Editorial Manager(tm) for Proceedings of the Institution of Mechanical Engineers, Part B, Journal of Engineering Manufacture
Manuscript Draft

Manuscript Number: JEM1255R1

Title: UNCERTAINTY DUE TO EXPERIMENTAL CONDITIONS IN COORDINATE MEASURING MACHINES

Article Type: Special Issue article

Keywords: Measurement; uncertainty; coordinate measuring machines

Corresponding Author: EngD Hugo Lobato,

Corresponding Author's Institution:

First Author: Hugo Lobato

Order of Authors: Hugo Lobato; Carlo Ferri; Julian Faraway; Nick Orchard

Abstract: Any measurement method of a physical quantity cannot provide an exact unequivocal result due to the infinite amount of information necessary to characterise fully both the physical quantity to be measured and the measuring process. A quantitative indication of the quality of a measurement result needs therefore to be given to enable its reliable use. Uncertainty is one such indication. Provision of incorrect uncertainty statements for measurements performed by a coordinate measuring machine (CMM) may leads to very serious economic implications. In this study, the uncertainty of CMM measurements is estimated by a single parameter accounting for both systematic and random errors. The effects that environmental conditions (temperature), discretionary set-up parameters (probe extension, stylus length) and measuring plan decisions (number of points) have on uncertainty of measurements is then investigated. Interactions between such factors were also shown to be significant.

UNCERTAINTY DUE TO EXPERIMENTAL CONDITIONS IN COORDINATE MEASURING MACHINES

Hugo Lobato [a] **, Carlo Ferri** [b] **, Julian Faraway** [c] **, Nick Orchard** [a]

[a] Rolls-Royce plc, P.O. box 31, Derby DE24 8BJ, UK

[b] University of Bath, Mechanical Engineering Dept., Quarry Road, Bath, BA2 7AY, UK

[c] University of Bath, Mathematical Sciences Dept., Quarry Road, Bath, BA2 7AY, UK

ABSTRACT

Any measurement method of a physical quantity cannot provide an exact unequivocal result due to the infinite amount of information necessary to characterise fully both the physical quantity to be measured and the measuring process. A quantitative indication of the quality of a measurement result needs therefore to be given to enable its reliable use. Uncertainty is one such indication. Provision of incorrect uncertainty statements for measurements performed by a coordinate measuring machine (CMM) may leads to very serious economic implications. In this study, the uncertainty of CMM measurements is estimated by a single parameter accounting for both systematic and random errors. The effects that environmental conditions (temperature), discretionary set-up parameters (probe extension, stylus length) and measuring plan decisions (number of points) have on uncertainty of measurements is then investigated. Interactions between such factors were also shown to be significant.

1 Introduction

During the last two decades coordinate measurement systems (CMS) have been assuming an increasingly predominant role in the verification of compliance to dimensional and geometrical specifications of manufactured parts in a number of industries (for instance aerospace and biomedical). The distinct advantage of coordinate measuring machines (CMM) over other inspection systems is their intrinsic versatility. This enables them to be deployed in a large variety of measurement tasks which are often very demanding.

Three standards are commonly used by CMM manufacturers to specify the performances of their machines, namely the ISO 10360-1:2001 [1], the ASME B.89 [2] and the VDI/VDE [3]. Measurements taken according to these standards tend to involve artefacts such as step gauges, length bars and gauge blocks. These produce an estimate of the machine performance in terms of a volumetric measuring uncertainty value also known as maximum permissible error (MPE).

Regardless of the standard method used, the evaluation of the machine specification is fully trustworthy only for the set of conditions under which the evaluation took place. The term "conditions" refers to all those factors that may have an effect on the measurement result. These factors could be the different types of probes (e.g. kinematic or piezoelectric), accessories (e.g. styli or probe extension), machine settings (e.g. measurement speed or measurement acceleration), sampling strategy (e.g. number of points and distribution [4]) and environment (e.g. temperature or vibration). Some of these factors may affect a measurement result only in terms of systematic error, others in terms of random error and others again in terms of both. In the "International Vocabulary of Basic and General Terms in Metrology" [5], also known as VIM and published by the International Organisation for Standardisation (ISO), Systematic error is defined as "mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand" [5]. The mean referred to in this definition is represented as $\overset{=}{x}_\infty$ in Figure 1. Random error, on the other hand, is the "result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions" [5]. The term result mentioned above is represent as $\bar{x}$ in Figure 1. In practice, neither can be known exactly but must be estimated. Error without any further specification is the sum of the systematic and the random error [5]. The term measurand refers to the quantity to be measured (e.g. the thickness of a metal sheet at a specified temperature), whereas true value (or simply value) of a measurand refers to that ideal value that completely fulfils the specification of the measurand (cf. annex D in [6]). Error, e, systematic error, e_s, random error, e_r and true value θ are illustrated in Figure 1.

[Figure 1 about here]

Errors in CMM's have been grouped by Hermann et al. [7] in three categories:

1-Geometric errors due to the individual machine components (e.g. scales, axes' motors).

2- Errors related to the stiffness of structural machine components (e.g. Z ram, bridge).

3- Errors due to thermal effects (internal or external).

For the systematic error, some of the factors having a significant effect on a CMM machine have been identified by Feng and Pandley [8].

For the random error, a number of factors significantly affecting a CMM while measuring circular features have been identified by Feng [9]. Miguel and Cauchick [10] demonstrated a technique for evaluating CMM touch trigger probes using a motorised traversing table coupled with a laser interferometer. The authors demonstrated how the random error of the probe head varied with the indexing angle and the stylus length.

