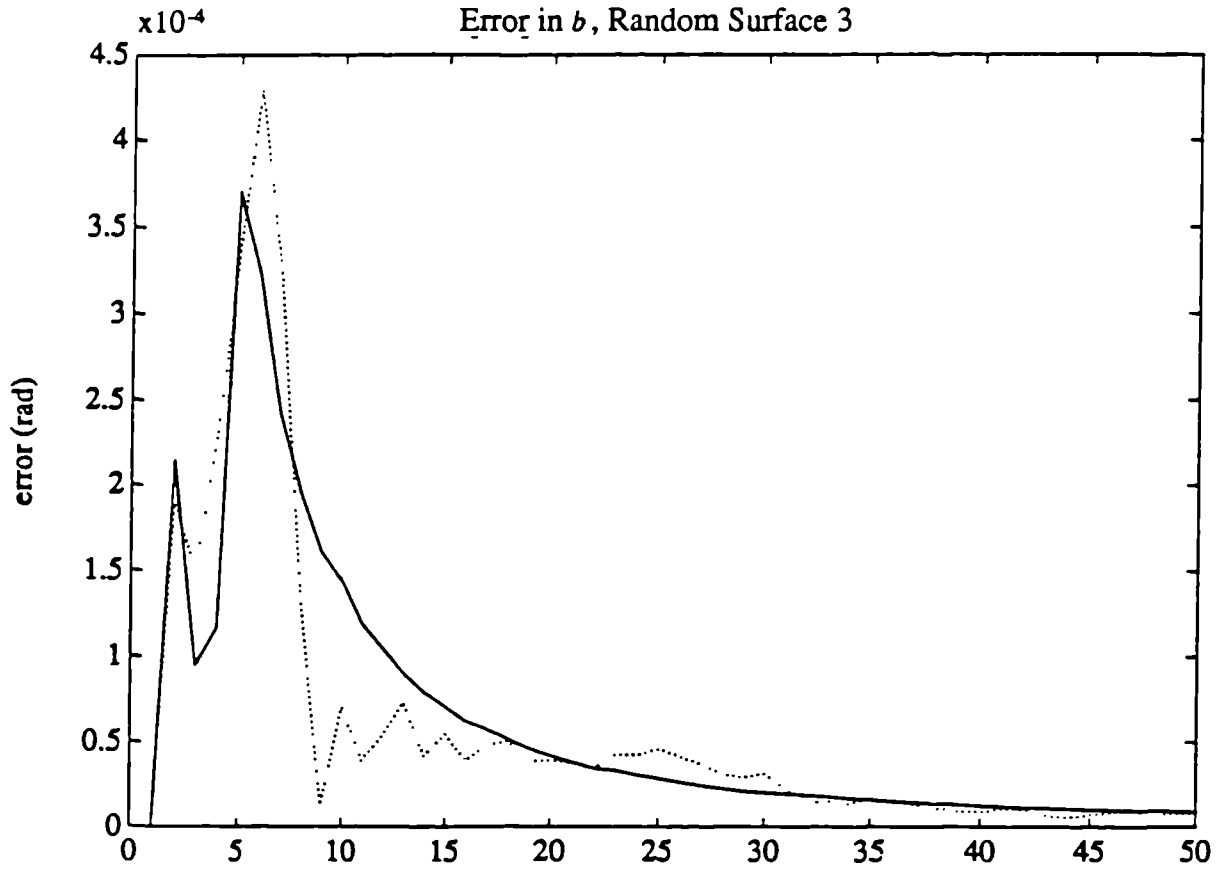
Graph 4.8 (b)



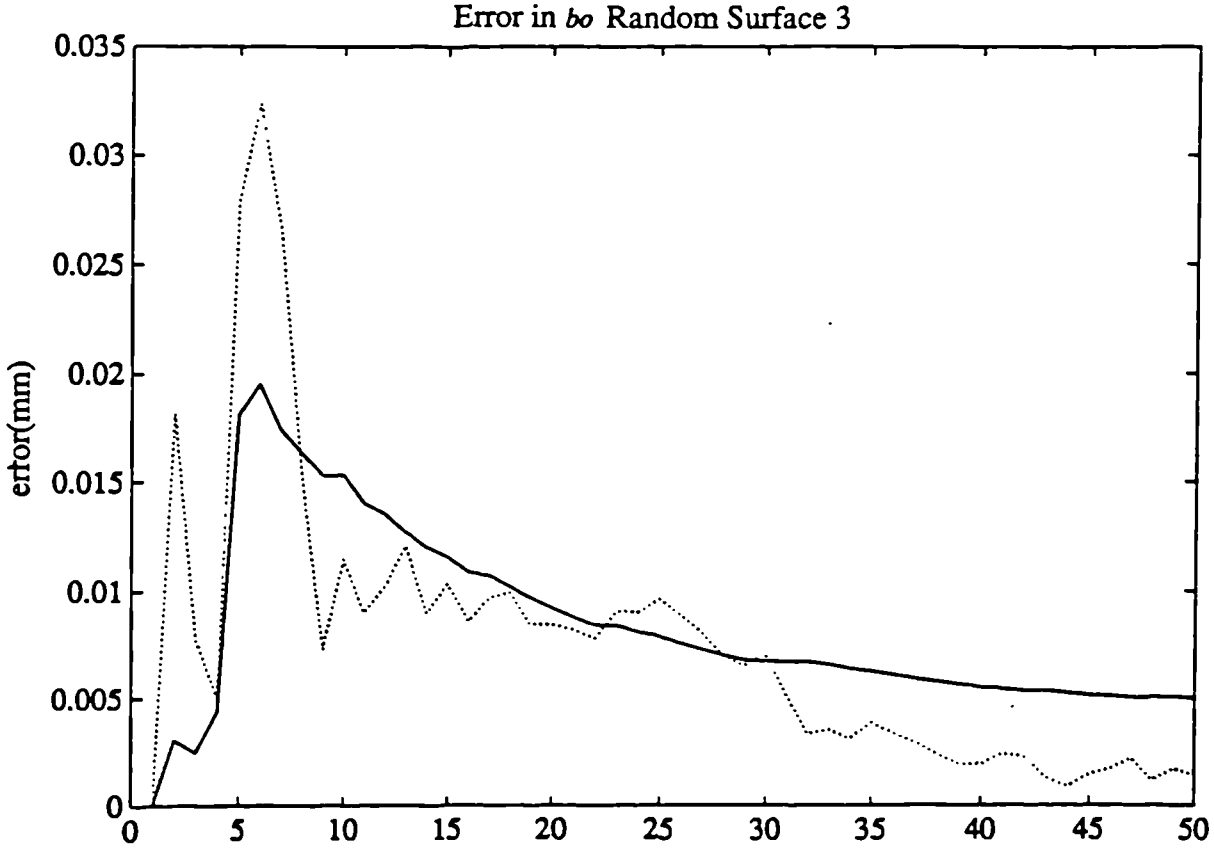
Graph 4.9 (a)



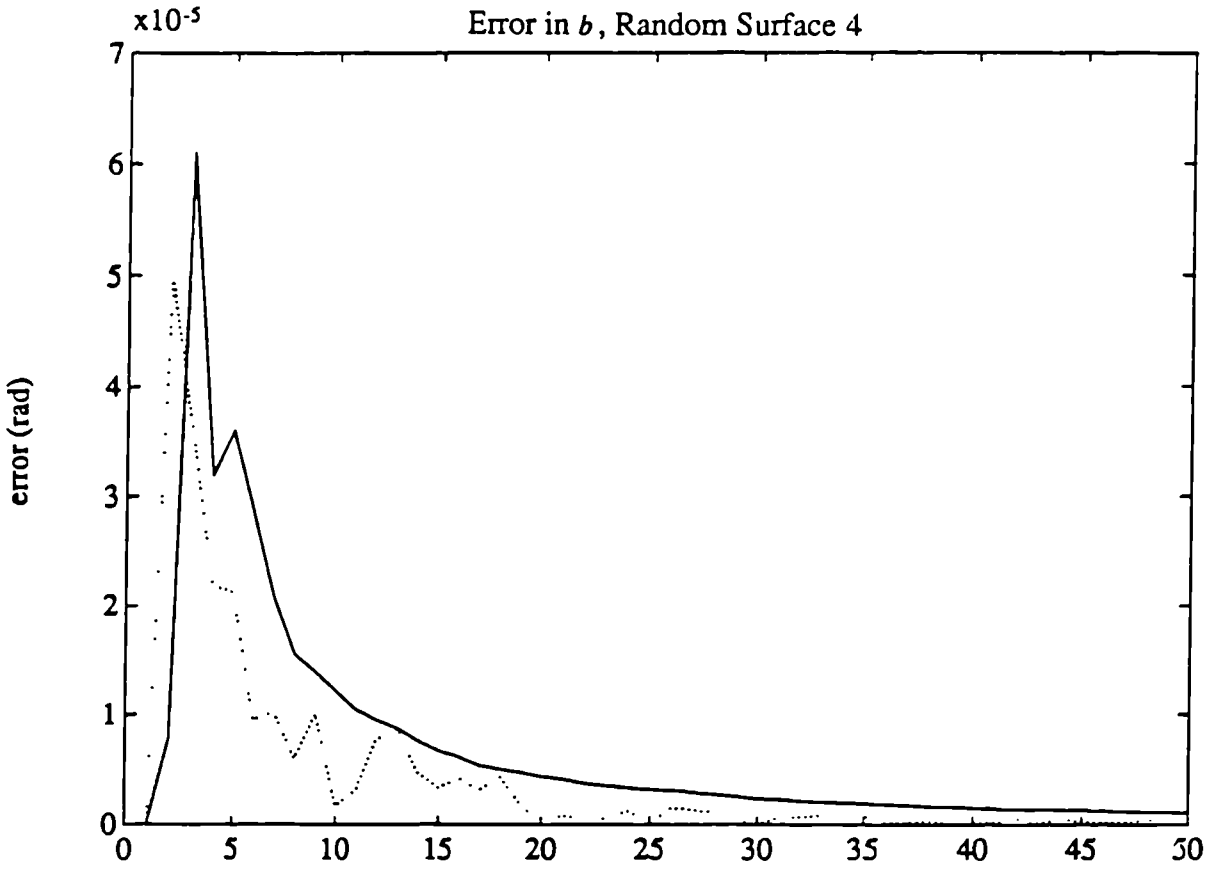
Graph 4.9 (b)



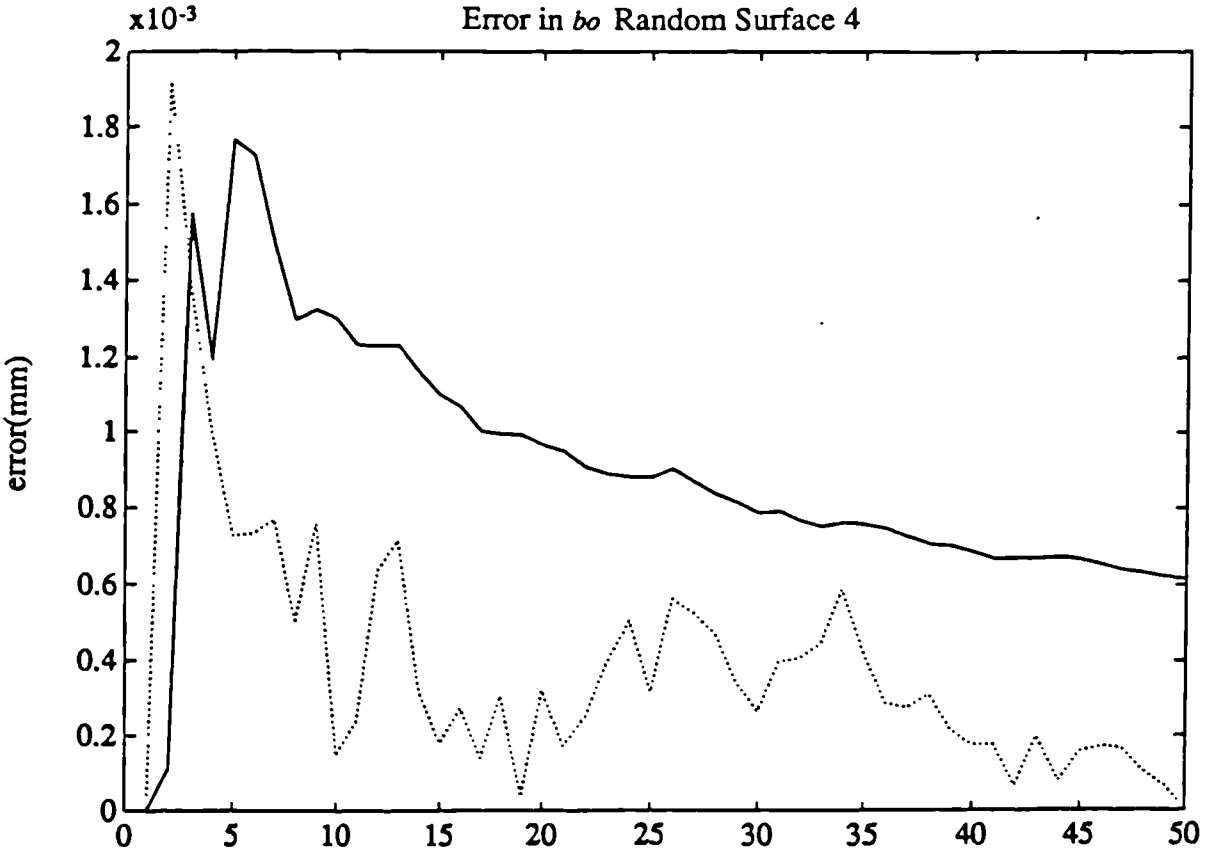
Graph 4.10 (a)



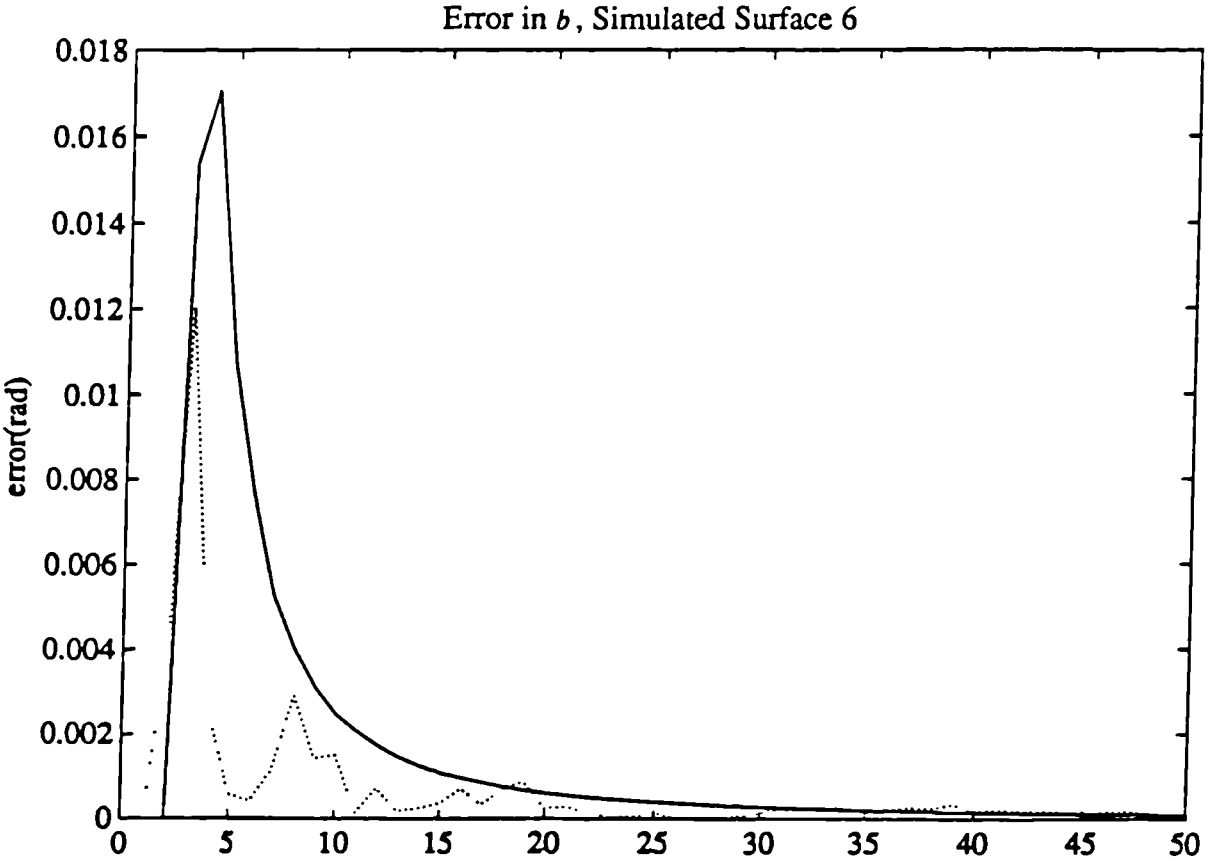
Graph 4.10 (b)



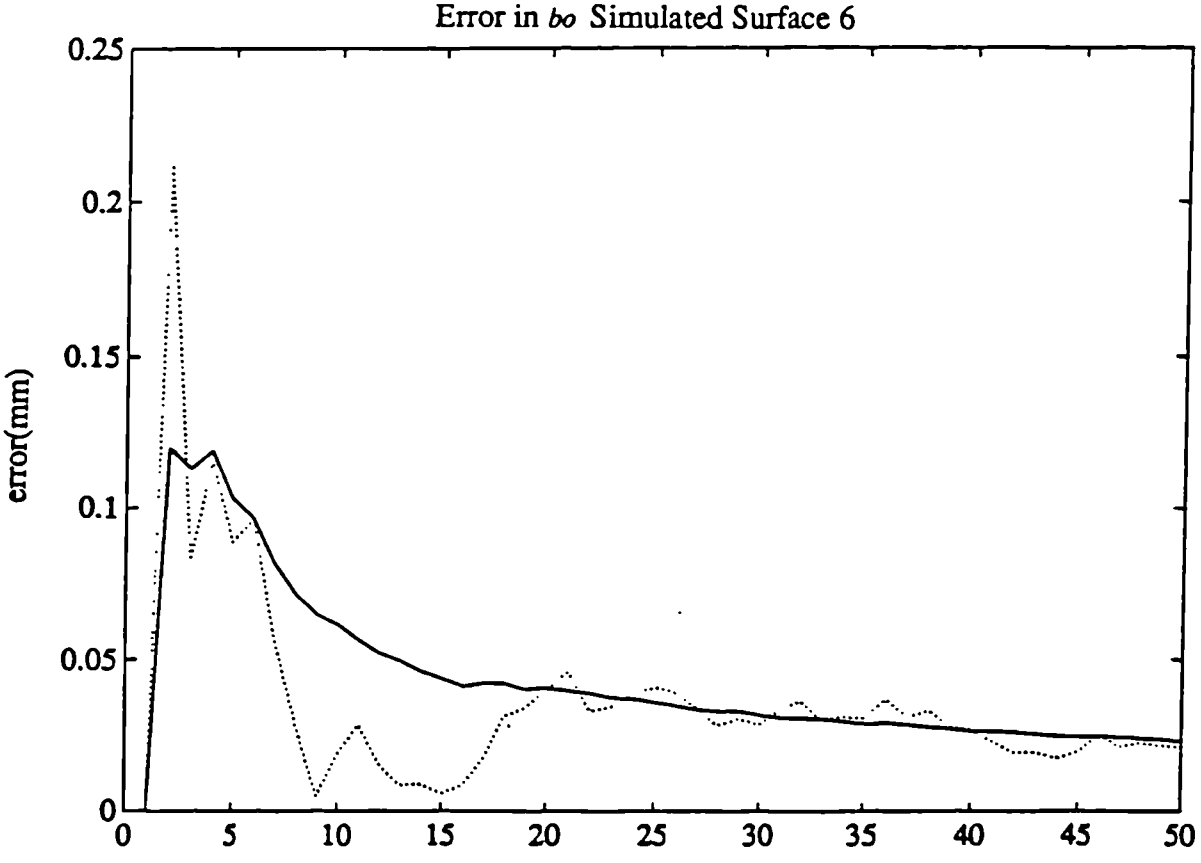
Graph 4.11 (a)



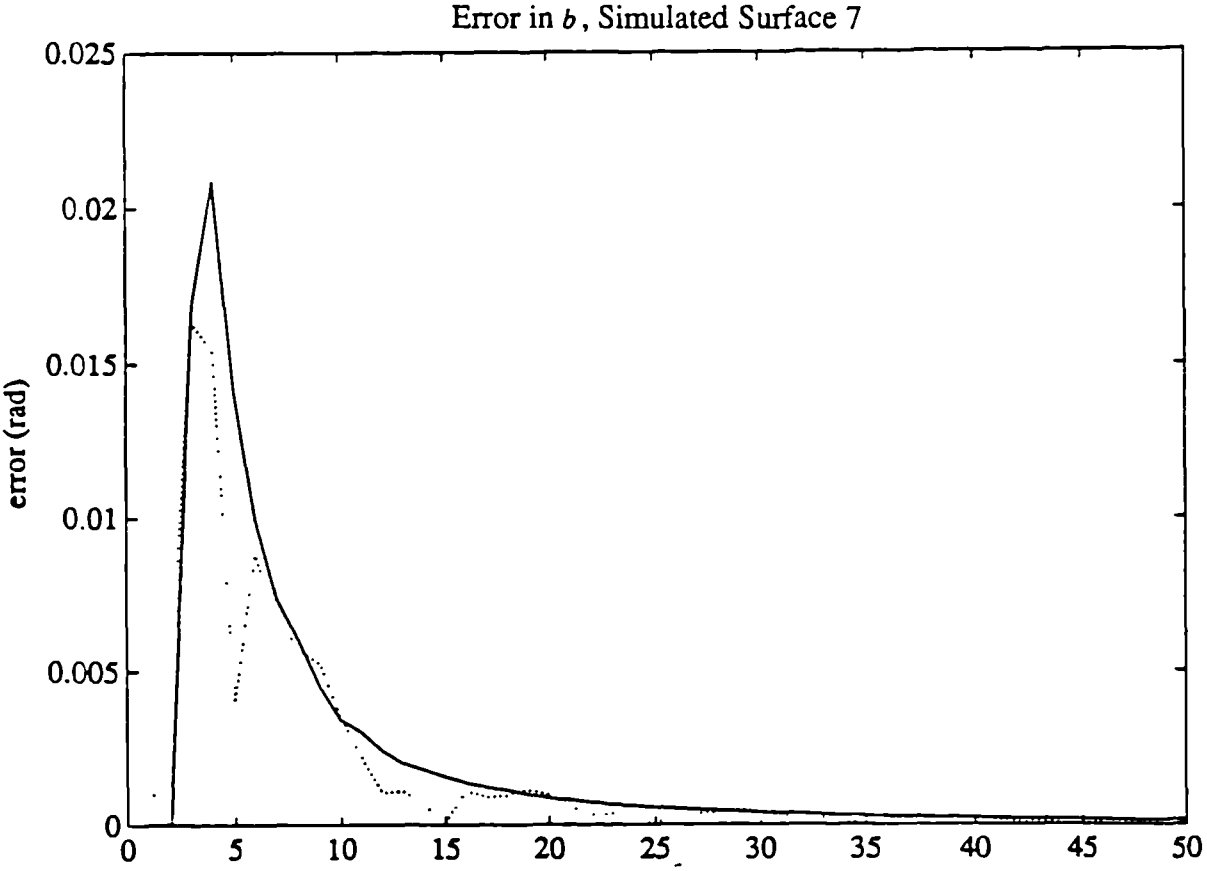
Graph 4.11 (b)



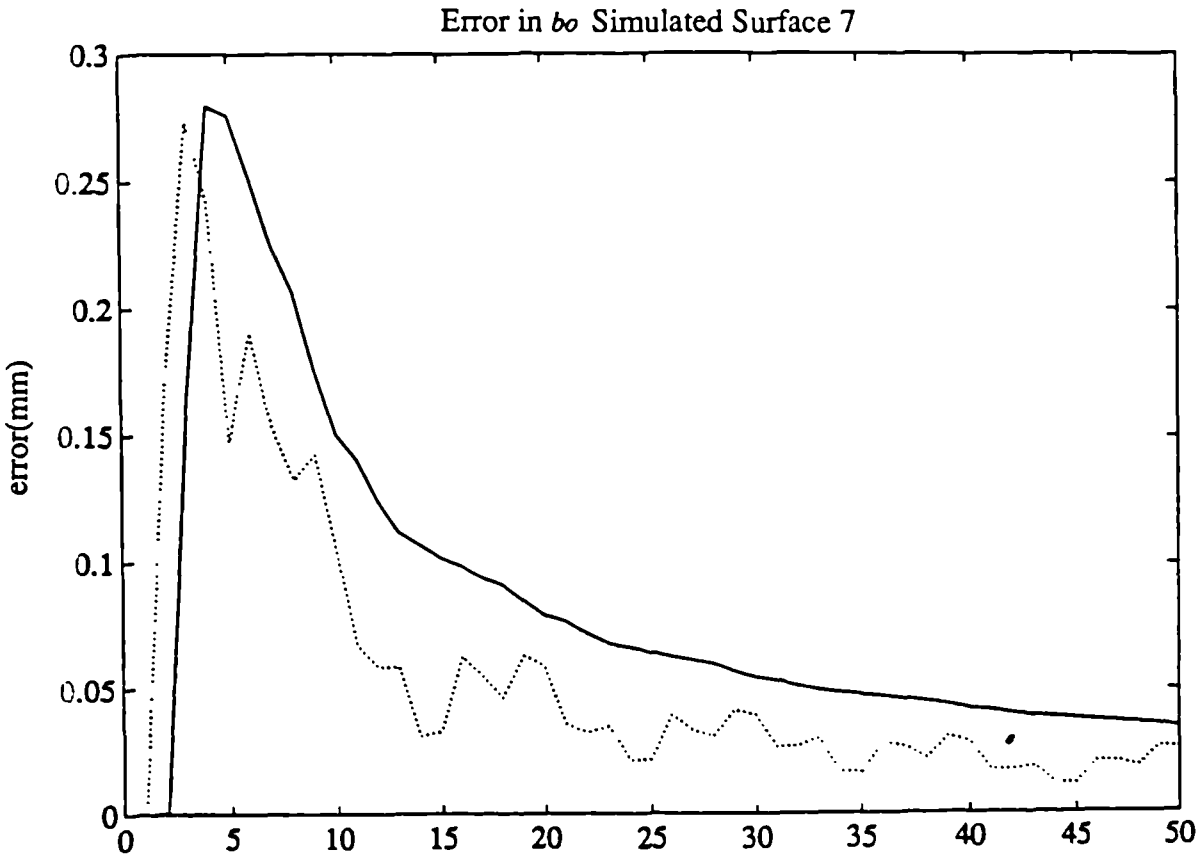
Graph 4.12 (a)



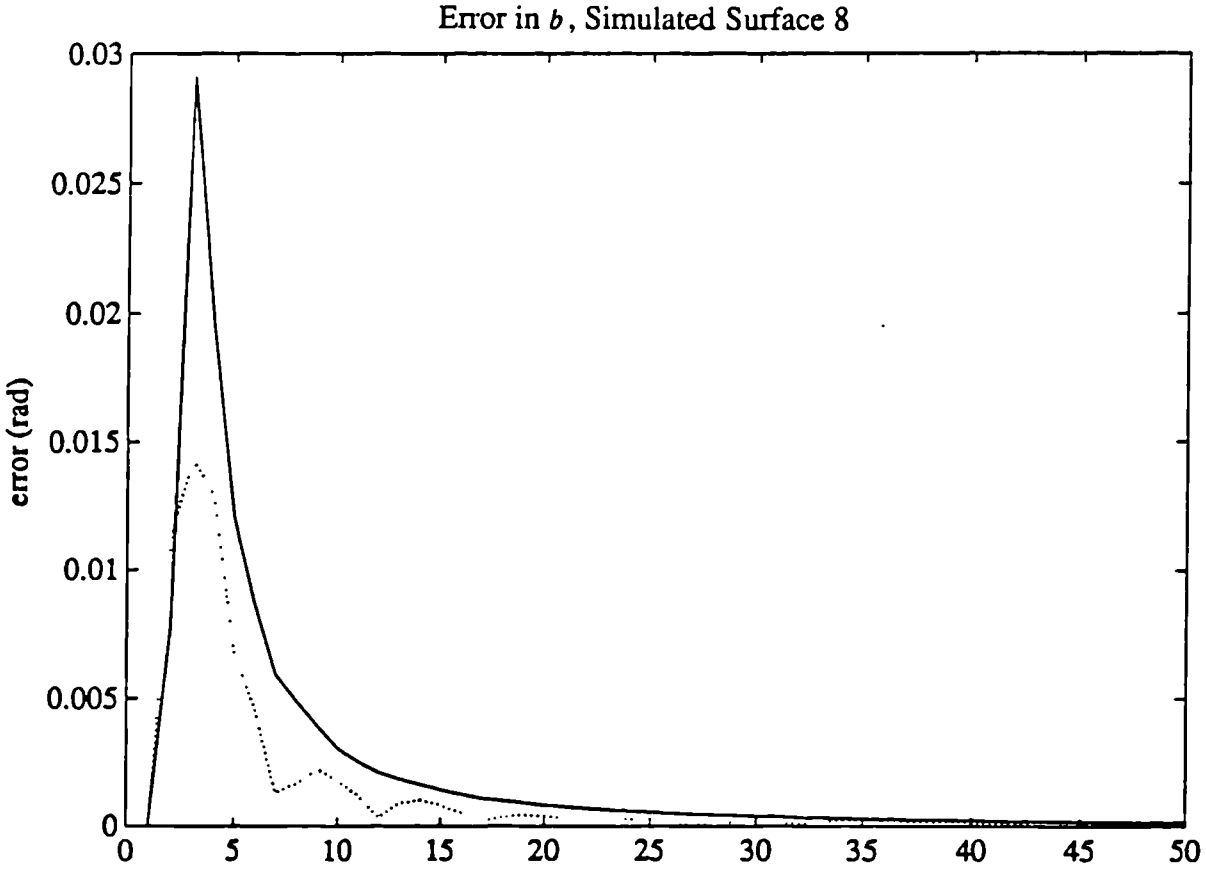
Graph 4.12 (b)



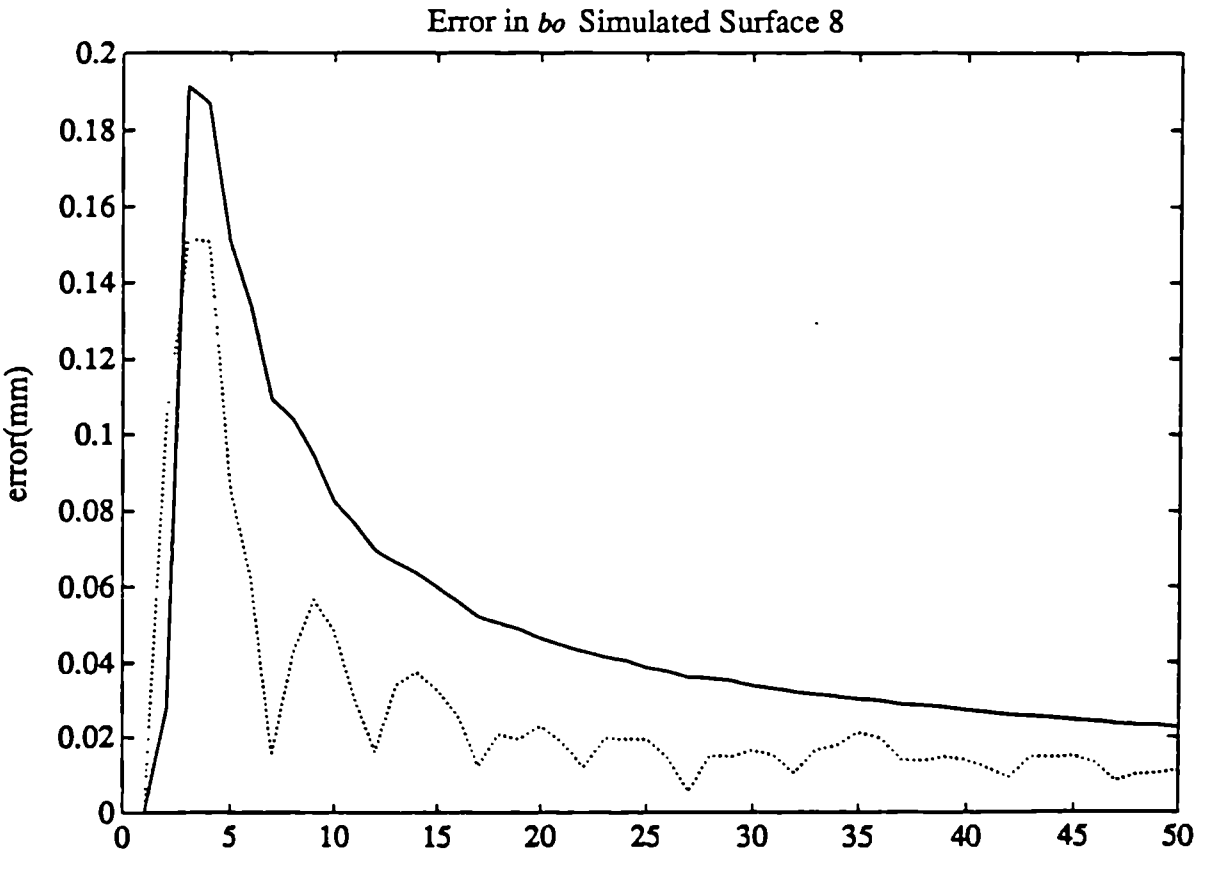
Graph 4.13 (a)



Graph 4.13 (b)



Graph 4.14 (a)



Graph 4.14 (b)

best way of achieving this would be to perform similar calculations to those carried out to obtain the results shown in graphs 4.1 to 4.15 , .i.e. plotting the results obtained using between two and the highest feasible number of data points (which would be dictated by the dimensions of the surface) and locating the coordinate on the function that corresponds to the required tolerance level and its corresponding number of data points. Alternatively, the analysis can be subjected to an iterative search process in n, minimising the function:

$$u = \min\left(tolerance - standard\ error\ of\ parameter\right) \tag{4.33}$$

5. Harmonic Analysis

If this method of quantifying sampling error is to be successfully applied to the inspection of engineering components, detailed harmonic information is required. There are two possible ways in which this information may be obtained.

The first possiblity is that detailed surface information could be gathered by the CMM itself, as many machines are now capable of profiling. This data can then be used in a Power Spectral Density Function or a Discrete Fourier Transformation to give information as to the harmonic content of the surface. However the P.S.D.F. has a disadvantage in that it relies upon the surfaces being well represented by a series of sinusoids, and that it contains no phase related information. Usually in surface metrology, adequate representation requires a very broad band spectrum, because the nature of the small scale surface effects are very complex, however, in the case of coordinate measurement far less detail is required as many of the small scale features are 'mechanically filtered' out by the geometry of the contacting stylus probe.

5.1. Computation of the Power Spectral Density Function

The computation of the power spectral density function can be achieved through many different numerical integration techniques, either effecting a transformation from the autocorrelation function or directly from surface data.

The most basic way of calculating the power spectral density function is based on an 'analysis of variance'. If the frequency content of a surface profile spans five octaves, say, and the lower frequencies correspond to wavelengths in the order of the length of traverse, the p.s.d.f is calculated by dividing the chart into 32 (2^5) sections , and the average worked out in each. The difference between the average values of consecutive sections are calculated over all 32 of the sections, these values are then squared and added. This represents an estimate within the 1/32 - 1/16 octave band. Consecutive sections are then added together, making a total of 16 sections, this process is repeated until only one section is left; which represents the whole record. When plotted, these numbers give a crude estimate of the spectrum (8).

The relationship described in diagram 4.1 still holds for the calculation of the p.s.d.f. in that it may be calculated from the autocorrelation (or autocovariance) function or directly from the surface data. In terms of the autocovariance function, the p.s.d.f may be defined thus:

$$G(\omega) = \frac{2}{\pi}\int_0^\infty R(\tau)\cos(\omega\tau)d\tau \qquad (4.34)$$

Using raw surface data, the p.s.d.f can be directly calculated by using the following equation:

$$G(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} z(x)\exp(-i\,\omega x)dx \qquad (4.35)$$

Which is solved by FFT methods.

An alternative approach for obtaining the harmonic data could possibly be in the use of knowledge of feed rates, cutting forces and the machine tool calibration data. A considerable amount of work has been done in this area, analysing how the aforementioned phenomena affect the characteristics of the P.S.D.F. (2).

6. Discussion

The analysis outlined in this chapter serves to illustrate the argument that sampling

errors in the estimation of geometric parameters can be quantified by using the autocorrelation function of a surface profile as a measure of spatial variation. Such an approach will be of increasing importance as the application of CMM algorithms using L_1 or L_∞ norms and functionally significant criteria become more widespread.

However, the harmonic analysis required prior to using the algorithm is computationally expensive and requires a considerable amount of data. As it stands, this algorithm would be suitable for application in batch sampling situations, where the errors incurred in the inspection of a particular component are thought to be similar to those incurred in a similar component manufactured by the same process, i.e. one component is very similar to another, this may be the case with N.C. machined artefacts (see Recommendations for Further Work).

RERERENCES

[1] F. EDSON & G. PARRY (1985): "Integrity of Software Associated with Coordinate Measuring Machines"; NPL Proc. Conf. Software for Coordinate Measuring Machines, 25/9/1985.

[2] H. HINGLE (1985): M. Phil. Thesis University of Warwick.

[3] D.J.WHITEHOUSE & M.J. PHILLIPS (1978): "Discrete Properties of Random Surfaces"; Phil. Trans. Roy. Soc., Vol. 290, pp267-298.

[4] J.S. BENDAT & A.G. PIERSOL (1971): "Random Data : Analysis and Measurement Procedures"; Wiley-Interscience.

[5] D.E. NEWLAND (1975): "An Introduction to Random Vibration and Spectral Analysis"; Longman .

[6] I. MILLER & J.E. FREUND (1977): "Probability and Statistics for Engineers"; Prentice-Hall International Editions.

[7] J. PEKLENIK (1967): "New Developments in Surface Characterisation and Measurements by Means of Random Process Analysis"; Proc. Instn. Mech. Eng,

Vol 182 pt 3K, pp 108-126.

[8] D.J. WHITEHOUSE (1987): "An Aspect of Errors Introduced into the Measurement of Form due to Miniturisation"; J. Phys. E: Sci. Instrum. Vol 16 pp:1076-1080.

[9] T.R. THOMAS (1982): "Rough Surfaces"; Longman.

[10] M.S. LONGUET-HIGGINS (1957): "The Statistical Analysis of a Random Moving Surface"; Phil. Trans. Roy. Soc., Vol 249, pp 321-387.

[11] D.J. WHITEHOUSE (1976): "Approximate Mehtods of Assessment of Surface Topography Parameters"; Annals of the CIRP Vol 25/1/1976.

[12] D.J. WHITEHOUSE & J.F. ARCHARD (1970): "The Properties of Random Surfaces of Significance in their Contact"; Proc. Roy. Soc. A316, pp 97-121.

[13] C.W. HORTON (1954): "On the use of Gram-Charlier Series to Represent Noise"; J. Appl. Phys, Vol. 27, No.4, pp 350-355.

[14] D.G. CHETWYND (1985): "On the Definition of Geometric Residuals"; NPL. Proc. Conf. Software for Coordinate Measuring Machines, 25/9/1985.

[15] DRAPER AND SMITH (1966): "An Introduction to Regression Analysis"; Johm Wiley and Sons Inc., New York.

Formulation of an Alternative Implementable Algorithm for the Quantification of Sampling Error, an Experimental Approach.

1. Introduction

For some time there has been a widely held belief that a significant proportion of the problems associated with current CMM practice are due to problems in the digital implementation of definitions of geometric form and the lack of a technique whereby the accuracy of the estimated geometric parameters can be known to some extent.

In both instances there is clearly a problem in the uncertainty of the estimates, hitherto there has been no way of knowing wether the estimates calculated are or are not within an acceptable 'tolerance' from the true value, i.e. as yet there is no implementable method for the validation of software. However, in the latter case the situation is exacerbated by difficulties and inconsistencies in the digital implementation of standards and definitions of geometric form. Existing standards refer to continua, but for application to coordinate metrology they should ideally relate to discrete sets of data. To achieve this transformation, Cox and Jackson (1) propose that standards should be expressed in terms of surfaces just containing best estimates calculated according to an appropriate mathematical criterion, and that references should be made to algorithms for estimating the uncertainty in the geometric parameters.

With the increasing development of coordinate inspection procedures such as soft-gauging (2) the need for the formulation of algorithms to quantify sampling error is becoming more urgent, as these routines cannot realistically compete with traditional

gauging methods until limits of accuracy can be firmly established, i.e. it must be possible to state the quantity and distribution of data required to satisfy comparable limits of accuracy.

It is clear that there is a need for an algorithm which has the capability to estimate the accuracy of the estimated parameters with respect to their true value. Provided that the true value can be taken as being that estimate calculated by the inclusion of all possible data in the estimation procedure, this uncertainty would effectively be the discrepancy between the estimate made using all data and that made using a particular number and distribution of data, i.e. sampling error. It is vital that the algorithm can be effectively implemented in the circumstances described above, to achieve this it is important to first of all establish a list of rigorous criteria that must be fulfilled.

2. Essential Criteria for Algorithm

If an algorithm for the quantification of sampling error is to be implemented in the circumstances described above, it must fulfill the following requirements:

(i) It must be possible to calculate/estimate the magnitude of the error incurred in least squares estimates.

(ii) The algorithm should be implementable in a procedure whereby it is possible to determine the quantity and distribution of data required to fulfill a specific accuracy requirement.

(iii) All computation must be sufficiently rapid, so that inspection speed, which is vital in coordinate measurement, is not significantly impaired.

(iv) The algorithm should not require information that can only be aquired by the use of equipment other than the CMM

(v) Ideally, no data should be required in excess of that used in the calculation of the geometric parameters.

(vi) The use of large amounts of computer memory should be avoided.

(vii) The structure of the algorithm should not involve the use of fixed models relating to surface profile, as this results in a greater degree of operator intervention and therefore increase the scope for error and reduce efficiency.

To satisfy these criteria it may be neccessary to make compromises. The results of chapter 4 indicate that to seek an exact solution, even using large amount of data, is probably futile. Therefore this chapter to seeks an approximate, but highly reliable algorithm for the quantification of sampling error that essentailly satisfies all of the above criteria.

3. Formulation of Algorithm

It was decided to concentrate attention upon the quantification of the sampling errors incurred in the estimation of the same two parameters as those considered in chapter 4. It is considered important to investigate more than one parameter in order to ensure that the algorithm derived is, to a large extent, generally applicable to the quantification of error in other geometric parameters and to identify the points at which the solutions may diverge.

For the formulation of the algorithms it was decided to use the unique derived expressions derived for the sampling error in slope eb and intercept ebo. The derivations are shown in chapter 4. The reasons for this decision are as follows:

No surface data models are required

(i) No assumptions involving the amplitude distribution of the surface data are made.

(ii) Using modern methods of computation, i.e. parallel and Fourier methods, the calculation of the autocorrelation function does not require large amount of memory as redundant data is discarded (3).

(iii) The calculation of the autocorrelation function does not require a large amount of data.

3.1. Uncertainties Associated with the Autocorrelation Function

However, the use of the autocorrelation function does present problems. The difficulties involved in making good use of digital surface data have long been recognised (4). When seeking to predict the behaviour of functional surface parameters, metrologists have found difficulty in knowing how waveform analysis relates to their measurements. These problems can be largely attributed to the fact that wave theory has been developed by electronic and communication engineers; the result being that the analysis is primarily concerned with the more temporal properties of the waveform, such as frequency and time series analysis rather than the shape of the waveform.

Much work has been done using the autocorrelation function to express surface characterisitcs (5,6). However, standard random process analysis presents an incomplete picture, to quote Whitehouse (4) "...it only gives the theoretical relationships between surface parameters and the autocorrelation function. It does not show the researcher how the measured parameters relate to the autocorrelation function."

Some work has been done to bridge the gap between theory and practice, Whitehouse and Archard (6) developed expressions for some surface parameters using an exponential autocorrelation function model, but, this implies Gaussian behaviour and therefore limits the applicability of their work. Although this situation has been improved by subsequent work relating to other types of surfaces (4) a well defined model is again used for the autocorrelation function, and this work would therefore be unsuitable in the formulation of a wholly implementable CMM algorithm. It is obvious that a neat, theoretical or modelled solution has yet to be found. The solution will therefore be sought through experimentation. The development of expressions for metrological parameters is in itself not a new idea (7), however, the following experiments for the formulation of an expression for sampling error in the estimates of b and b_o are original, and the resulting algorithm is entirely unique.

3.2. Experimental Method and Development of Algorithm

The development of a fast, implementable algorithm for the quantification of sampling error in the least squares estimates of slope (b) and intercept with the Y axis (bo) will begin with the unique theoretical expressions for their sampling error (eb) and (ebo) respectively, developed in chapter 4.

$$VI.(b) = eb = \left[\frac{1}{\sum(X_i - \bar{X})^2} \sum_{i=1}^{n} \frac{(R_i(\tau) - R_1(\tau))^2}{n^2} \right]^{1/2} \tag{5.1}$$

Similarly for the intercept b_o:

$$VI.(b_o) = eb_o = \left[\left(\frac{\bar{X}^2}{\sum(X_i - \bar{X})^2} \right) \frac{\sum(R_i(\tau) - R_1(\tau))^2}{n^2} \right]^{1/2} \tag{5.2}$$

These expressions are to be manipulated by systematic experimentation in a computer simulation to derive reliable expressions for the sampling error in both parameters. All simulations and analysis were performed on a SUN 3/50 computer using the MATLAB sofware package.

3.2.1. Data used in the Experiments

The first experiments used simulated, periodic data of known form so that the discrepancies between theoretical and actual sampling error could be observed in a systematic manner. Using data of this type it is possible to fully control the experiments as every input parameter, either constant or variable is known and it is possible to analyse the behaviour of sampling error with respect to them. Experiments began using data consisting of a single sinusoid, and the complexity of the data was increased until it comprised all of the elements expected of real surface data in its most complex form, i.e. multiple periodic and random characteristics. The developed algorithm was then tested upon data taken from typical engineering surfaces, collected by the Rank Taylor Hobson 'Form Talysurf'

The 'Form Talysurf' shown in photograph 5.1 is a device capable of measuring surface finish and form, it has a ball stylus of radius $397\mu m \pm 1.5\mu m$, traversing over a maximum distance of 120mm, the profile sensor uses a laser interferometric transducer. Measuring force is between 15 an 20 mN, varying less than 10% throughout its range of movement, measuring speed is in the order of 0.5mm/s (8).

The data output from the 'Form Talysurf' is not strictly 'true' surface data, similarly the data output from the CMM may not be a perfect representation of the component surface because of the following phenomena:

i) Mechanical filtering effect of the contacting stylus geometry.

ii) Quantisation interval of the machine

iii) Electrical noise

iv) Mechanical noise

v) Vibration

vi) Electrical filtering

v) Deterioration in machine calibration

3.2.1.1. Mechanical Filtering

This occurs from several sources, but it is mainly due to the fact that the penetration of the surface features is limited by the geometry of the stylus.

The consequence of this is that peaks and valleys on the component surface cannot be followed exactly, so that a distorted record of the profile is recorded (9), valleys may be represented as cusps, and the radius of curvature of some of the peaks may be exaggerated when sylii of relatively large dimension are used. Some surface characteristics may be filtered out altogehter. Where the stylus geometry is spherical, the effective profile will correspond to the E envelope system.

There will also be distortion due to the stylus 'ploughing' through surface features when the measuring force is comparatively great. The stylus may also suffer

deformation in this way.

3.2.1.2. Quantisation Interval of the Machine

This refers to the accuracy of the digitisation of each individual element of surface data, the 'count' of surface characterisics will be affected by this value, i.e. when the difference in consecutive data is less than the quatisation interval it is effectively missed. (9)

3.2.1.3. Electrical noise from the Machine

Electrical noise usually has a broad band spectrum and is random in nature, it can originate from various sources:

Johnson Noise:

This is generated by the resistive part of any source impedance and is proportional to the temperature of the source.

Shot Noise:

This noise is due to the non-smooth flow of a current around a circuit i.e. a nominally steady current has normally or Gaussianly distributed fluctuations. Johnson and Shot noise are usually in the form of broad band flat frequency spectrum noise, or 'white noise' (10).

Additionally, resistive impedances suffer from flicker or 1/f noise. They are subject to fluctuations in resistance depending on the material used in their manufacture. External sources of electrical noise also contribute to the problem.

3.2.1.4. Mechanical Noise from the Machine

This can result from friction and deformity in machine components, particularly bearings.

3.2.1.5. Electrical Filtering

Electrical filtering is used to extract the measured signal from the total recorded signal, which is done by removing certain frequencies from the recorded values, this is dangerous as some of the surface data may be compromised.

3.3. Experimental Procedure

The derived values for the errors in slope and intercept were calculated according to the expressions for *eb* and *ebo* using a between 2 and 50 data points inclusive spaced at at particular sampling interval (5mm to begin with). The result was plotted against the number of data points used in the calculation. The actual sampling error was calculated by using the same number and distribution of data to calculate the appropriate parameter, calculating the same parameter using approx 500 data points distributed over the same profile length, the latter estimate was then subtracted from the former to give the sampling error. The result was plotted in a similar manner to the derived values.

The resulting plots were compared and the discrepancy between them expressed in terms of a factor of multiplication whereby the application of this factor to the derived expression for sampling error would result in the that plot equalling/enveloping the plot of actual sampling error. These discrepancies were plotted against every concievable parameter, including mean amplitude of data, mean value of autocorrelation function and sum of residuals from estimated line, none of which offered a direct relationship. Plotting the discrepancies against the reciprocal of the derived value of sampling error that produced the discrepancy in the first instance produced more promising results. The clarity of the relationship improved slightly when the natural log of the reciprocal derived expression value was taken, but produced the most distinct relationship was seen by plotting the double natural log against the factor of multiplication. The result of this is seen in graph 5.1 a

From this relationship it is now possible to predict the factor of multiplication that

would, as far as possible, quantify the difference between the actual and theoretical samping error in b and b_o respectively:

$$cb_1 = = \frac{((ln(ln(eb^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \tag{5.3}$$

$$cbo_1 = \frac{((ln(ln(ebo^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \tag{5.4}$$

Then the sampling error incurred in the estimation of slope by any number and distribution of data may be expressed as

$$eb_2 = eb\left[\frac{((ln(ln(eb^{-1}))) - 2.1491}{3.316 \times 10^{-3}}\right] \tag{5.5}$$

and for intercept with the Y axis

$$ebo_2 = ebo\left[\frac{((ln(ln(ebo^{-1}))) - 2.1491}{3.316 \times 10^{-3}}\right] \tag{5.6}$$

This relationship remained valid until the magnitue of eb and ebo fell below particular levels, i.e. $eb \leq 10^{-7}$ and $ebo \leq 10^{-5}$. It was then found that when these criteria were met, an additional term was required, such that if

$$eb \leq 10^{-7}$$

$$eb_3 = eb\left[\frac{((ln(ln(eb^{-1}))) - 2.1491}{3.316 \times 10^{-3}}\left[5\,(2.30258\,\ln(eb^{-1}) + 5)(2.30258\,\ln(eb^{-1}) + 6)\right]\right] \tag{5.7}$$

and

$$ebo_3 = ebo\left[\frac{((ln(ln(ebo^{-1}))) - 2.1491}{3.316 \times 10^{-3}}\left[10\,(2.30258\,\ln(ebo^{-1}) + 5)(2.30258\,\ln(ebo^{-1}) + 4)\right]\right] \tag{5.8}$$

Errors were then calculated using data equispaced at different sampling intervals which were altered systematically. It was subsequently discovered that a further term was needed, it was required to pre-multiply the expressions for sampling errors by (5/dx).

Two sets of result were calculated, comparing the theoretical sampling errors calculated as per the above algorithms with the actual sampling errors. For the first set of results, the autocorrelation function was calculated using all of the available surface data, either from the simulation or the Form Talysurf output. In each case approximately 1500 data points were used.

The simulated data used in the first set of experiments is described in table 5.1, and where real surface data has been used a profile of this data precedes the results that it produces.

Simulated data (Set 1)		
Surface	Surface Parameters	Sampling Interval(mm)
1	0.2000 sin ($2\pi 0.0700x$) 0.1000 sin ($2\pi 0.2800x$)	5
2	0.0060 sin($2\pi 0.0633x + \pi/2$) 0.0300 sin ($2\pi 0.3400x + \pi/12$) 0.0081 sin ($2\pi 4.6620x - \pi/16$ 0.0010 sin ($2\pi 0.0223x - \pi/8$ 0.0044 sin ($2\pi 1.1300x + \pi/66$	5
3	0.0090 sin ($2\pi 0.0080x + \pi/7$) 0.0800 sin ($2\pi 0.4800x + \pi/19$) 0.0081 sin ($2\pi 1.6640x - \pi/40$) 0.0030 sin ($2\pi 0.0442x - \pi/8$) 0.0074 sin ($2\pi 2.1200x + \pi$) rand, std=0.0143	10
4	0.0090 sin ($2\pi 0.0080x + \pi/7$) 0.0800 sin ($2\pi 0.0600x + \pi/19$) 0.0081 sin ($2\pi 1.6640x - \pi/40$) 0.0030 sin ($2\pi 0.0442x - \pi/8$) 0.0074 sin ($2\pi 2.1200x + \pi$) rand, std=0.0143	7

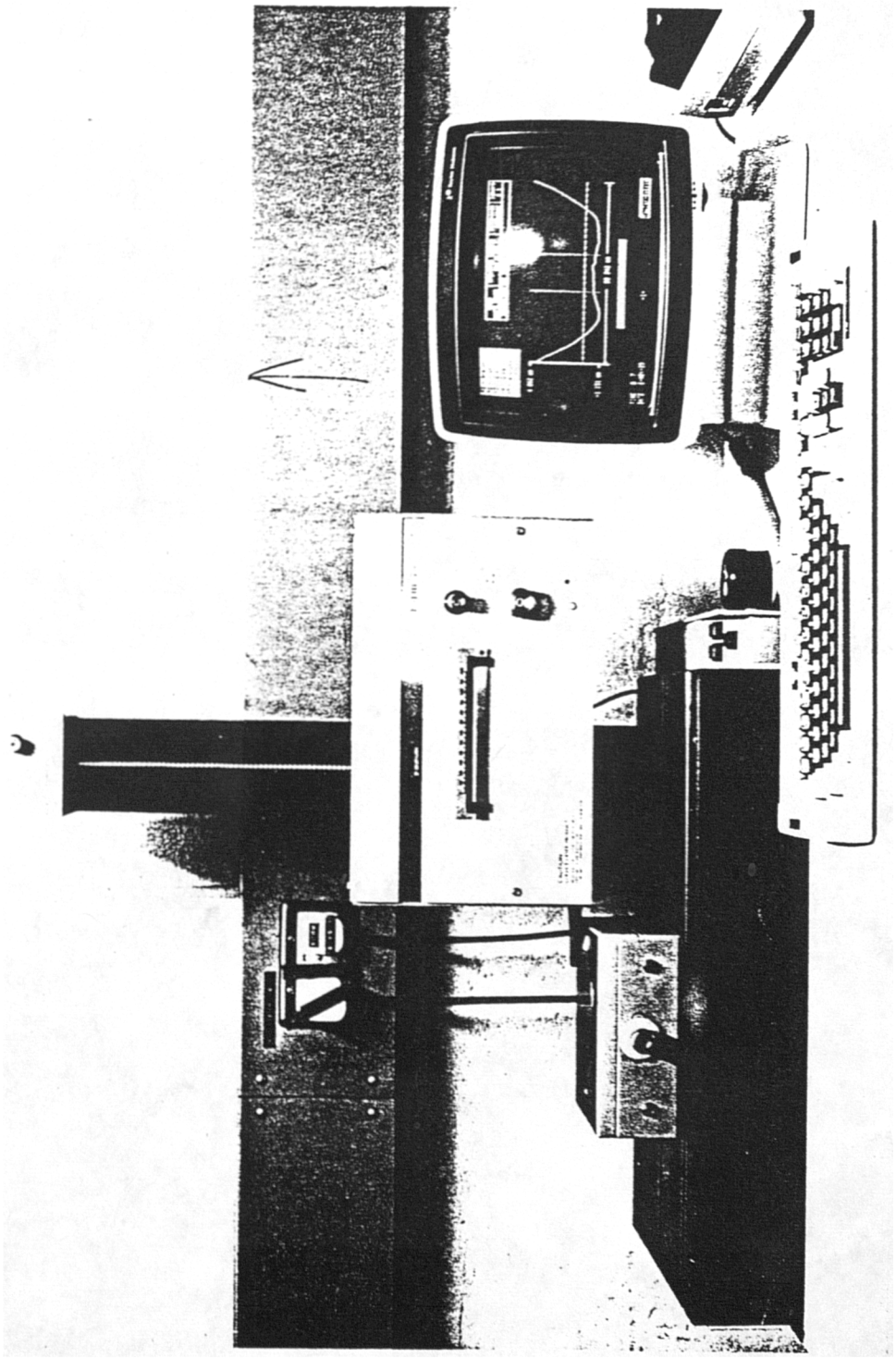
Table 5.1

The results of this set of experiments are displayed in graphs 5.2 to 5.17.

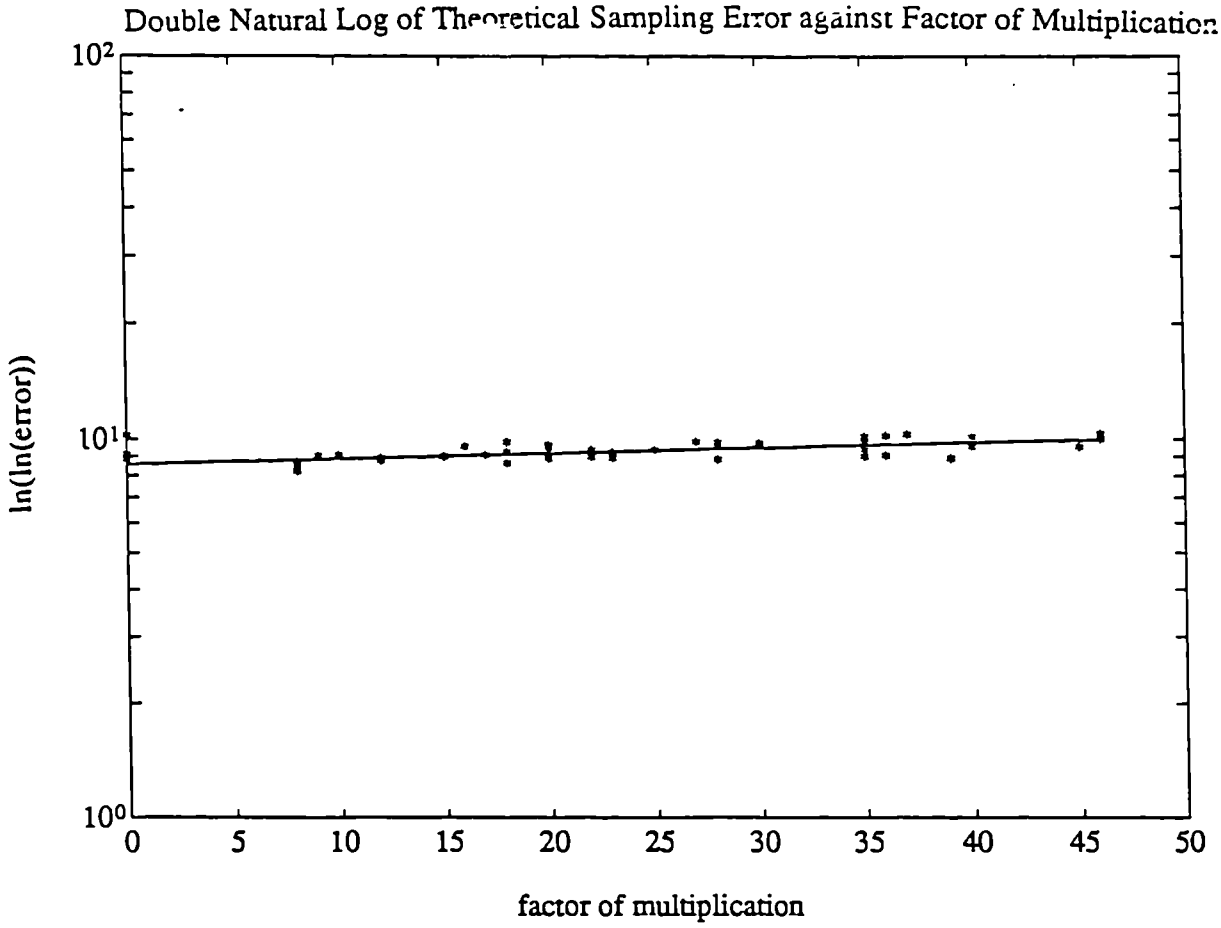
Key

———————:Theoretical Sampling Error

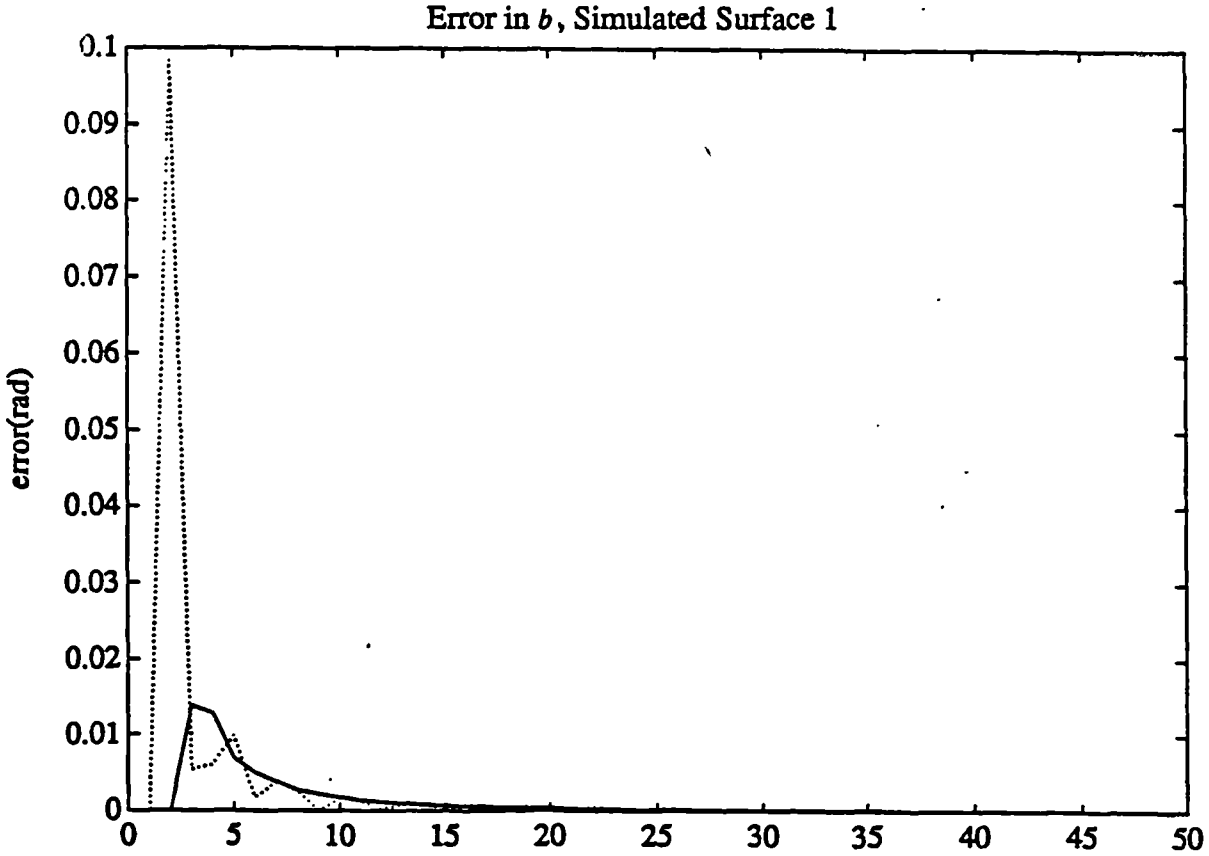
.....:Actual Sampling Error



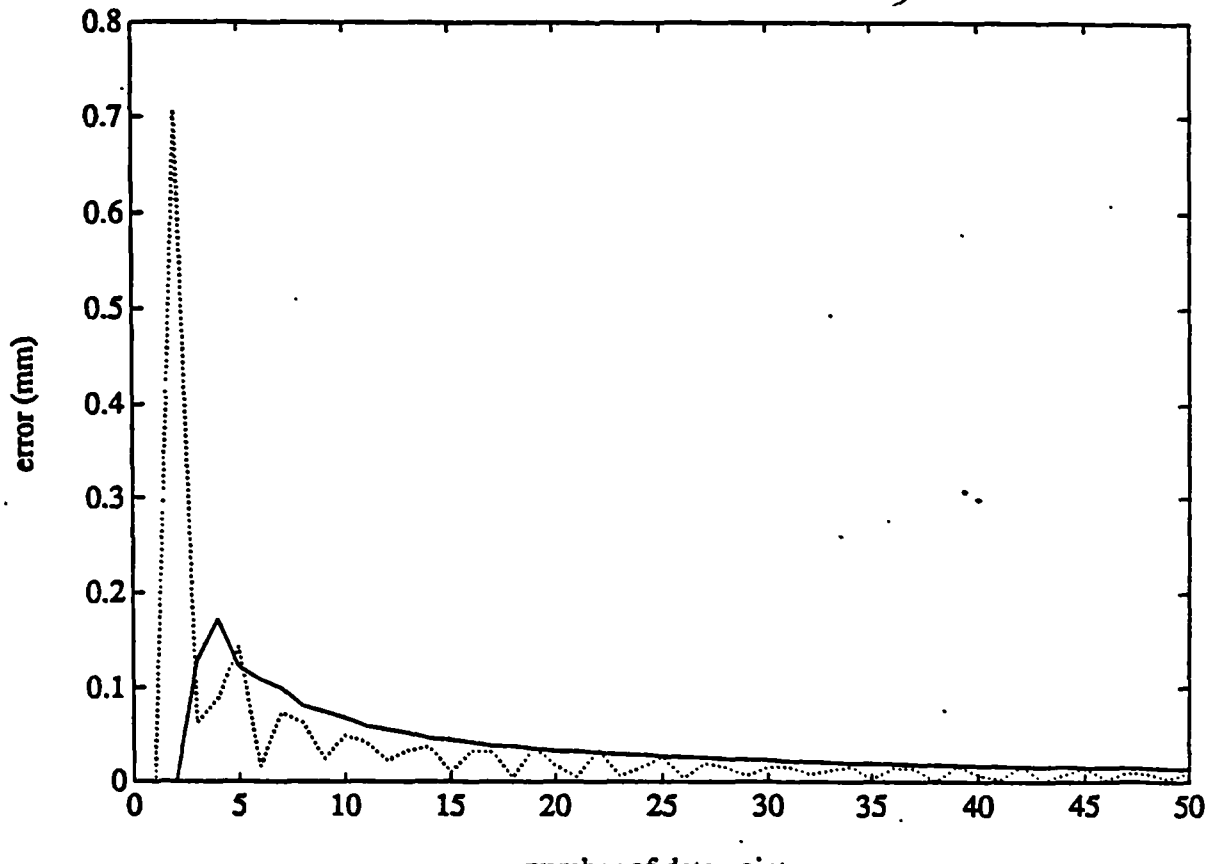
Photograph 5.1


Double Natural Log of Theoretical Sampling Error against Factor of Multiplication
ln(ln(error))
10^2
10^1
10^0
0
5
10
15
20
25
30
35
40
45
50
factor of multiplication

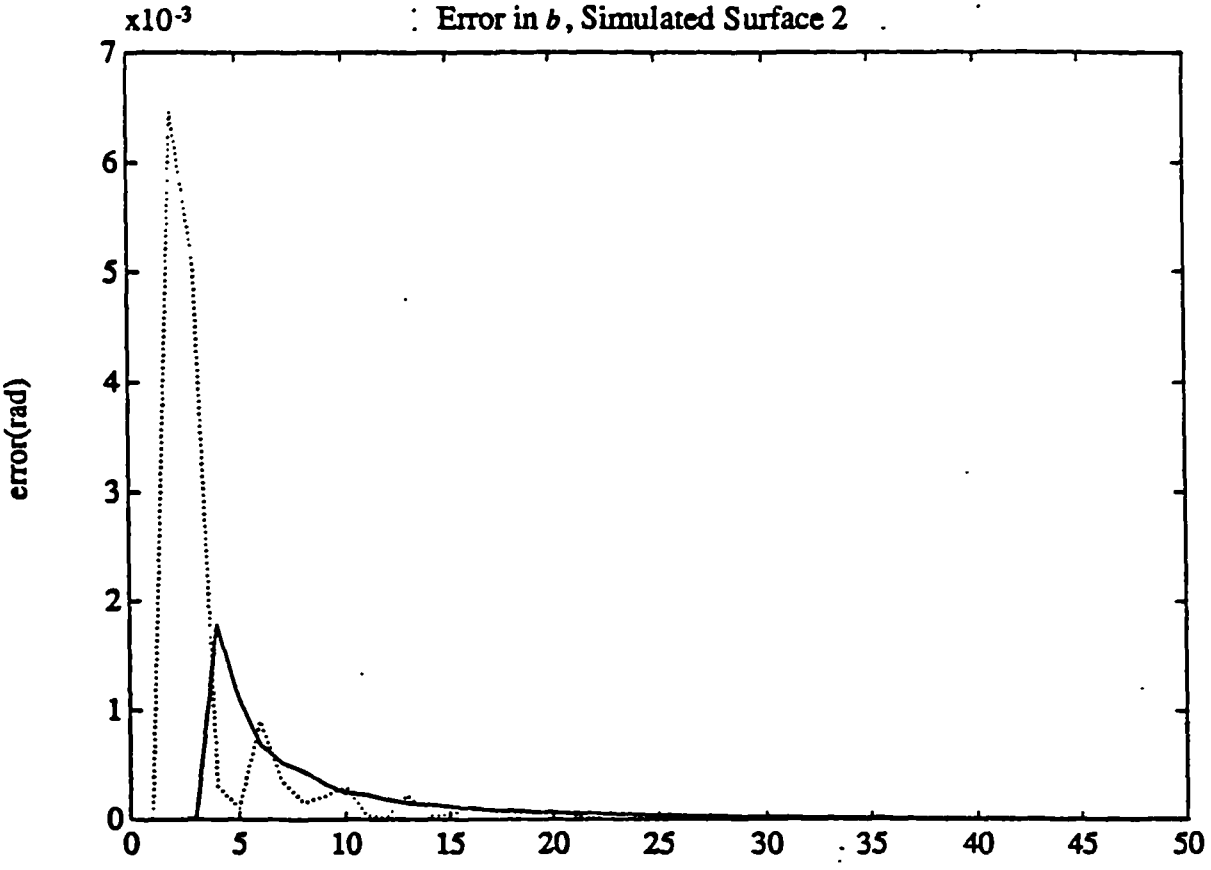
Graph 5.1 (a)



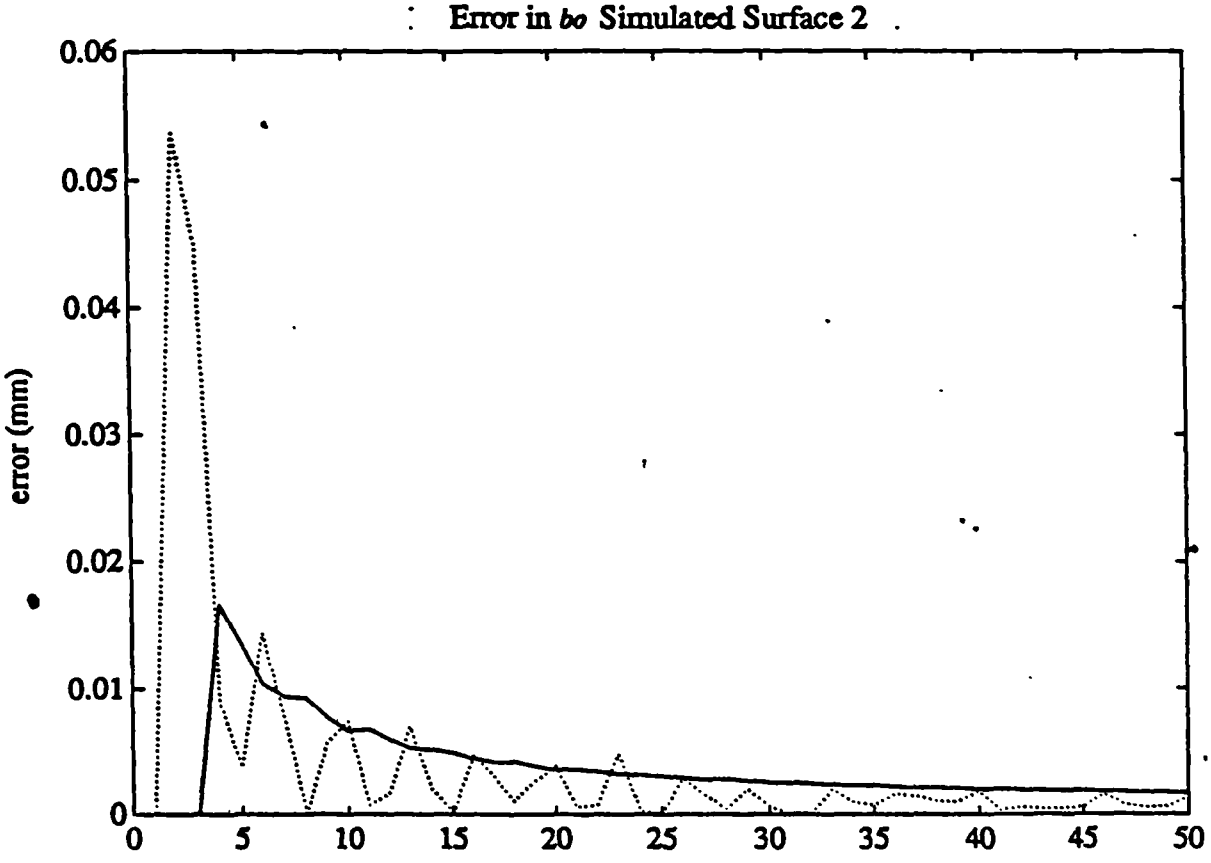
Graph 5.2 (a)



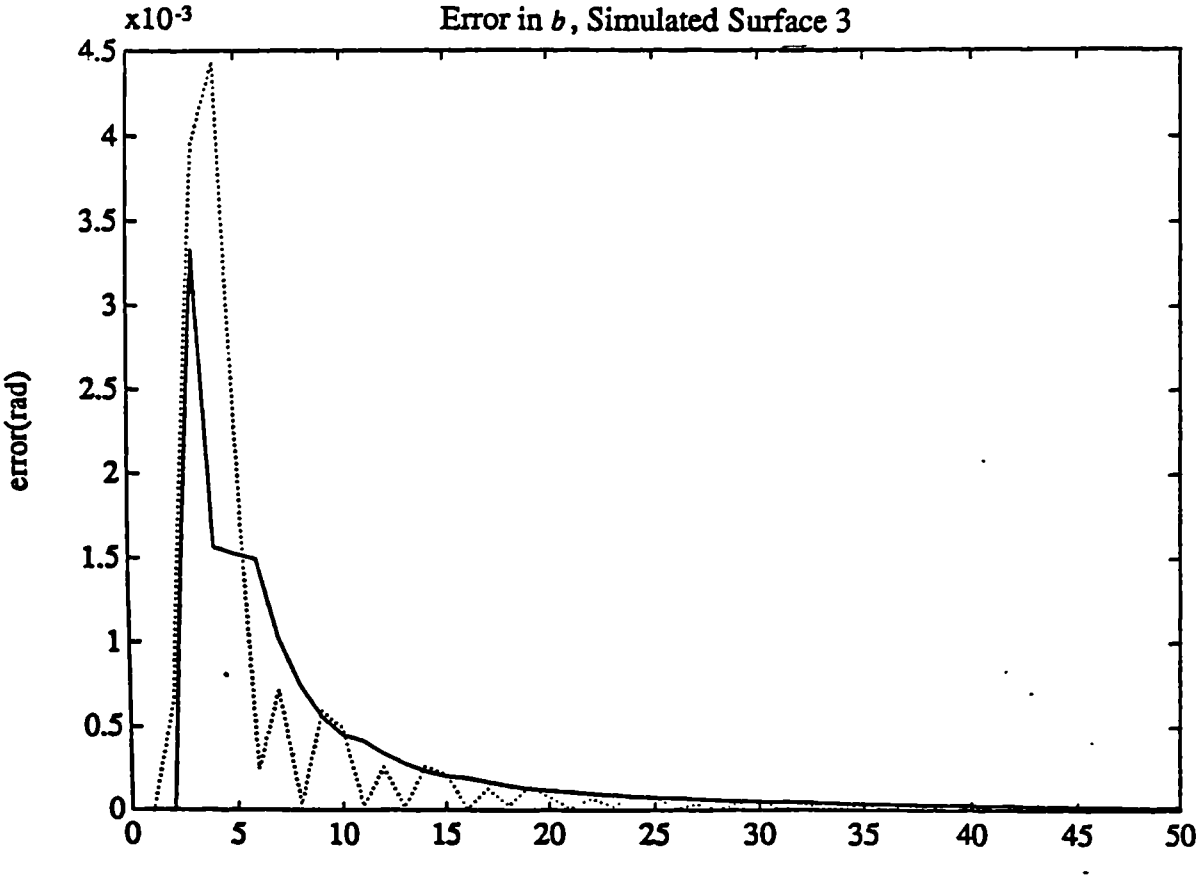
Graph 5.2 (b)



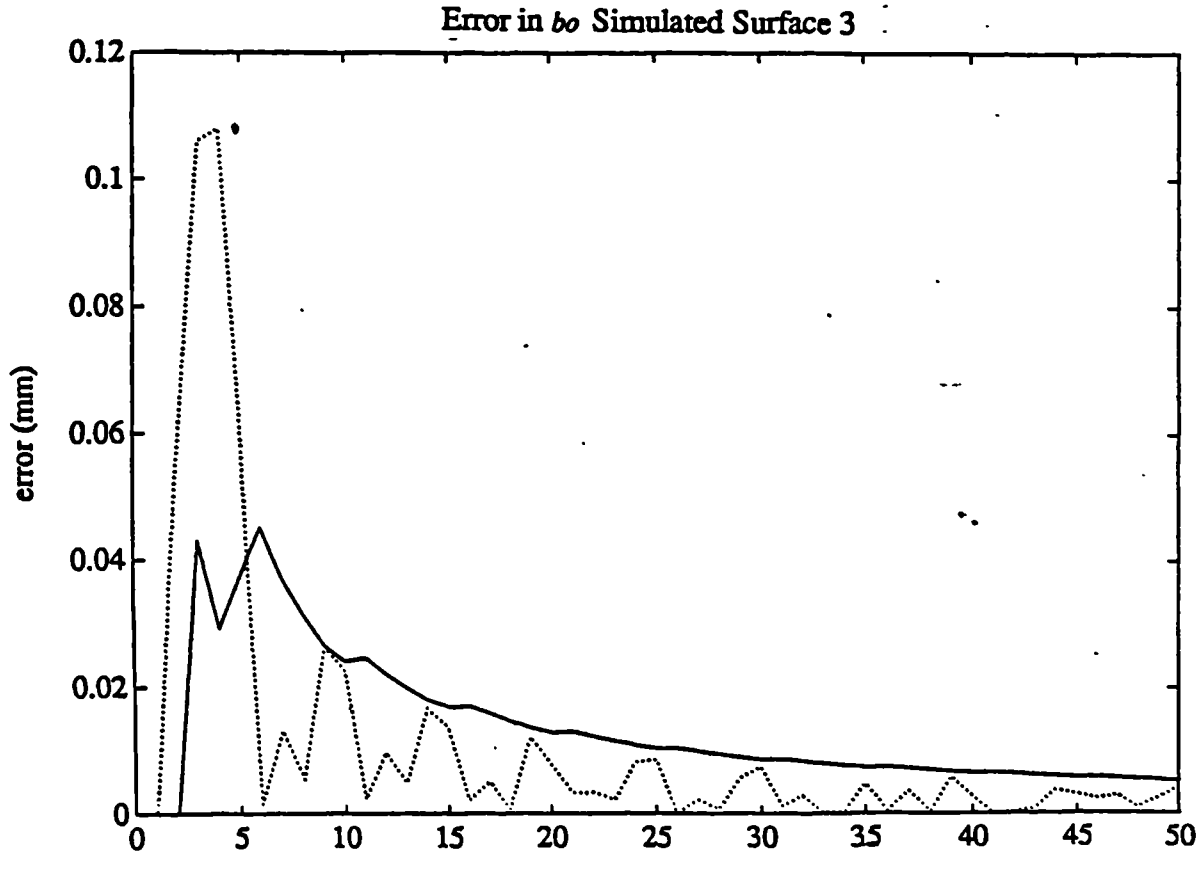
Graph 5.3 (a)



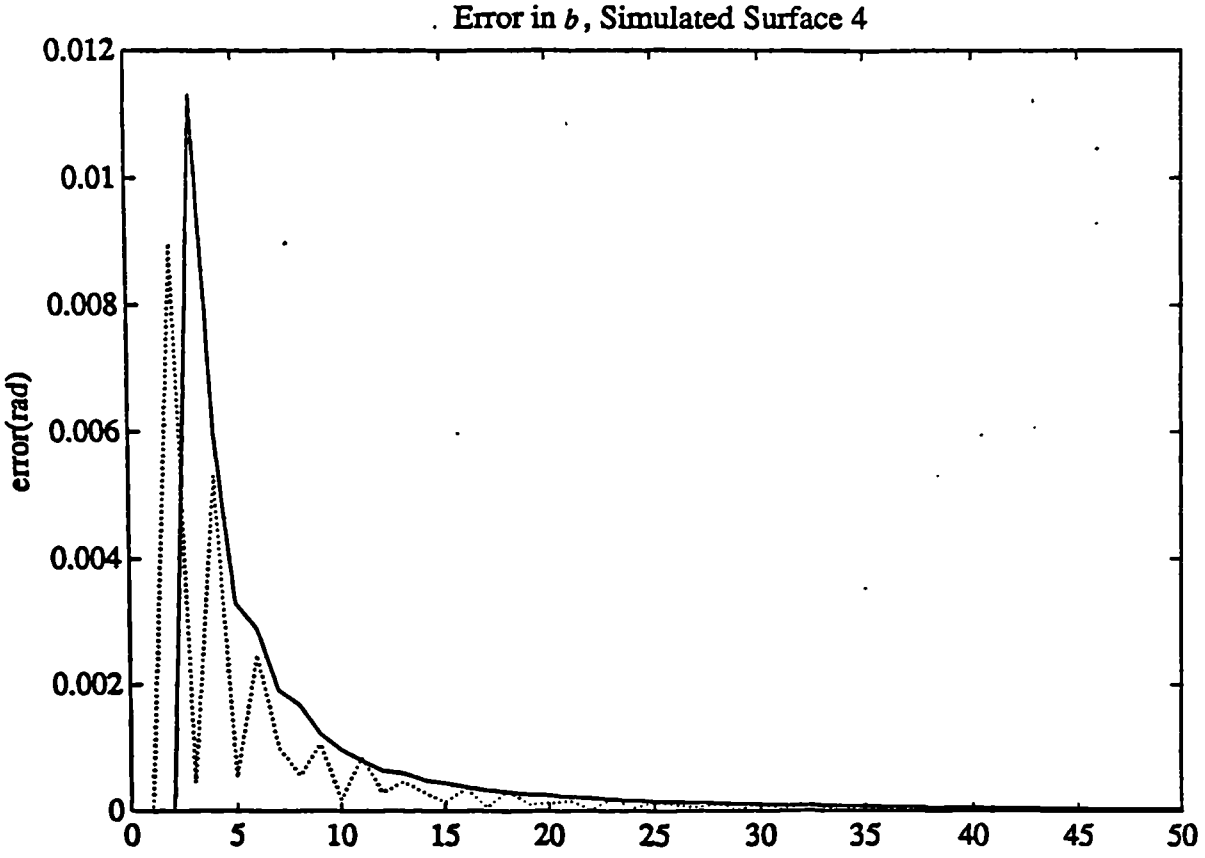
Graph 5.3 (b)



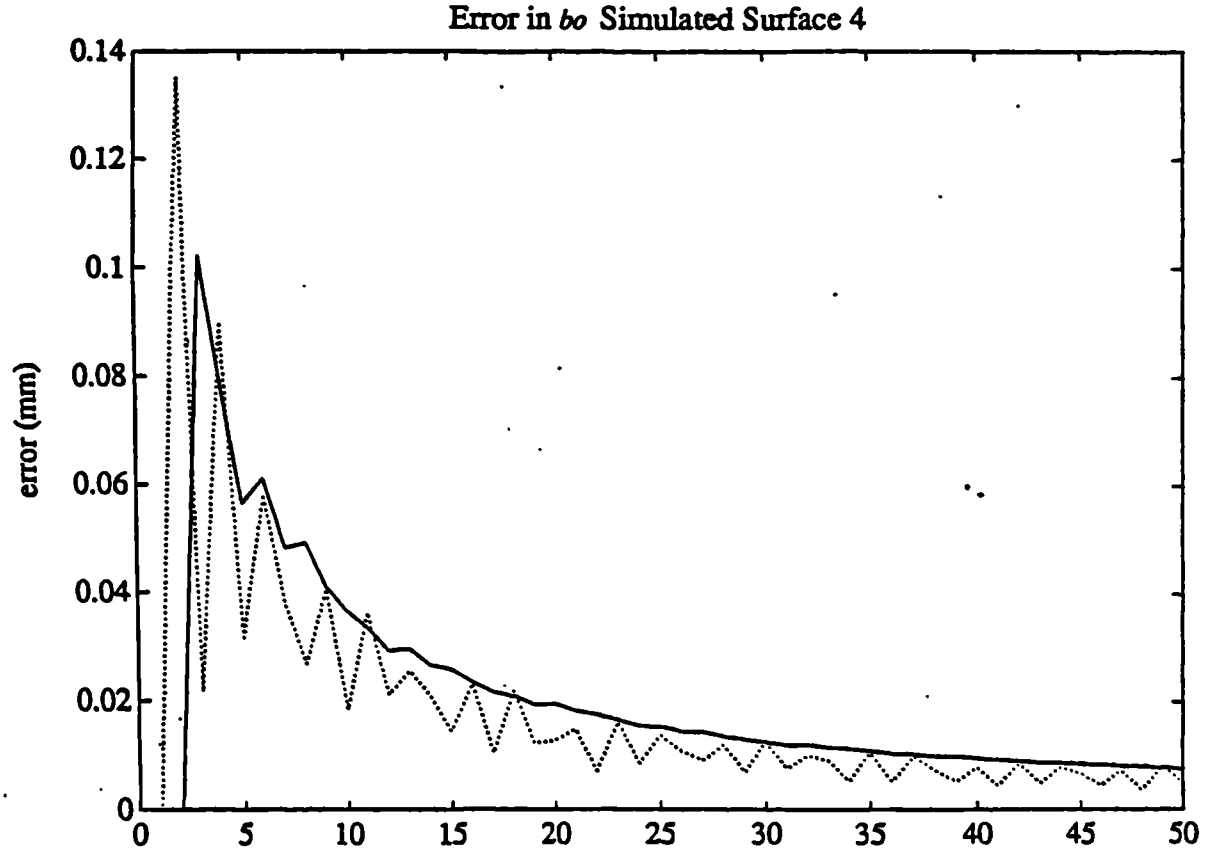
Graph 5.4 (a)



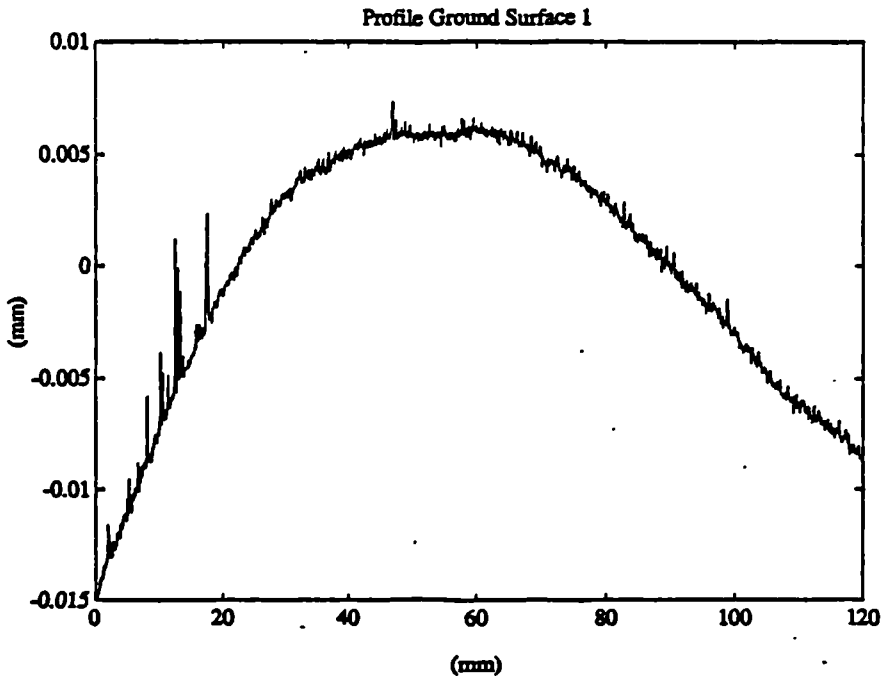
Graph 5.4 (b)