A first objective of this investigation is to provide a single performance parameter that jointly accounts for both systematic and random errors of a coordinate measuring machine. However, characterizing the concept of error, both systematic and random using estimation procedures founded on experimental activities and subsequent statistical analyses of the results does not enable the investigator to reach conclusions that are certain. A doubt about how adequately a measurement result represents the value of the quantity undergoing measurement is apparent (cf. section 0.2 in [6]).

To account for this impossibility of reaching conclusions that cannot be doubted, the concept of uncertainty was introduced and detailed in the "Guide to the expression of uncertainty in measurement" (GUM) published by ISO [6]. Adopting this perspective, thereafter the term uncertainty is preferred to the term error. In the GUM [6], the word uncertainty conveys two different meanings. The first is the generic concept of "doubt about the validity of the result of a measurement". The second is the specific concept of “parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand”.

The specific concept will be used throughout this investigation unless otherwise stated. The root mean squared error, $rmse$, has been selected as the parameter mentioned in the uncertainty definition. Such a quantity represents the dispersion of a series of n measurement results, $\bar{x}_i$, from the value of the measurand, θ, namely:

$$rmse\left(\bar{x}_i,\theta,n\right)=\sqrt{\frac{\sum_{i=1}^{n}\left(\bar{x}_i-\theta\right)^2}{n}} \qquad (1)$$

In equation 1 the measurement result of the *i-th* measurement task is represented with $\bar{x}_i$ to denote the fact that in this study it has been computed using the average of three test results performed in repeatability conditions (cf. section 3.6 in the VIM [5] for a definition of repeatability conditions). This is equivalent to saying that a measurement task encompasses three measurement tests. The $rmse$ is expressed in the same unit as the measurement result and the value of the measurand (e.g. metres for a length).

Equation 1 can then be rearranged to $rmse\ \left(\bar{x}_i,\theta,n\right)=\sqrt{\frac{\sum_{i=1}^{n}\left(\bar{x}_i-\bar{\bar{x}}\right)^2}{n}+\frac{\sum_{i=1}^{n}\left(\bar{\bar{x}}-\theta\right)^2}{n}}$, which shows that $rmse$ is given by the square root of two additive terms (further detail can be found for instance in [11]). The first term is the square of the bias, which accounts for the systematic error. The second term is the sample variance that represents the dispersion of the series of measurements about their mean and that therefore expresses the random error. These considerations enable the selected uncertainty parameter to be identified as completely fulfilling the first objective of this study. They also reveal the existence of some similarities between the approach based on *rmse* presented in this investigation and the method for determining the uncertainty of measurement illustrated in the technical specification ISO/TS 15530-3 [12].

A second objective of this investigation, is the identification of experimental conditions (i.e environment, probe extension) that may significantly affect the *rmse* and that are likely to be encountered in the large variety of measurement tasks that the machine can perform. It is believed that pursuing this second objective may contribute to raising the awareness of the practitioners regarding the detrimental effect that uncontrolled or uncontrollable experimental conditions may exert on the machine specification. Moreover, it enables set-up parameters to be chosen so that the resulting uncertainty of measurement quantified by *rmse* is lower. The method presented in this study can then be adapted to suit the specific environment conditions and needs of the measurement tasks of interest to a specific organisation.

2 Experimental set up

A commercially available CMM was used for the experimental study. The machine was a moving bridge with a specification MPE=(3.5+L/250)μm (L being a length in mm) according to ISO 10360-1:2001 [1]. The experimental set-up is shown in Figure 2.

[Figure-2 about here]

The machine was located in a temperature controlled room where the temperature can be set at a pre-specified reference value within an uncertainty of +/- 1 ˚C at 95 % significance level. Therefore, by setting different levels of room temperature it is possible to simulate measurement tasks performed in workshop environments where the temperature may vary considerably throughout a working day during normal operating conditions. In environments that lack temperature control, the temperature is an uncontrolled nuisance factor, whose effects on the uncertainty of measurement expressed in terms of *rmse* it is believed sensible to investigate.

In this investigation, two levels of room temperature were selected, 21 and 24 °C, respectively, and no temperature compensation settings were enabled on the CMM throughout the whole experimental activity. The stability of the machine temperature at each of the two levels of air temperature considered was monitored using K type thermocouples applied in a number of points of the machine.

Another factor that may have a significant effect on the uncertainty of measurement of a CMM is the geometric characteristic (form and dimensions) of the parts to be measured. In fact, performing a measurement task on parts with different dimensions engages each of the axes of motion of the CMM in different ways. Moreover, the extension of motion of each of them is expected to be different. Similar considerations apply for parts of comparable dimensions but with different form. To represent the variety of parts that have different geometrical characteristics and that can be measured using a CMM, two different features were selected for this study: a ring gauge (R) and a sphere (S) to represent two and three dimensional features, respectively. In both cases, the measurand was defined as the diameter of the part at each of the two examined levels of air temperature. As shown in equation 1 the $rmse$ is also a function of the value of the measurand, θ, that cannot be completely known because the measurand itself cannot be completely identified without an infinite amount of information (cf. section D1.1 in the GUM [6]). Therefore an estimate of θ, i.e. $\hat{\theta}$, is needed in order to have an estimate of the $rmse$, i.e. $r\hat{m}se$. By using a certified reference material (CRM, cf. section 6.14 in the VIM [5] for a definition) encompassing both a ring gauge and a sphere, not only is it possible to have an estimate of the value of the measurand, i.e. $\hat{\theta}$, accompanied by a standard uncertainty value (cf. section 2.3.1 in the GUM [6] for a definition), but also traceability to the official realisation of the unit of measurement of the measurand is established. That is, for CRM of length, an official realisation of the definition of the metre.