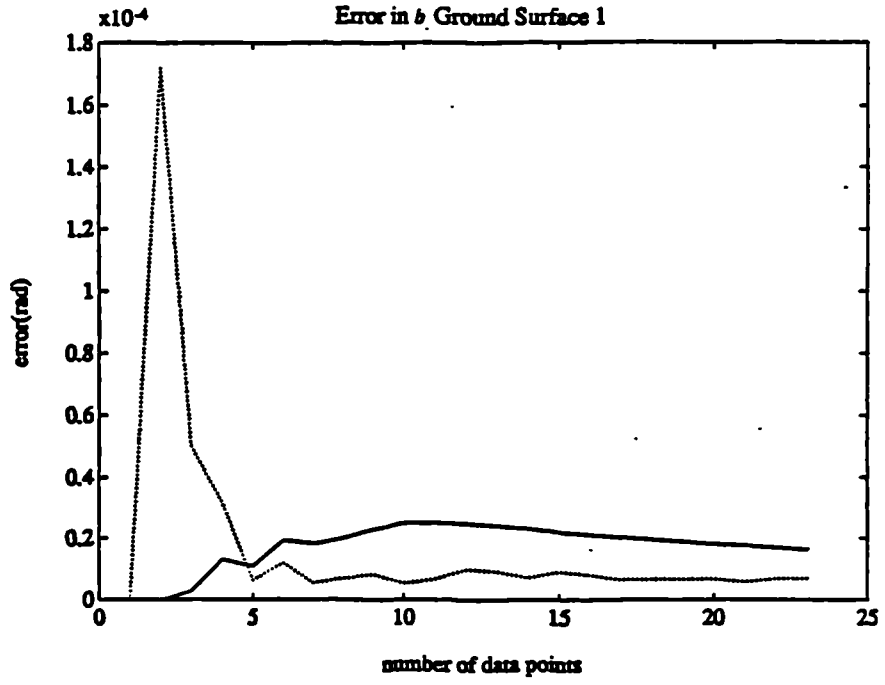


Graph 5.5 (a)

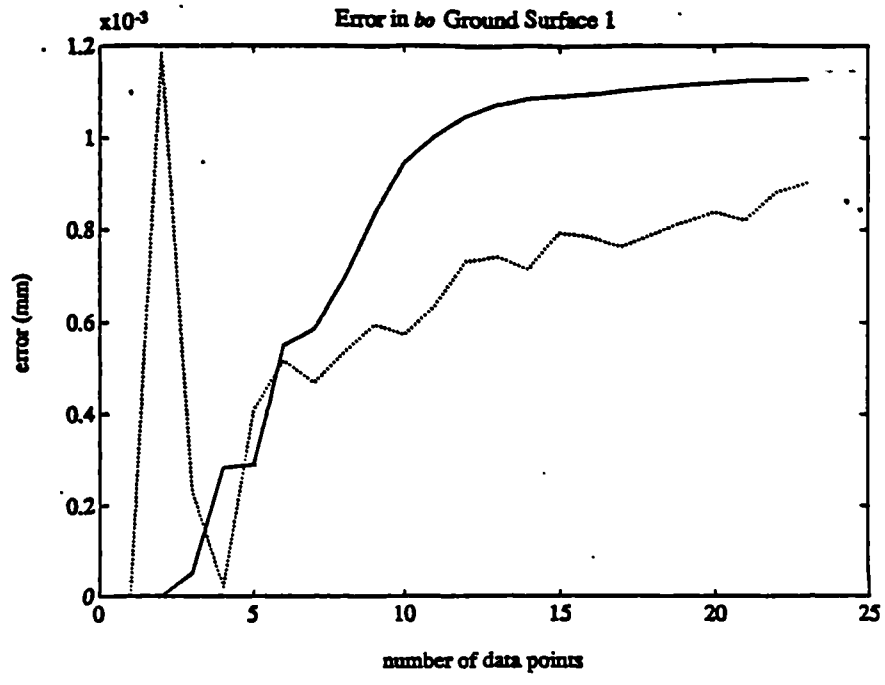


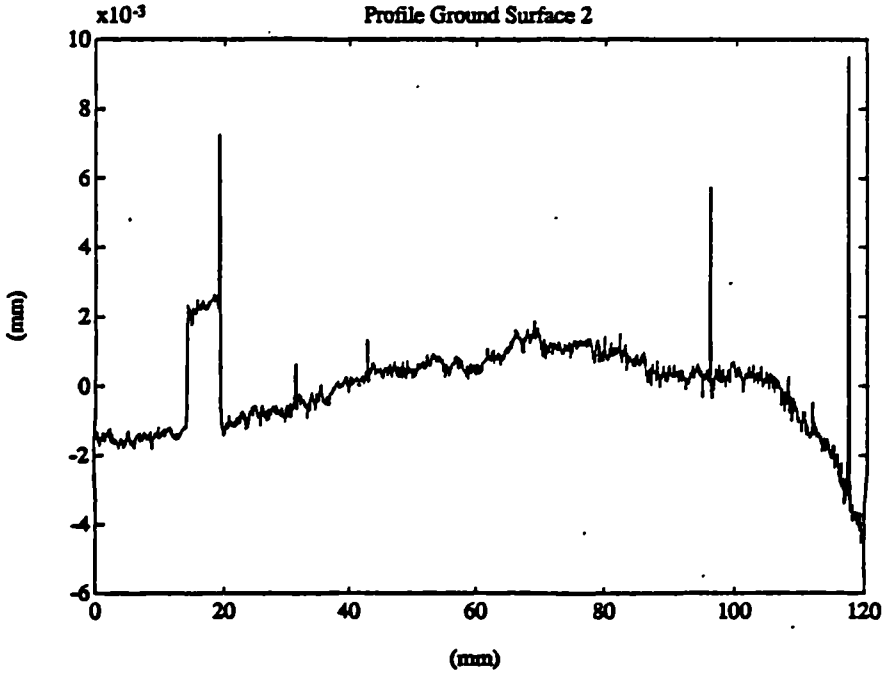
Graph 5.5 (b)

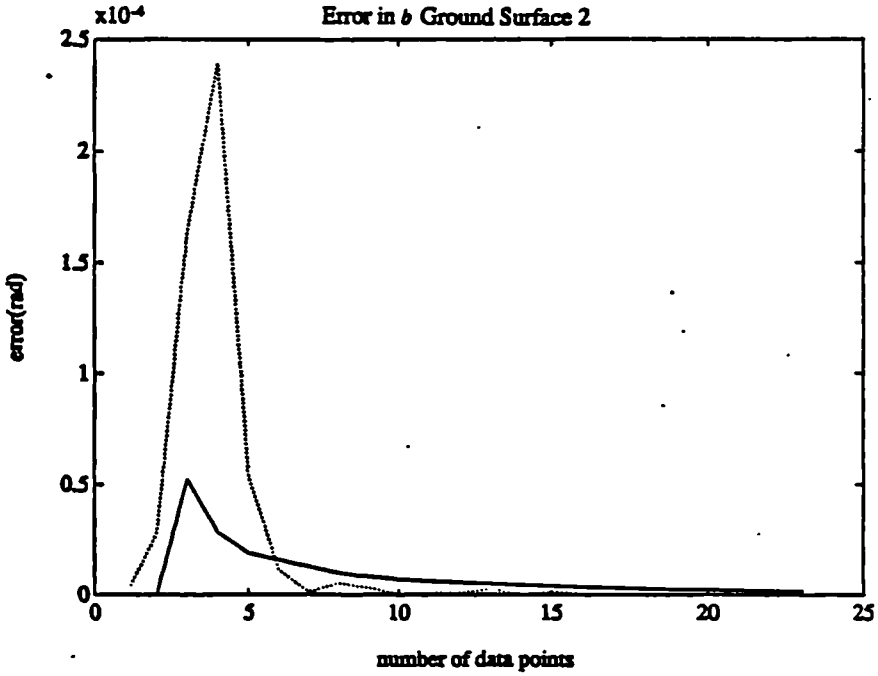

Profile Ground Surface 1
0.01
0.005
0
-0.005
-0.01
-0.015
(mm)
0
20
40
60
80
100
120
(mm)


x10-4
Error in b Ground Surface 1
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
error(rad)
0
5
10
15
20
25
number of data points

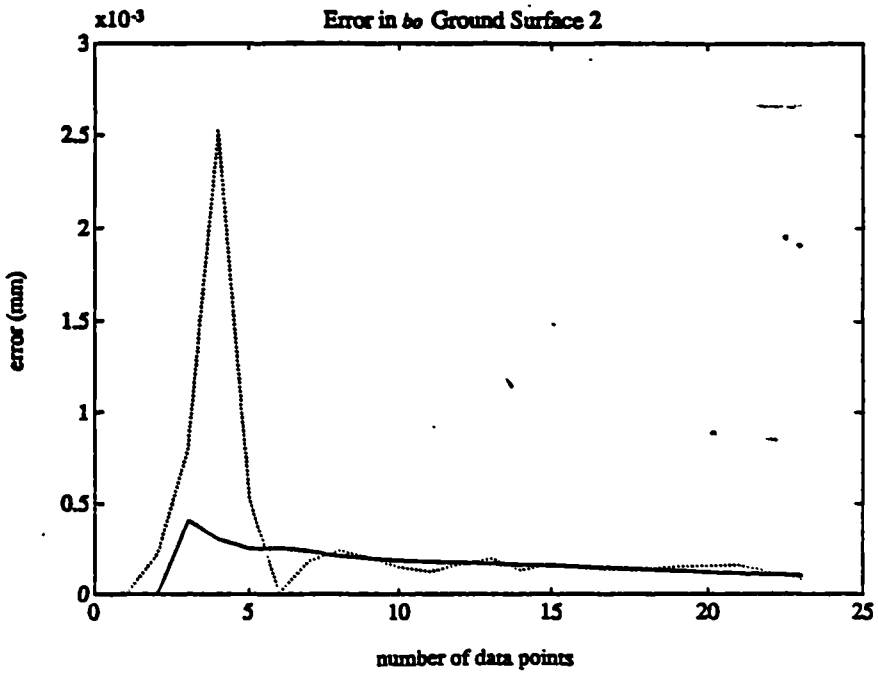
Graph 5.6 (a)


x10-3
Error in bo Ground Surface 1
1.2
1
0.8
0.6
0.4
0.2
0
error (mm)
0
5
10
15
20
25
number of data points

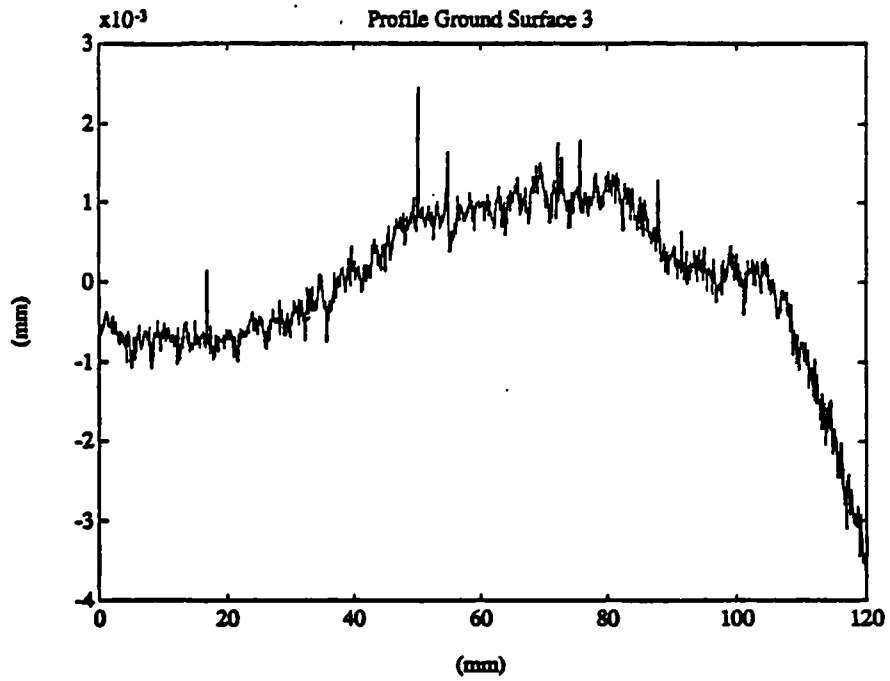


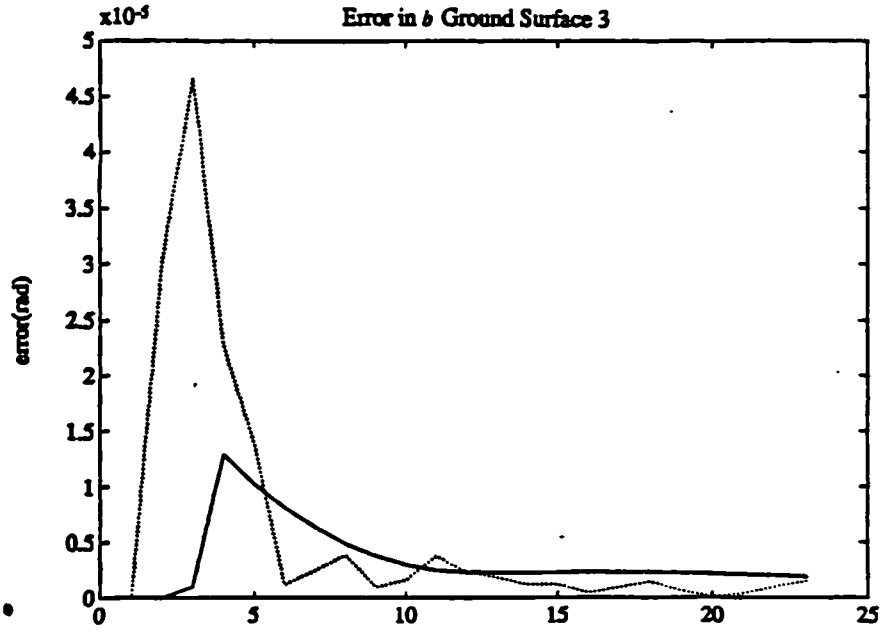


Graph 5.7 (a)

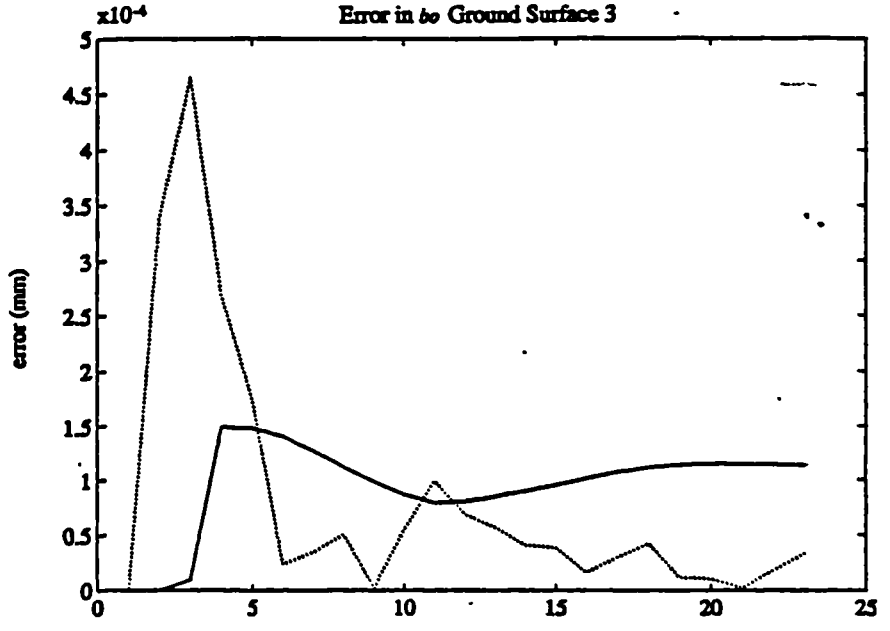


Graph 5.7 (b)

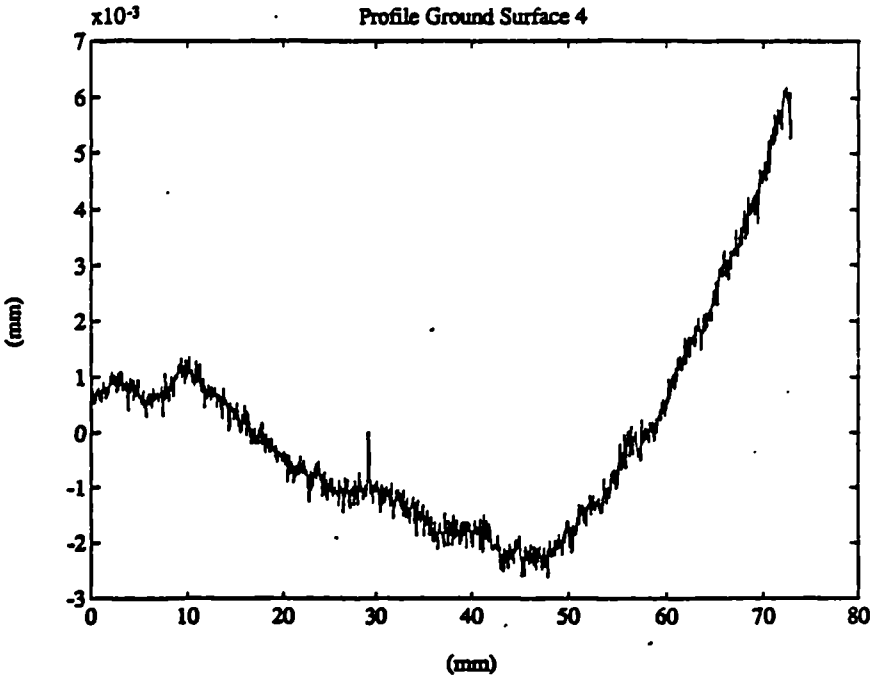


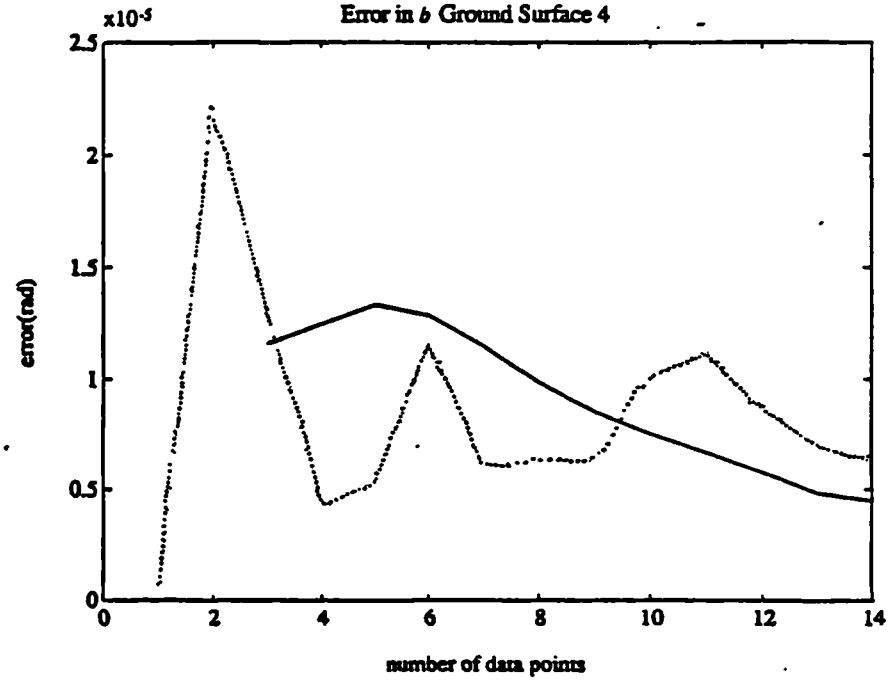


Graph 5.8 (a)

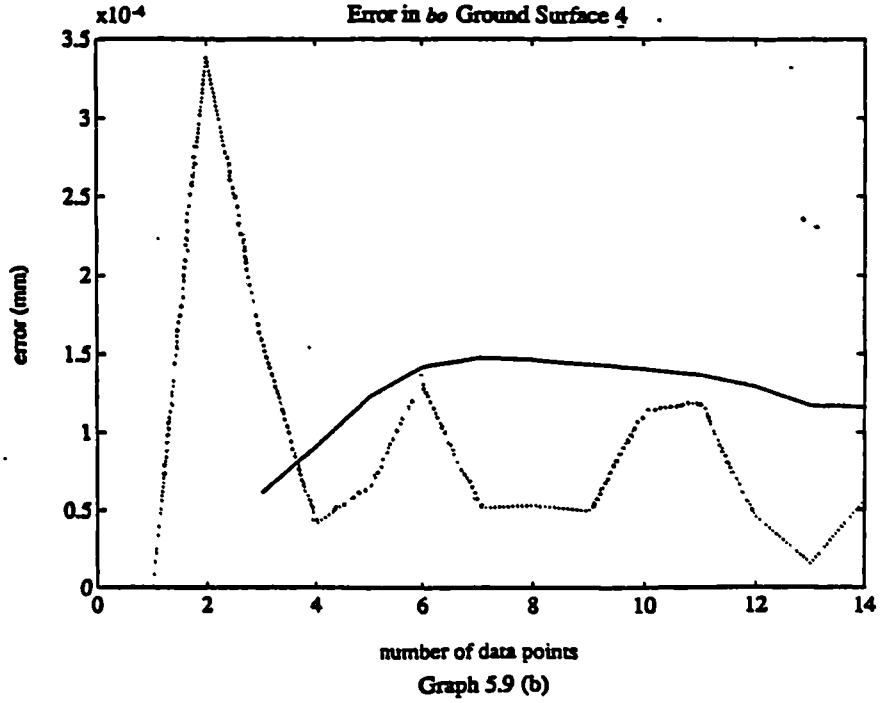


Graph 5.8 (b)

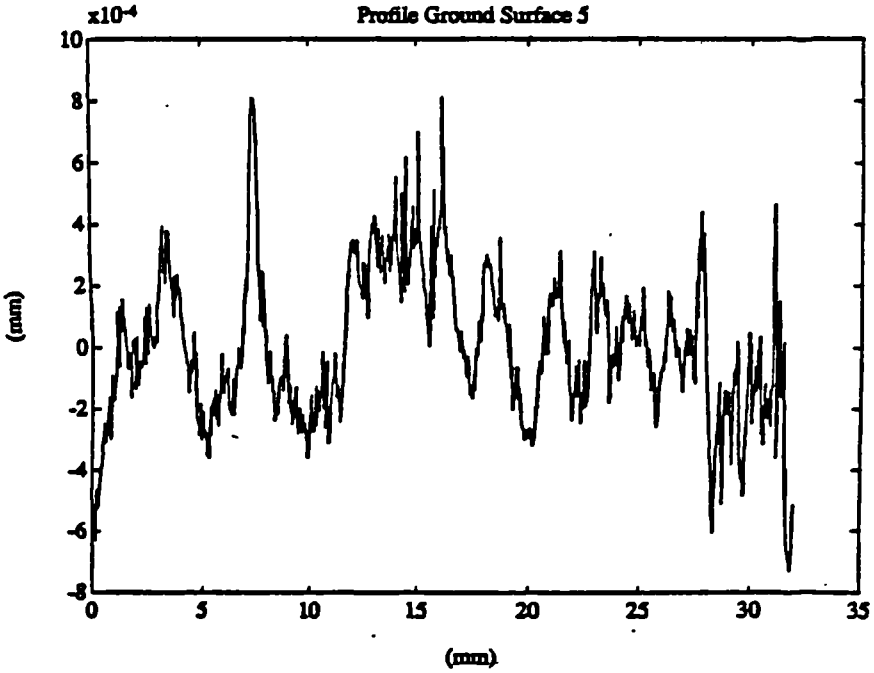


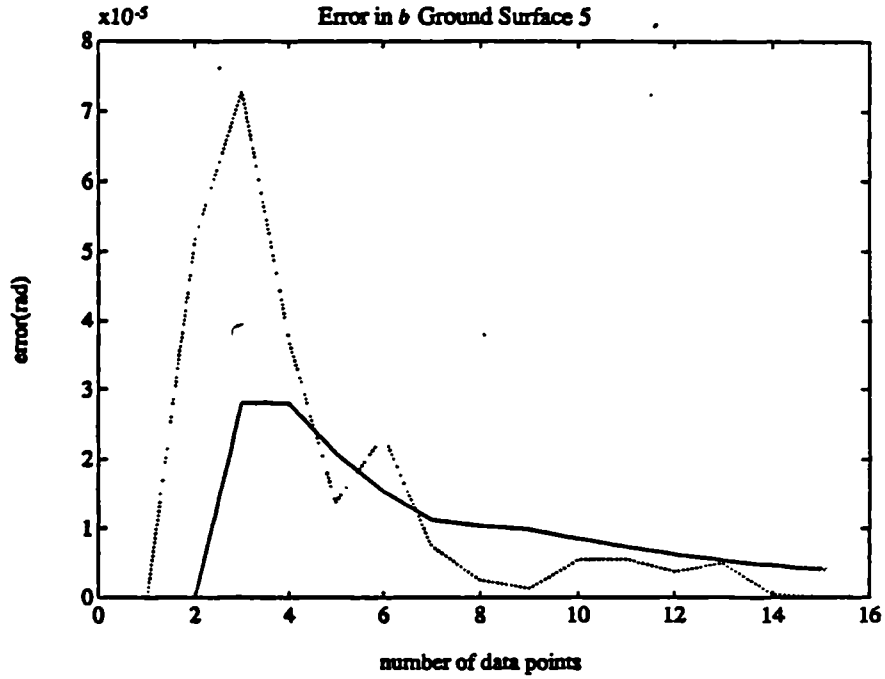


Graph 5.9 (a)

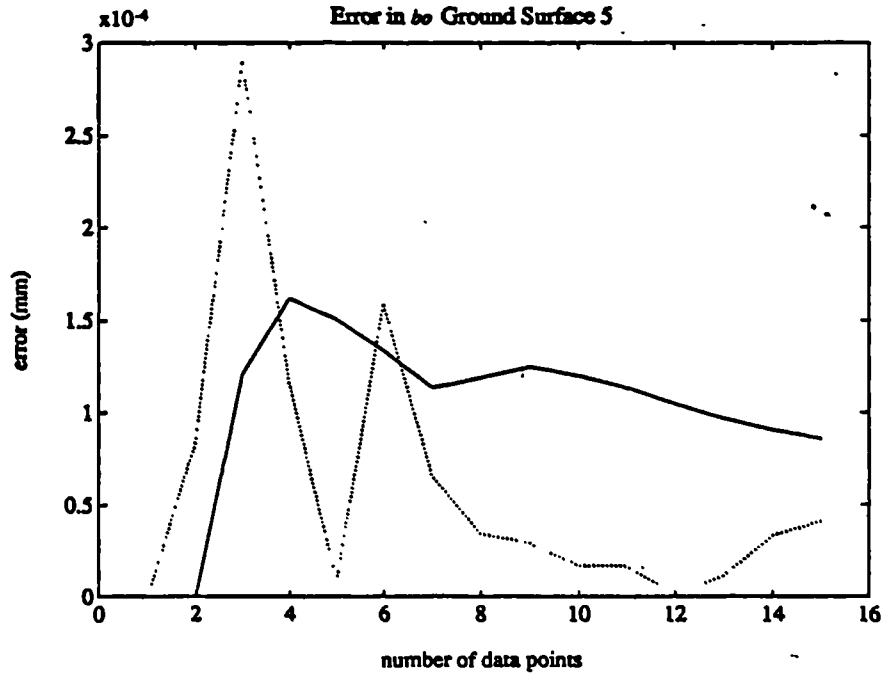


Graph 5.9 (b)

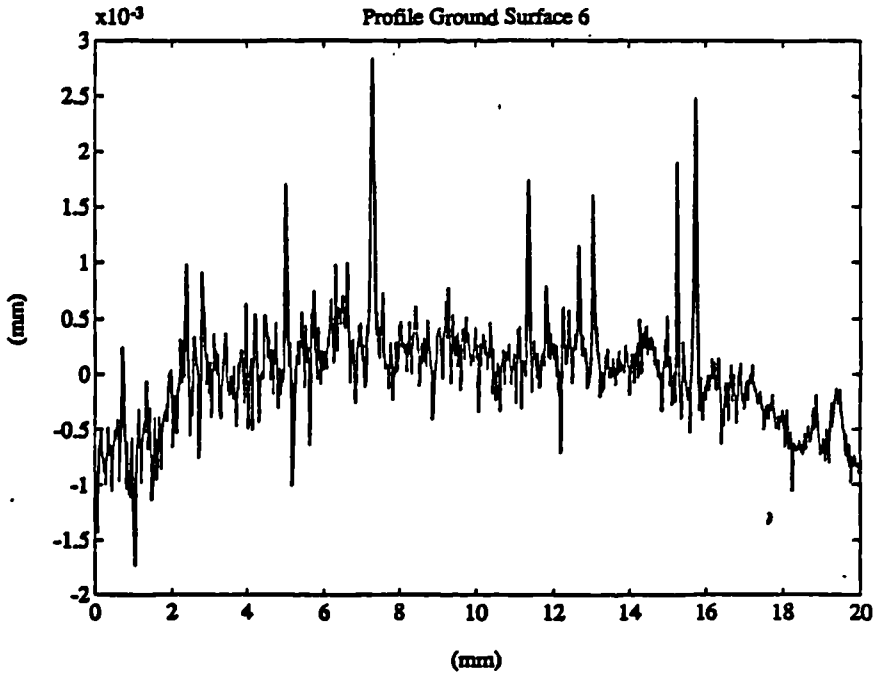


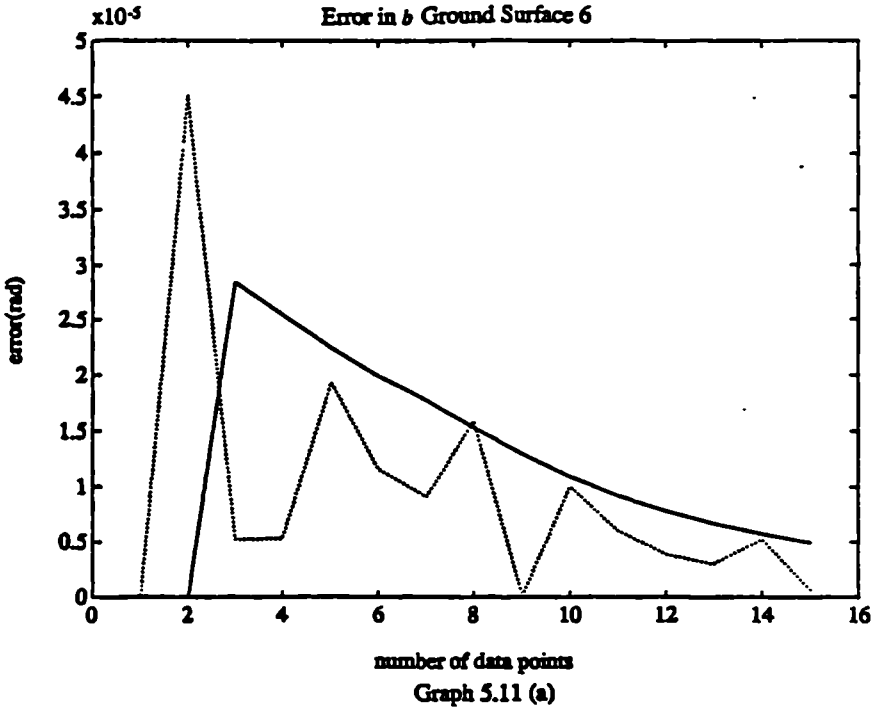


Graph 5.10 (a)

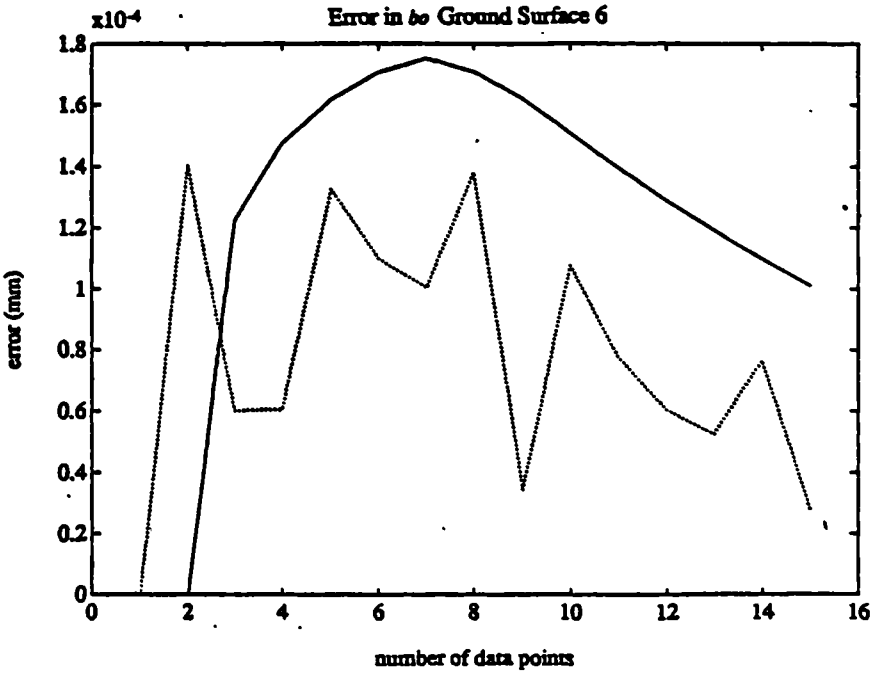


Graph 5.10 (b)

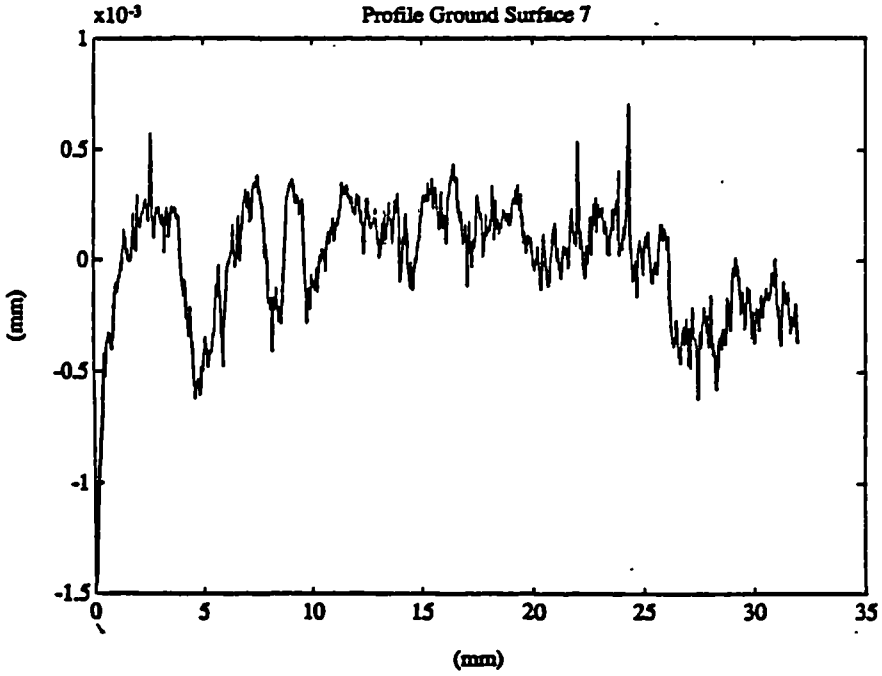


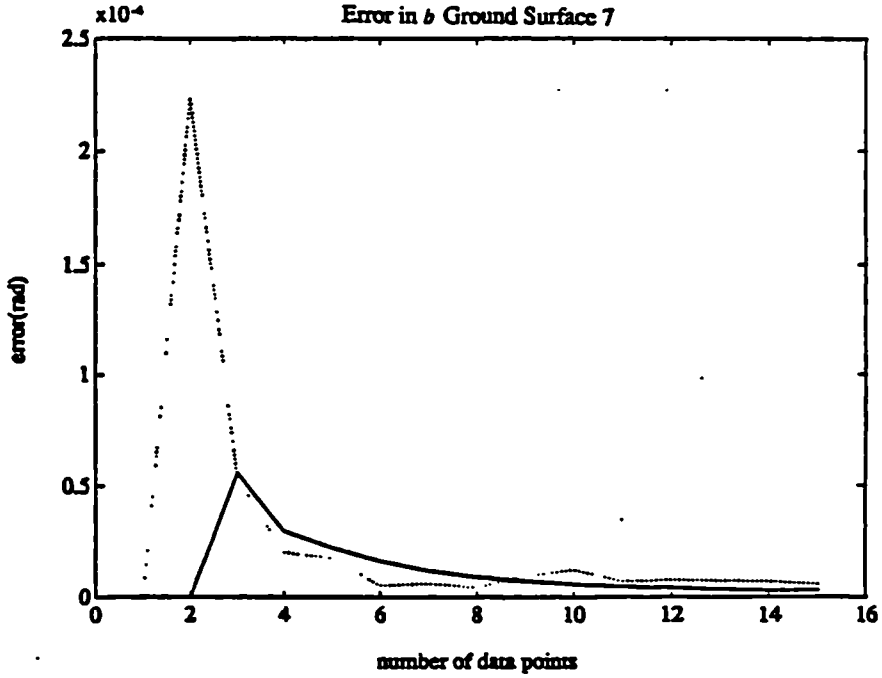


Graph 5.11 (a)

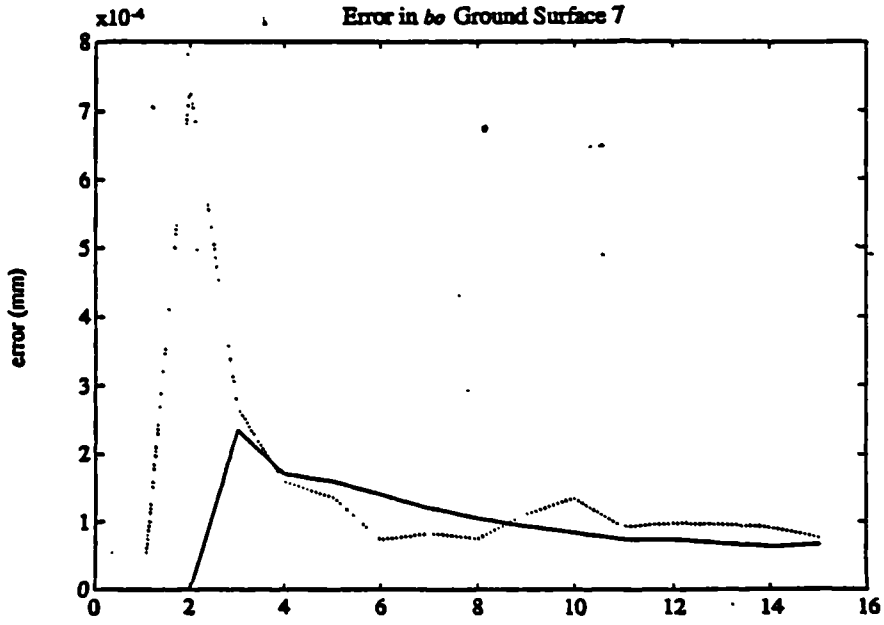


Graph 5.11 (b)

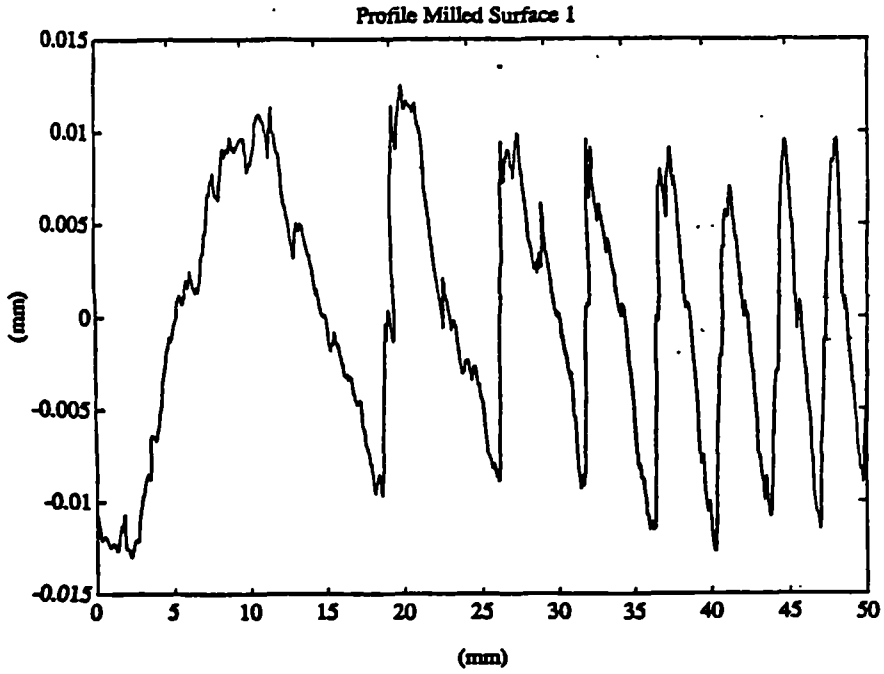


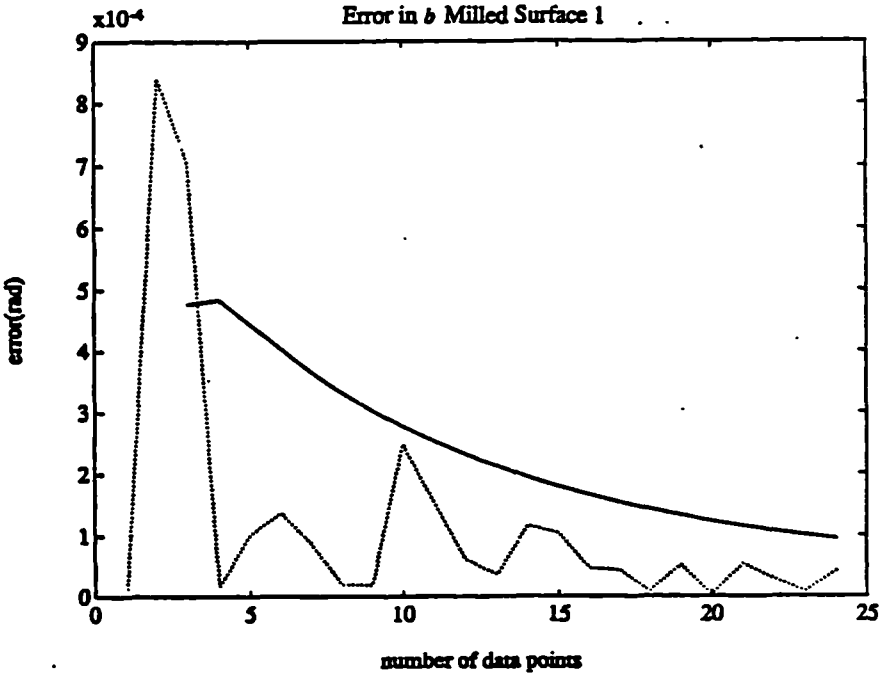


Graph 5.12 (a)

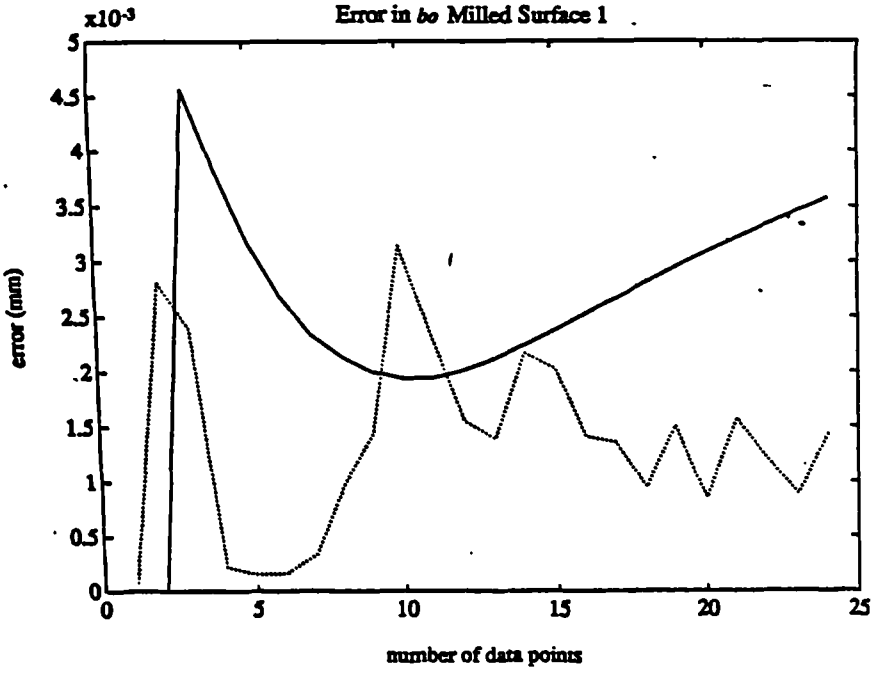


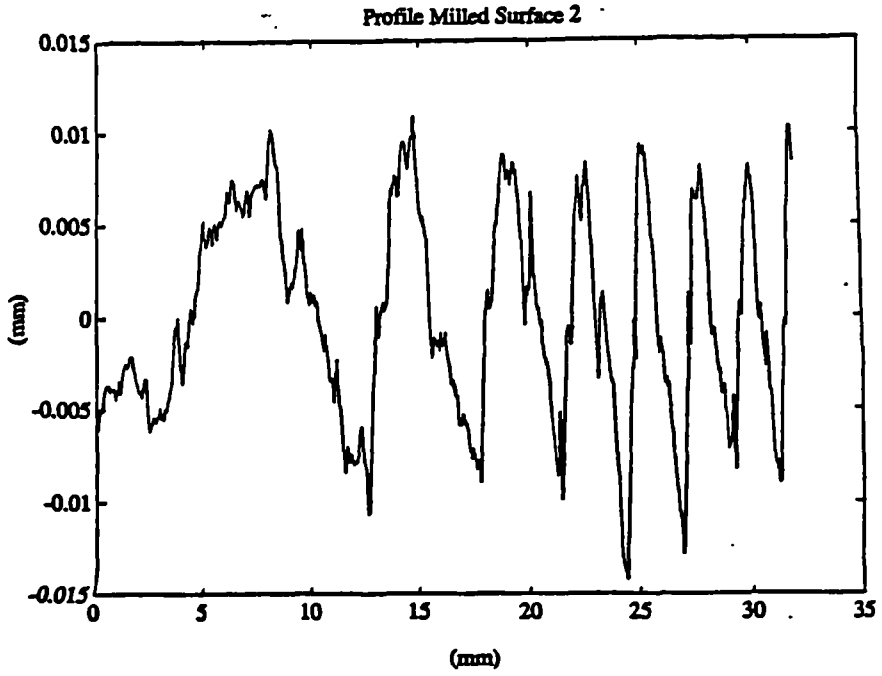
Graph 5.12 (b)

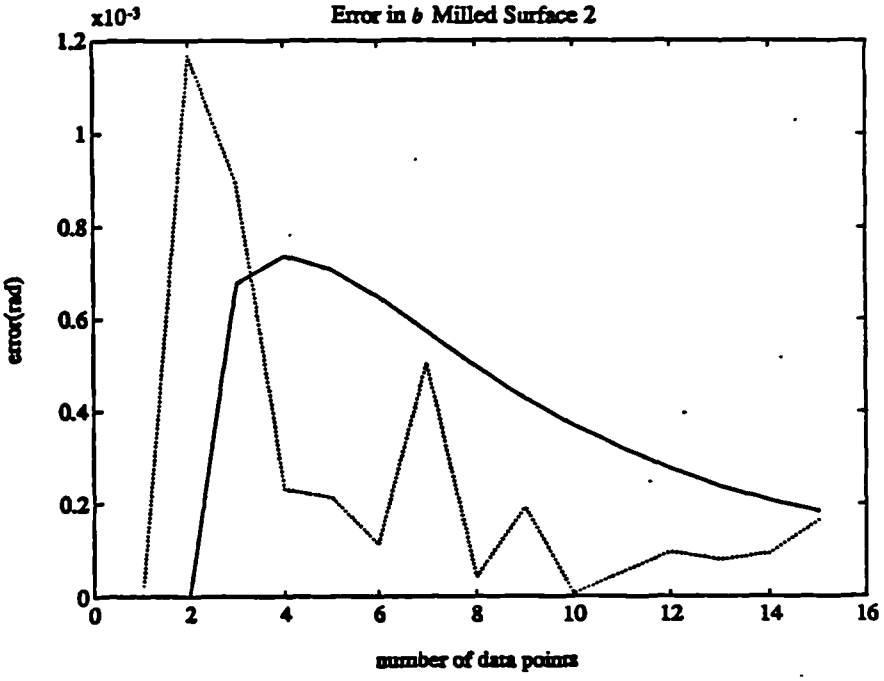

Profile Milled Surface 1
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
(mm)
0
5
10
15
20
25
30
35
40
45
50
(mm)


x10⁻⁴
Error in b Milled Surface 1
error(rad)
0
1
2
3
4
5
6
7
8
9
0
5
10
15
20
25
number of data points

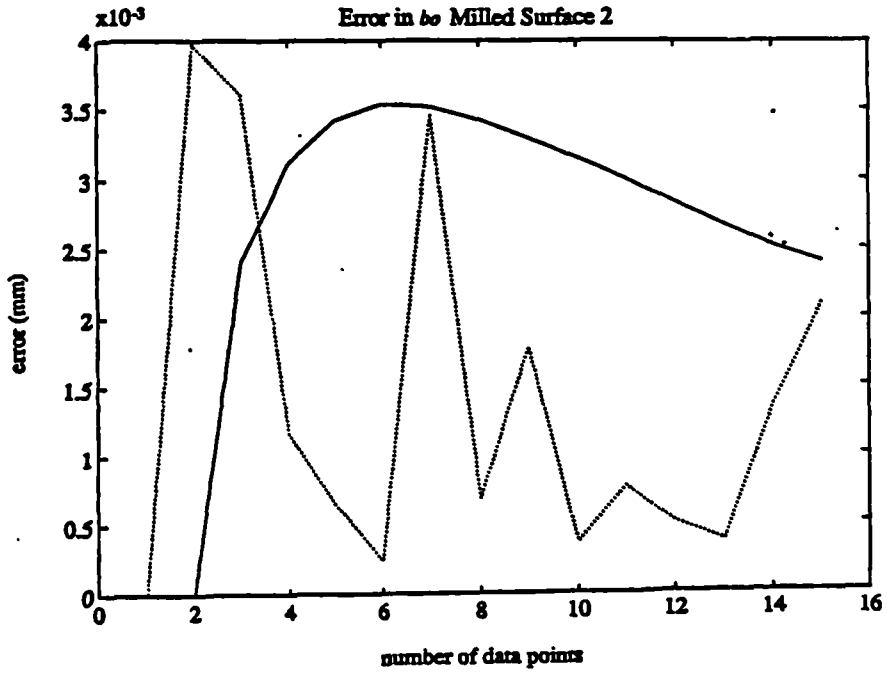
Graph 5.13 (a)


x10⁻³
Error in bo Milled Surface 1
error (mm)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
5
10
15
20
25
number of data points

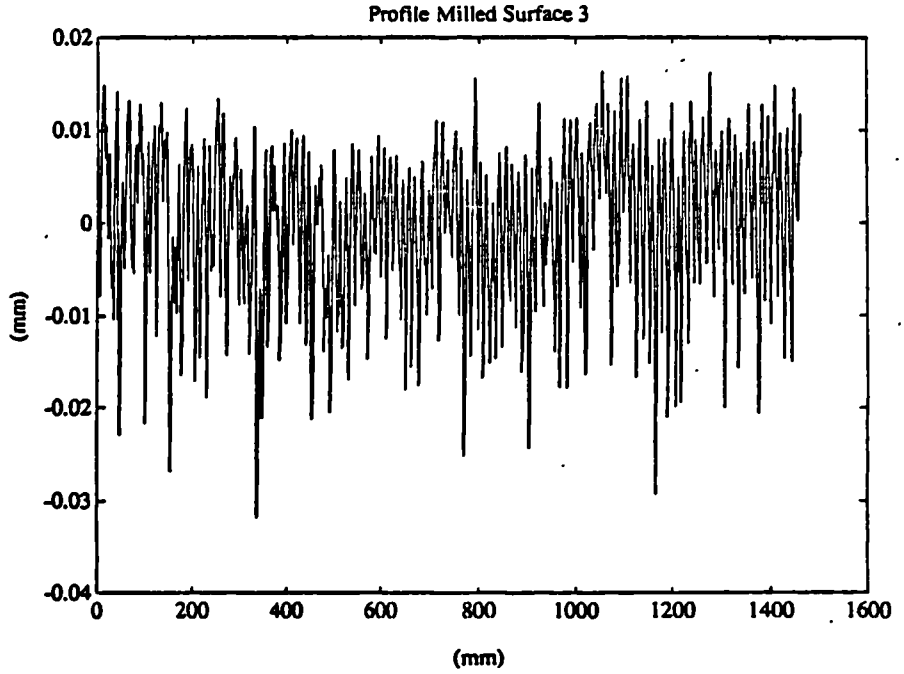


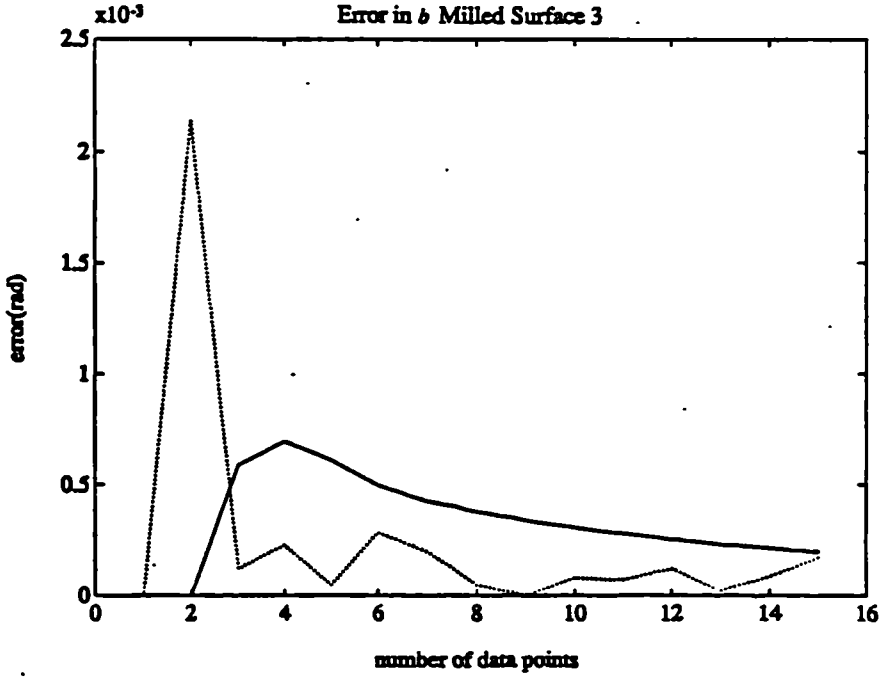


Graph 5.14 (a)

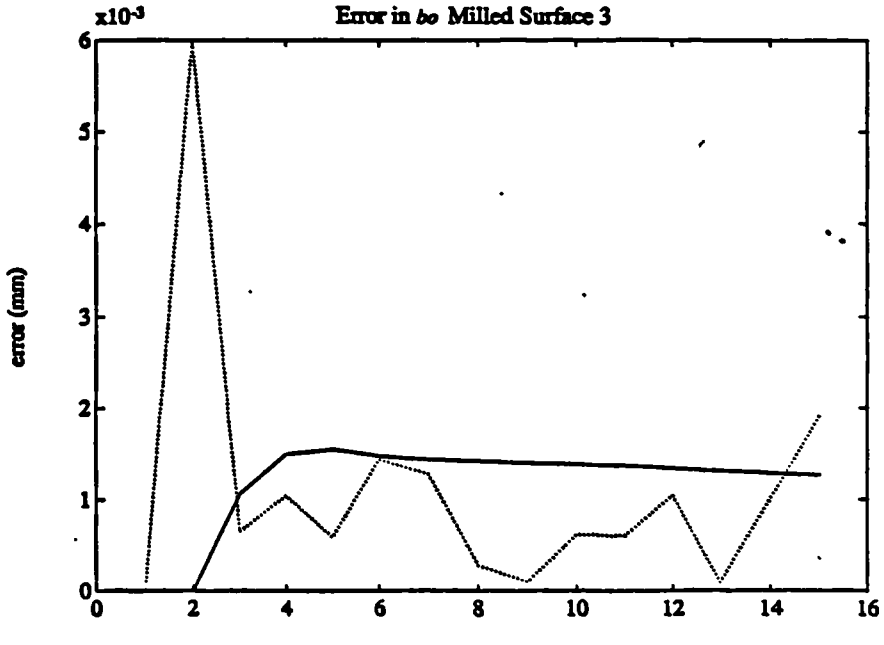


Graph 5.14 (b)

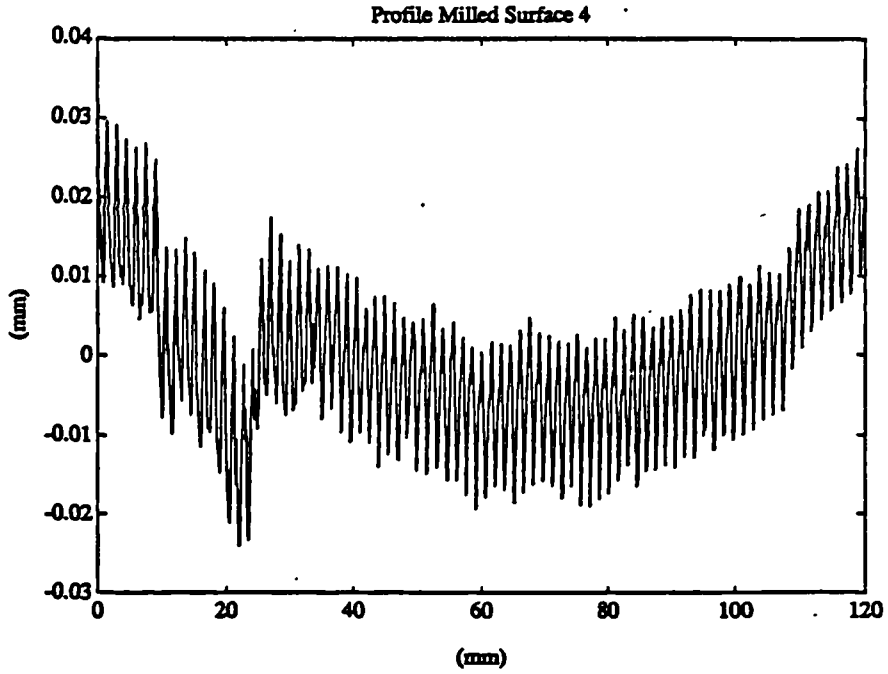


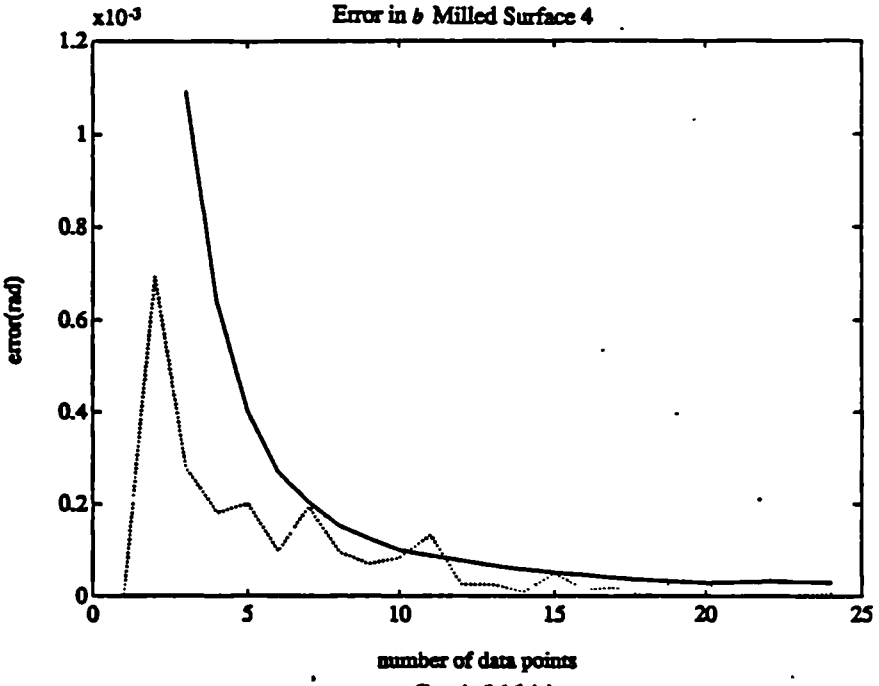


Graph 5.15 (a)

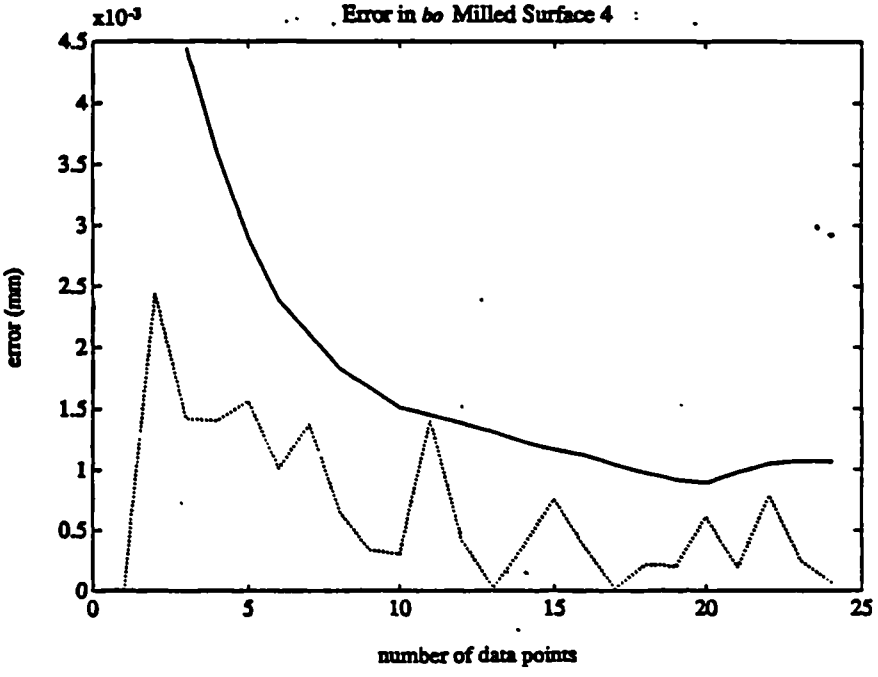


Graph 5.15 (b)

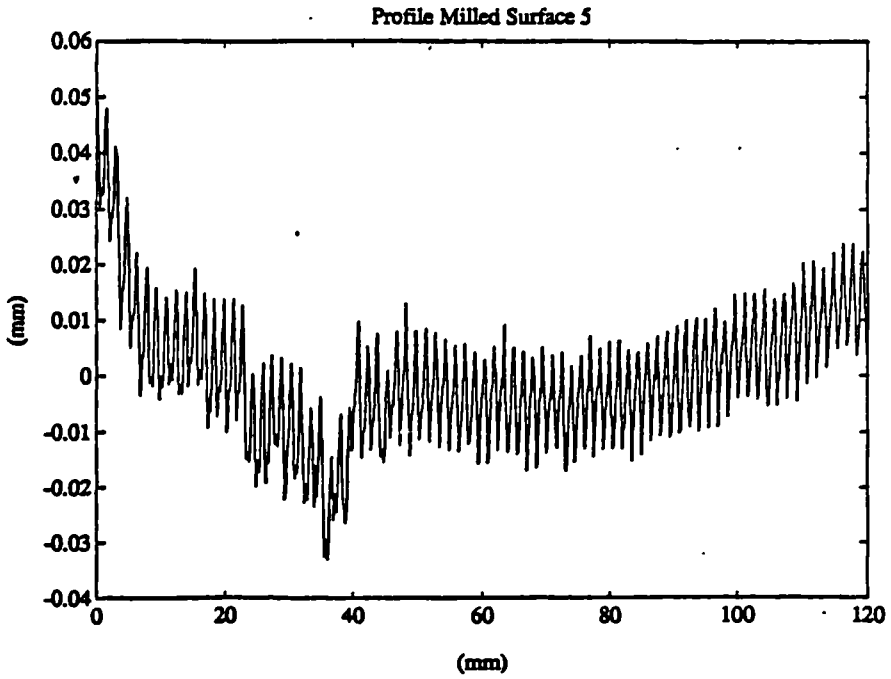


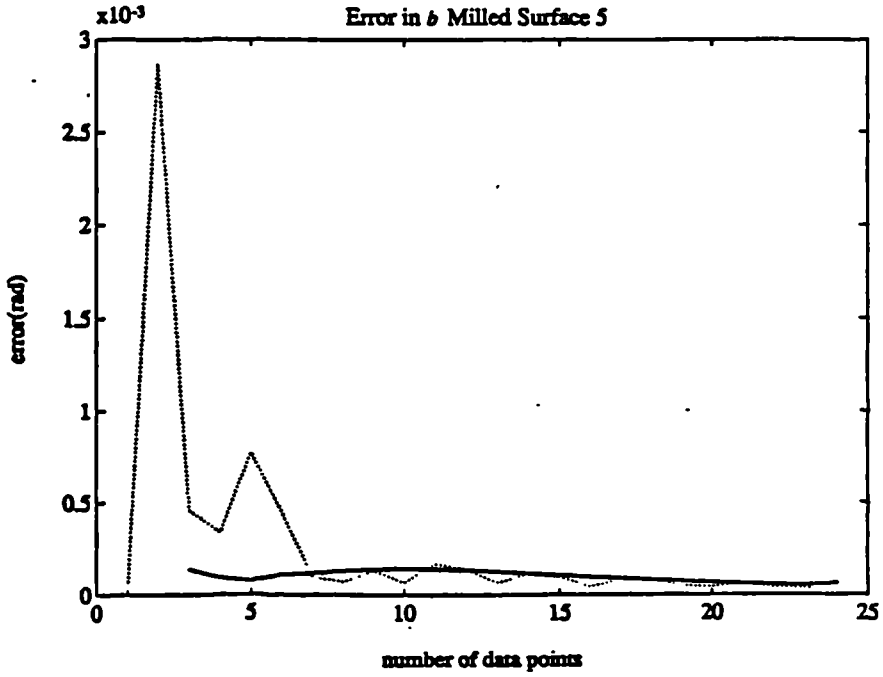


Graph 5.16 (a)

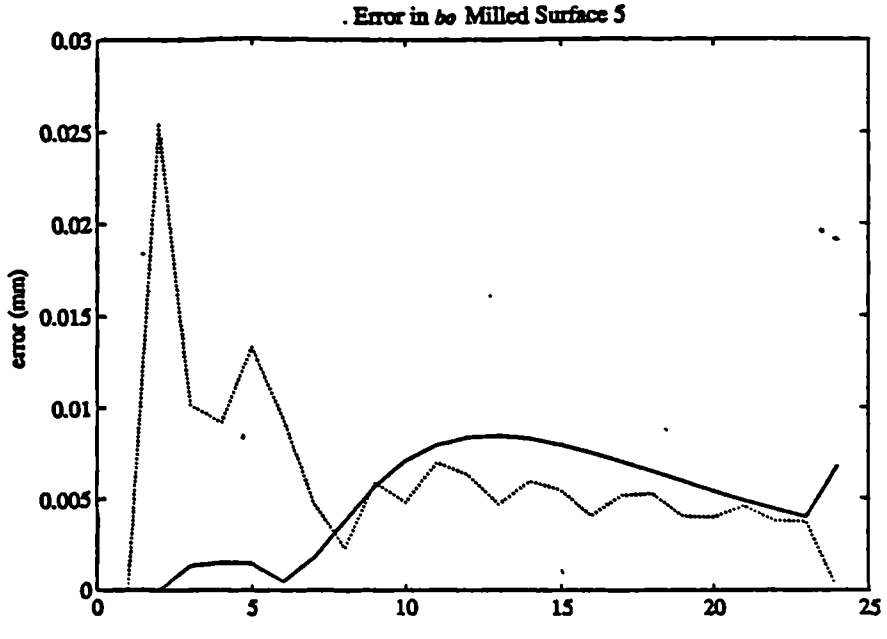


Graph 5.16 (b)

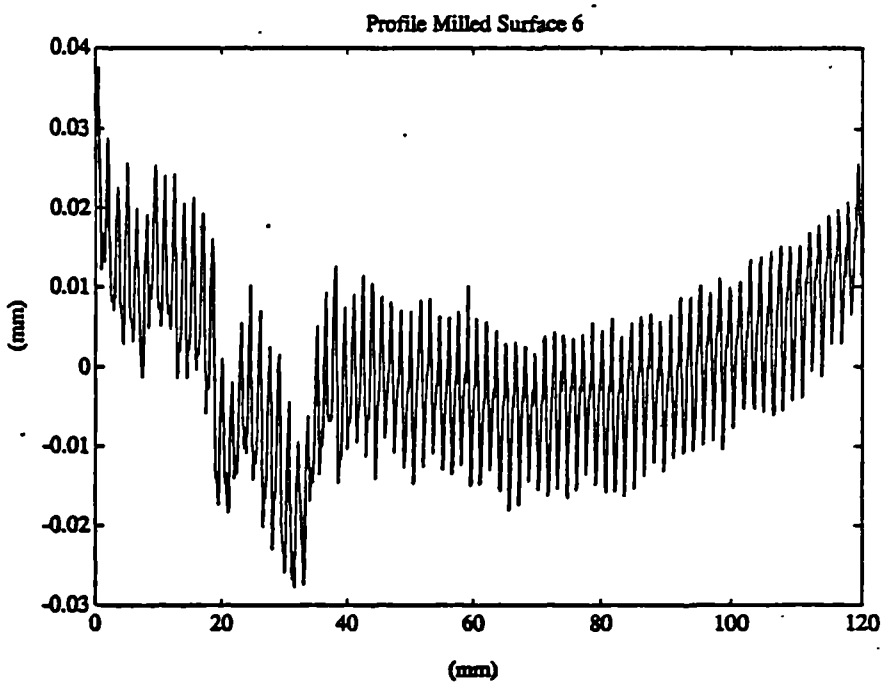


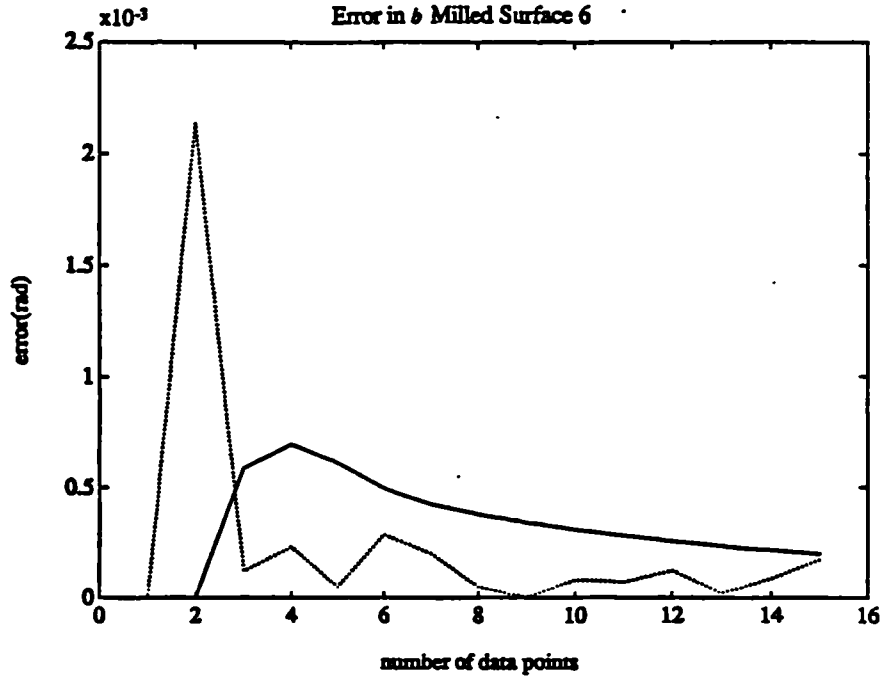


Graph 5.17 (a)

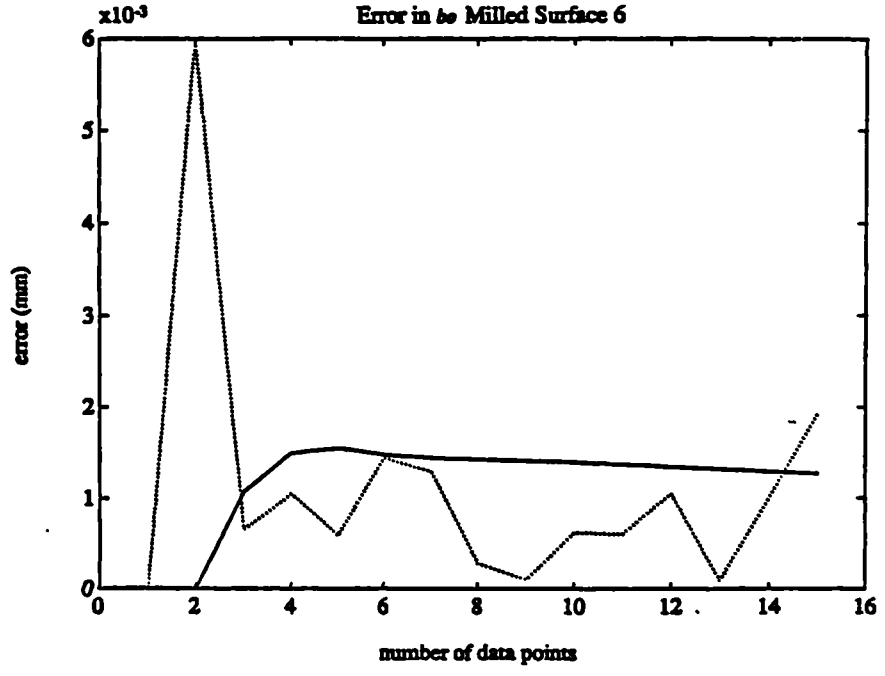


Graph 5.17 (b)


Profile Milled Surface 6
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
(mm)
0
20
40
60
80
100
120
(mm)


x10-3
Error in b Milled Surface 6
2.5
2
1.5
1
0.5
0
error(rad)
0
2
4
6
8
10
12
14
16
number of data points

Graph 5.18 (a)


x10-3
Error in bo Milled Surface 6
6
5
4
3
2
1
0
error (mm)
0
2
4
6
8
10
12
14
16
number of data points

3.4. Procedure for the Estimation of Sampling Error Requiring No Redundant Data

Similar experiments were then performed, but using only that data used in the estimation of the parameters in order to calculate the autocorrelation function. Early results showed that the only required ammendment to the algorithms in order to facilitate this change was to multiply the expressions for the sampling error in both parameters by one half. The final algorithm for requiring no excess data can be summarised thus,

if

$$eb > 10^{-7}$$

$$eb_4 = \frac{2.5}{dx} eb \left[\frac{((ln\,(ln\,(eb^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \right] \tag{5.9}$$

if

$$eb \leq 10^{-7}$$

$$eb_4 = \frac{2.5}{dx} eb \left[\frac{((ln\,(ln\,(eb^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \left[25\,(2.30258\,\ln(eb^{-1}) + 5)(2.30258\,\ln(eb^{-1}) + 6) \right] \right] \tag{5.10}$$

and for intercept with the Y axis

if

$$ebo > 10^{-5}$$

$$ebo_4 = \frac{2.5}{dx} ebo \left[\frac{((ln\,(ln\,(ebo^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \right] \tag{5.11}$$

if

$$ebo \leq 10^{-5}$$

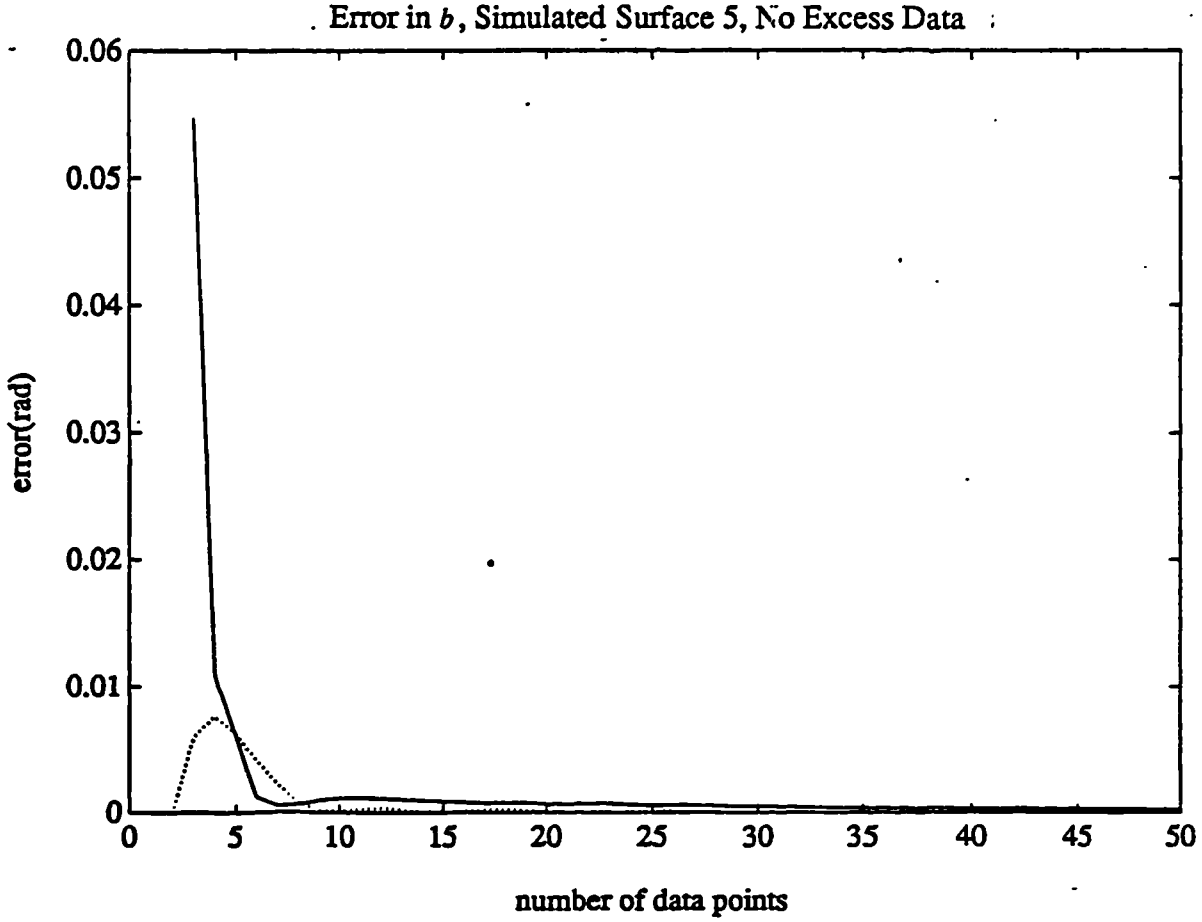
$$ebo_4 = \frac{2.5}{dx} ebo \left[\frac{((ln\,(ln\,(ebo^{-1}))) - 2.1491}{3.316 \times 10^{-3}} \left[25\,(2.30258\,\ln(ebo^{-1}) + 5)(2.30258\,\ln(ebo^{-1}) + 6) \right] \right] \tag{5.12}$$

The simulated data used in these experiments is described in table 5.2, where real surface data is used a profile of this data precedes the corresponding results.

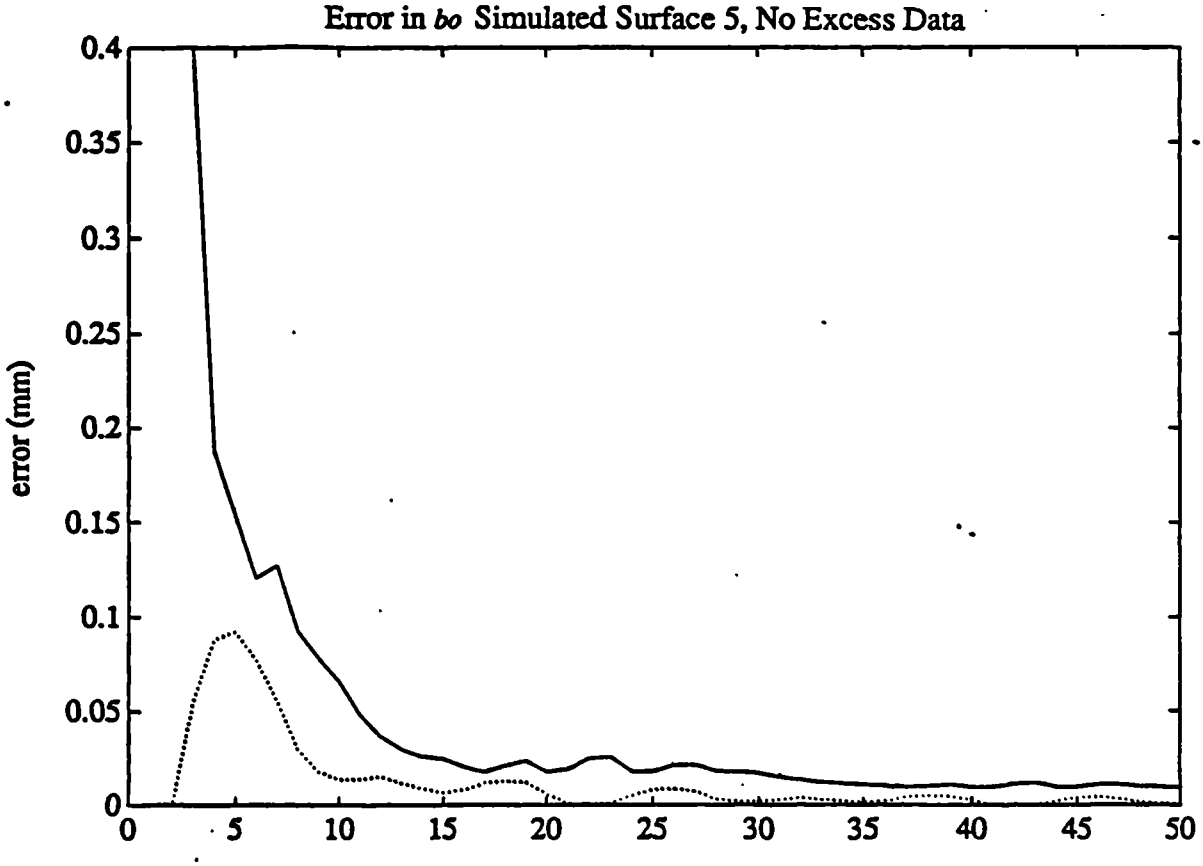
Simulated data (Set 2)		
Surface	Surface Parameters	Sampling Interval(mm)
5	0.0500 sin (2π0.0200x + π/10) 0.0070 sin (2π0.2800x + π/30) 0.0600 sin (2π0.0300x - π/6) 0.0900 sin (2π0.6600x - π/2)	5
6	0.0090 sin (2π0.0080x + π/7) 0.0800 sin (2π0.3600x + π/19) 0.0081 sin (2π1.6640x - π/40) 0.0030 sin (2π0.0442x - π/8) 0.0074 sin (2π2.1200x + π) rand, std=0.0143	5
7	0.0090 sin (2π0.0080x + π/7) 0.0800 sin (2π0.0400x + π/19) 0.0081 sin (2π1.6640x - π/40) 0.0030 sin (2π0.0442x - π/8) 0.0074 sin (2π2.1200x + π) rand, std=0.0114	7

Table 5.2

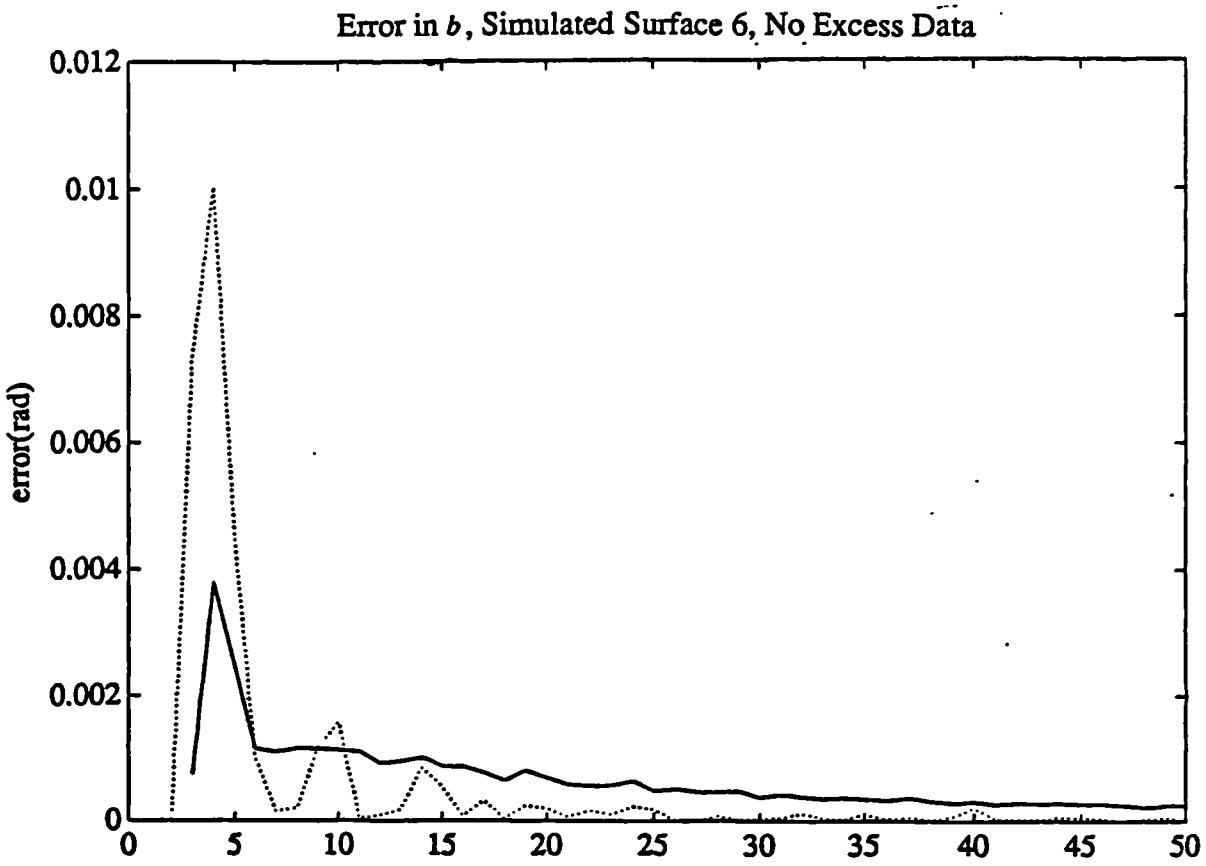
The results of this final algorithm are presented in graphs 5.18 to 5.33, where a similar key to the first set of results applies.