These characteristics of a CRM are summarised in a document called a calibration certificate. Notwithstanding, the values of both the measurands provided in this document are valid at a reference temperature T_{ref} that is also stated in the certificate. For the measurand in this study, as is typical with any length, $T_{ref} = 20\,°\mathrm{C}$. Thermal expansion for the sphere (external feature) and thermal contraction for the ring gauge (internal feature) is expected to affect the values provided by the certificate when the operating temperature of the CRM is higher than T_{ref}, as in this study. Consequently, new estimates $\hat{\theta}_T$'s for the values of the measurands valid when the temperature of the measurand is T were produced using the following equation, under the assumption of linear thermal expansion of the CRM:

$$\hat{\theta}_T = \hat{\theta} + \alpha_{T_{ref}} \cdot \left(T - T_{ref}\right) \cdot \hat{\theta} \tag{2}$$

In equation 2 $\alpha_{T_{ref}}$ is the coefficient of linear thermal expansion when the CRM is at the temperature T_{ref}. The temperature T of the CRM when the air temperature was set at 21 and 24 °C respectively, was monitored attaching K type thermocouples to the CRM at a number of points. The average of these measured values of temperature was used in equation 2. Some of the information available on the calibration certificate of the CRM used have been summarised in Table 1.

[Table 1 about here]

Ultimately, an estimate $r\hat{m}se$ of a series of measurement results taken in the *i-th* experimental condition is obtained using $\hat{\theta}_T$ from equation 2. The series of measurements has been taken in repeatability conditions. The results $\bar{x}_i$ and $\bar{x}_{i+1}$ have not been obtained one after the other in a temporal sequence, but have been assigned to the run order by randomly selecting them from all the measurements in all the investigated experimental conditions at a pre-specified temperature. Differently stated, the measurements results are replicates and not repetitions of the measurement process.

When setting up a CMM for a specific already assigned measurement task, it often appears that the operator may be left with some discretionary decisions to take regarding the set-up of the machine and/or the planning of the measurements. Some attempt to automate this decision making has been investigated by Zhang et all [13]. In this investigation, it appeared reasonable to ascertain whether some of these decisions may have a significant effect on the uncertainty of measurement expressed in terms of $rmse$.
The set-up parameters chosen as discretionary factors were the probe extension, the stylus length and the number of probing points. For the probe extension, three different set-ups of the analysed CMM were considered: without any probe extension, with probe extensions of length 100 mm and 200 mm. Three styli of the same type and geometrical characteristics (e.g. material, tip size), but with lengths 20, 60 and 110 mm, respectively, were chosen. Regarding the planning of the measurements, the potential effects on the uncertainty of measurement due to two different numbers of probing points (seven and eleven) were examined.

A kinematic probe with a standard force module was used throughout this experiment. The factors examined in this study with their levels are displayed in Table 2.

[Table-2 about here]

3 Randomisation issues

A fully randomized experimental design with three factors at two levels each and two factors at three levels each identifies 72 different experimental conditions, henceforth also referred to as treatments or cells of the design. Three replicates of the design were considered, i.e. $r = 1,2,3$. This resulted in an overall experimental effort of 216 measurement tasks, i.e. 648 measurement tests.

In the experimental set-up examined, it is not practical to assign randomly a measurement task in a pre-specified experimental condition to the run order, due to the fact that some of the considered factors are hard-to-change. In particular, the air temperature cannot be changed easily. So all the measurement tasks at one level of temperature were carried out first, and then all the others were performed at the remaining level of temperature investigated.

Therefore, if some nuisance factor occurred while performing the measurement task at a certain temperature, it would lead the experimenter to attribute incorrectly such effects on the response variable ($r\hat{m}se$) to the temperature.

Accounting for such possibility, would require the experimenters to replicate all the measurement tasks performed in one day at a certain temperature a number of times (i.e. a number of days) sufficiently large to estimate the variability of the response variable from day to day. This would dramatically increase the experimental burden in a way which is inconsistent with the main objectives of this investigation.

Only one day at each level of temperature was therefore considered. This is the reason why, the reader must exert caution and not to neglect the possibility that the effects attributed to the temperature are in reality due to some lurking nuisance factor. However, in the authors' point of view, the extremely controlled conditions in which the experiment was carried out makes unlikely that such nuisance factors would have indeed occurred.

The two types of the features, ring and sphere, were not randomly assigned to the run order. In fact, the sequence of measurement tasks was constructed as a sequence of pairs, each consisting of one measurement of the ring and one of the sphere in identical experimental conditions.

This experimental strategy was adopted with the intent of counteracting the potential presence of nuisance factors that increase the variability of the response variable, thus making it more difficult to identify any significant effect on the response variable due to the type of the feature measured.

Once, the room temperature was set and the constraint on the run order for the type of features was introduced, all the others combinations of factors were randomly assigned to the sequence of the measurement tasks.

It is worth mentioning that when changing the probe extension or the stylus length a calibration procedure was run. Consequently, the random assignment of the experimental conditions to the order of the measurement tasks may result some times in a calibration procedure being run, but in some other time in no calibration procedure being run. The last circumstance happens when the probe extension or the stylus length are not changed between two consecutive conditions. This is considered acceptable because this experiment is meant to be representative of the actual operational conditions in which the measuring system is used. In such circumstances, the random sequence of calibration and non-calibration is most likely to happen depending on the variety of measuring tasks performed.