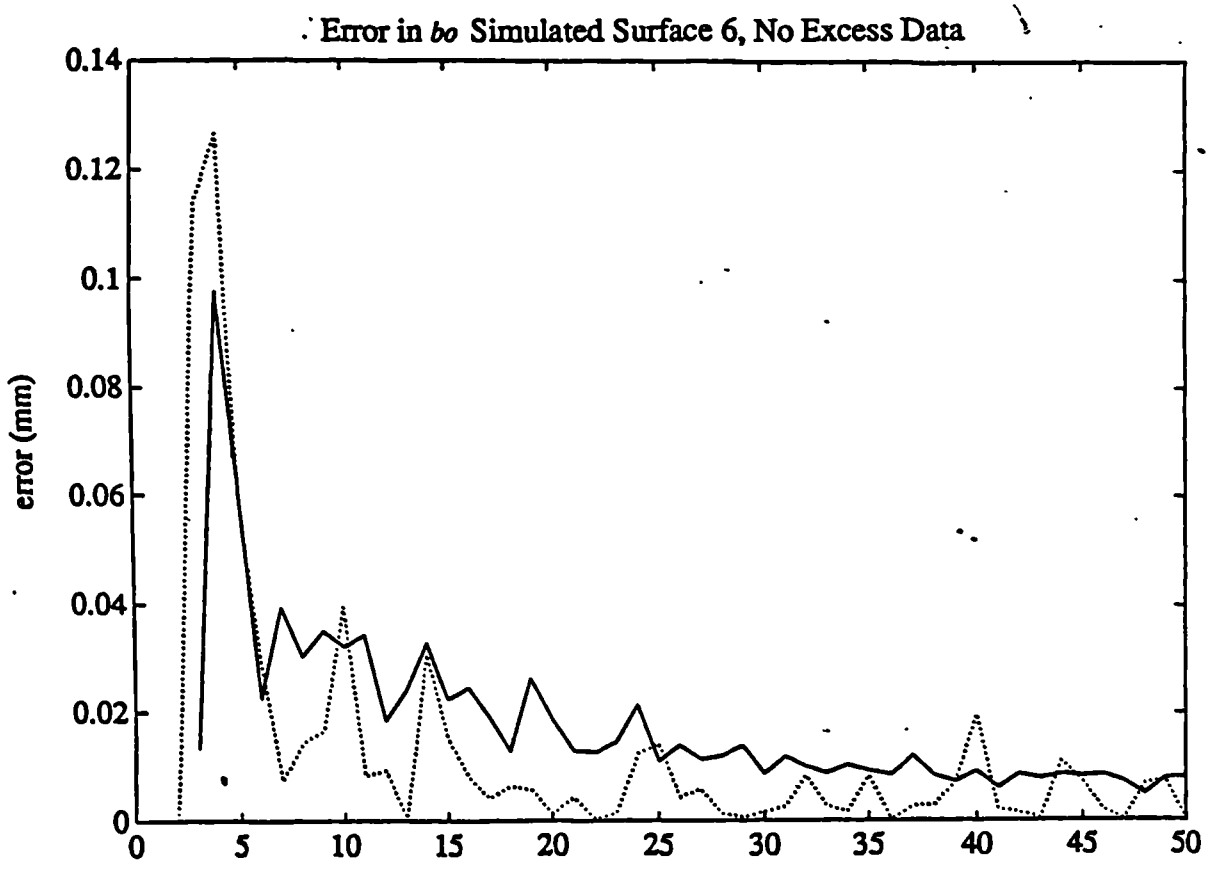
Graph 5.19 (a)



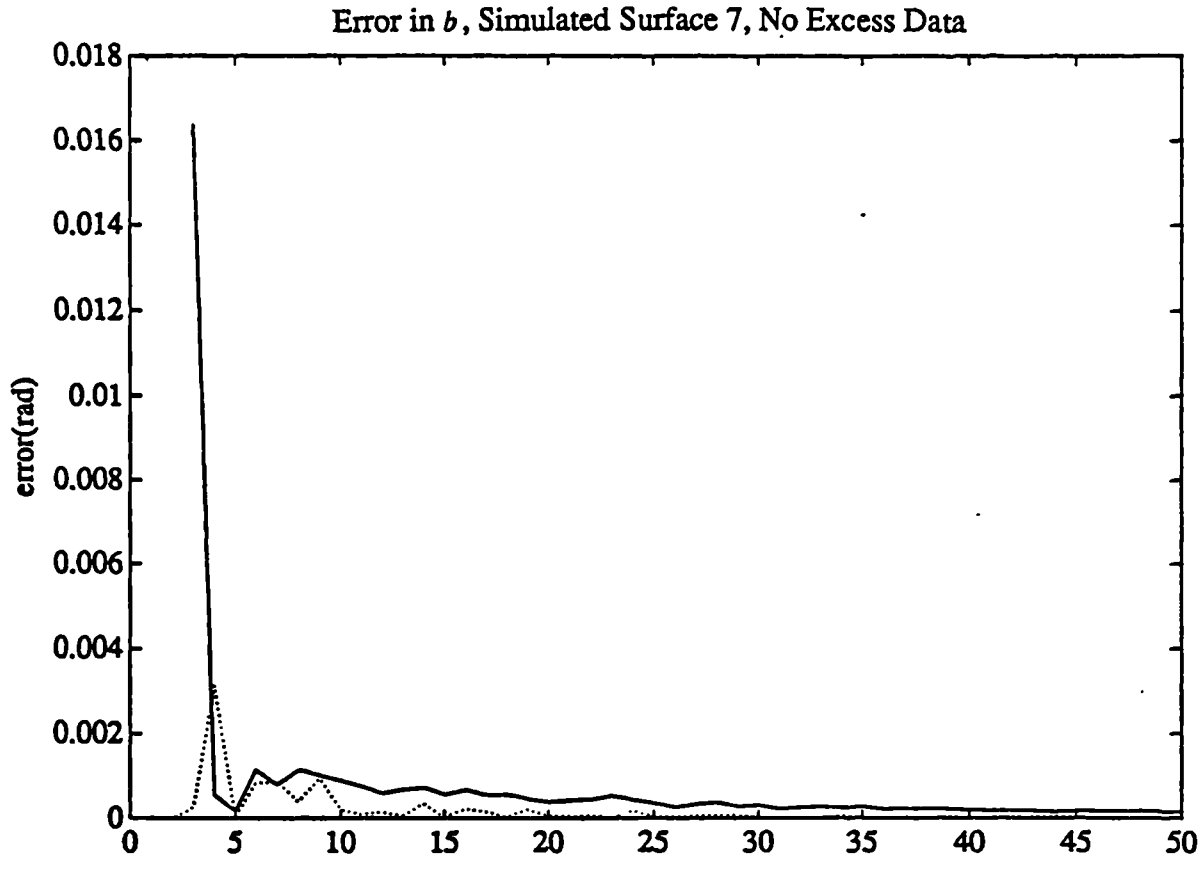
Graph 5.19 (b)



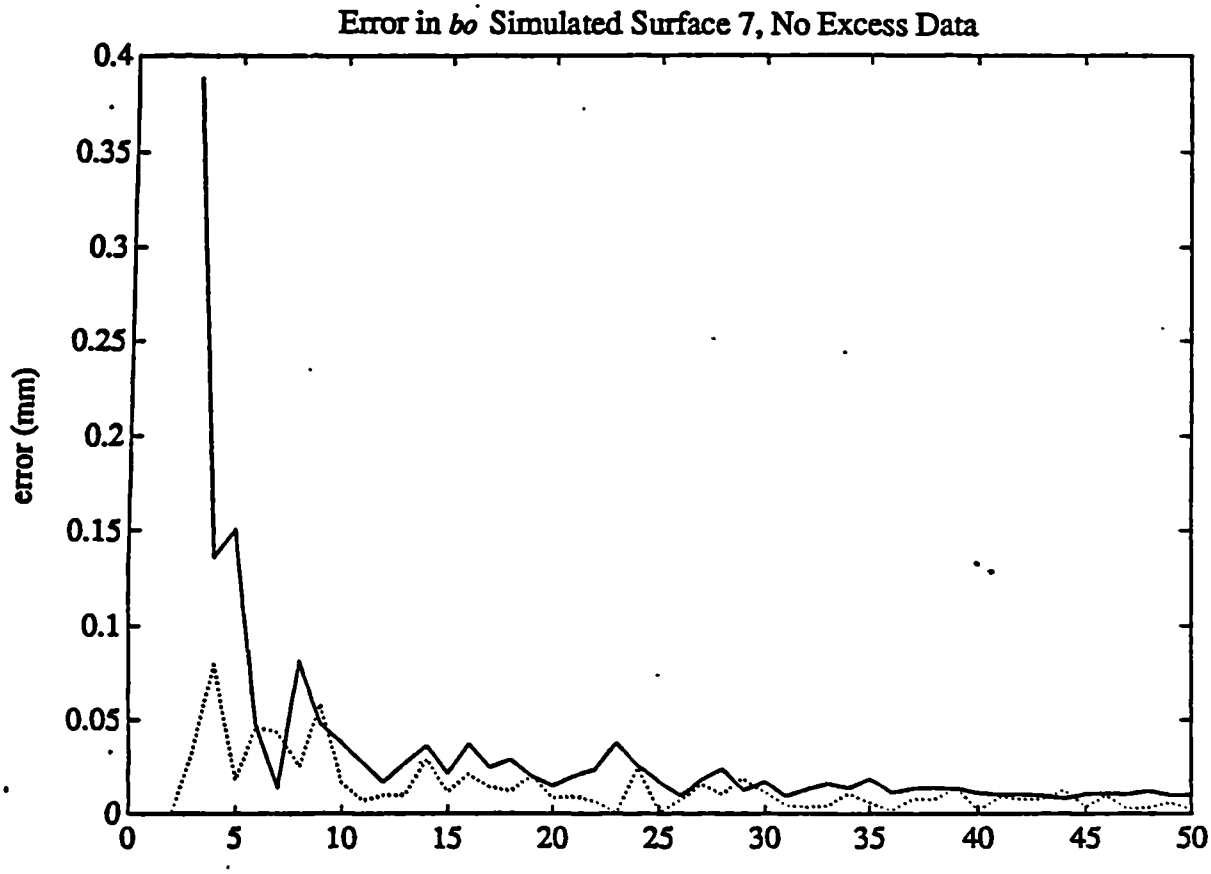
Graph 5.20 (a)



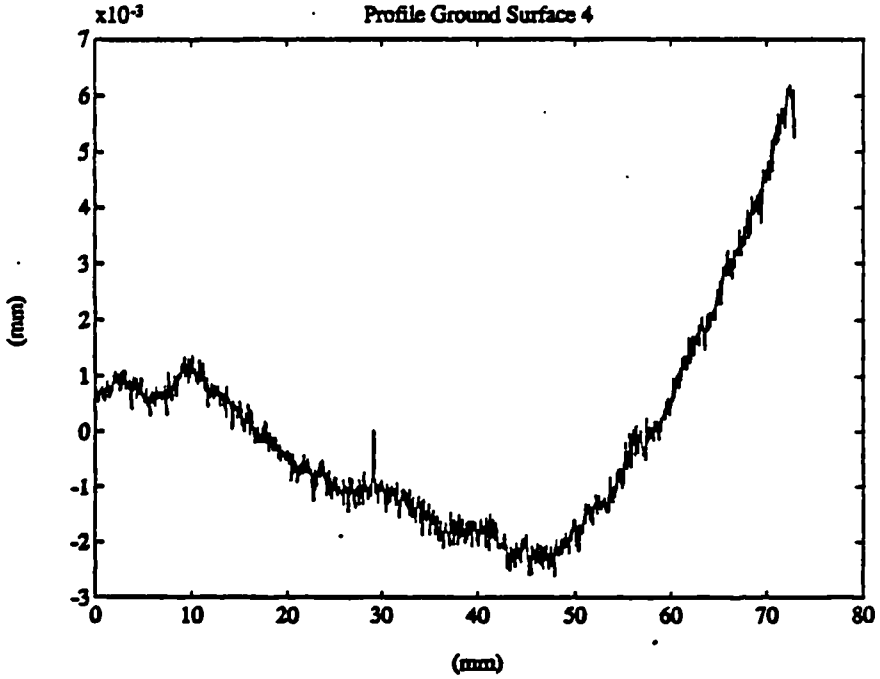
Graph 5.20 (b)

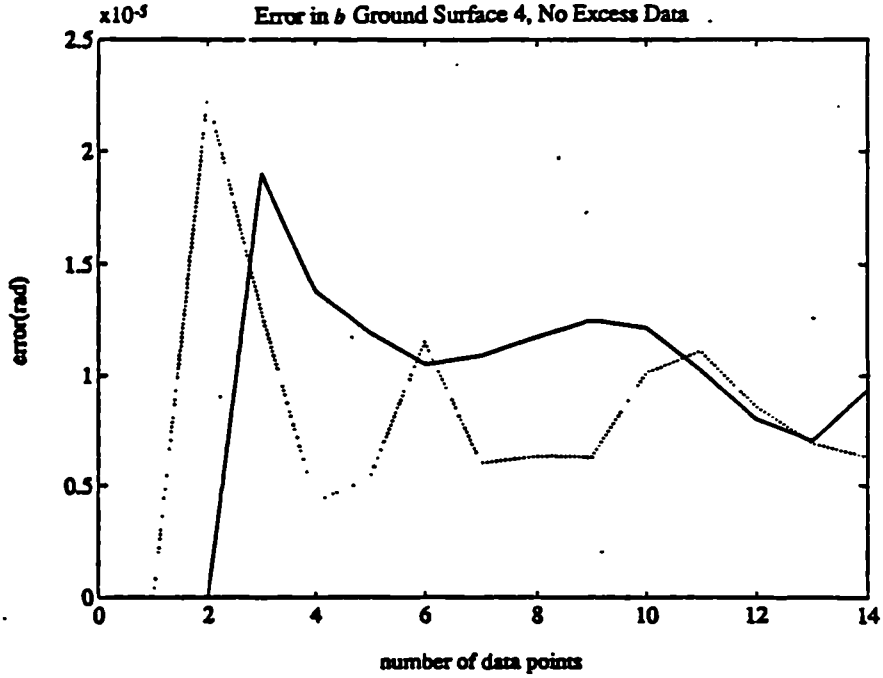


Graph 5.21 (a)

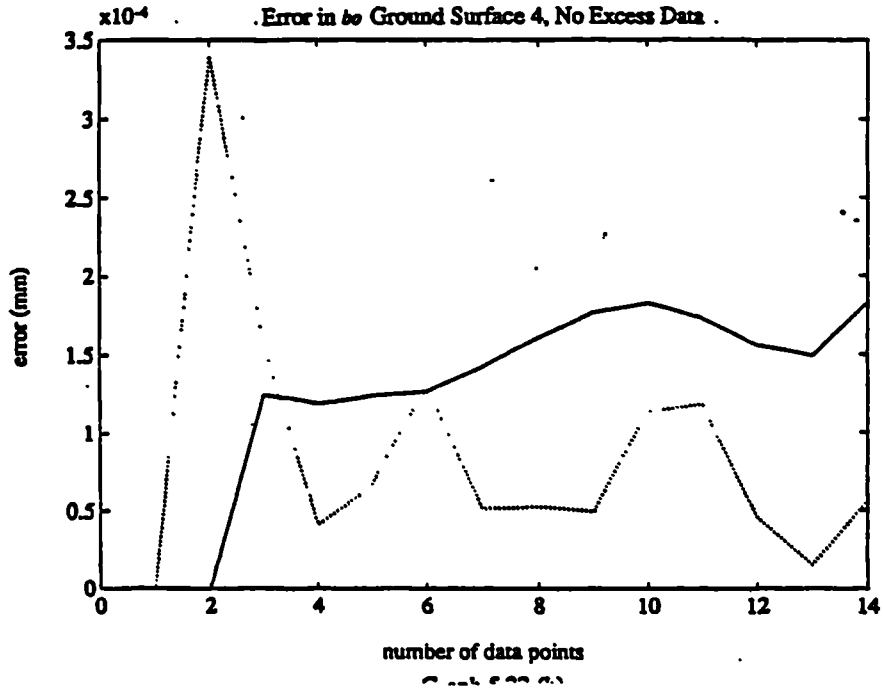


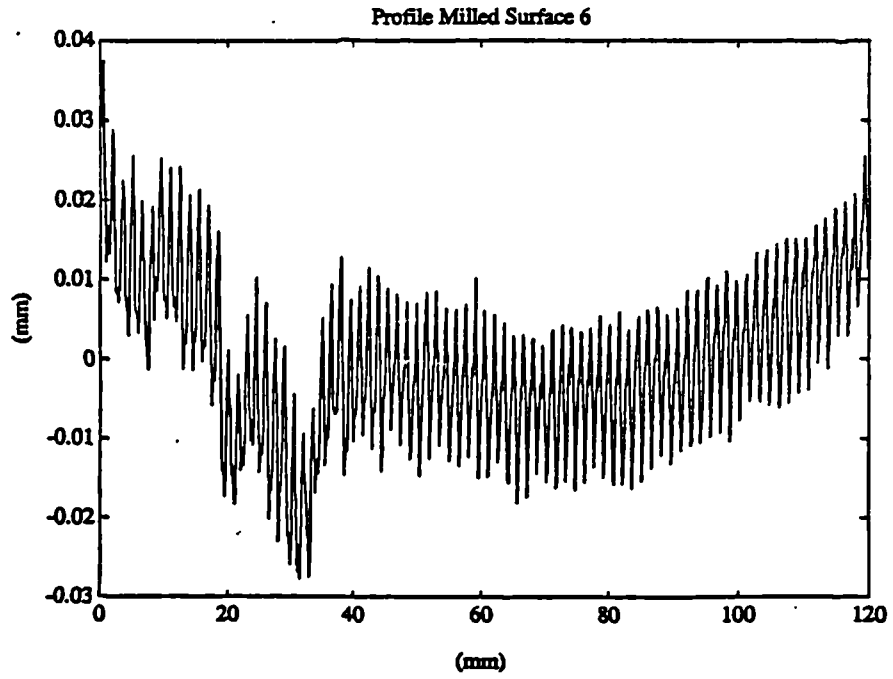
Graph 5.21 (b)

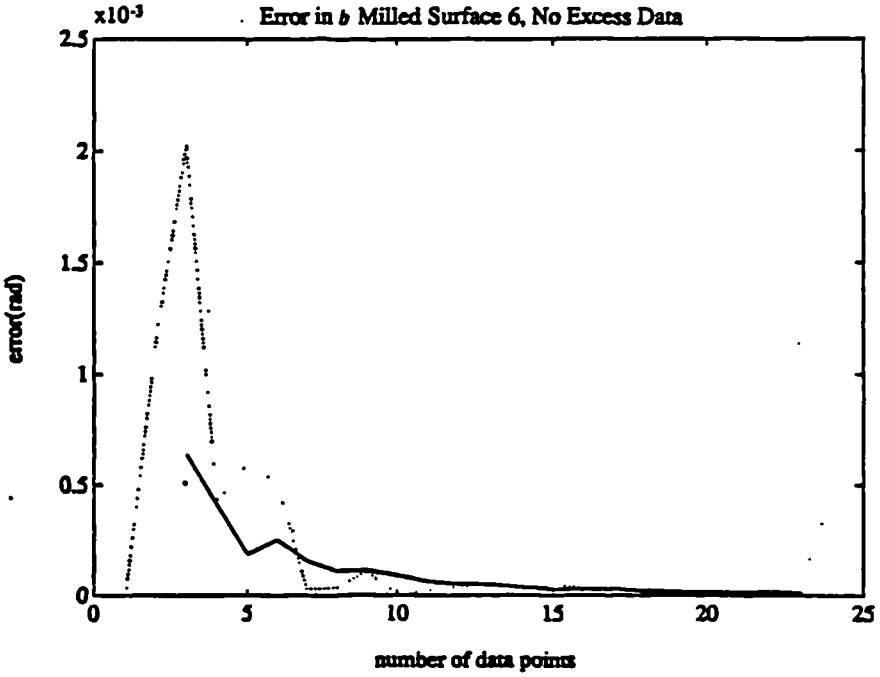

x10-3
Profile Ground Surface 4
(mm)
(mm)


x10-5
Error in b Ground Surface 4, No Excess Data
error(rad)
number of data points

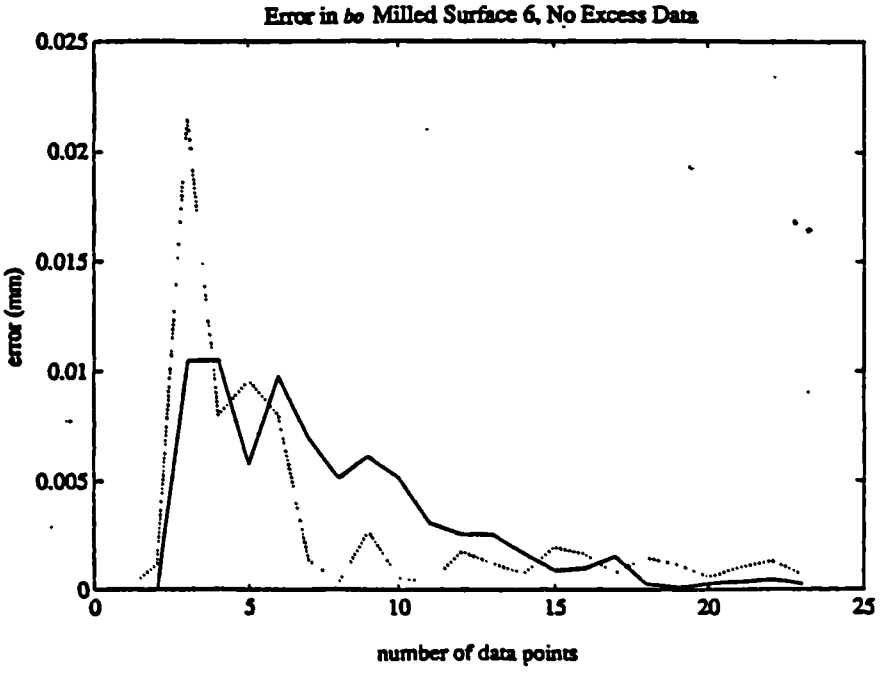
Graph 5.22 (a)

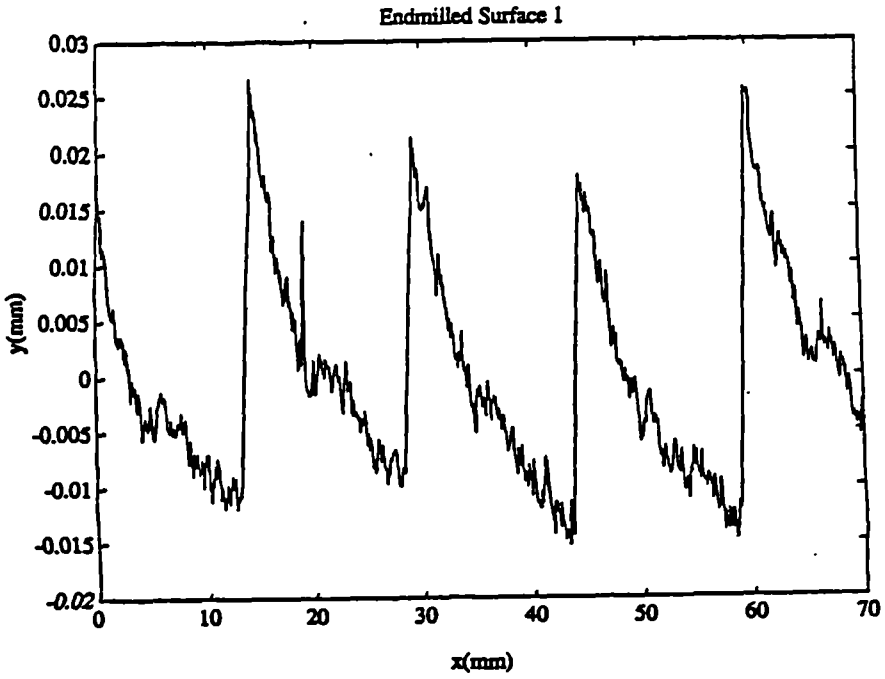

x10-4
Error in bo Ground Surface 4, No Excess Data
error (mm)
number of data points

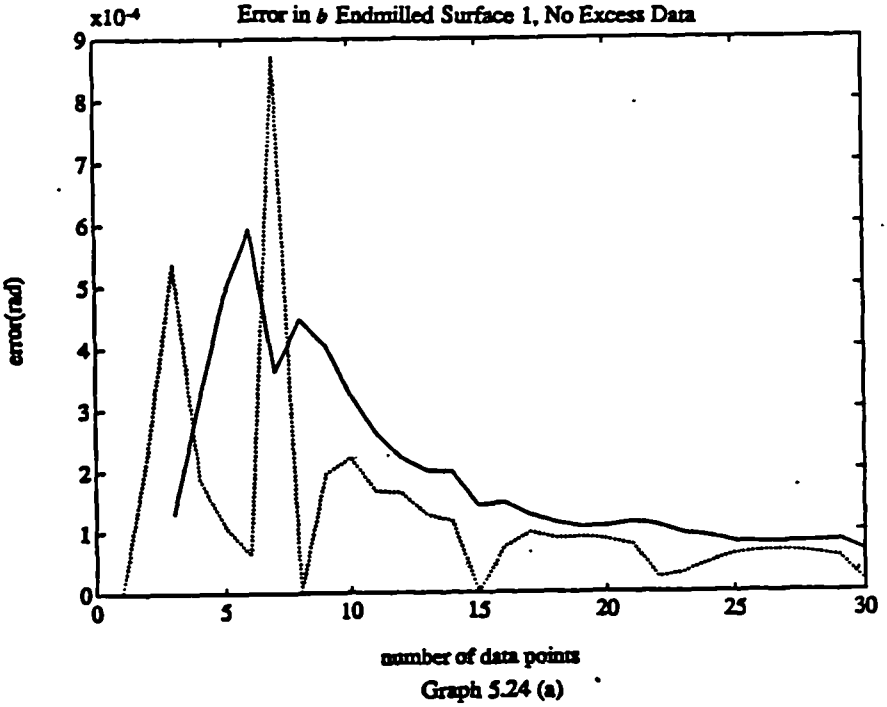

Profile Milled Surface 6
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
(mm)
0
20
40
60
80
100
120
(mm)


x10-3
Error in b Milled Surface 6, No Excess Data
2.5
2
1.5
1
0.5
0
error(rad)
0
5
10
15
20
25
number of data points

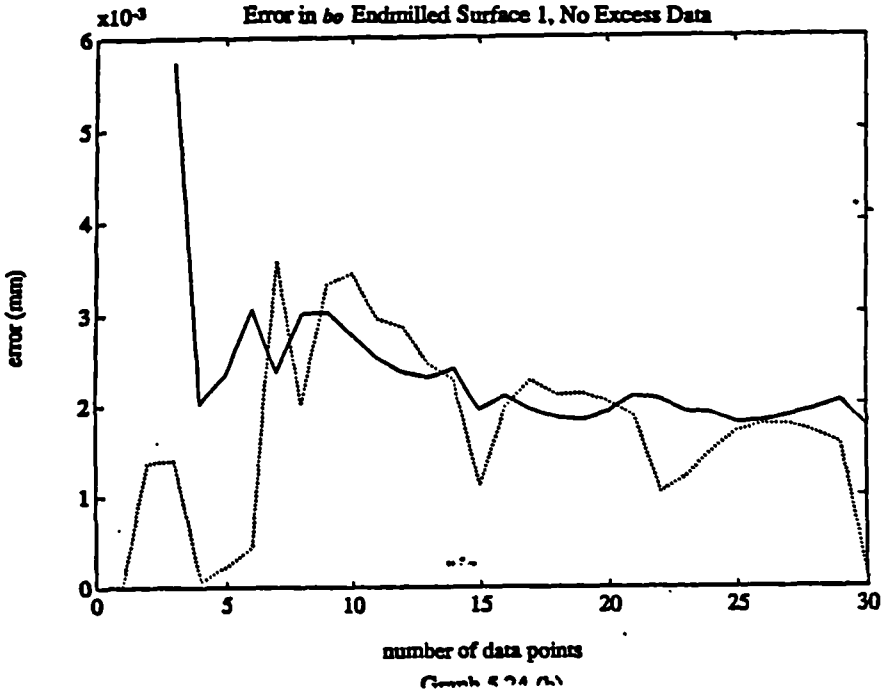
Graph 5.23 (a)


Error in bo Milled Surface 6, No Excess Data
0.025
0.02
0.015
0.01
0.005
0
error (mm)
0
5
10
15
20
25
number of data points

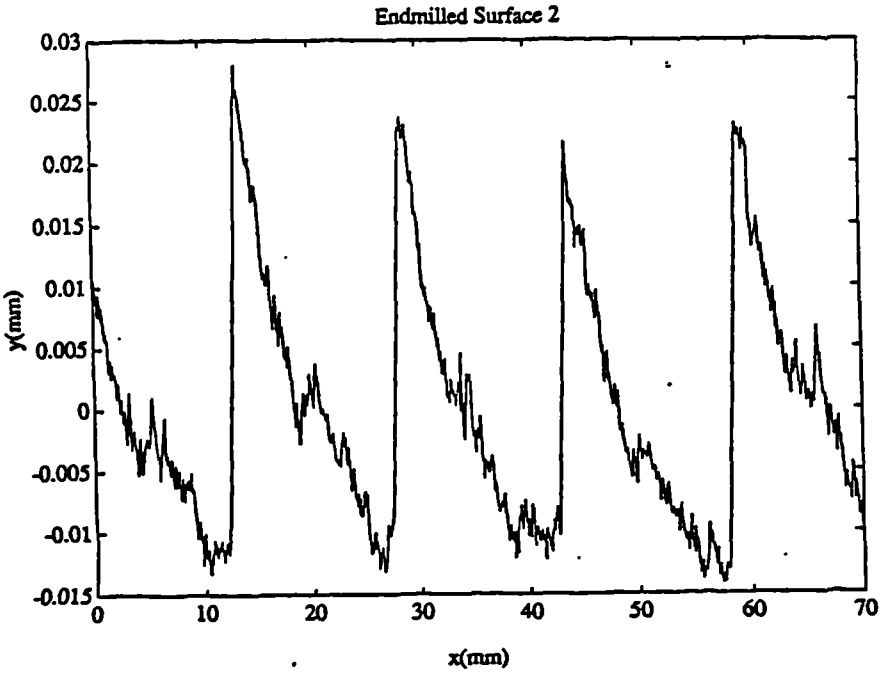


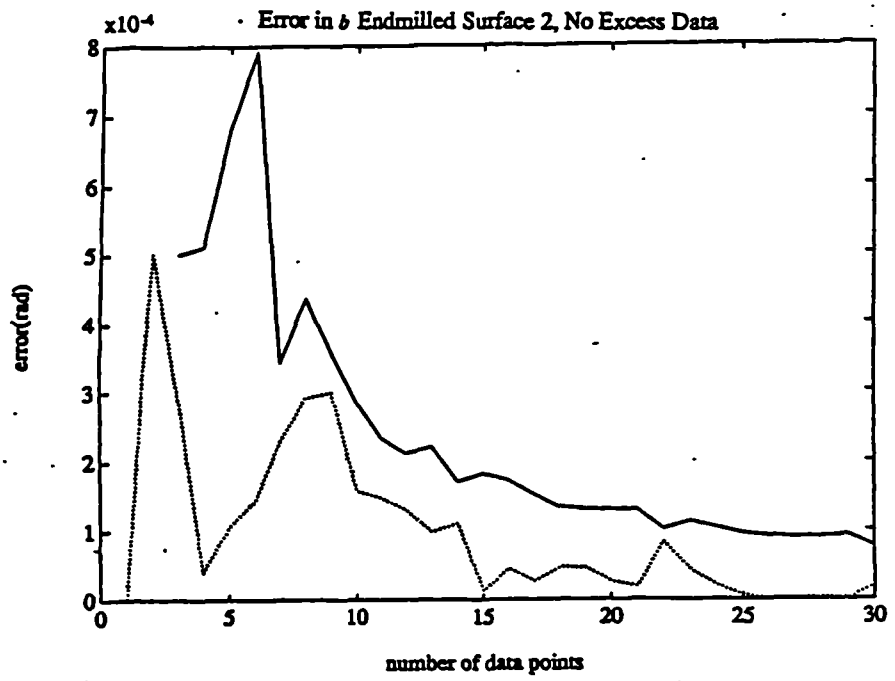


Graph 5.24 (a)

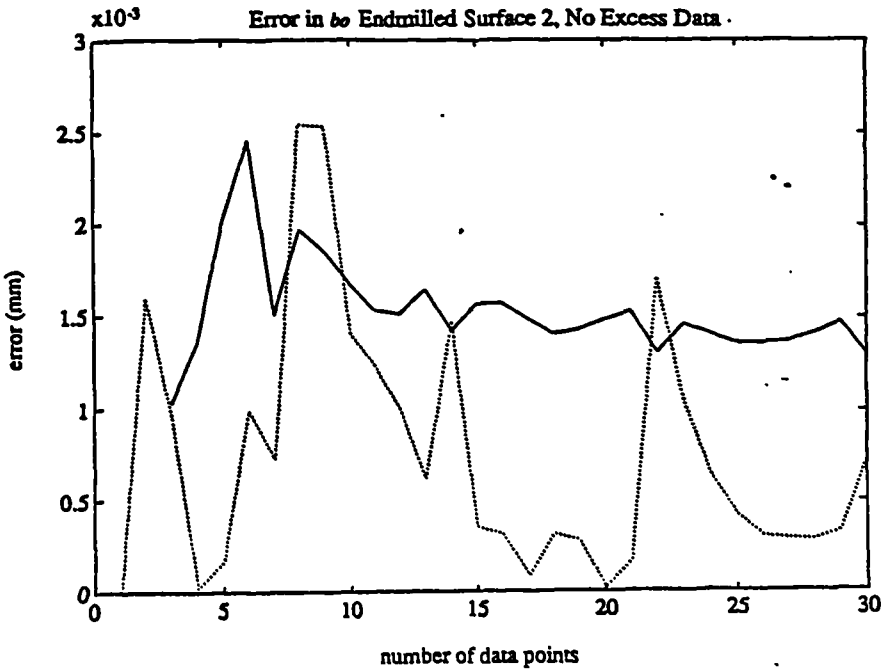


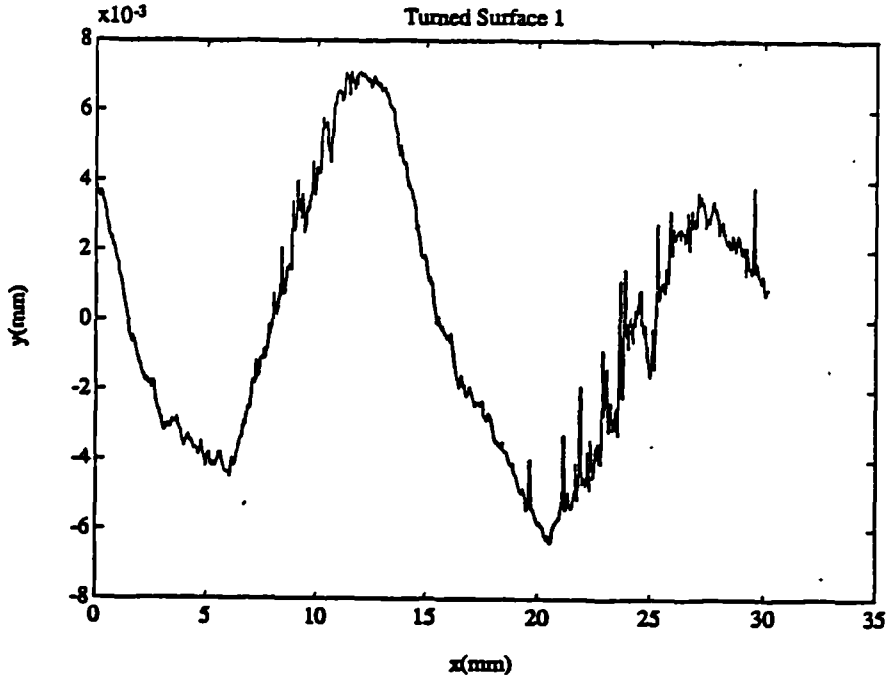
Graph 5.24 (b)

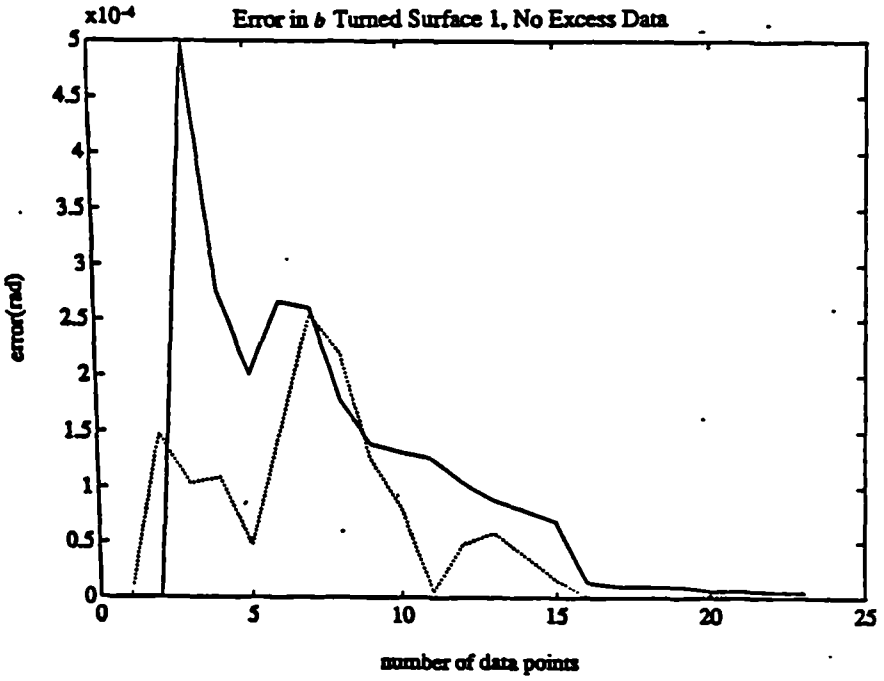

Endmilled Surface 2
y(mm)
x(mm)
0.03
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
0
10
20
30
40
50
60
70


x10-4
Error in b Endmilled Surface 2, No Excess Data
error(rad)
number of data points
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30

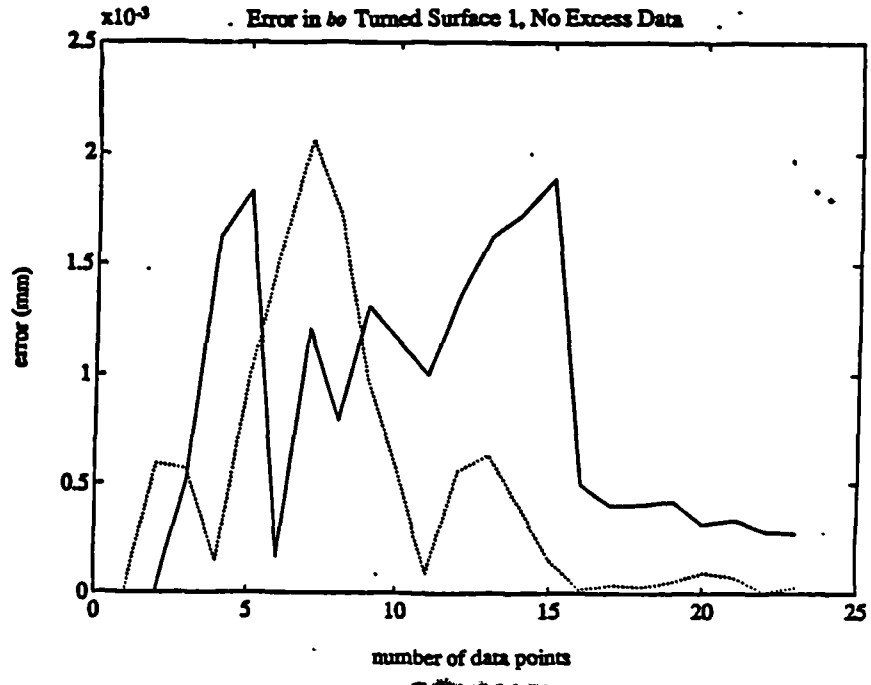
Graph 5.25 (a)

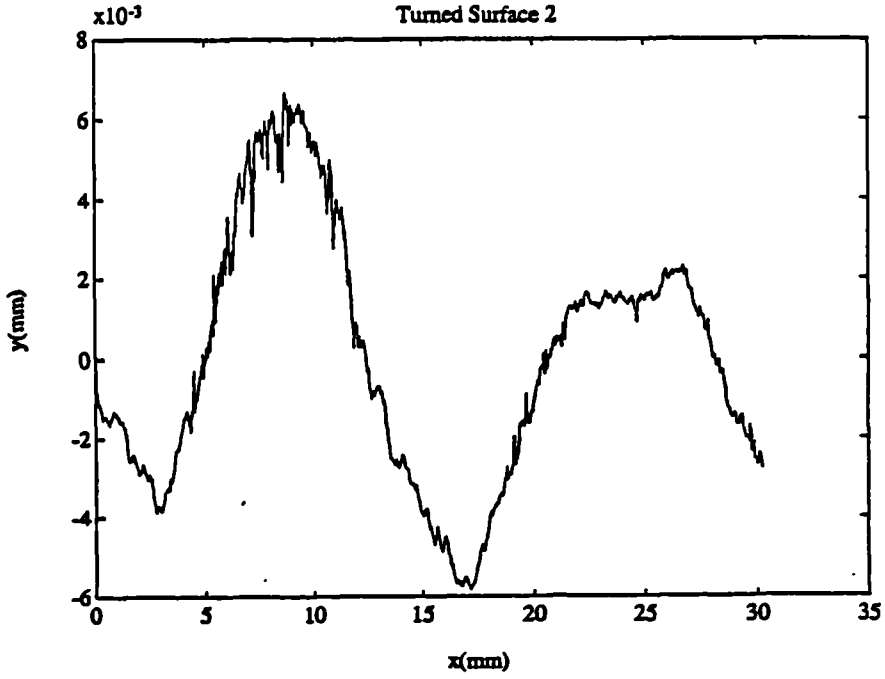

x10-3
Error in bo Endmilled Surface 2, No Excess Data
error (mm)
number of data points
3
2.5
2
1.5
1
0.5
0
0
5
10
15
20
25
30

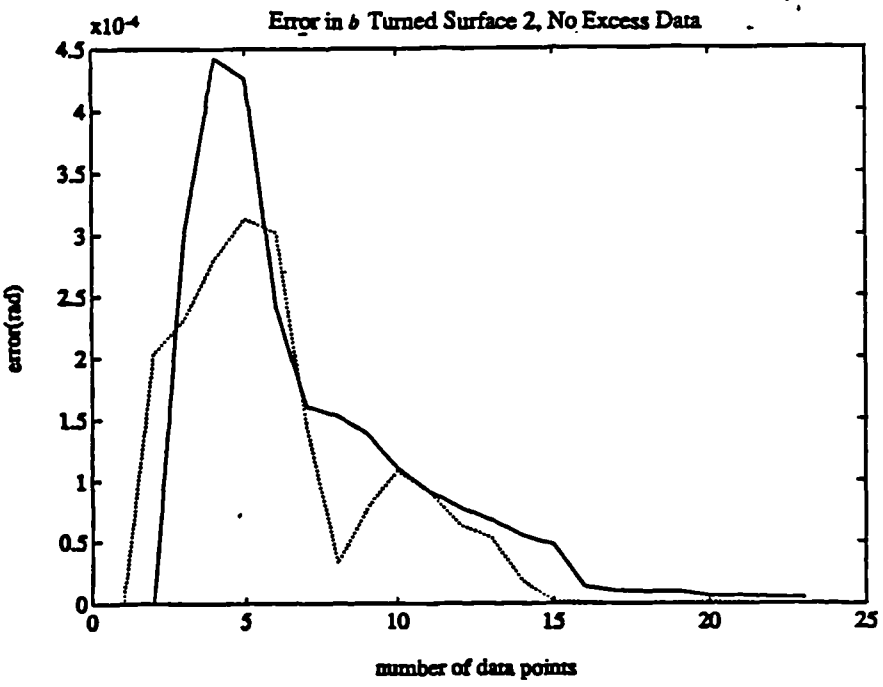

x10-3
Turned Surface 1
y(mm)
x(mm)


x10-4
Error in b Turned Surface 1, No Excess Data
error(rad)
number of data points

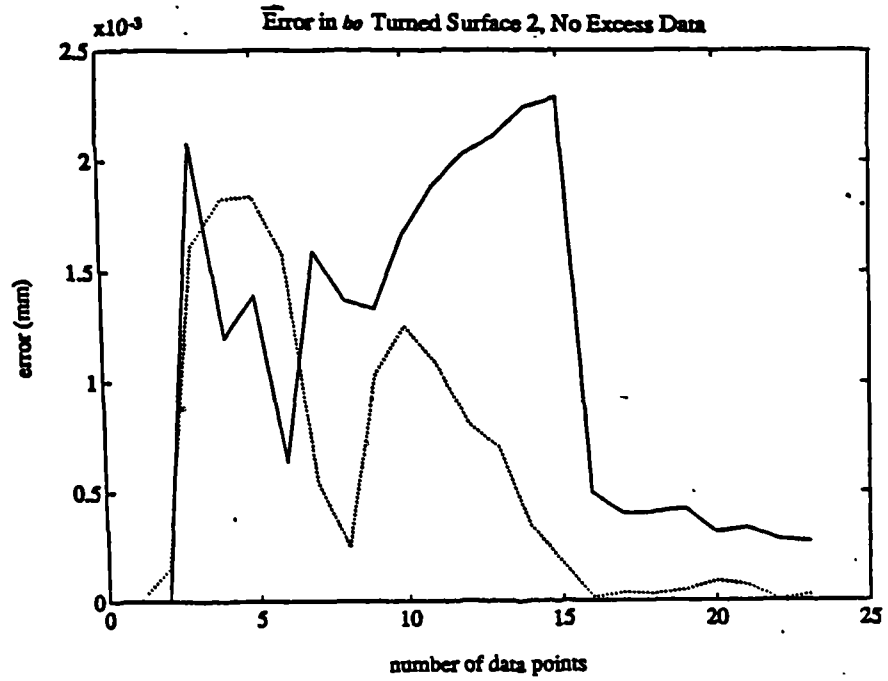
Graph 5.26 (a)

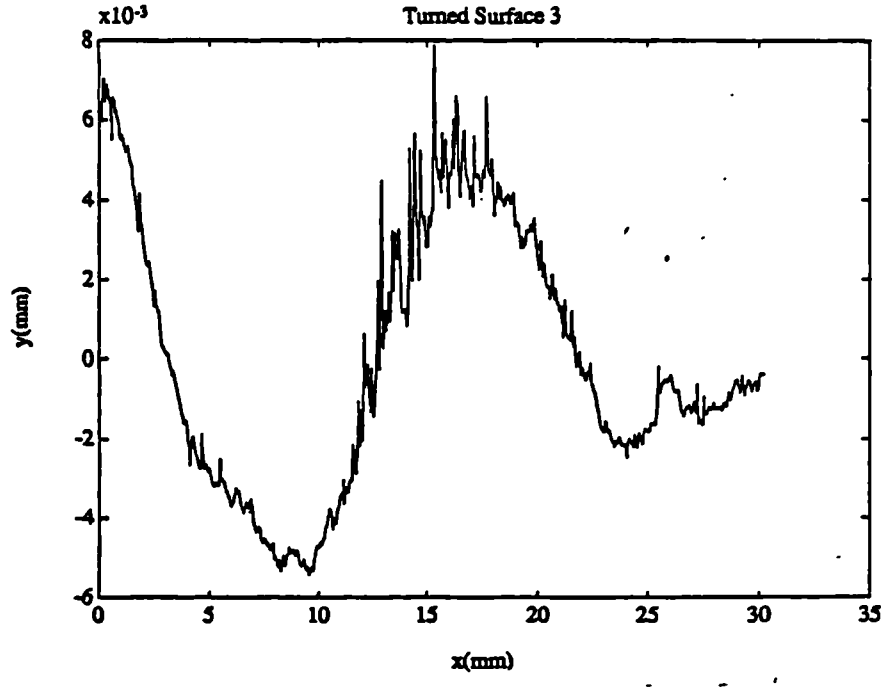

x10-3
Error in bo Turned Surface 1, No Excess Data
error (mm)
number of data points