It is moreover argued that performing calibration procedures during the experiment may increase the overall measured uncertainty of the system in comparison with ideal laboratory conditions.

4 Exploratory data analysis

The $r\hat{m}se$ is composed of two additive terms of equal importance: an estimate of the variance of a measurement result and an estimate of its bias. Each of the experimental factors considered in this study may affect differently each of the two components. The effects of the room temperature are graphically examined to demonstrate this observation.

The estimated bias at each temperature for both the ring gauge and sphere is displayed in Figure 3, which as the following figures was obtained in `R`, a language and environment for statistical computing and graphics [14]. The temperatures displayed on the abscissa do not represent actual values of the temperature at which each measurement result was obtained. They are instead obtained by artificially adding to the original categorical abscissae (21 and 24 °C) an horizontal random component to reduce the occurrence of overlapping points and so to enhance the clarity of the figure (this technique is called jittering).

At the lower level of temperature, the bias distribution for both features is centred close to zero, whereas at the higher level of temperature there is a positive shift of the bias only for the ring. This induces a strong suspicion that there is a significant interaction effect of temperature and feature on the estimated bias. From a practitioner's perspective, this means that uncontrolled variations of air temperature may induce negligible bias on parts with some specific geometric characteristics but a very large bias on others.

On the other hand, this interaction effect may also be attributed to the inadequacy of thermal expansion model expressed in equation 2 when applied to the ring gauge. Further measurement tests involving a ring gauge calibrated at the investigated air temperature would be needed to clarify the matter, but they would be beyond the scope of this investigation.

In Figure 4 the sample standard deviations of the measurement results are displayed. For both the ring and the sphere, no significant effect of the air temperature on the variability of the results is apparent from an examination of this figure. In fact, while considering increased air temperature, only a mild increment in the average standard deviation of the measurement results grouped by temperature is observed in Figure 4.

Under these circumstances, it is therefore argued that, when increasing the air temperature, the significant increment of $r\hat{m}se$ displayed in Figure 4 for the ring only can be mainly attributed to the bias.

In Figure 4 it can be noticed that the $r\hat{m}se$'s are not symmetrically distributed around their average values, when grouped by the temperature. More data points are apparent in the region between zero and the group averages, i.e. the end points of the two continuous segments, than in the region above such group averages. The skewness of the distribution of the $r\hat{m}se$'s is independent from the way they are grouped and has implications on the formulation of plausible statistical models for the experimental data. These implications are discussed in the next section.

5 Statistical model

A first attempt model that could be considered suitable to describe the experimental results is as follows:

$$r\hat{m}se_j = \mu + temp_{(j)} + fea_{(j)} + pe_{(j)} + sl_{(j)} + np_{(j)} + temp : fea_{(j)} + \ldots + er_j \qquad (3)$$

In equation (3), the symbol μ represents the mean of the response variable $r\hat{m}se_j$ over all the experiment and $j = 1,\ldots,72$ is the index associated with each of the experimental conditions. The meaning of the other symbols is summarised in Table 2, whereas the colon is used to identify an interaction effect on the response variable due to the factors it divides. The parenthesised subscripts map the rows in the data to the levels of the factor used in that row. For example, $temp_{(j)}$ corresponds to the temperature used for that j. For brevity, the ellipsis stands for all the remaining possible second order interactions. Interactions of higher order, i.e. involving more than two factors, were not considered because it is difficult to foresee how the experimental conditions considered could possibly cause them. Moreover, from a practitioner's point of view, it is also difficult to see how the awareness of the significance of a third, fourth or fifth order interaction could enrich the knowledge of the measuring system investigated. The terms er_j's are random variables that, without losing generality, are assumed to be independent and identically distributed with mean zero and constant variance σ_{er}^2. If they are also normal statistical inferences regarding the parameters of the model is facilitated.

In the previous section it was observed that the realisations of $r\hat{m}se_j$ are distributed asymmetrically. This circumstance makes it very unlikely that the errors of the model to follow a symmetrical distribution such as the normal. For this reason, it would make the inferential process easier if the response variable were transformed in such a way to assume a more symmetrical distribution. A transformation that appears to suits this purpose is the logarithm. Therefore the following model was considered:

$$\log\left(r\hat{m}se\right)_j = \mu + temp_{(j)} + fea_{(j)} + pe_{(j)} + sl_{(j)} + np_{(j)} + temp_{(j)} : fea_{(j)} + \ldots + er_j \tag{4}$$

Equation (4) represents a multiplicative model in the domain of the untransformed response variable. It can in fact be rewritten as in its equivalent form:

$$r\hat{m}se_j = e^{\mu} \cdot e^{temp_{(j)}} \cdot e^{fea_{(j)}} \cdot e^{pe_{(j)}} \cdot e^{sl_{(j)}} \cdot e^{np_{(j)}} \cdot e^{temp_{(j)}:fea_{(j)}} \cdot e^{temp_{(j)}:pe_{(j)}} \cdot \ldots \cdot e^{er_j} \tag{5}$$

This model was fitted to the experimental data using the ordinary least squares method (OLS) as implemented in `R` [14]. A large number of two-way interactions were found not to be statistically significant resulting in the following final model:

$$\begin{aligned}\log(r\hat{m}se)_j = \mu + temp_{(j)} + fea_{(j)} + pe_{(j)} + sl_{(j)} + np_{(j)} + \\ + temp_{(j)} : fea_{(j)} + pe_{(j)} : sl_{(j)} + fea_{(j)} : np_{(j)} + er_j\end{aligned} \quad (6)$$

The coefficient of determination (R^2), was equal to 40.9 %. This means that about 60% of variability of the response variable is not accounted for by this model and must be due to other unknown sources.