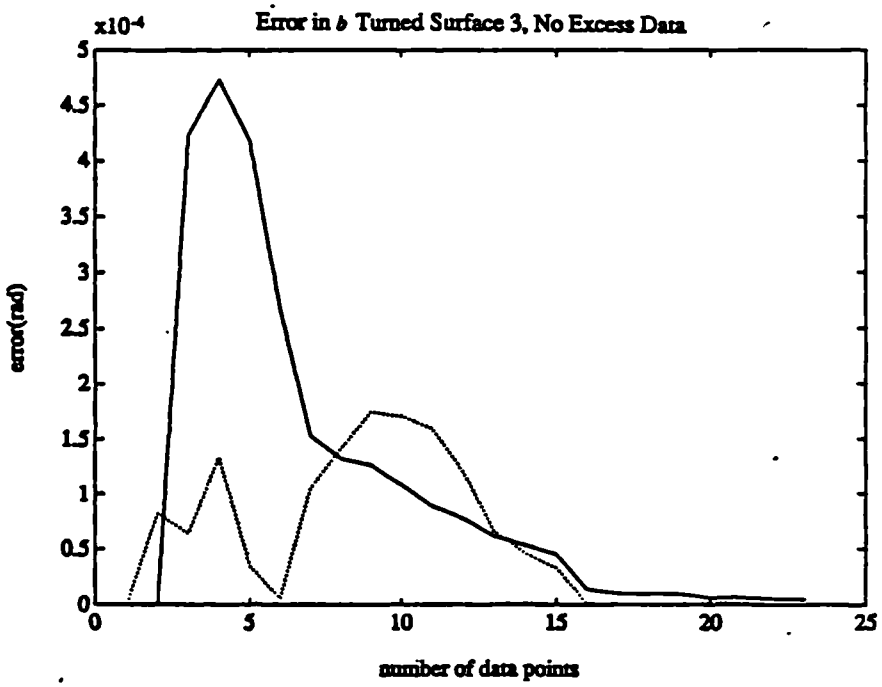

x10-3
Turned Surface 2
y(mm)
x(mm)


x10-4
Error in b Turned Surface 2, No Excess Data
error(rad)
number of data points

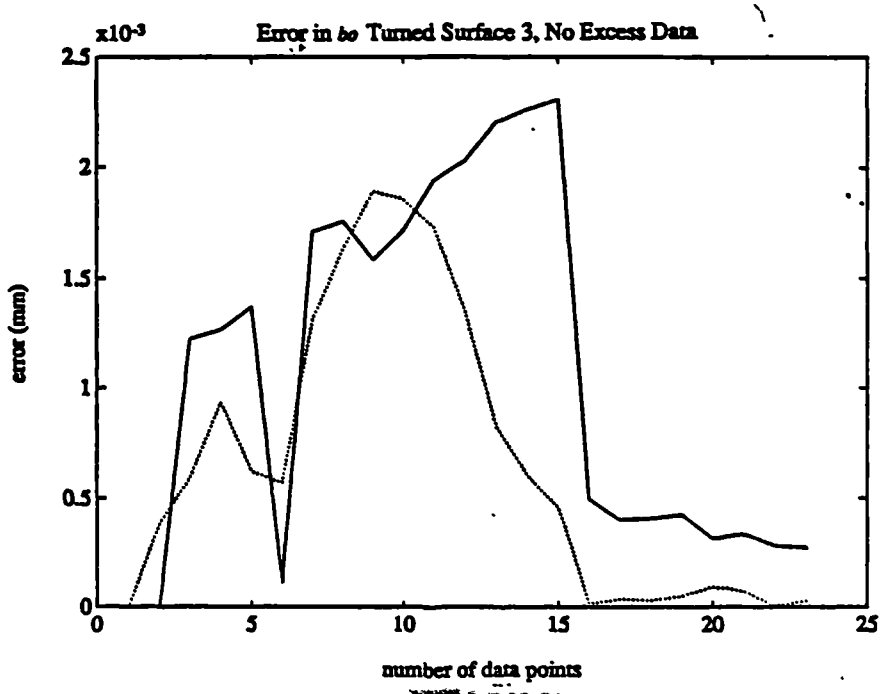
Graph 5.27 (a)

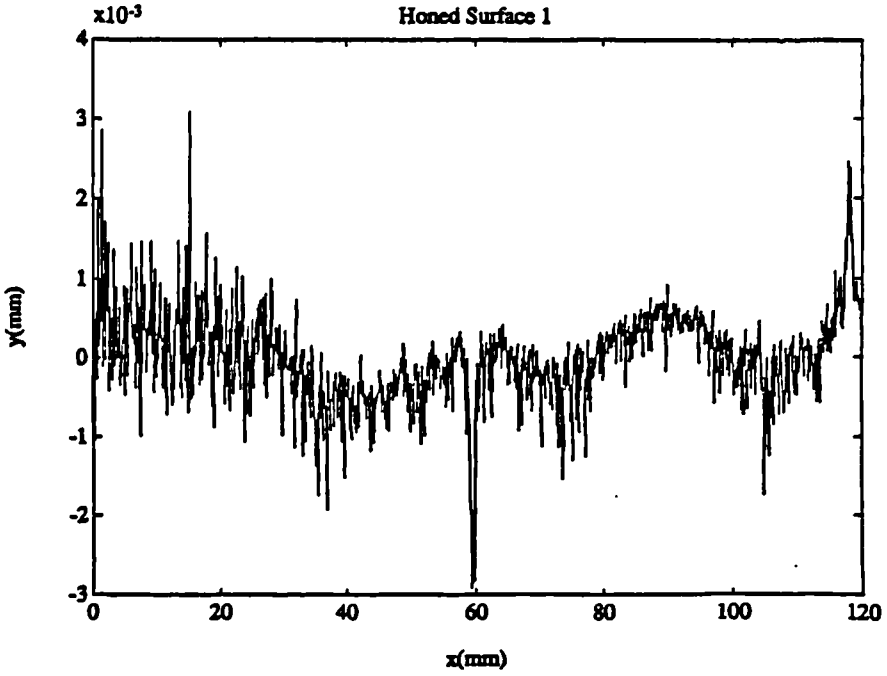

x10-3
Error in bo Turned Surface 2, No Excess Data
error (mm)
number of data points

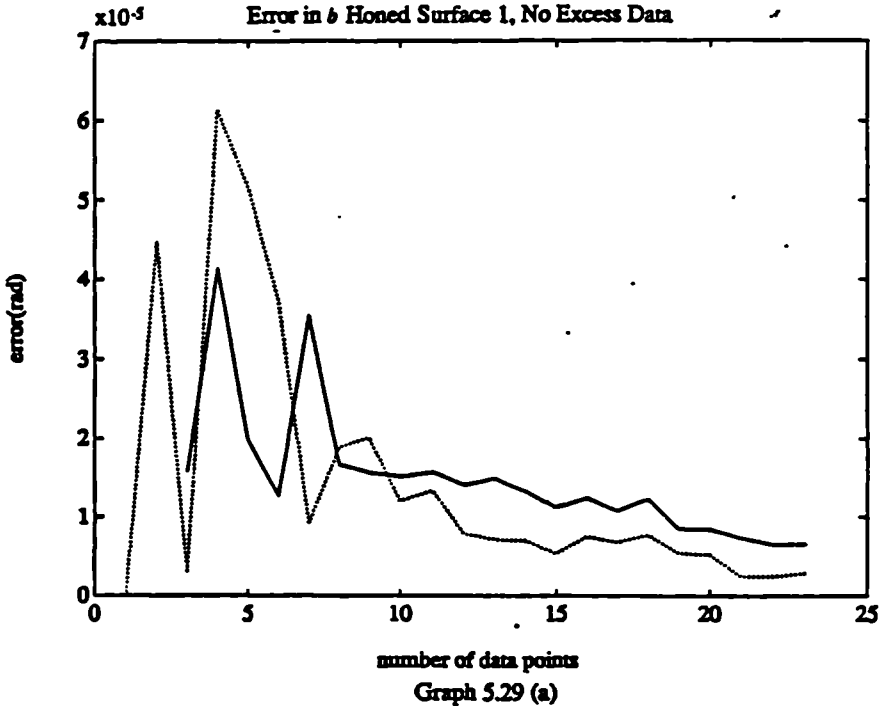

x10-3
Turned Surface 3
y(mm)
x(mm)


x10-4
Error in b Turned Surface 3, No Excess Data
error(rad)
number of data points

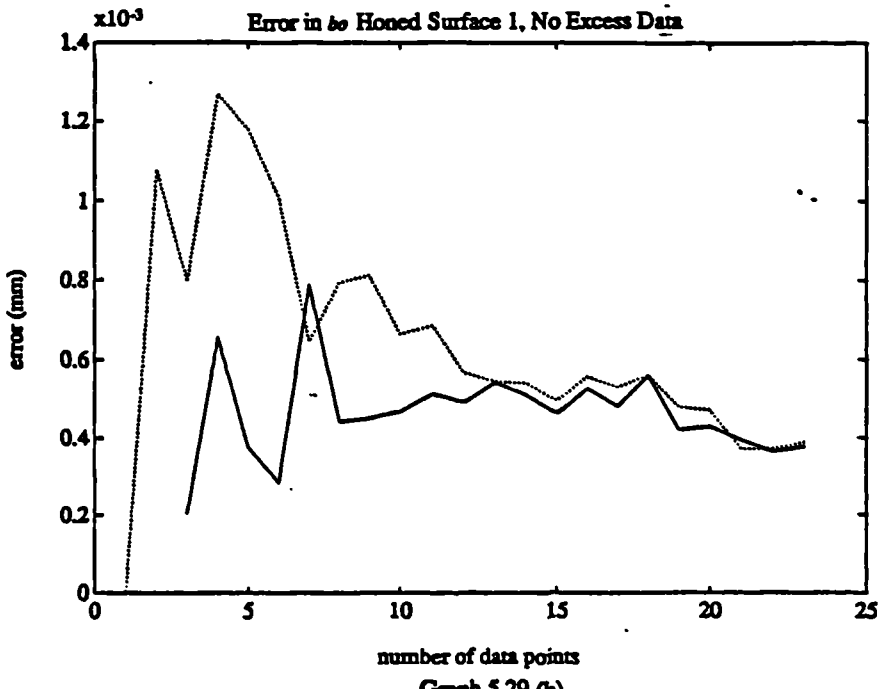
Graph 5.28 (a)


x10-3
Error in bo Turned Surface 3, No Excess Data
error (mm)
number of data points

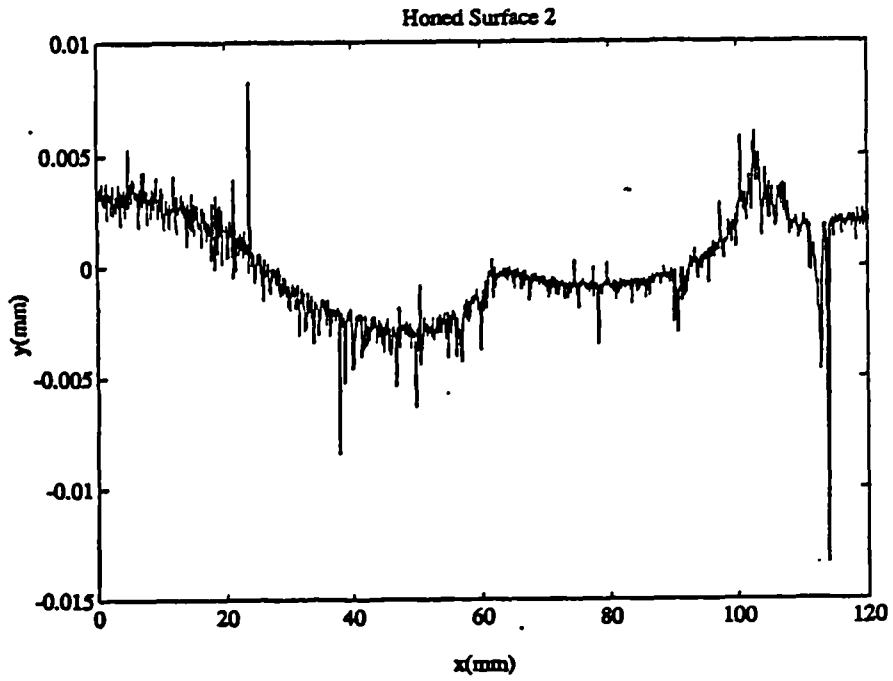


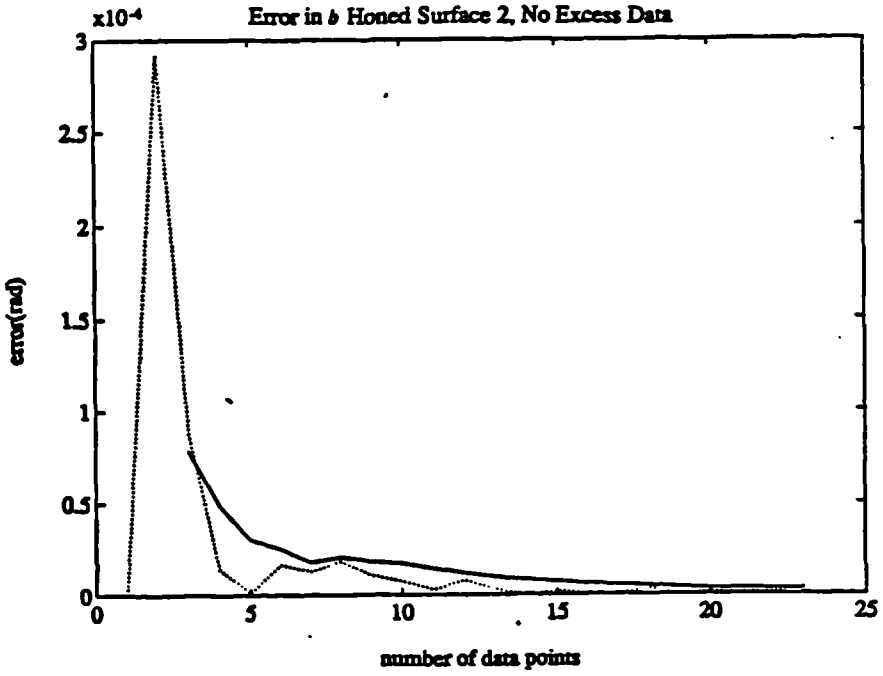


Graph 5.29 (a)

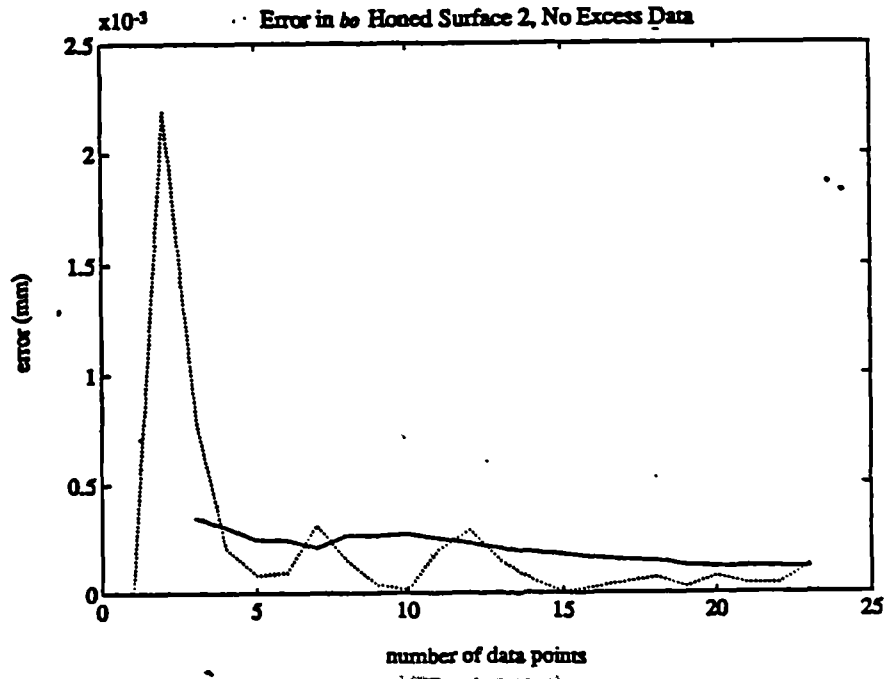


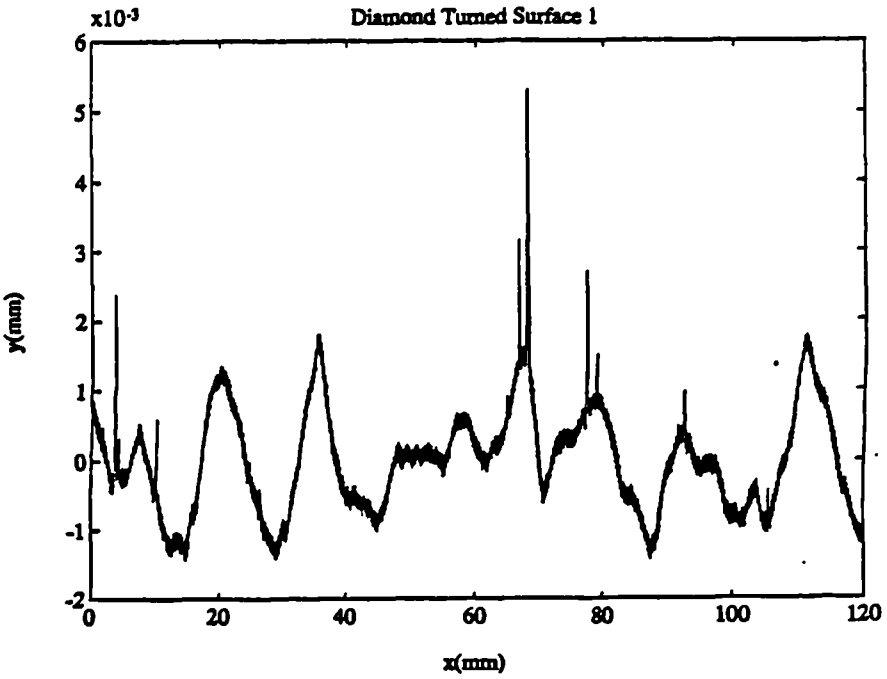
Graph 5.29 (b)

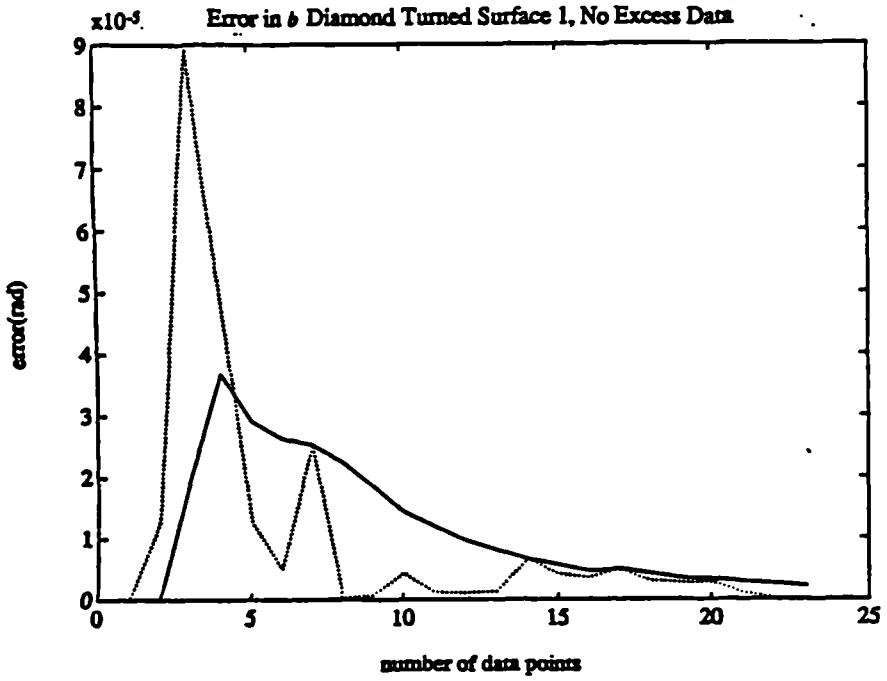

Honed Surface 2
0.01
0.005
0
-0.005
-0.01
-0.015
y(mm)
0
20
40
60
80
100
120
x(mm)


x10-4
Error in b Honed Surface 2, No Excess Data
3
2.5
2
1.5
1
0.5
0
error(rad)
0
5
10
15
20
25
number of data points

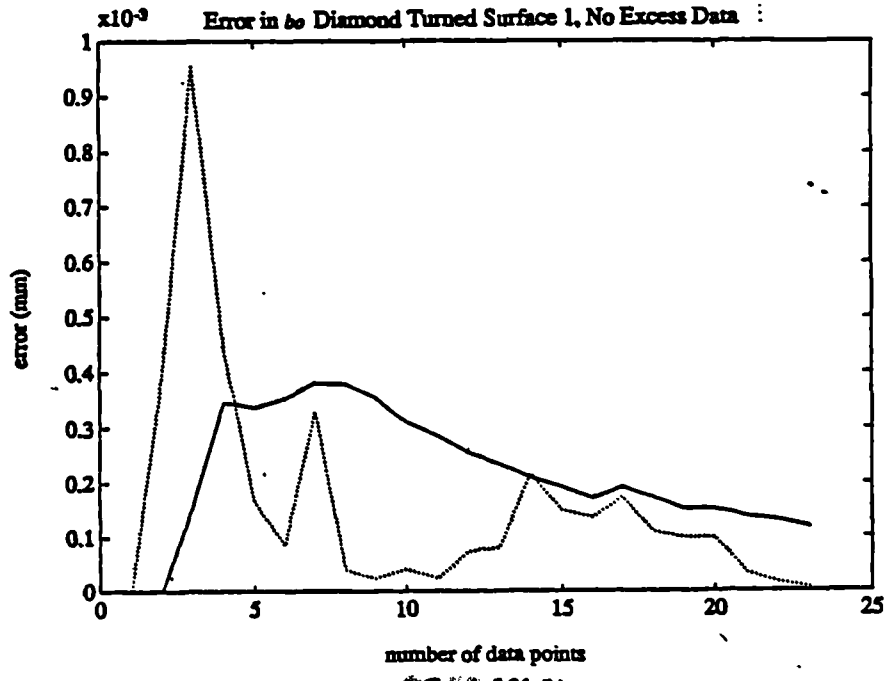
Graph 5.30 (a)

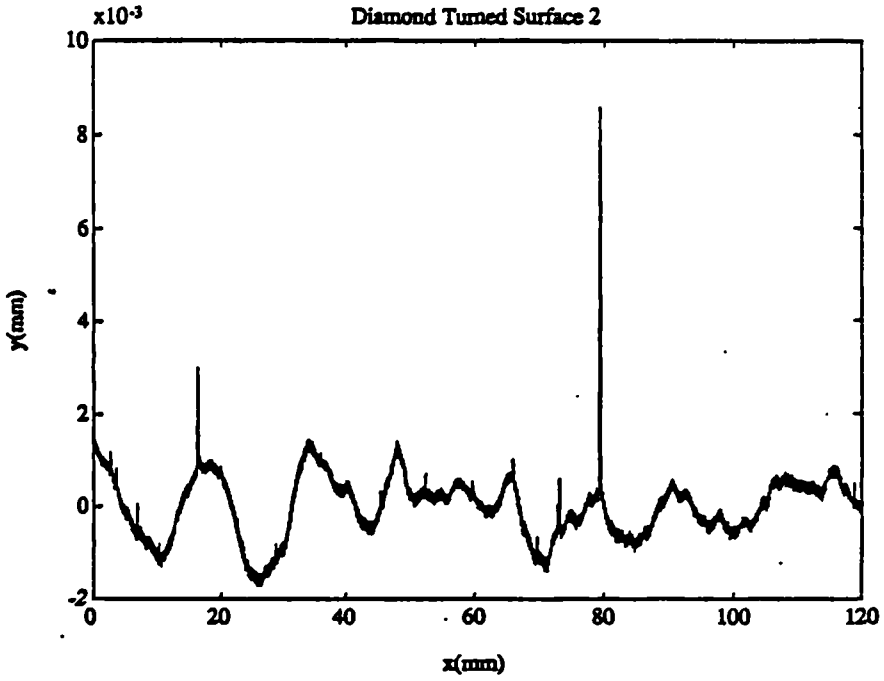

x10-3
Error in bo Honed Surface 2, No Excess Data
2.5
2
1.5
1
0.5
0
error (mm)
0
5
10
15
20
25
number of data points

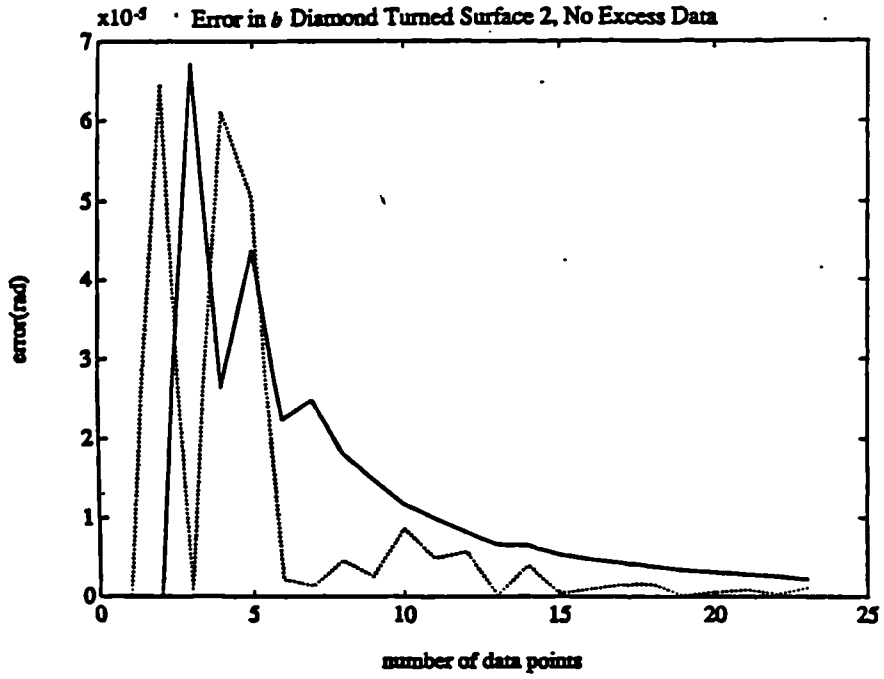

x10-3
Diamond Turned Surface 1
y(mm)
x(mm)


x10-3
Error in b Diamond Turned Surface 1, No Excess Data
error(rad)
number of data points

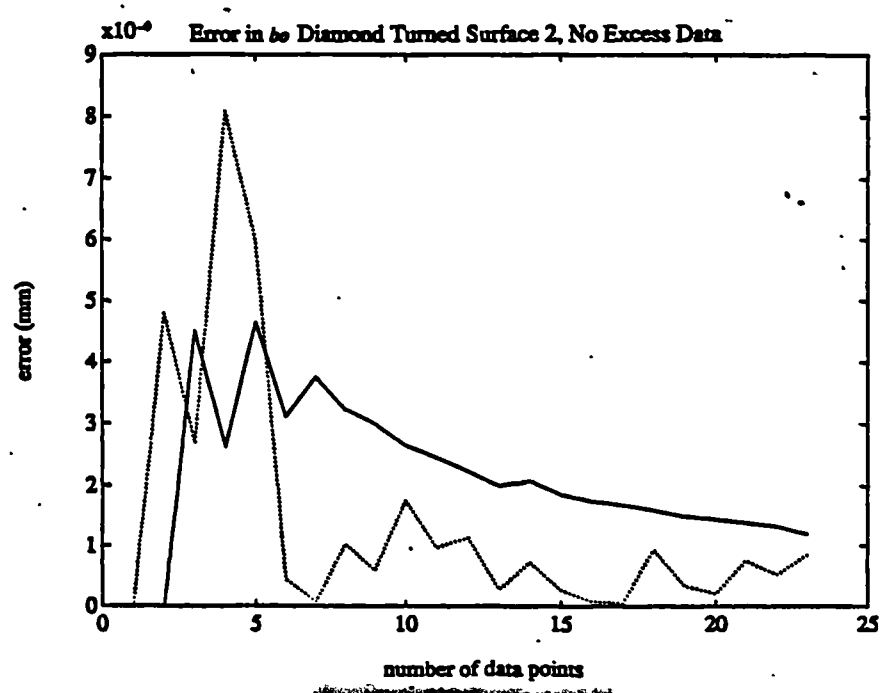
Graph 5.31 (a)

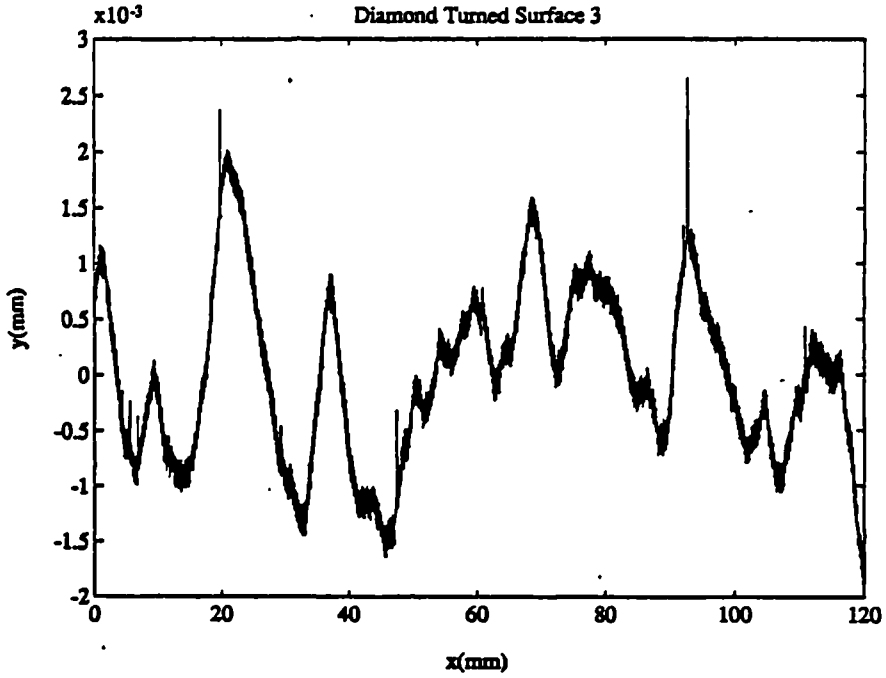

x10-3
Error in bo Diamond Turned Surface 1, No Excess Data
error (mm)
number of data points

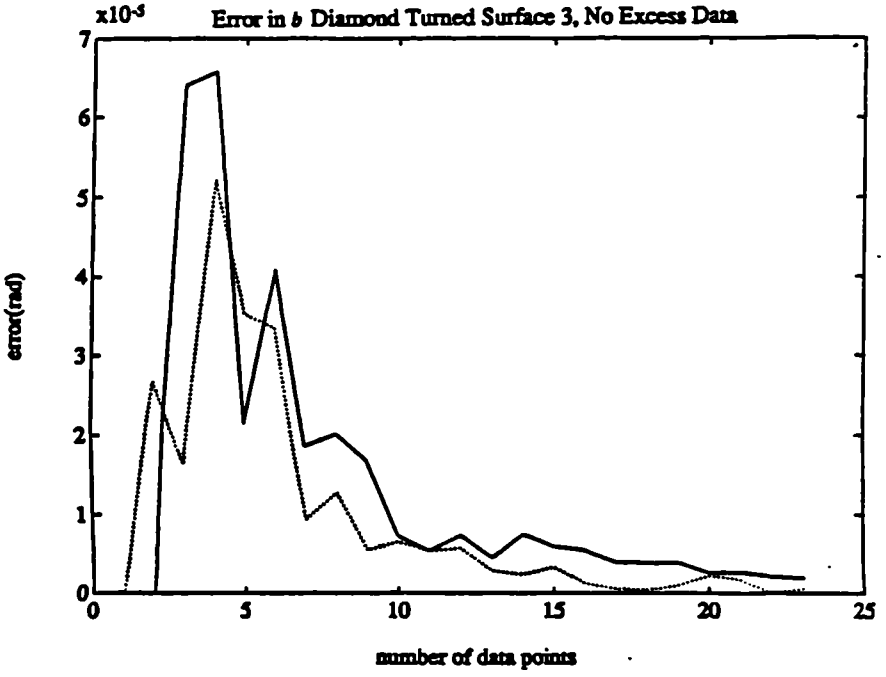

x10-3
Diamond Turned Surface 2
y(mm)
x(mm)


x10-5
Error in b Diamond Turned Surface 2, No Excess Data
error(rad)
number of data points

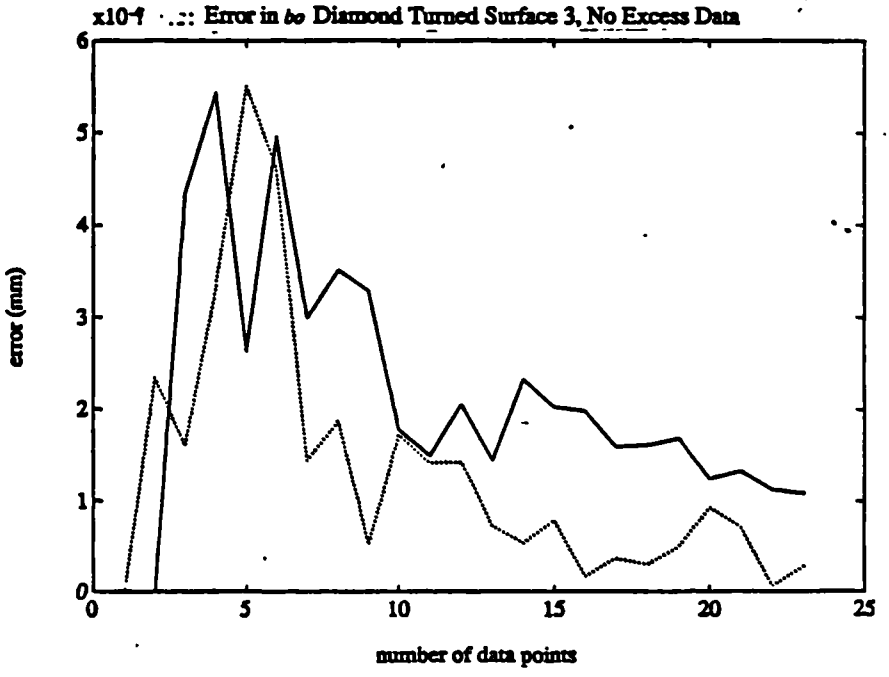
Graph 5.32 (a)


x10-6
Error in bo Diamond Turned Surface 2, No Excess Data
error (mm)
number of data points





Graph 5.33 (a)



Graph 5.33 (b)

The possible reasons for the discrepancies between the two set of data (that using all surface data and that using only a small amount of data to calculate the autocorrelation function) could be in the 'completeness' of the function. To explain, consider the auto-correlation function

$$R(\tau) = \lim_{L\to\infty} \frac{1}{L}\int_0^L z(x)z(x+\tau)dx$$

$R(\tau)$ approaches the exact autocorrelation function as $L\to\infty$ (3), using a relatively small amount of widely spaced surface data in the manner outlined above, implies that L is small and so the function is effectively less complete than if a large quantity of closely sampled data had been used.

Additionally, where the surface has a predominantly periodic character, aliasing may occur. This phenomenon occurs when a signal or surface is sampled at a relatively infrequent rate, which can result in confusion between the harmonic elements in the data (11).

4. Comparison of Sampling Errors and Typical Machine Errors

In order to obtain an idea of the significance of sampling error in coodinate measurement practice it was decided to compare them to typical machine errors. For convenience the geometric machine errors measured in chapter one are re-iterated in table 5.3 and the expected range of sampling errors are shown in table 5.4. it must be borne in mind that these machine errors were taken form a machine that had undergone no corrective treatment. It has been shown that calibration and error compensation techniques can reduce machine errors significantly, typically, linear displacement errors can be reduced to the order of ½μm (12). If similar improvements an be expected on rotational errors, then the error values shown in table 5.3 could realistically be divided by any figure between 4 and 40. Even withour error compensation it can be seen that ground surfaces can incurr errors in the estimates of b that are in the same order of magnitude as the rotational machine errors, and the other surface investigated incurr errors that are orders of

magnitude higher. The significance of the sampling errors would obviously increase as the machine errors are suitably reduced by error compensation techniques.

Approximate Sampling Error Range No Redundant Data		
Surface	b	b_o
Ground	0.7×10^{-5} to 1.5×10^{-5} rad	1.2×10^{-4} to 1.7×10^{-4} mm
	≈0 to 5×10^{-4} rad	0.001 to 0.001 mm
Endmilled	9×10^{-5} to 6×10^{-4} rad	2×10^{-3} to 3×10^{-3} mm
	9×10^{-5} to 4×10^{-4} rad	1.5×10^{-3} to 2×10^{-3} mm
Turned	1×10^{-5} to 2.5×10^{-4} rad	2.5×10^{-4} to 2×10^{-3} mm
	≈0 to 3×10^{-4} rad	3×10^{-4} to 2.2×10^{-3} mm
	1×10^{-5} to 1.5×10^{-4} rad	3×10^{-4} to 2.2×10^{-3} mm
Honed	9×10^{-6} to 4×10^{-5} rad	4×10^{-4} to 7×10^{-4} mm
	1×10^{-5} to 6×10^{-5} rad	1×10^{-4} to 4×10^{-4} mm
Diamond Turned	2×10^{-5} to 3.5×10^{-4} rad	1.2×10^{-4} to 3.5×10^{-4} mm
	2×10^{-6} to 7×10^{-5} rad	1.2×10^{-4} to 4×10^{-4} mm
	2×10^{-6} to 5×10^{-5} rad	1×10^{-4} to 5.6×10^{-4} mm

Range of Approximate Machine Errors		
Pitch	X	≈0 to 10×10^{-5} rad
	Y	0.4 to 7×10^{-5} rad
	Z	≈0 to 1.5×10^{-4} rad
Yaw	X	0.4 to 4×10^{-5} rad
	Y	0.1 to 5×10^{-5} rad
	Z	0.3 to 2.5×10^{-5} rad
Roll	X	−2 to 3×10^{-5} rad
	Y	2 to 7×10^{-5} rad
Displ.	X Y Z	22μm 18μm 30μm

Table 5.4

5. Discussion

Hitherto is has not been possible to obtain any idea of the accuracy of the least squares estimates of any given parameter. The use of the algorithm developed in this chapter enables CMM practitioners to assure the accuracy of the estimates of two prime parameters. This effectively means that it is now possible to quantitatively validate these

procedures regardless of the nature of the surface profile from which the coordinate data are taken. Additionally, with regard to the digital implementation of definitions of geometric form, manipulation of the algorithm would provide a method whereby it is possible to determine the quantity of data that would be required in order to satisfy predetermined limits of accuracy. To achieve this, the algorithm could be used interactively; as the CMM gathers the data, a calculation of the accuracy of the parameter estimate could be made between the collection of each point, the process stopping when a required accuracy is achieved. Standards could refer to algorithms of this type to ensure that definitions of form were being interpreted with controlled accuracy, as Cox (1) stipulated they should. The optimum distribution of data (the most judicious placing of data in order to achieve maximum accuracy of estimate) could also be found by using the algorithms in a constrained (by dimension of component surface and max. feasible number of data points) iterative minimisation routine.

However, this algorithm is merely the begining, the results of the investigation in chapter 3 illustrate that sampling error behave very similarly for the least squares estimates of the parameters of all of the geometric elements investigated. These results, but more importantly, the parallel development of the algorithm for the expression for the sampling errors in b and bo , strongly indicate that similar solutions for the quantification of sampling error in the estimation of these other parameters would be along very similar lines. This work has shown catagorically that the use of the autocorrelation function provides a feasible way forward in the development of a whole species of algorithms for the validation of all least sqaures inspection processes. The crux of this claim is that it is now proved that the raw (unmodelled) autocorrelation function calculated from a relatively sparse data set can provide a sufficient measure of spatial variation for these purposes.

The continuation of this work is vital if the present difficulties in CMM inspection are to be alieviated. The formulation of similar algorithms for other geometric parame-

ters would result in the ability to quantify the errors incurred in virtually all inspection processes, and the replacement of traditional gauging methods by the ability to inspect to within specific tolerances.

Some CMM manufacturers evidently took the view that their software inaccuracy problems were to be solved by the introduction of CMMs having servos with the capability of profiling. This is an extremely inefficient method of inspection because if profiling is to be undertaken in a manner that will not either result in inaccuracies in the data or significant stylus damage (13), the speed of profiling may be such that inspection speed, which is of utmost importance in coordinate measurement, is seriously impaired. This problem would also be exacerbated by the processing of extremely large quantities of data. Even so, for all practical purposes, it is impossible to profile over an entire surface, and the question of how many profile traces would be neccessary to achieve the estimation of a three dimensional parameters with a required accuracy.

It may also be circumspect to consider the uniquness of approach in this algorithm, and some of the possible implications in other aspects of Metrology.

Firstly, this is the first instance (to the author's knowledge) that the autocorrelation function has been used to bridge the gap between theory and practice in its raw state, i.e. no ideallised models have been used to predict the behaviour of surface borne parameters. This implies that models may not always be necessary in the assessment or characterisation of surface parameters, and that the manipulation of the actual autocorrelation function may be an alternative, less limiting approach.

Additionally, the fact that a working algorithm was possible using only very few data points indicates another significant discovery in that a remarkably good indication of spatial variation can be achieved using only very few data points.

References

[1]: M.G. COX & K. JACKSON (1983): "Algorithms and Software for Metrology":

A Statement of Need; NPL Report MOM65, June 1983.

[2]:W.H. RASNICK & N.E. ZURCHER (1986): "Soft Functional Gauging on Coordinate Measuring Machines"; Pre-print for submission to Prec-Eng Conf. Thresholds for Precision Engineering, Dallas, Texas, November 1986.

[3] J.S. BENDAT & A.G. PIERSOL (1971): "Random Data : Analysis and Measurement Procedures"; Wiley-Interscience.

[4] D.J.WHITEHOUSE & M.J. PHILLIPS (1978): "Discrete Properties of Random Surfaces"; Phil. Trans. Roy. Soc., Vol. 290, pp267-298.

[5] J. PEKLENIK (1967): "New Developments in Surface Characterisation and Measurements by Means of Random Process Analysis"; Proc. Instn. Mech. Eng, Vol 182 pt 3K, pp 108-126.

[6] D.J. WHITEHOUSE & J.F. ARCHARD (1970): "The Properties of Random Surfaces of Significance in their Contact"; Proc. Roy. Soc. A316, pp 97-121.

[7] H. ISHIGAKI & I. KAWAGICHI (1981): "Effect of a Skid on the Accuracy of Measuring Surface Roughness"; Wear, Vol. 68 , pp 203-211.

[8] The Form Talysurf, (Manual) Rank Taylor Hobson, Leiscester, England. .

[9] T.R. THOMAS (1982): "Rough Surfaces"; Longman.

[10] H. HINGLE (1985): M. Phil. Thesis University of Warwick.

[11] D.E. NEWLAND (1975): "An Introduction to Random Vibration and Spectral Analysis"; Longman .

[12]:H. KUNZMANN & F. WALDELE (1985): "Two Numerical Error Correction Methods for Coordinate Measuring Machines"; NPL Conference Proc. on Software for Coordinate Measuring Machines.

[13] D.J. WHITEHOUSE (1988) : "A Revised Philosophy of Surface Measuring Systems"; Proc. Instn. Mech. Engrs. Vol 202 no C3 pp 169-189.

Mechanical Filtering Effects of Probe Stylus

1. Introduction

The contacting stylus on a CMM is not an infinitely sharp mathematical point, but a nominal sphere of finite dimension. Therefore, a complete penetration of the surface irregularities is not possible and the profile obtained is effectively an envelope of the profile of the true surface characterisitics (1), (the phenomenon is illustrated in figure 6.1.) As the radius of the probe increases, the minute irregularities present in the profile gradually disappear and when the stylus is of such dimension that only the peaks of the surface are contacted, the profile is transformed into a curve made up of circular arcs.

The intent of this chapter is to investigate the effects these phenomena upon the estimation of sampling error, and to ultimately illustrate how an algorithm can be developed to specify optimum probe dimension with respect to the minimisation of these errors.

The effects of these distortions upon the measurement of surface texture parameters have been investigated (2,3,4) , and it was implied that this effect was a function of the radius of the contacting member, which is either a stylus or a skid. From these findings it may be tentatively assumed that mechanical filtering will have some significant effect upon the estimation of other parameters, and that this effect is in some way dependent upon the radius of the contacting probe.

2. Experimental Procedure

In order to assess the effects of mechanical filtering profile distortions upon the estimation of sampling error it is first neccessary to calculate the profiles that would be

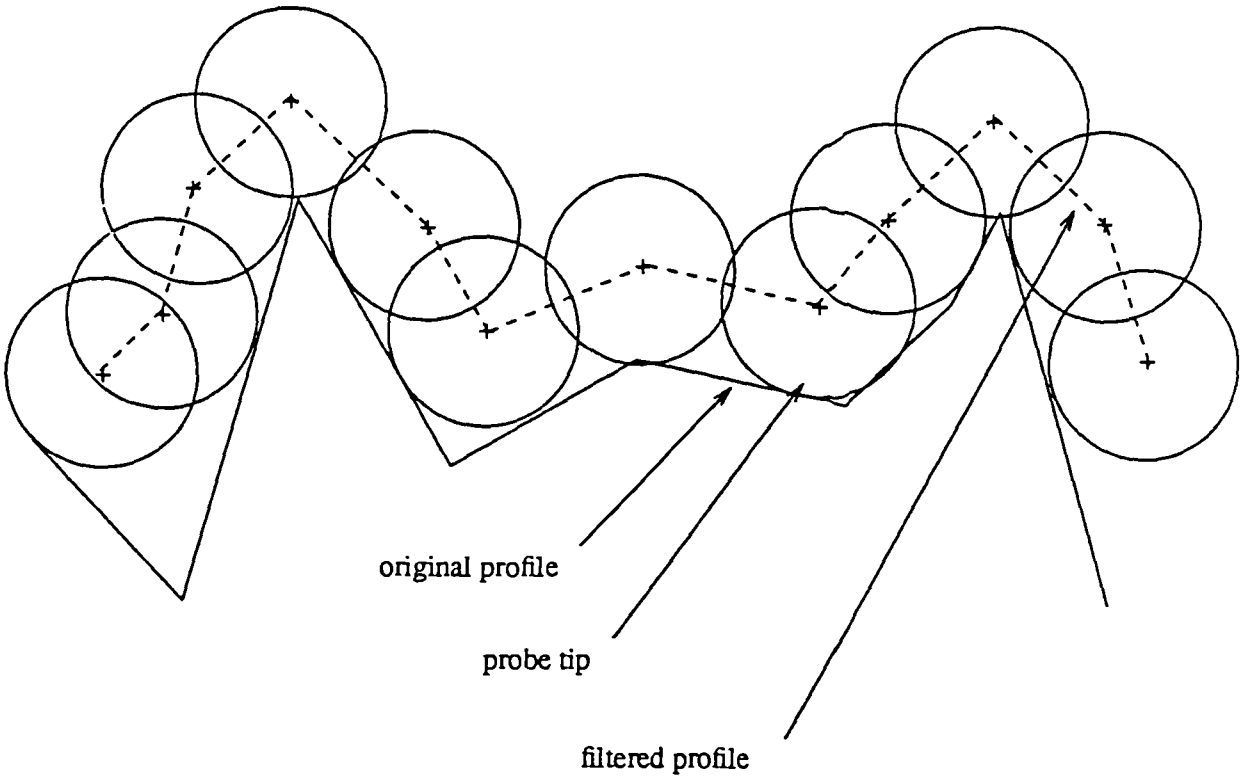


Figure 6.1

obtained by different probe radii. Several established methods for doing this exist, but are unsiutable for this particular application for the reasons outlined below.

2.1. Mechanical Filtering

Whitehouse (5) outlines a method whereby the averaging effect is simulated by replacing point to point variations of the surface by a set of averaged measurements, each average measurement requires a probability distribution and a mean height equal to the profile. Just enough averaged measurements have to be taken to encompass the tip dimension and shape, e.g for a circle this would be three. The surface has effectively been repalced by a three element Markov chain. This method is only applicable if the autocorrelation function of the surface or its probability of amplitude distribution may be assumed. In cases whereby the stylus is negligibly small compared to machine tool striations, this method can be implemented as the contacting surface can then be assumed Gaussian, but this is not the case for the dimension of probe commonly used on a CMM.

An alternative strategy was used by Radhakrishnan (2) to simulate mechanical filtering when investigating the effect of stylus radius on roughness measurement. In this method, the peaks of the profile were taken as a series of starting points. The enveloping circle ordinates for the stylus tip circle of the required radius were then computed. This envelope circle segment was placed on one of the peaks and the new ordinates around the apex of the peak were found. The segment was then placed on each successive ordinate to the right of the starting peak until a valley ordinate was reached, each time retaining only the highest ordinate obtained at a particular horizontal location. The procedure was repeated on the left hand side of the peak. The next starting peak was then taken and the entire process repeated. This algorithm was tried on surfaces taken by the Rank Taylor Hobson 'Form Talysurf' for an appropriate range of CMM probe dimensions, but with poor results. The profiles supposedly representing traverses made by probes between 1 and 2.5mm were still similar to the original data. The main reason for this is probably due to the nature of the surface data. Surface profiles as measured by surface measuring

instruments are not represented as a continuum but by a discrete set of data, approximately 1400 data points in the case of the Form Talysurf which are distributed over surfaces of 20 to 120 mm , such that the data are very sparse compared with those used in the shorter traverses common in surface texture assessment, which implies that the intervals between the data are comparatively large. This gives the problem that the height of the enveloping circle at the adjacent horizontal location to the starting peak will be a considerable distance below the surface height at the same ordinate. This problem could have been resolved by interpolation of the surface data, this algorithm was tried but required far too much computation time.

Due to the drawbacks outlined above an original approach was sought. As the analysis is in two dimensions only, the probe was modelled as a circular data set with its centre coordinates corresponding to each horizontal ordinate and at a vertical location 0.75 times the radius below the highest point on the surface data sub-set placed centrally about the circle centre and of a length twice the radius of the probe. The centre of this sphere was then moved upwards incrementally (50 nm), until all of the surface data was outside the set described by the sphere (fig 6.2), i.e :

all

$$\varepsilon_i \leq 0 \tag{6.1}$$

where

$$\varepsilon_i = r - ((x_i - a)^2 + (y_i - b)^2)^{1/2} \tag{6.2}$$

where x_i , y_i are the coordinates of each of the surface data subset, and a, b are the coordinates of the centre of the circle. The process was repeated with the centre of the probe at horizontal locations corresponding to each of the surface data ordinates.

All of the simulation and calculation procedures were performed on a SUN 3/80 computer using the MATLAB software package.

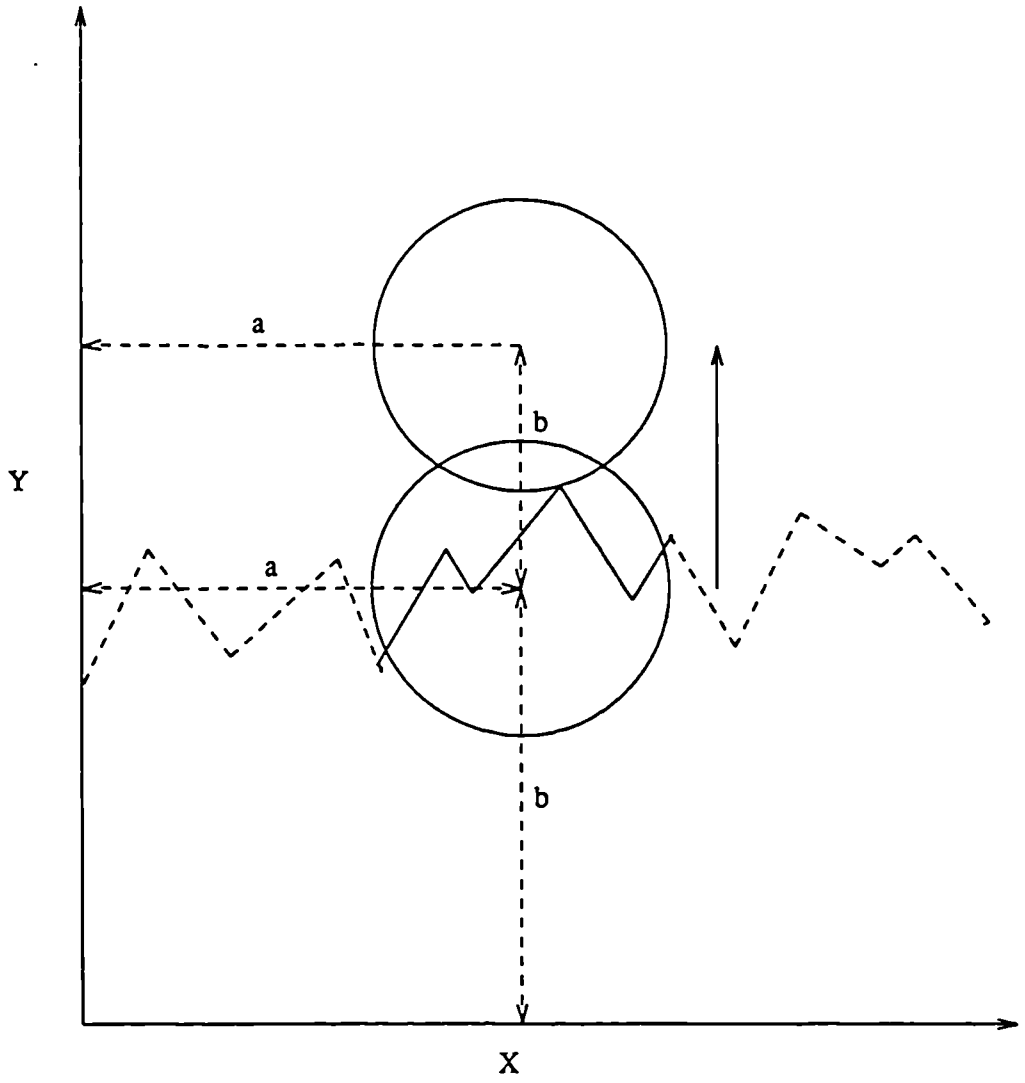
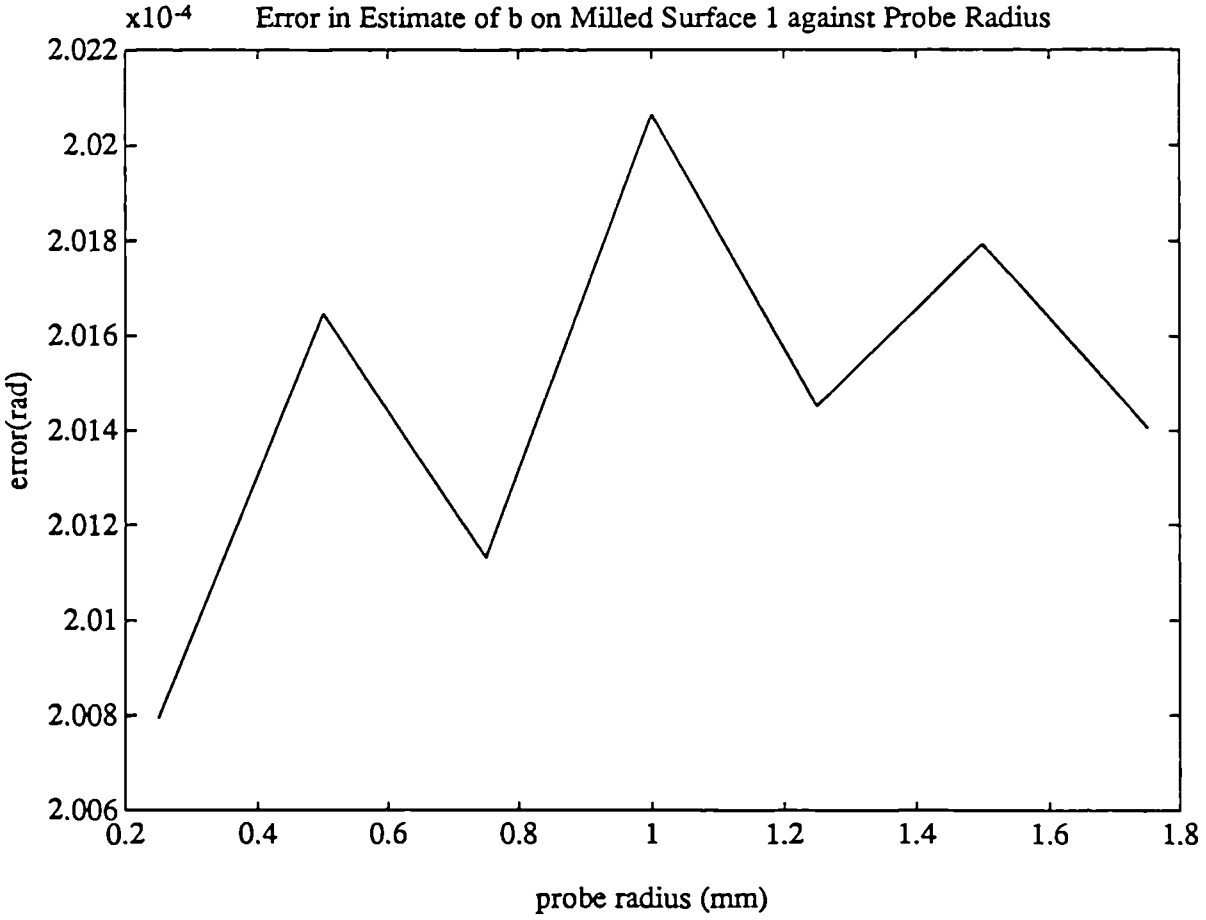
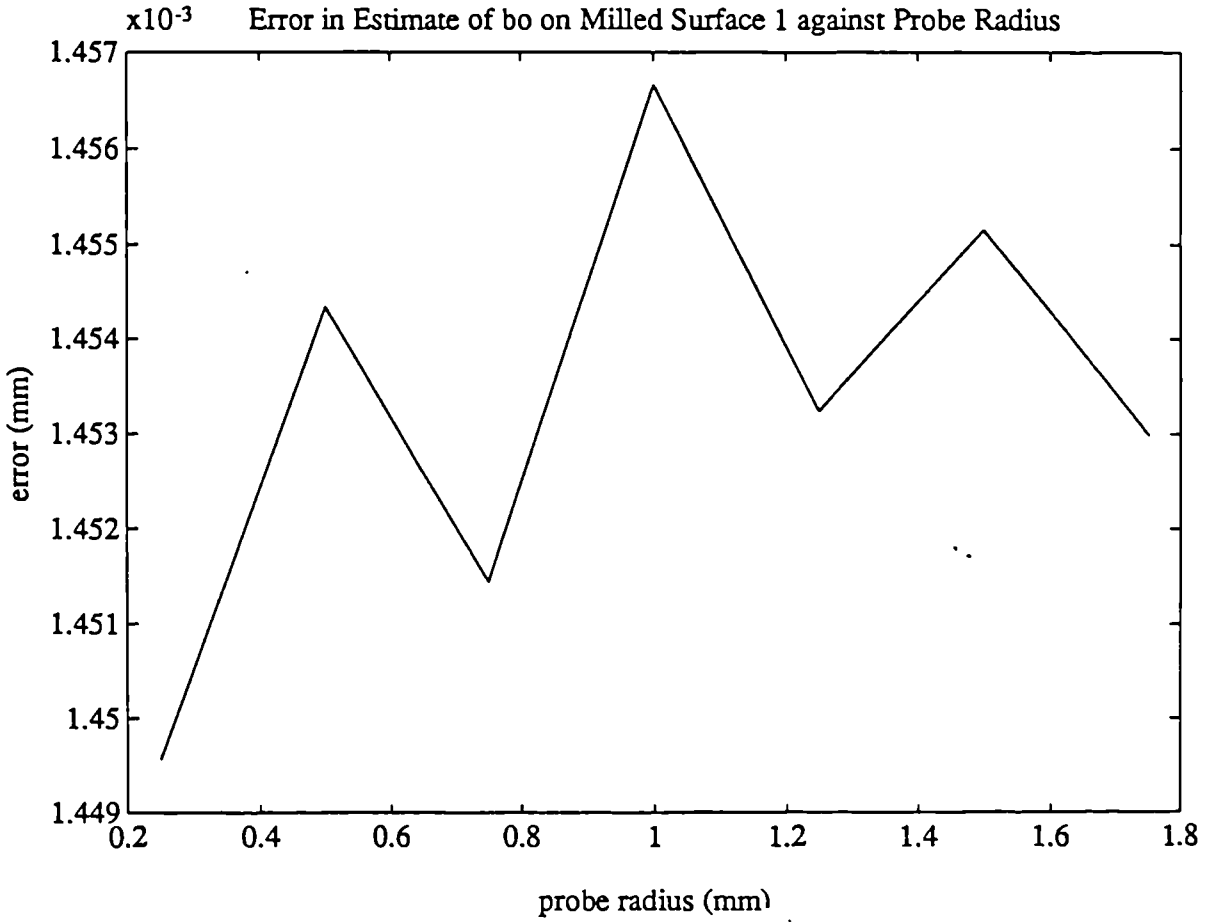


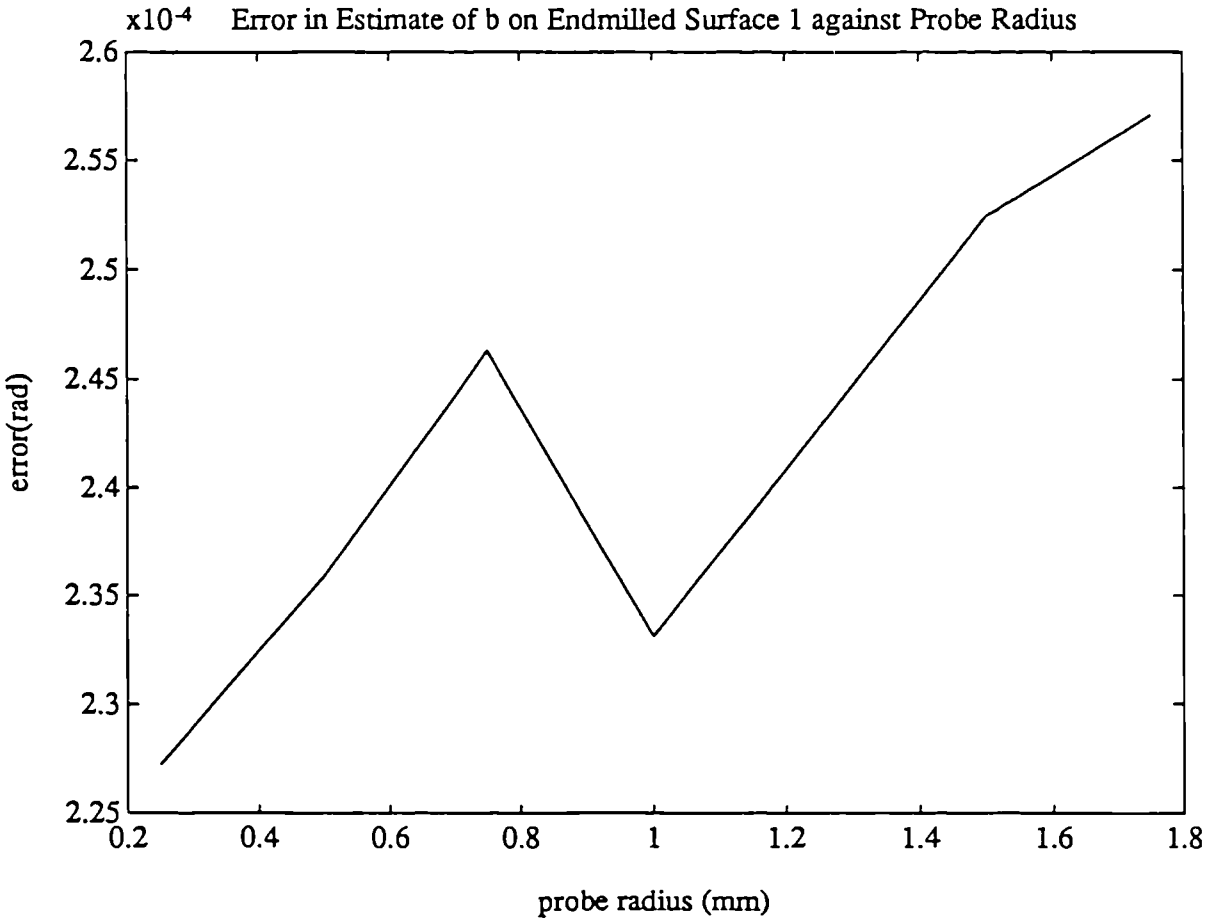
Figure 6.2



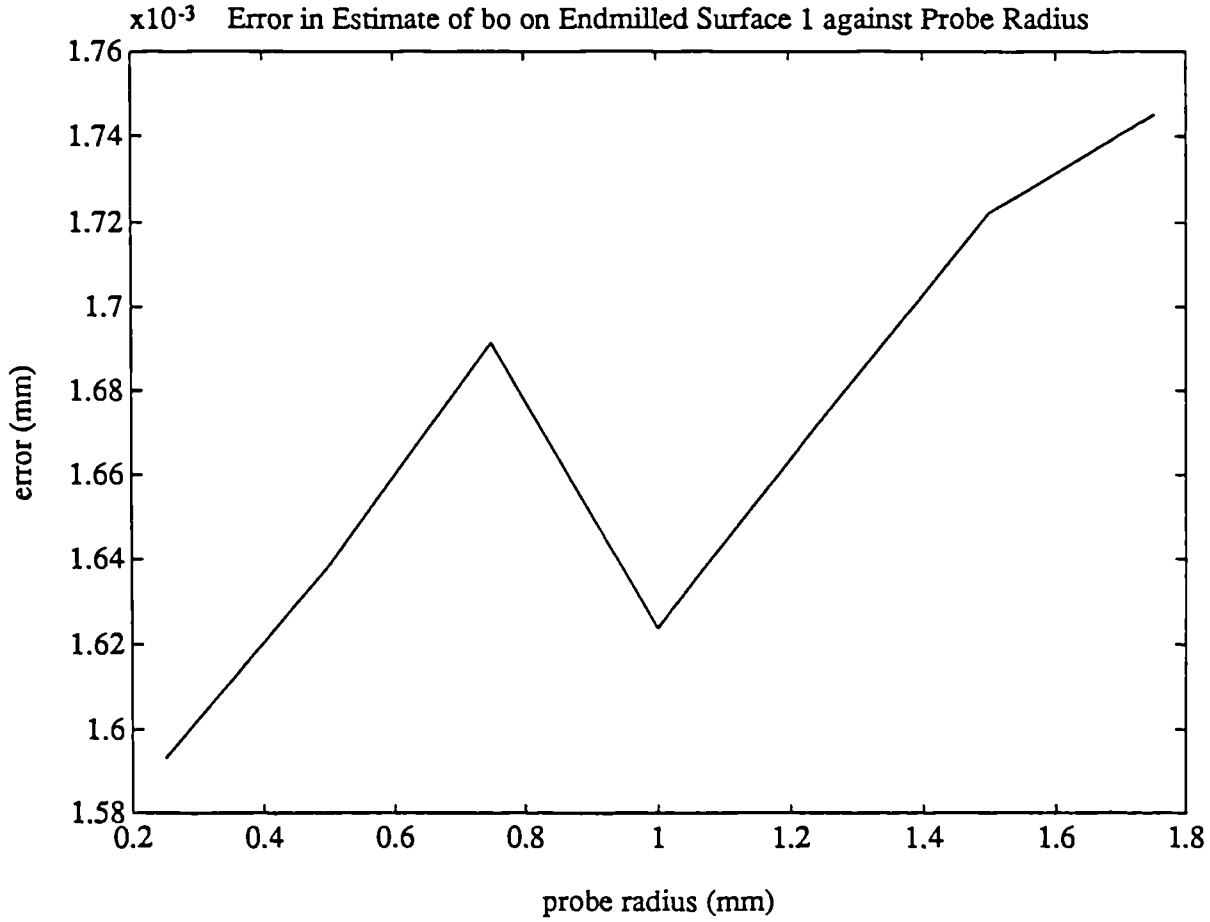
Graph 6.1 (a)



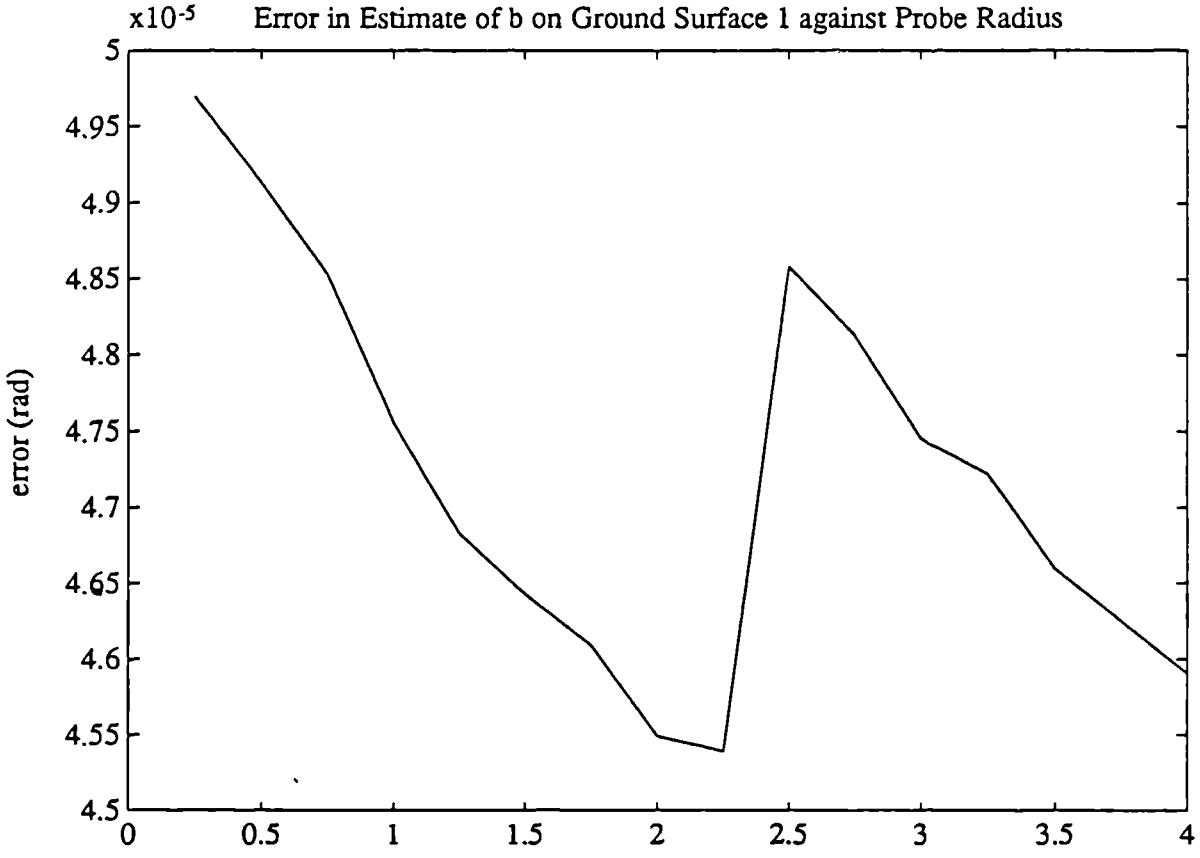
Graph 6.1 (b)



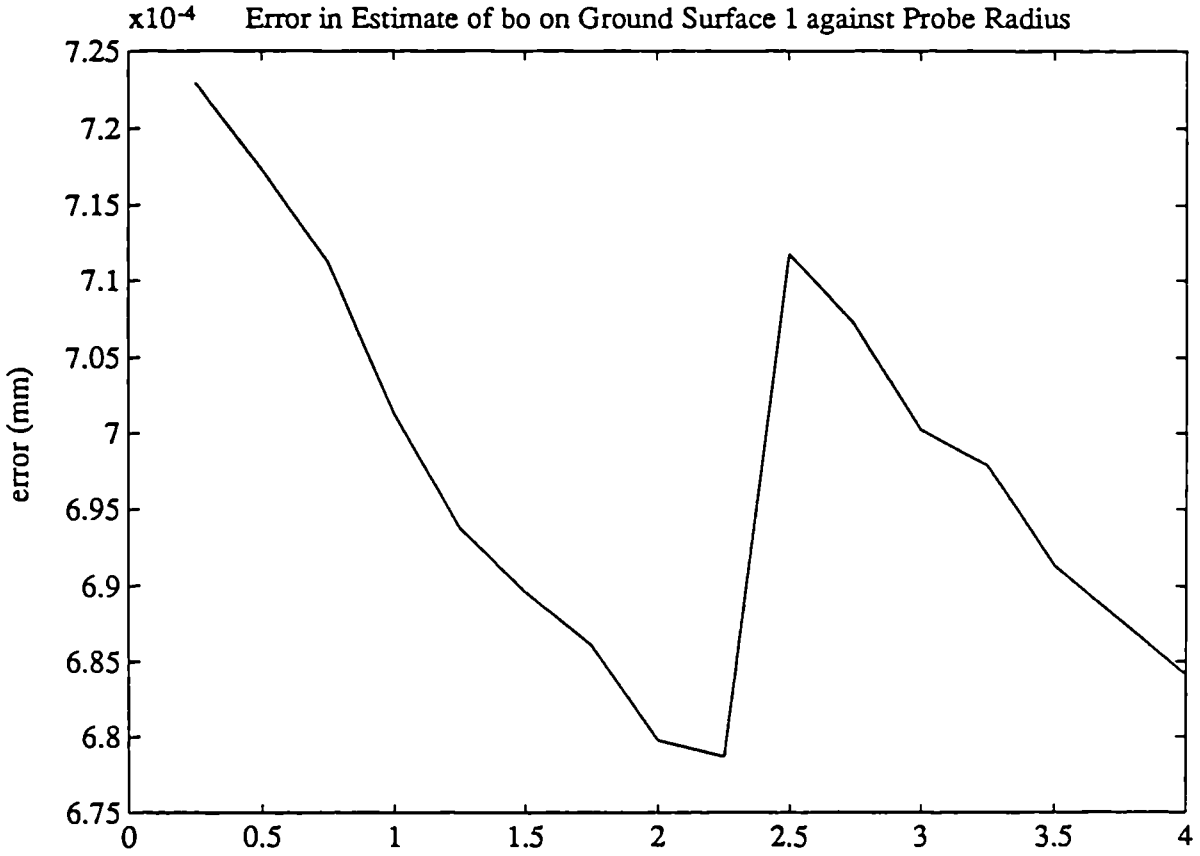
Graph 6.2 (a)



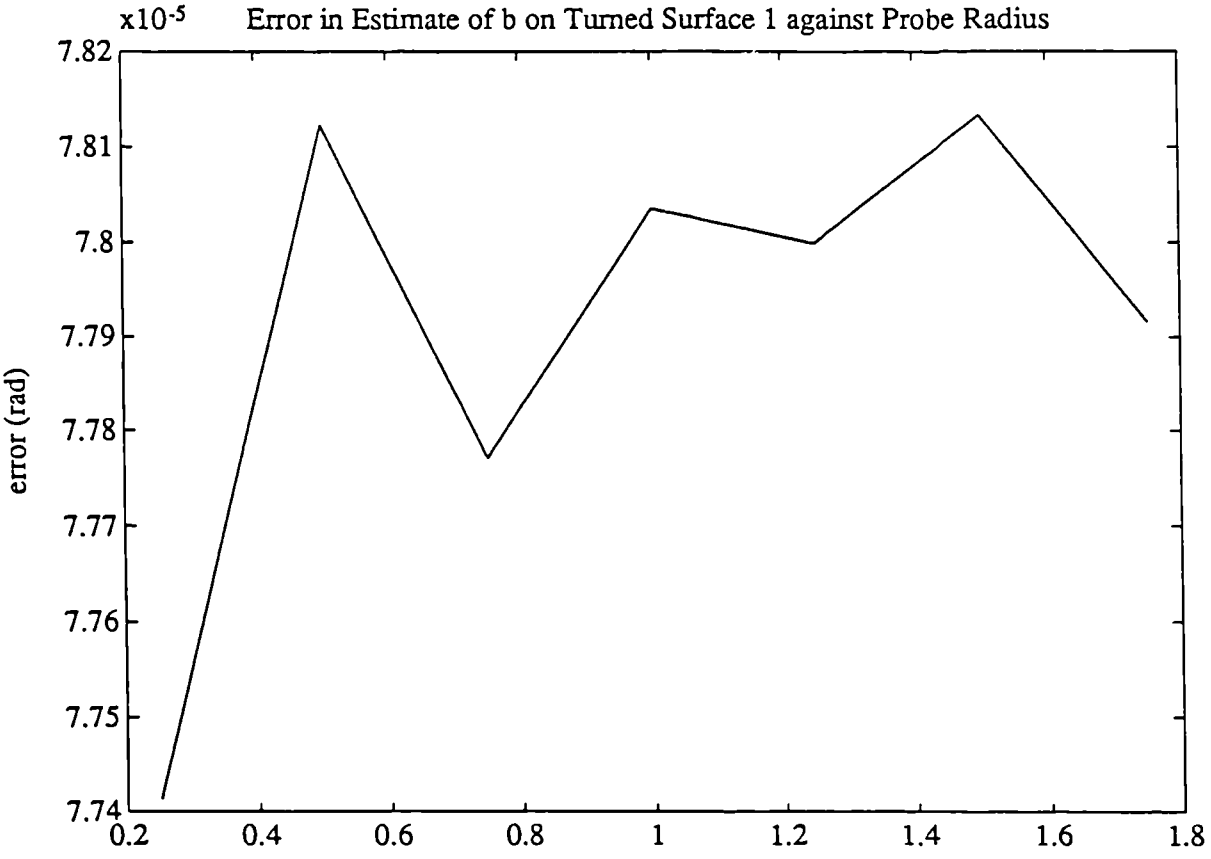
Graph 6.2 (b)



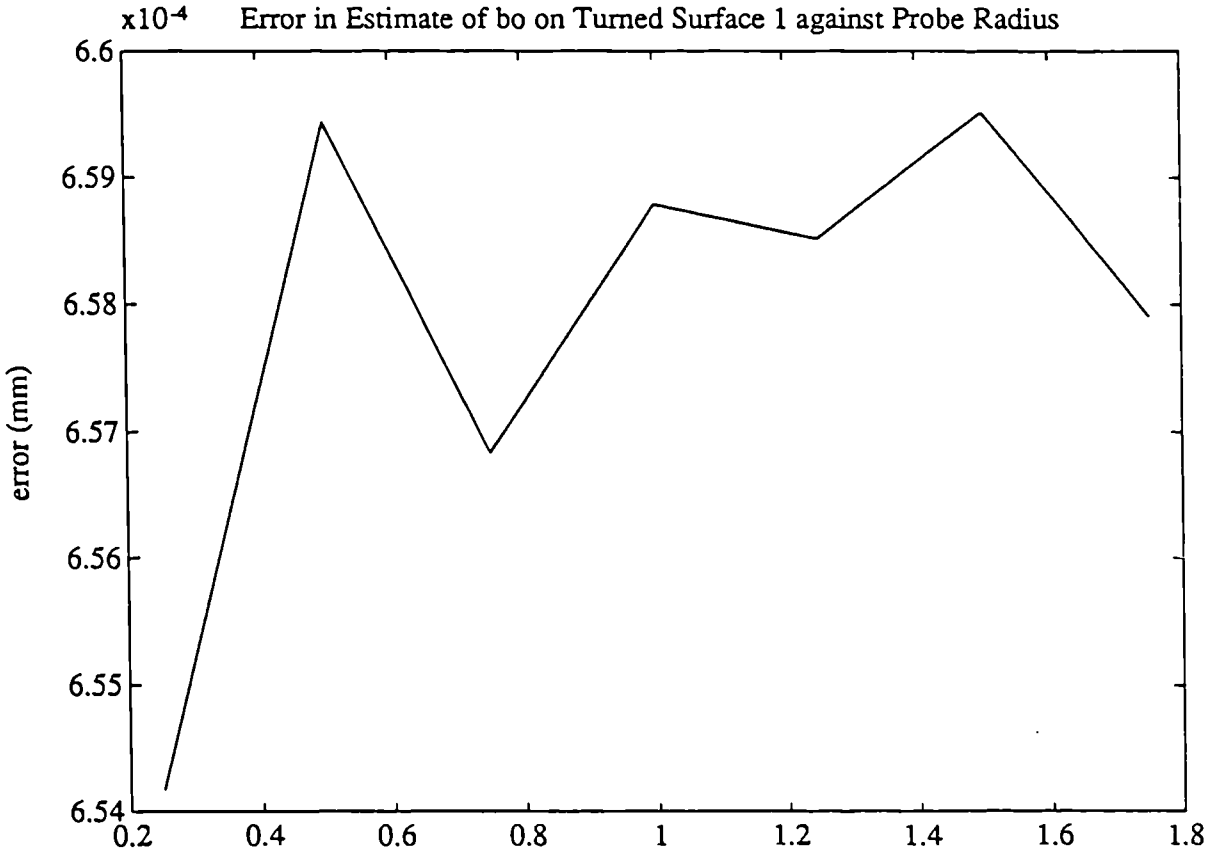
Graph 6.3 (a)



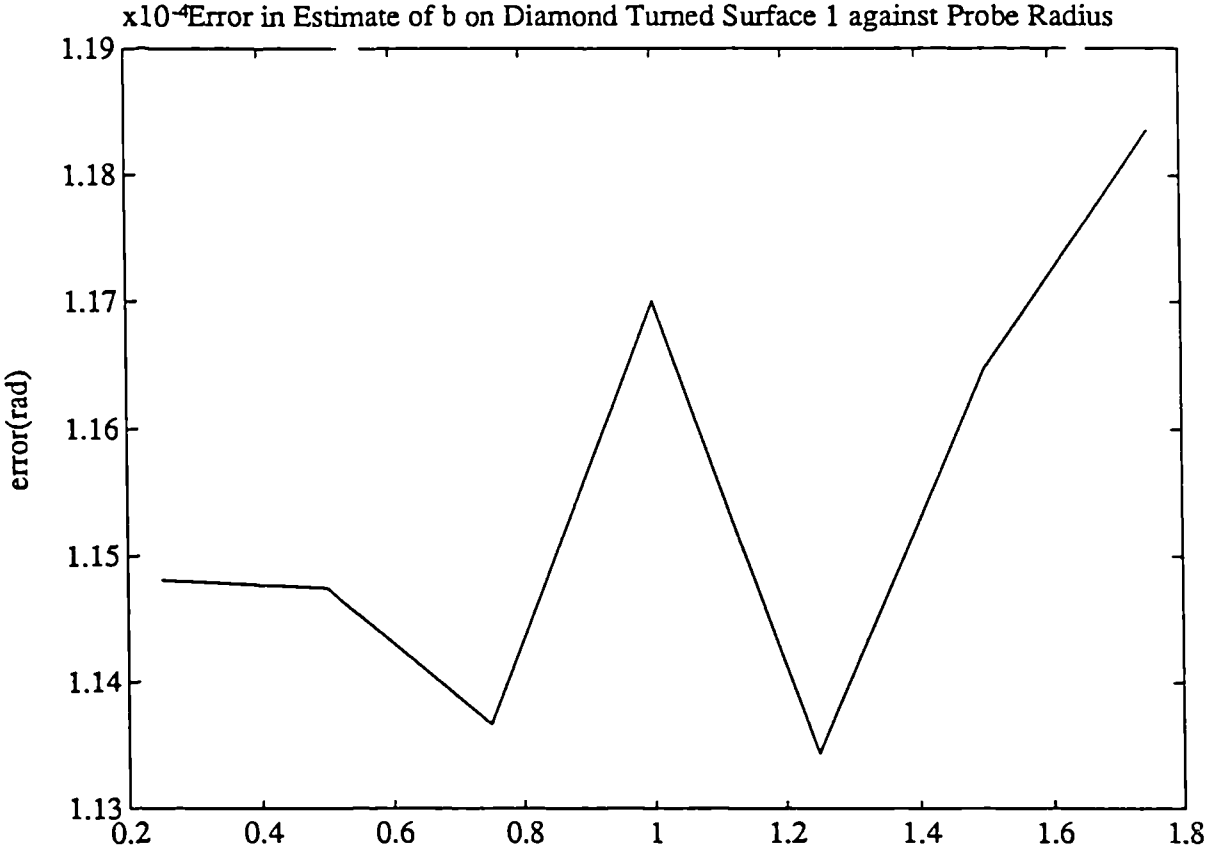
Graph 6.3 (b)



Graph 6.4 (a)



Graph 6.4 (b)



Graph 6.5 (a)

Graph 6.5 (b)

2.2. Calculation of Sampling Errors

Sampling errors in the least-squares estimates of slope and intercept were calculated according to the algorithm developed in chapter 5 using no redundant data with the parameters outlined in table 6.1, where dx:denotes the sampling interval and n is the number of data points used.

Data and Parameters				
Surface	length (mm)	dx (mm)	n	radii (mm)
Milled 1	49.983	2	12	0.25-1.75
Endmilled 1	69.955	2	12	0.25-1.75
Ground 1	119.950	5	12	0.25-4
Turned 1	30.246	2	12	0.25-1.75
Diamond Turned 1	119.951	5	12	0.25-4

Table 6.1

2.3. Discussion of Results

It is interesting to note that changes in the data have a similar effect on the estimates of each parameter. This property is extremely useful as the optimum probe geometry for a particular surface would remain the same for the calculation of each parameter. Had this result proved differently,then a method for optimal probe selection/design would, for all practical purposes have been unworkable as different probe geometries would be suited to the collection of data for the calculation of each different parameter.

3. Development of Algorithm for Minimisation of Sampling Error by Optimum Probe Dimension

Figures 6.3 to 6.7 clearly illustrate that the change in sampling error is not a direct function of probe radius. These results imply that in itself, probe dimension is not an reli-

able descriptor of the degree of mechanical filtering incurred by the surface data and cannot therefore be used as a parameter against which sampling error can be related. It is therefore used as the basis of an algorithm by which sampling error can be minimised by optimum probe dimension selection minimising sampling error. It is likely that the nature of the surface characteristics should also be taken into account. It would therefore be appropriate to relate the sampling error to an alternative parameter encompassing the combined interactive effects of the probe and the surface profile. This parameter would have to be an expression of the differences in the sets of filtered data in comparison to the original data, a suitable term to employ would be the maximum value of the cross correlation function of the two data sets . This term has the additional advantage of being readily calculable by several methods.