The ANOVA table that shows the significance of each of the factors included in equation (6) is displayed in Table 3.

[Table 3 about here]

Figure 6, 7 and 8 show interaction plots corresponding to the three significant interaction effects in the final model. These show the mean $r\hat{m}se$ for each combination of the interacting factors and are useful in interpreting the combined effect of these factors.

The significance of the interaction between temperature and type of feature that was expected by the observation of Figure 5 is confirmed by Figure 6 and it has already been discussed in the exploratory data analysis section.

Figure 7 shows that in the selection of the stylus length to obtain improved uncertainty performance, the probe extension must be also considered. For different probe extensions, different styli may be preferable from the point of view of limiting the uncertainty. Stylus length and probe extension should therefore be chosen together. In Figure 7, this is demonstrated observing that with the same probe extension of length 200 mm, uncertainty of measurement can be greatly improved if the stylus length is carefully chosen (stylus length 60 mm). Moreover, the same figure suggests that a set-up that does not make use of any probe extension can produce measurement results with improved uncertainty, independently from any specific stylus length. It also appears that the stylus with length 60 mm has superior uncertainty performance in absolute terms and also in terms of robustness to changes of probe extension.

Figure 8 supports the intuitive idea that in the selection of the number of probing points the type of feature to be measured has a part in affecting the uncertainty of measurement that will be achieved. The same number of probing points that provides satisfactory uncertainty on a specific feature may lead to deteriorated uncertainty performance when different type of features are measured.

The assumptions underlying the model of equation 6 are graphically tested by examining the residuals. Figure 9 shows the exponentiated residuals against the exponentiated fitted values. Both the residuals and the fitted falues have been exponentiated to convert them back to the micron scale. This also means that the residuals must be interpreted multiplicatively (equation 5). Hence, a value of one indicates a perfect fit to the model whereas the few larger residuals indicate observed errors about 5 or 6 times larger than expected. Most importantly, we see no association with the fitted values.

In Figure 10 the realisations of the exponentiated residuals are grouped by type of feature. If the assumptions of independence and identical distribution of the errors is satisfied, the exponentiated residuals should not exhibit any pattern or difference in behaviour however they are grouped. The fact that no differences are apparent in Figure 10 supports the conjecture that all the effects caused by the type of feature on the response variable are correctly captured by the considered model. Therefore, the type of feature does not appear to have any effect on the realisations of the exponentiated residuals.

Figure 11 shows a Q-Q plot to assess the normality of the errors, which seems to be confirmed.

One final concern is the lack of a full randomisation in our experimental design due to practical considerations. In particular, for each setting of the experimental factors, we perform the ring and sphere measurements together. We can modify our model to take account of this as follows:

$$\begin{aligned}\log\left(r\hat{m}se\right)_j = \mu + temp_{(j)} + fea_{(j)} + pe_{(j)} + sl_{(j)} + np_{(j)} + \\ + temp_{(j)} : fea_{(j)} + pe_{(j)} : sl_{(j)} + fea_{(j)} : np_{(j)} + pa_{(j)} + er_j\end{aligned} \tag{7}$$

The term $pa_{(j)}$ is a random effect with mean zero and some variance to be estimated. There will be one such term for each pair of a ring and sphere measurements, i.e. 36 pairs in total. Such a model is called a mixed effects model and is described in [15]. The hypothesis that the variance of $pa_{(j)}$ is zero can be tested using a parametric bootstrap method as long as we assume normality of the random effect. In this case, the term is found not be statistically significant (p-value=0.46). Thus, it is reasonable to conclude that there is no association between these pairs of measurements and that the lack of a full randomisation has had no consequence. Nevertheless, it is wise for experimenters to investigate these concerns in similar designs where practicality precludes a full randomization.

6 Conclusions

Often in industrial environments the adequacy of the measurement system to perform a measurement task may be assessed solely on the basis of random error evaluations. Repeatability studies can be considered among this kind of approaches. The state of calibration of the measurement system should instead provide assurance of the lack of systematic error (bias) when performing a measurement task in the same conditions for which the instrument was calibrated. $rmse$ was analysed as a single parameter that provides the practioner with a tool to monitor the performance of the measurement system in terms of both random and systematic error. The effect of the environment temperature, feature type, probe extension, stylus length and number of probing points on $rmse$ were considered by fitting a linear random effect and a linear mixed-effect statistical model to the experimental results. All these five factors were found to be statistically significant. The significance of all the second order interactions of these factors was also considered and only three of them were found to be statically significant (temperature with feature type, probe extension with stylus length and number of probing points with feature type).

The nominal performance of a CMM is evaluated in a pre-specified allowable range of experimental conditions. Even when the machine is meant to be deployed within such a range, the degrees of freedom left to the operators when setting-up the machine or preparing a measurement plan, may lead to significantly deteriorated performance with detrimental effects on the pertinent costs. Among these, there are for example the costs sustained for unnecessary reworking, the costs for the rejection of good parts and costs due to increased failure rate of the final products caused by the acceptance of defective components.

Performance of a CMM should therefore be evaluated in experimental conditions as close as possible, ideally identical, to those in which the machine is meant to be actually deployed. Such experimental conditions should encompass both uncontrollable factors (temperature and parts, in this study) and controllable (settings such as probe extension, stylus length and number of probing points, for example).