3.1. The Cross-Correlation Function

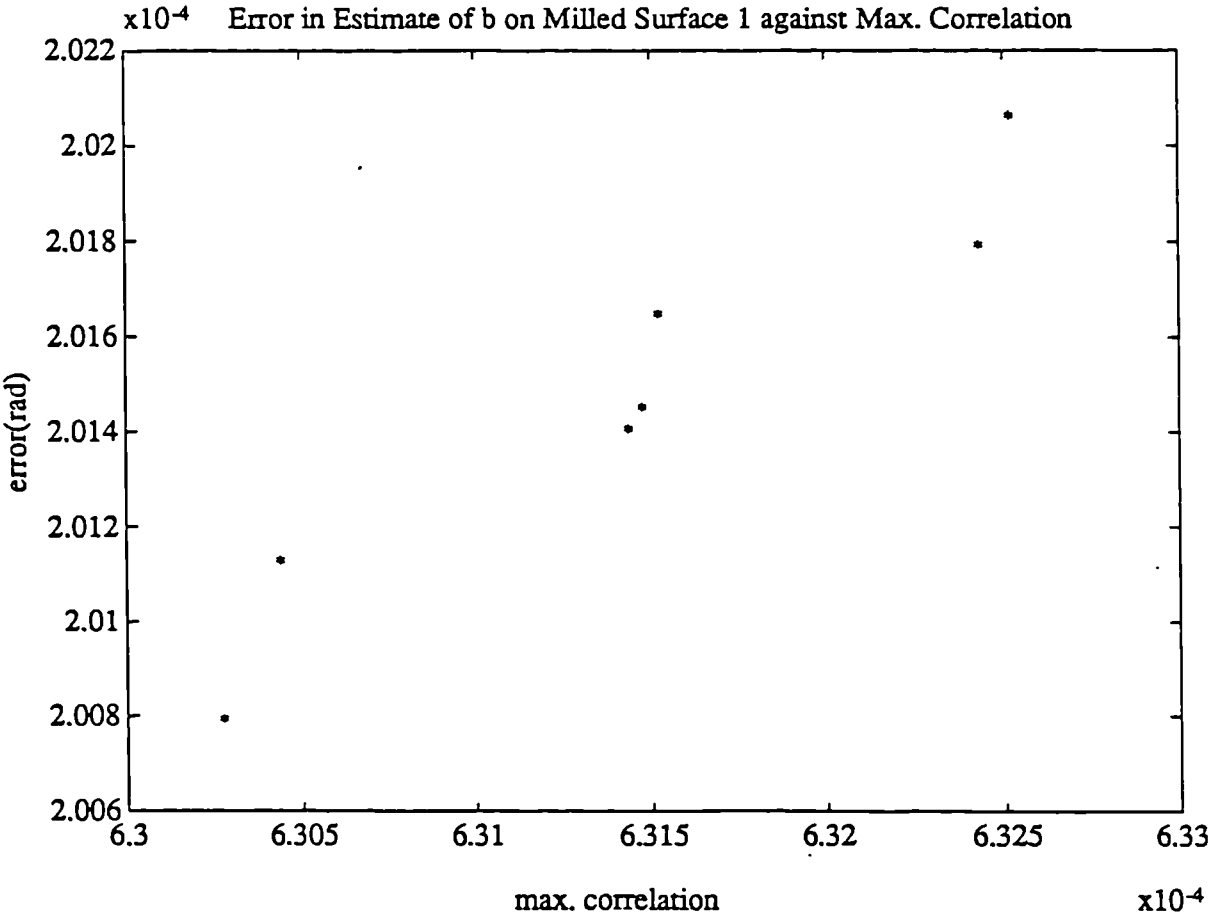
The cross correlation function of two data sets describes the general dependence of the values of one set of data on the other. If a pair of surface profiles $z_1(x)$ and $z_2(x)$ are considered, an estimate for the cross correlation function of the values of $z_1(x)$ at horizontal location X and $z_2(x)$ at a position of $(X + \tau)$ may be obtained by taking the average product of the two values over the sample length L, exactly as is done for the autocorrelation function. The resulting average product approaches an exact cross correlation as $L \to \infty$ i.e.

$$R_{z_1 z_2} = \lim_{L \to \infty} \int_0^L z_1(X) z_2(X + \tau) dX \tag{6.3}$$

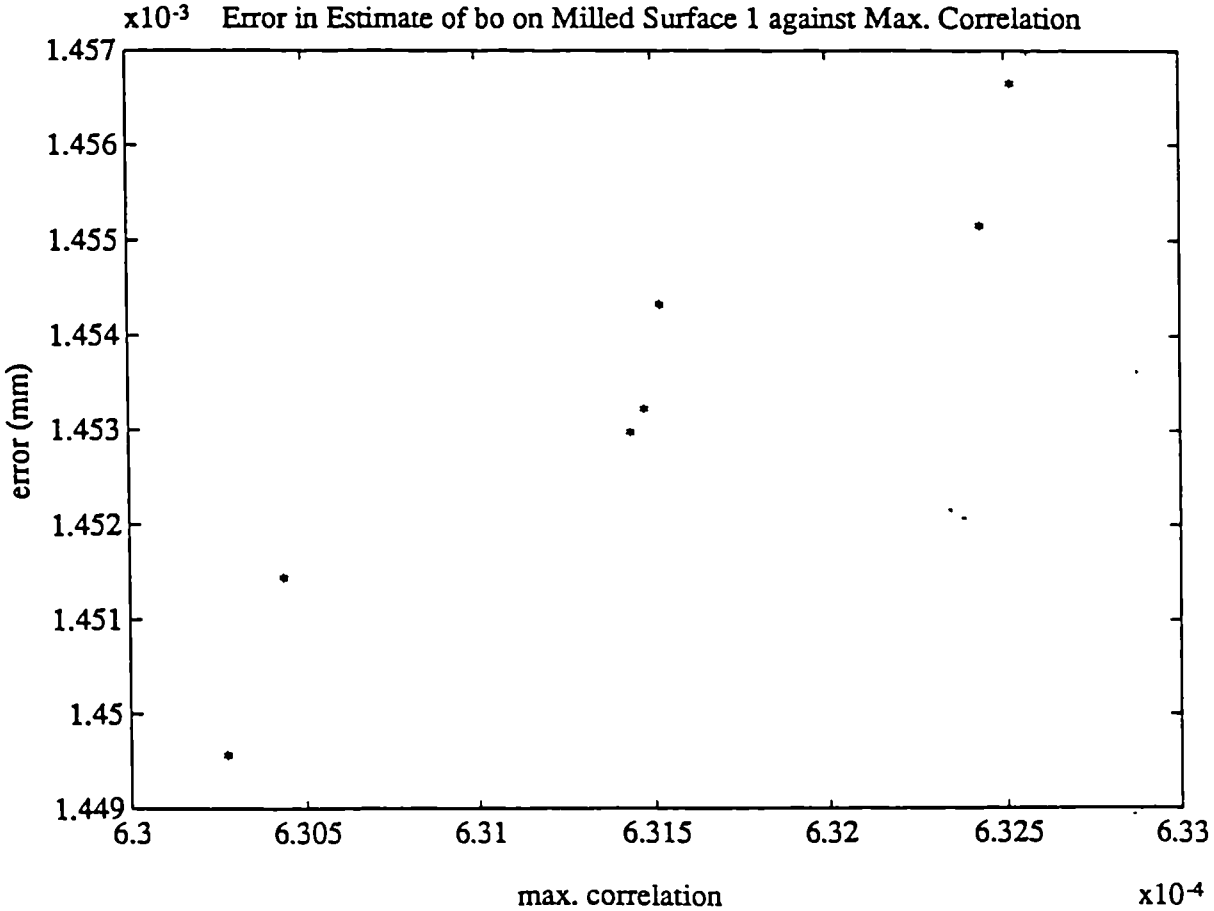
This function is always real valued and may be either positive or negative. Furthermore, $R_{z_1 z_2}(\tau)$ is not an even function as is the case for the autocorrelation function. However, a symmetry is displayed when z_1 and z_2 are interchanged such that :

$$R_{z_1 z_2}(-\tau) = R_{z_2 z_1}(\tau) \tag{6.4}$$

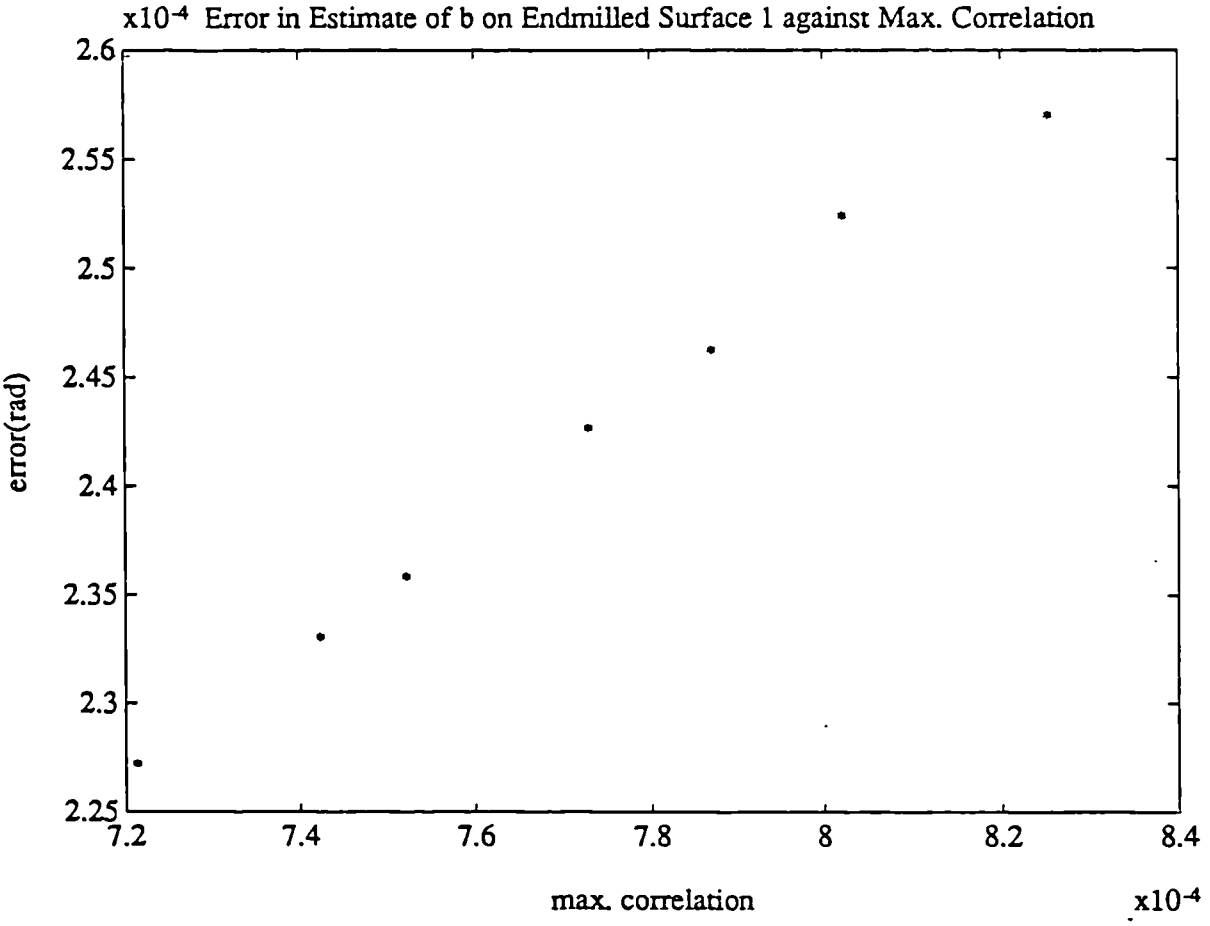
when $R_{z_1 z_2} = 0$, $z_1(X)$ and $z_2(X)$ are uncorrelated. If $z_1(X)$ and $z_2(X)$ are stastistically independent then $R_{z_1 z_2}(\tau) = 0$ for all X (6).



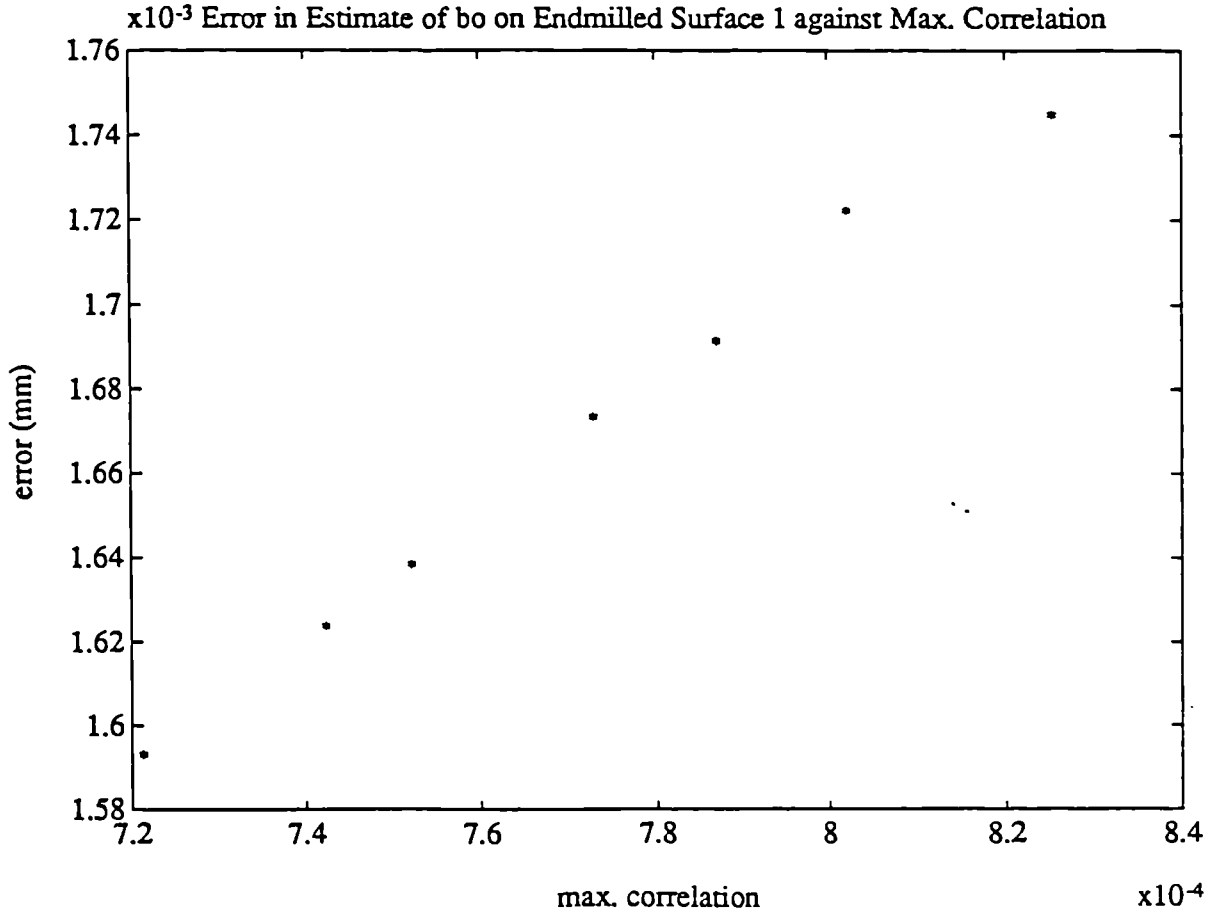
Graph 6.6 (a)



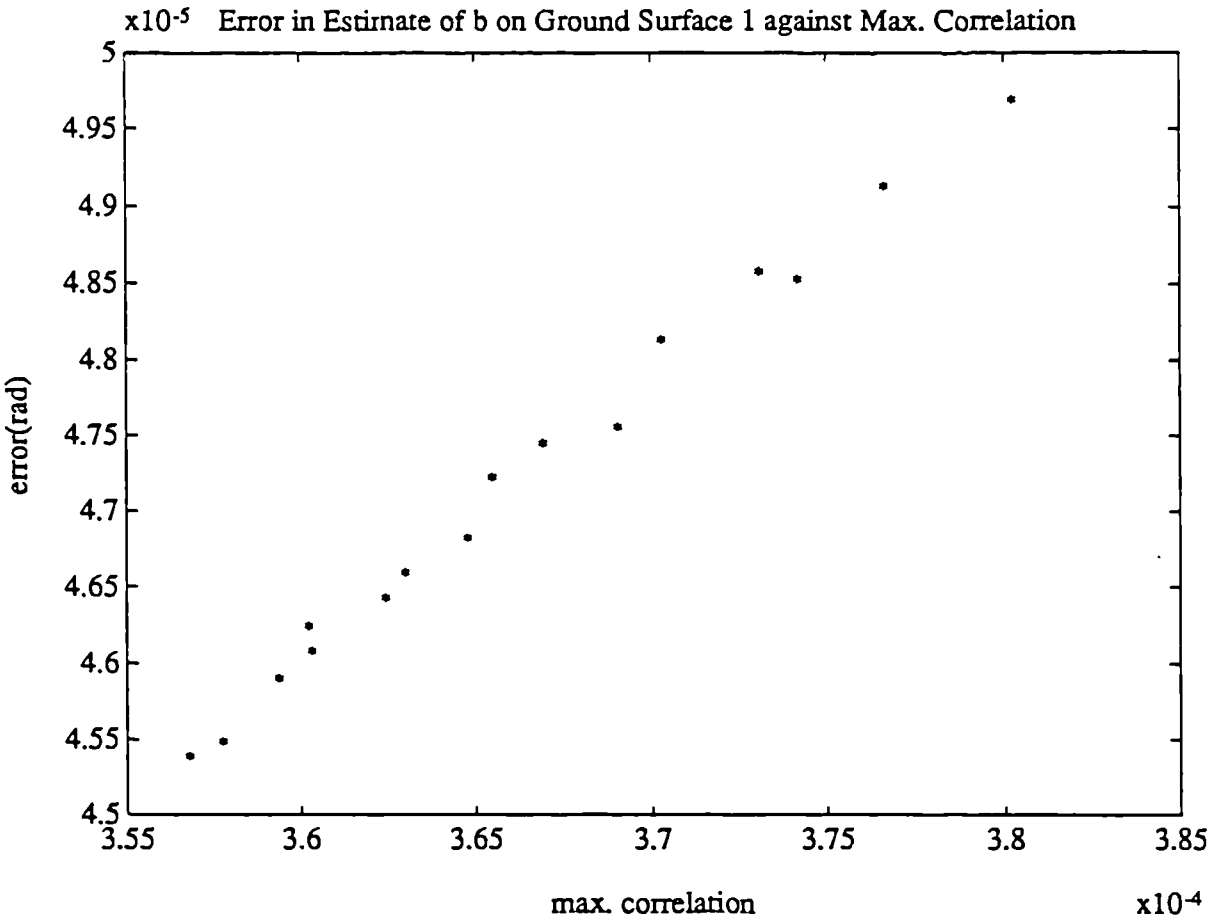
Graph 6.6 (b)



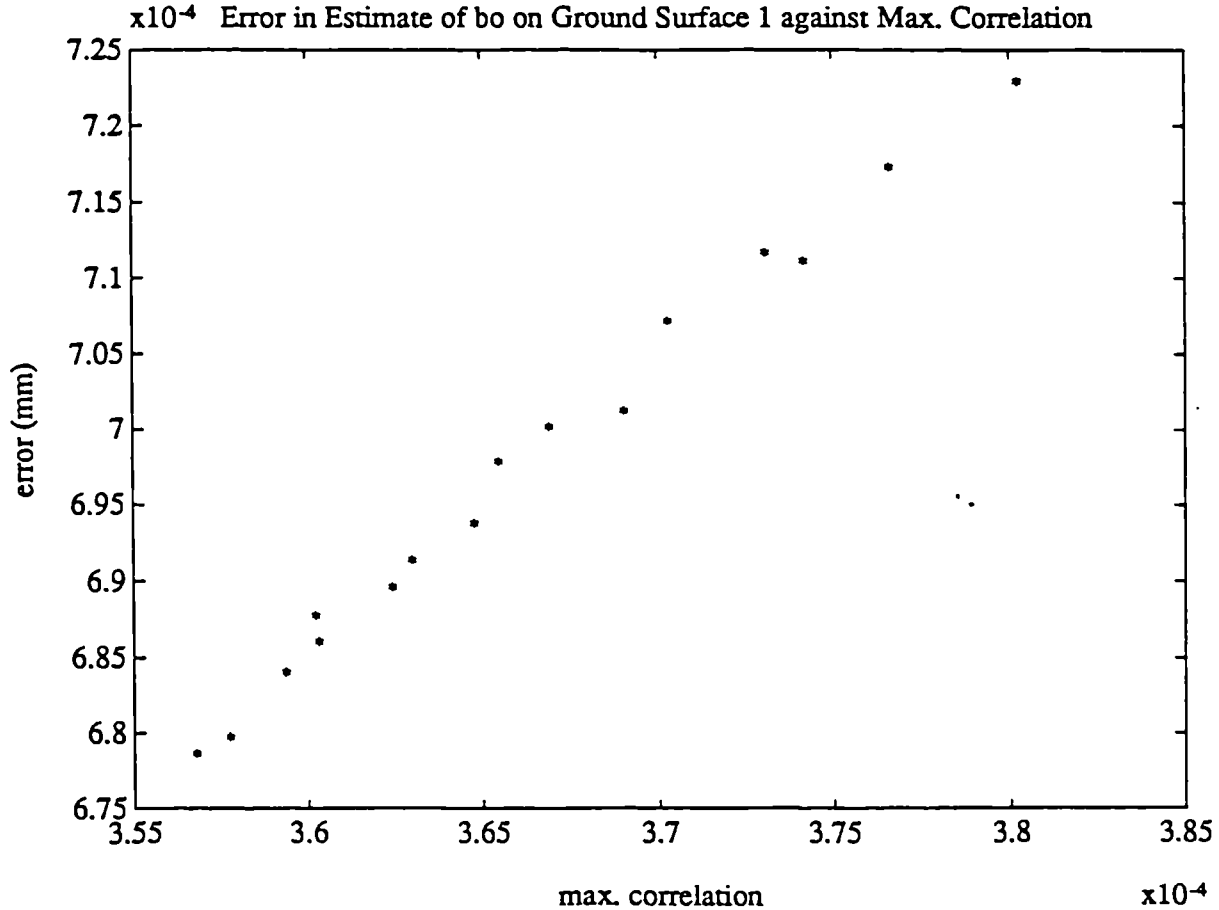
Graph 6.7 (a)



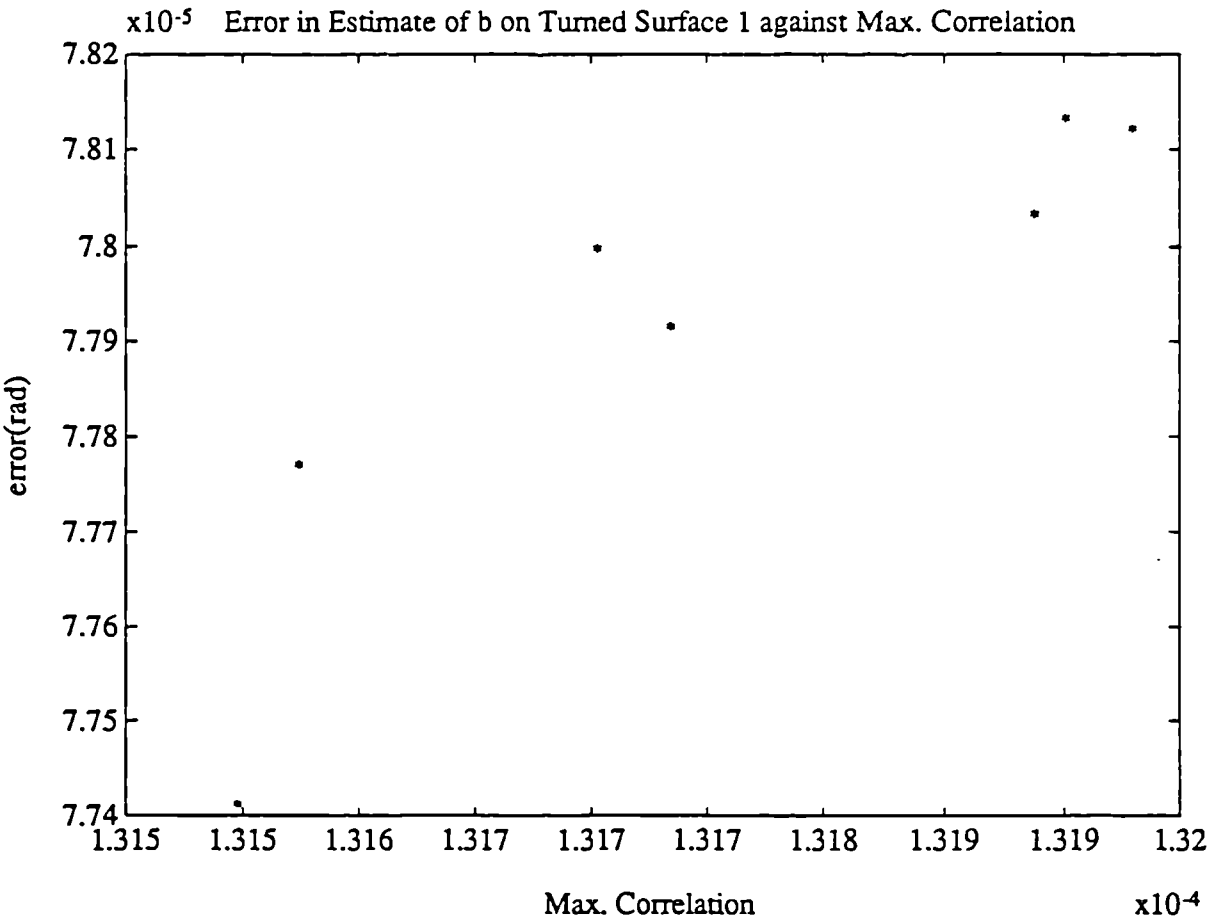
Graph 6.7 (b)



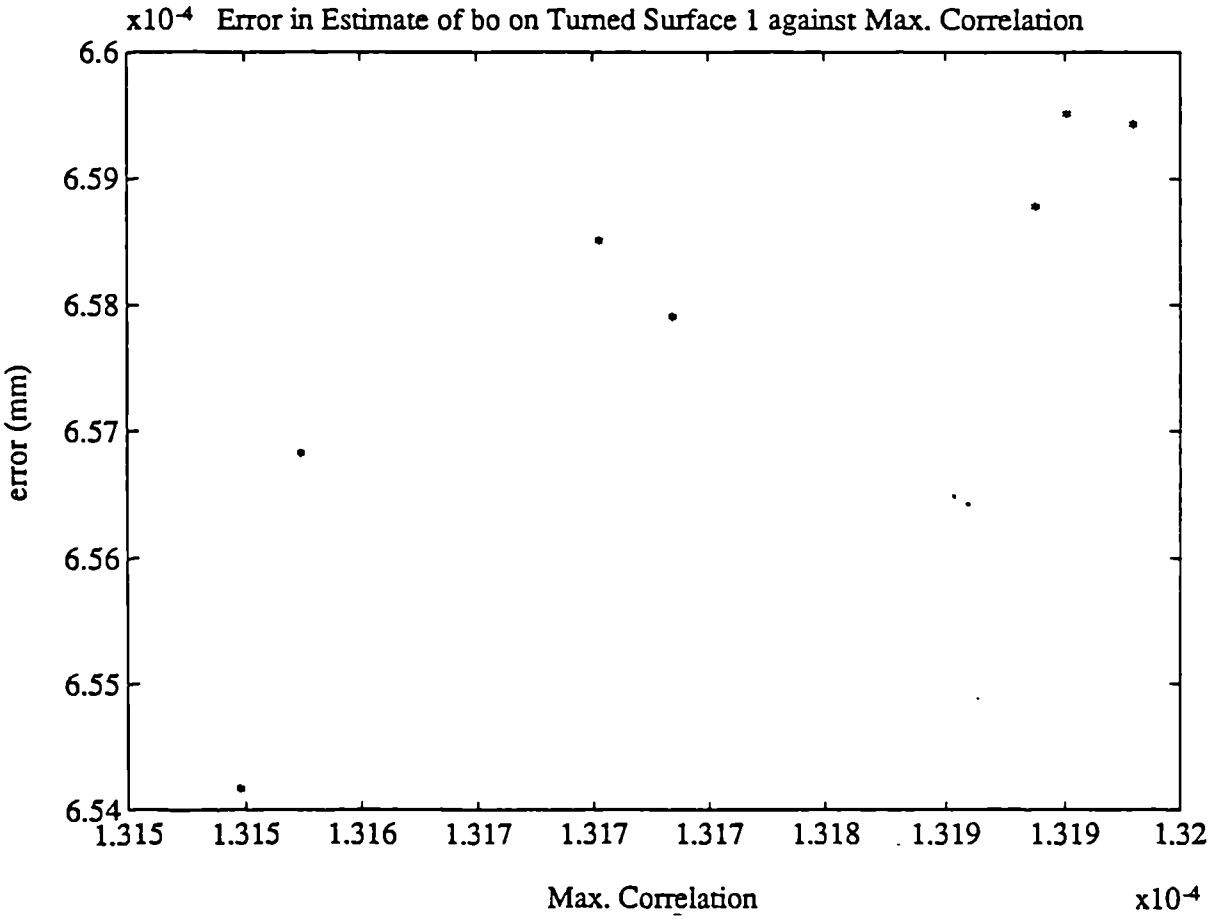
Graph 6.8 (a)



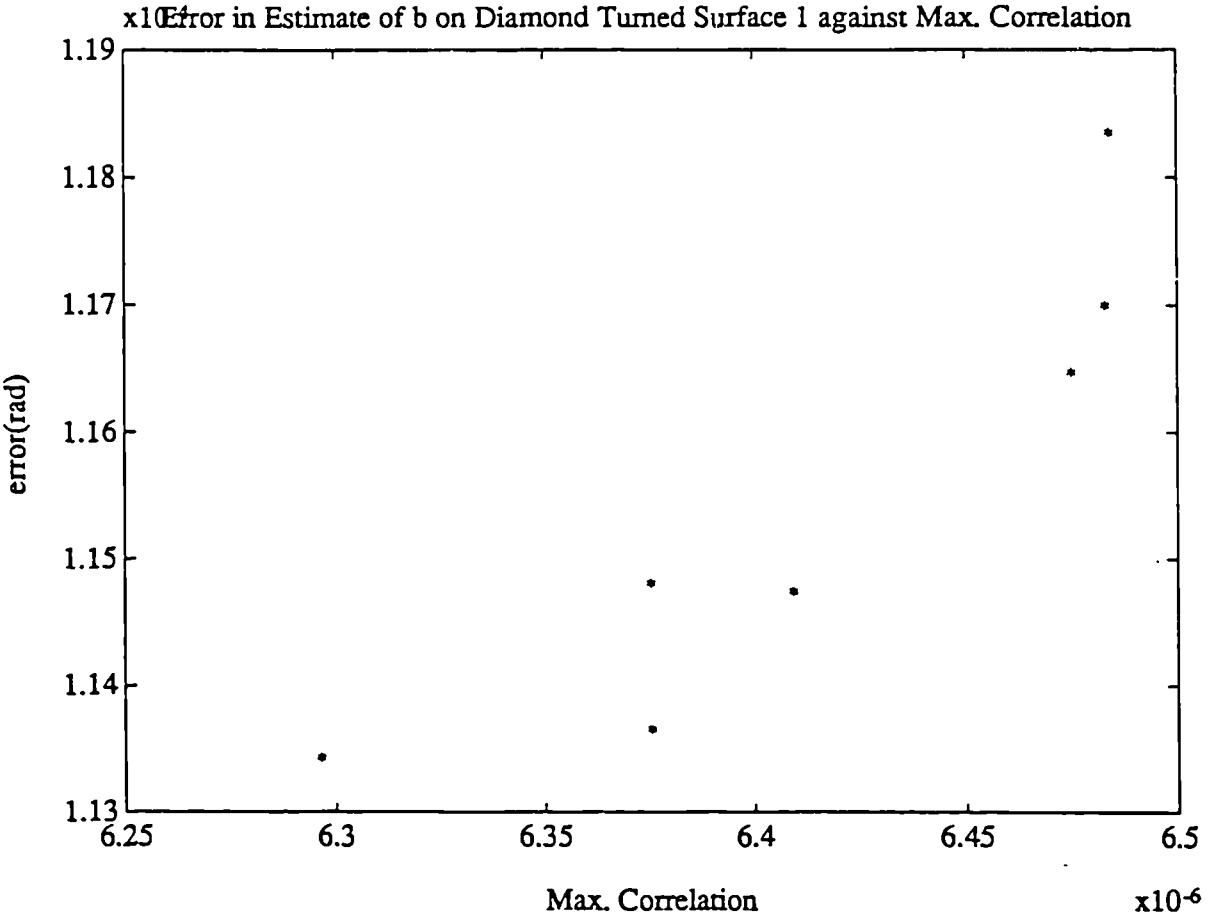
Graph 6.8 (b)



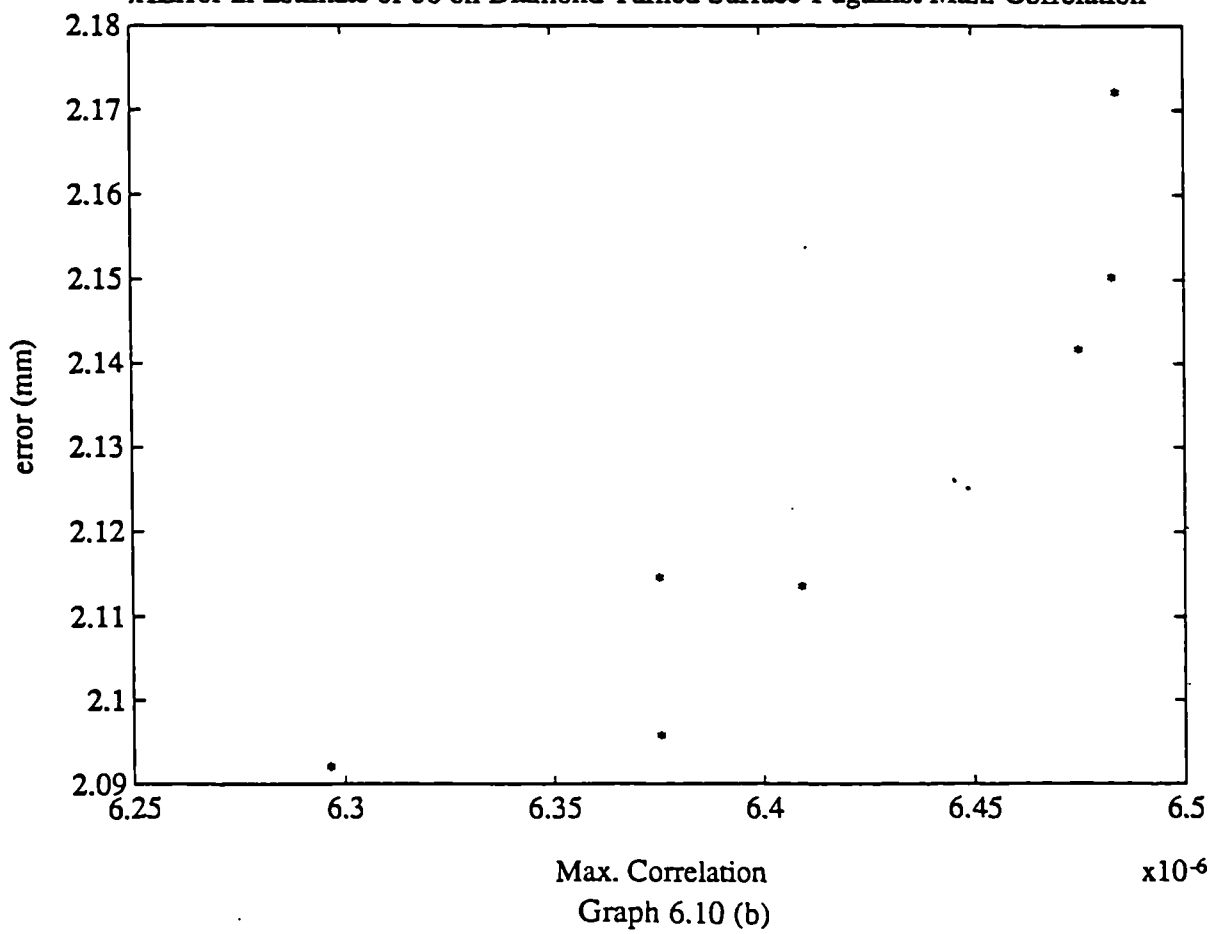
Graph 6.9 (a)



Graph 6.9 (b)



Graph 6.10 (a)



Graph 6.10 (b)

3.2. Results

The data used these experiments was the same data as that used in the previous plots of sampling errors against probe radii. Figures 6.8 to 6.13 clearly show that there is a direct relationship, generally almost linear between the change in sampling error and the maximum cross-correlation between the filtered and original data sets. This implies that the higher this cross-correlation, the higher the sampling error. (Higher maximum cross-correlation values imply that the mechanical filtering effects are less, i.e. the data is relatively unchanged.) Consequently, it is now possible to be able to specify the probe radius that would be the most appropriate for the minimisation of sampling error, without having to calculate it directly. It would only be neccesssary to model the probe radii of varying dimension and compute the maximum value of the cross-correlation function of the filtered and unfiltered data. It is the dimension of probe that renders the least value that is the most appropriate for the minimisation of sampling error.

4. Discussion

When selecting a probe it is important to consider other factors, most importantly the fidelity with which the envelope represents the actual surface, the degree of uncertainty associated with the precise point of contact of the probe and the fact that wear may alter the geometry of the probe. Each of these considerations is now discussed in turn.

4.1. Surface Representation

The mechanical filtering effect on the surface data has resulted in a degree of distortion. The sampling error calculated can no longer be taken as being the true sampling error of the original, functional component surface. The total problem is therefore not only concerned with pure sampling error, but also of surface misrepresentation, i.e. the probe mechanism indicates that contact with the surface at a horizontal location x has been established at a vertical position y_m (measured surface point), but the actual vertical height of the surface at x is y_a, for most points of contact:

$$y_a \neq y_m \tag{6.3}$$

The maximum correlation between the filtered and unfiltered data sets could again be used, but as a measure of the similarity of the data; a higher cross- correlation corresponds to a more accurate surface data representation. When selecting an appropriate probe size, it may be desirable to strike a balance between the minimisation of sampling error and the fidelity of surface representation (the weight awarded to each factor being based on functional considerations), this process would now be a simple case of selecting the radius rendering a maximum cross-correlation that is an appropriate proportion of the way between the maximum and minimum values, this proportion corresponding to the ratio of the respective weights imposed upon each factor.

4.2. Measuring Uncertainty

The spherical geomtery of the probe tip presents a complication in that contact is assumed to occur at the pole of the sphere, but there is no real way of knowing where contact really occurs, this uncertainty is again a function of probe dimension and surface profile.

A method for the computation of the probability (of contact) density for contacting members of relatively large radii does exist (3) but it relies upon the assumption that the surface distributions are nominally Gaussian, which cannot be held to be true for large surfaces for the reasons outlined in chapter 4. A non-Gaussian solution would be extremely complex and beyond the scope of this chapter.

4.3. Stylus Damage

It must also be bourne in mind that the geometry of the contacting probe stylus suffers a degree of variation due to wear. This rate of wear is at present largely unpredictable as the mechanisms involved are varied and each is influenced by different conditions e.g. measuring forces, stiffness of the probe mechanism and the nature of the contacting surface irregularities. It is interesting to note that the Al_3O_2 (synthetic ruby)

tips used on CMMs and for some applications of the Form Talysurf, have a particular problem in that many components inspected are made of aluminium, the similarity in chemical composition results in a substantial degree of adhesion , a particularly vociferous wear mechanism which is now being recognised as a significant source of error.

References

[1] T.R. THOMAS (1982): "Rough Surfaces"; Longman.

[2] V. RADHAKRISHNAN (1970): Wear,Vol. 16, pp 325-335.

[3] D.J. WHITEHOUSE (1982): "Surface Texture Assessment Errors of Finishing Processes Caused by Skid Distortion"; J. Phys. E. Sci. Instrum., Vol. 15, pp 1337-1340.

[4] H. ISHIGAKI & I. KAWAGICHI (1981); "Effect of a Skid on the Accuracy of Measuring Surface Roughness"; Wear, Vol. 68 , pp 203-211.

[5] D.J. WHITEHOUSE: "Theoretical Analysis of Stylus Integration", Primary Paper.

[6] J.S. BENDAT & A.J. PIERSOL (1971): "Random Data : Analysis and Measurement Procedures"; Wiley-Interscience.

Recommendations for Further Work

The algorithms for the quantification of sampling error developed in this thesis are of ext reme importance for the applications outlined in their respective chapters and will significantly contribute to the advancement of coordinate inspection techniques towards greater reliability and improved accuracy. Further research should be directed towards developing algorithms of this type for an increased range of inspection purposes. The following section is an outline of proposed further research that would enable practitioners to develop a full set of algorithms for the calculation of the errors incurred in the least-squares estimates of parameters of two and three dimensional geometric elements.

The first restriction upon the applicability of these algorithms that should ideally be eradicated is the use of equispaced data. Restricted access may prohibit the use of data spaced at exactly equal intervals and it has been thought for some time that the use of non-equispaced data in the estimates of parameters would improve their accuracy (1). The primary reason for this belief is that an improved representation of spatial variation may be achieved because equi-distant sampling may mask underlying surface periodicities if the sampling interval is either equal to or a multiple of the wavelength of one of the periodic elements. It would therefore be circumspect to investigate sampling error incurred in the use of data distributed in this way. Additional complication would result as each permutation of data would produce a different result.

The efficiency of the process of determining the number and distribution of data required to satisfy a particular limit of accuracy could be further increased by the undertaking of statistical analysis in order to discover the extent to which the data requirement for the estimation of a parameter remains constant for a given manufacturing process (

criteria being based on feed rate, machining forces etc) on the same machine (machine wear and misalignment can result in particular surface characteristics.) If indeed the data requirement does remain constant to within acceptable limits, it would be worth setting up a knowledge base of this data. If successful, a project of this nature would speed-up and simplify the process of determining the number and distribution of data required to maintain a particular accuracy of inspection.

Another direct application of the algorithms developed in chapters 4 and 5 is in probe design. Chapter 6 briefly outlines a way in which its use can be applied to the optimal selection of existing shperical probes, but through computer simulation methods different probe geometries could be investigated in order to determine an optimum probe design specific to functional requirements.

Developement of Algorithm for 3 Dimensional Surfaces

The most significant development of the sampling error quantification algorithms would be to extend the analysis to three dimensions. The first problem encountered is how to decide upon a suitable method of surface characterisation that could ultimately lead to the formulation of a similar algorithms for the assessment of three dimensional geometric elements. Geometric elements consisting of surface profiles are relatively simple to assess in that only one measure of spatial variation is required to describe the data set of which they effectively consist. However, three dimensional surfaces require more careful consideration. Surface types vary, but may be very broadly catagorised into three types (2), those having:

(i) Pronounced directional characteristics (strongly anisotropic).

(ii) Less pronounced directional characteristics (weakly anisotrpoic)

(iii) Virtually no directional lay (isotropic)

A data sampling distribution would therefore have to effectively represent the surface characteristics oriented in both directions. Peklenik characterised surfaces by the

maximum cross correlation between surface profiles and the distance between them. This would appear to be a suitable way forward in the development of algorithms for the assessment of plane parameters, as it is ideally suited to the data placement scheme recommended in chapter 3. In the computer generation of surfaces (3), Whitehouse relies upon the description of the surface being effected by a combination of the two autocorrelation functions in the X and Y directions, where the surface is deemed to have a directional characterisitc in both the X direction (s(x)), and the Y direction (t(y)). In this study, the combination is used to generate the autocorrelation function coefficients of a diagonal trace

$$A(\alpha,\beta) = \frac{1}{L^2}\int_0^L\int_0^L s(x).t(y)s(x+\alpha)t(y+\beta)dxdy$$

If the data is oriented at an angle of 45 degrees it could be taken as being equally representative of each directional characteristic, but if a similar distribution of data were oriented at any other angle, the directional patterns would not be equally represented. However, in plane fitting algorithms this pattern of data distribution (i.e. one straight diagonal) would result in the problem becoming ill-conditioned, and would therefore not be suitable. In cylinders, cones, etc., this would correspond to a helical or spiral arrangement which would not constitute ill conditioning, and this approach would therefore be suitable. This data should then be used to calculate an autocorrelation function which is then used to formulate a sampling error quantification algorithm in a similar manner to that used in the assessment of alignment error.

Application to other elements

The results of the investigations carried out in chapter 3 into the sampling errors incurred in the estimates of various geometric elements and the parallel development of an algorithm for the quantification of errors in the intercept with the Y axis show that great similarities exist in the behavior of these errors between one geometric parameter and another. These findings imply that formal solutions for various geometric elements

would be along broadly similar lines as that developed for the assessment of alignment error. This will be possible providing an explicit expression for the parameters exists and no assumptions regarding Gaussian behavoir are made in its derivation. Parameters having iterative solutions may be difficult and computationally expensive to assess. However, the probability of quantifying the errors in non-linear solutions may not merit investigation if the errors in the far more efficient linear solutions are known, indeed these rather clumsy and computationally expensive routines could possibly be replaced provided functional considerations permit this action.

References

[1] I. KASA (1976); "A Circle Fitting Procedure and its Error Analysis"; IEEE Trans. Inst. & Meas.

[2] J. PEKLENIK (1967): "New Developments in Surface Characterisation and Measurements by Means of Random Process Analysis"; Proc. Instn. Mech. Eng, Vol 182 pt 3K, pp 108-126.

[3]: D.J. WHITEHOUSE (1983): "The Generation of 2-D Random Surfaces Having a Specified Function"; Annals of the CIRP, Vol 32, January (1983).

Appendix A2

Vector and Matrix Norms

The norm referred to in this thesis are the one, two and ∞ norms. The differences between these are consisely described by considering the p norm of a vector x of length n :

$$\| x \|_p = \left[\sum_{i=1}^{n} |x_i|^p \right]^{\frac{1}{p}} \qquad \text{(A2.1)}$$

The p norm satisfies the following properties:

(i) $\| x \| \geq 0$ for all vectors x.

(ii) $\| x \| = 0$ only if x is a zero vector.

(iii) $\| \delta x \| = |\delta| \| x \|$ if δ is a real number.

(iv) $\| x + y \| \geq \| x \| + \| y \|$ for any two vectors x,y. This is called the triangle inequality.

The norms of matrices are also expressed as per (A2.1) and where A and B are matrices, the following properties are satisfied:

(i) $\| A \| \geq 0$ for all matrices A.

(ii) $\| A \| = 0$ only if A is a zero matrix

(iii) $\| \delta A \| = |\delta| \| A \|$.

(iv) $\| A + B \| \leq \| A \| + \| B \|$

Matrix norms can also be conveniently defined in terms of vector norms. The vector norms one two and ∞ induce the correspoding matrix norms for the $(m \times n)$ matrix.

$\| A \|_1 = \max_{1 \le j \le n} \left[\sum_{i=1}^{m} | a_{ij} | \right]$, which is the absolute column sum.

$\| A \|_2 = (\lambda_{max}[A^T A])^{1/2}$, the square root of the largest eigenvalue of $A^T A$.

$\| A \|_\infty = \max_{1 \le i \le m} \left[\sum_{j=1}^{n} | a_{ij} | \right]$, the maximum absolute row sum.

Appendix A4

A suitable algorithm for the nullification of the 'Wrap Around' Effect in the calculation of the autocorrelation function is as follows,

(1) g = the raw autocorrelation function of data set y.

(2) l = length of data set y.

(3) do for i = 1:l

(4) if i is less than l/2 then the autocorrelation function value corresponding to g(i) is (1/i)× g(i).

(5) if i is greater than l/2 then the autocorrelation function value corresponding to g(i) is (1/(l-i))× g(i).

(6) end