ACKNOLEDGEMENTS

The authors gratefully acknowledge Rolls-Royce plc for their financial contribution and the Interdisciplinary Research Centre (IRC) in Materials for High Performance

Applications at the University of Birmingham for granting one of authors the use their facilities.

REFERENCES

[1] ISO 10360-1 (2001), Geometrical Product Specifications (GPS)-Acceptance and reverification tests for coordinate measuring machines (CMM). International Organisation for Standardisation.

[2] ANSI/ASME B89.4.1 (1997), Methods for performance evaluation of coordinate measuring machines (CMM).

[3] VDI/VDE 2617-3 (1989), Accuracy of coordinate measuring machines (CMM), components of measurement deviation of the machine.

[4] Collins, C E. Fay, E B. Aguirre-Cruz, J A. and Raman, S. Alternate methods for sampling in coordinate metrology. *Proc. ImechE, Part B: Journal of Engineering Manufacture,* 2007, 221, 1041-1052.

[5] International Vocabulary of Basic and General Terms in Metrology, reproduced verbatim in PD 6461-1 (1995). General metrology – Part 1: Basic and general terms (VIM). BSI – British Standards Institution.

[6] Guide to the expression of uncertainty in measurement, reproduced verbatim in PD 6461-3 (1995). General metrology – Part 3: Guide to the expression of uncertainty in measurement (GUM). BSI – British Standards Institution. 1995.

[7] Hermann, G. Geometric Error Correction in Coordinate Measurement. *Acta Polytechnica Hungarica*, 2007, 4(N1), 47-61.

[8] Feng, C-X J. Pandey, V. Experimental study of the effect of digitising parameters on digitising uncertainty with a CMM. *International Journal of Production Research*, 2002, 40(N3), 683-697.

[9] Feng, C-X J. Saal, A. Salsbury, J. and Ness, A. Lin, G. Design and analysis of experiments in CMM measurement uncertainty study. *Journal of Precision Engineering*, 2007, 31(N2), 94-101.

[10] Miguel, P. King, T. and Abackerli, A. CMM touch trigger performance verification using a probe test apparatus. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2003, 25(N2), 1-16

[11] Mood, A. Graybill, F. and Boes, D. Introduction to the theory of statistics. 3rd edition, McGraw-Hill, 1974

[12] ISO/TS 15530-3 (2004) Geometrical Product Specifications (GPS) – Coordinate measuring machines (CMM): technique for determining the uncertainty of measurement. Use of calibrated workpieces or standards. International Organisation for Standardisation.

[13] Zhang, S G. Ajmal, A. Wootton, J. and Chisholm, A. An automated inspection planning system for coordinate measuring machines. *Proc. ImechE, Part B: Journal of Engineering Manufacture,* 2001, 215, 419-424.

[14] R Development Core Team. A Language and environment for statistical computing. R Fundation for Statistical Computing, Vienna, Austria. ISSB: 3-900051-07-0. URL: www.r-project.org . 2008.

[15] Jose' C. Pinheiro and Douglas M. Bates. Mixed-Effects Models in S and S-plus. Springer. 2000.

APPENDIX

Notation

$temp_{(\ )}$ Room temperature ($°C$).

$\alpha_{T_{ref}}$ Coefficient of linear thermal expansion at T_{ref} .

$fea_{(\ )}$ Feature measured, i.e. ring or sphere.

$pe_{(\ )}$ Probe extension (mm),.

$sl_{(\ )}$ Styli length (mm).

$np_{(\ )}$ Number of probing points.

θ Value of a measurand alias true value of a measurand.

$\hat{\theta}$ Estimate of the value of a measurand with the certified reference material at T_{ref} .

μ Overall mean or intercept in a linear statistical model.

$\hat{\theta}_T$ Estimate of the value of a measurand with the certified reference material at T .

$\hat{\sigma}$ Estimate of the standard deviation of a single test.

σ_{er}^2 Variance of the errors of a statistical model.

e Error of measurement.

e_r Random error.

e_s Systematic error.

er_x Random errors in a statistical model indexed by the series of subscripts x .

$rmse$ Root mean squared error.

$r\hat{m}se$ Root mean squared error.

S^2 Sample variance of a series of tests.

T Generic temperature.

T_{ref} Reference temperature stated in the calibration certificate.

$\bar{x}_i$ i-th measurement result in a series of n measurements.

$\bar{x}$ Generic measurement result of a measurement task calculated as average of a series of measurement tests.

$\bar{\bar{x}}$ Average of large number of generic measurements ($\bar{x}$).

$\bar{\bar{x}}_\infty$ Average of an infinite number of generic measurements ($\bar{x}$).

LIST OF TABLES

LIST OF FIGURES

Table 1

FEATURE	CALIBRATED VALUE (mm)	UNCERTAINTY (mm)	COEFFICIENT OF THERMAL EXPANSION (pp/mC)
Ring Gauge	49.9994	0.4	11.5
Sphere	29.9992	0.4	5.5

Table 2

FACTORS	LABELS	LEVELS		
Room temperature ($^\circ$C)	$temp_{(j)} \quad j = 1,\ldots,72$	20	24	
Feature	$fea_{(j)} \quad j = 1,\ldots,72$	Ring (R)	Sphere (S)	
Probe extension (mm)	$pe_{(j)} \quad j = 1,\ldots,72$	0	100	200
Styli length (mm)	$sl_{(j)} \quad j = 1,\ldots,72$	20	60	110
No. of probing points	$np_{(j)} \quad j = 1,\ldots,72$	7		11

Table 3

	Degrees of freedom	Sum of squares	Means of squares	F value	Pr(> F)
Temperature $temp_{( )}$	1	5.78	5.78	30.43	$8.41 \cdot 10^{-7}$
Probe extension $pe_{( )}$	2	4.06	2.03	10.67	$1.13 \cdot 10^{-4}$
Stylus length $sl_{( )}$	2	2.28	1.14	5.99	$4.31 \cdot 10^{-3}$
Type of feature $fea_{( )}$	1	0.879	0.879	4.63	$3.56 \cdot 10^{-2}$
Number of probing points $np_{( )}$	1	0.783	0.783	4.12	$4.70 \cdot 10^{-2}$
$temp_{( )} : fea_{( )}$	1	3.57	3.57	18.8	$5.83 \cdot 10^{-5}$
$pe_{( )} : sl_{( )}$	4	4.52	1.13	5.95	$4.39 \cdot 10^{-4}$
$np_{( )} : fea_{( )}$	1	1.18	1.18	6.22	$1.55 \cdot 10^{-2}$
Residuals	58	11.0	0.190		

Figure 1
Click here to download high resolution image

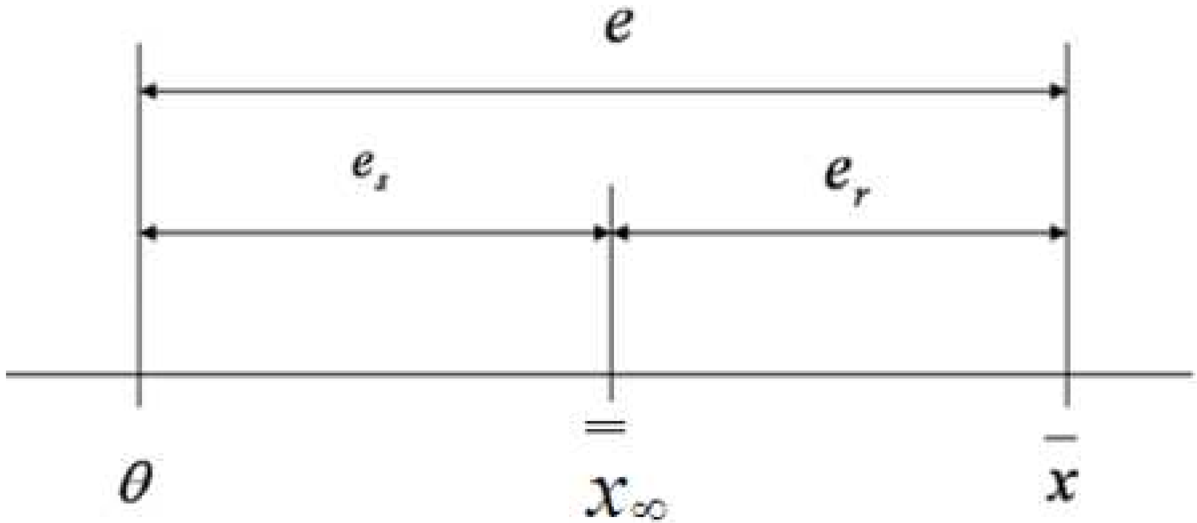

Figure 2
Click here to download high resolution image

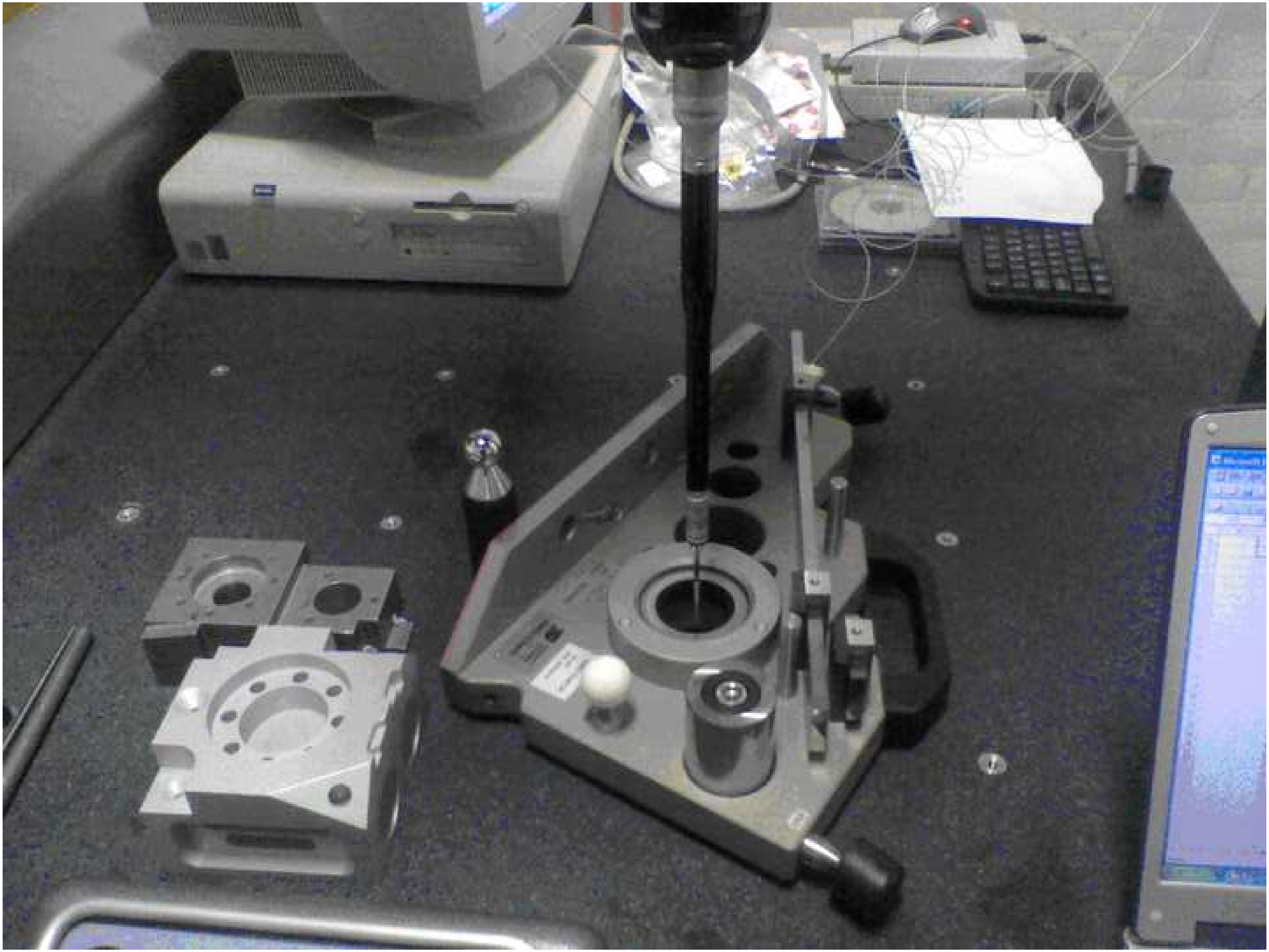

Figure 3
Click here to download high resolution image

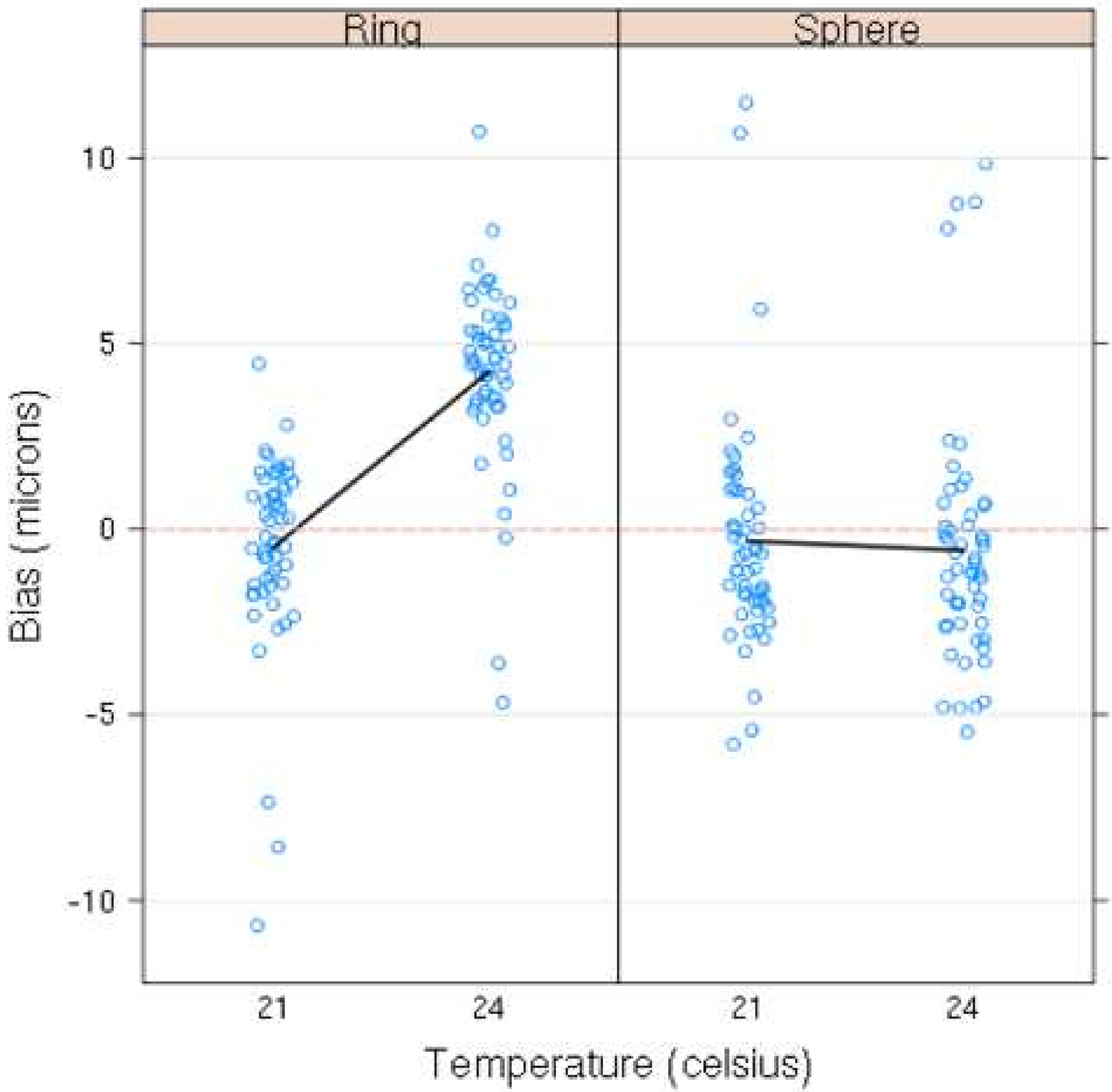

Figure 4
Click here to download high resolution image

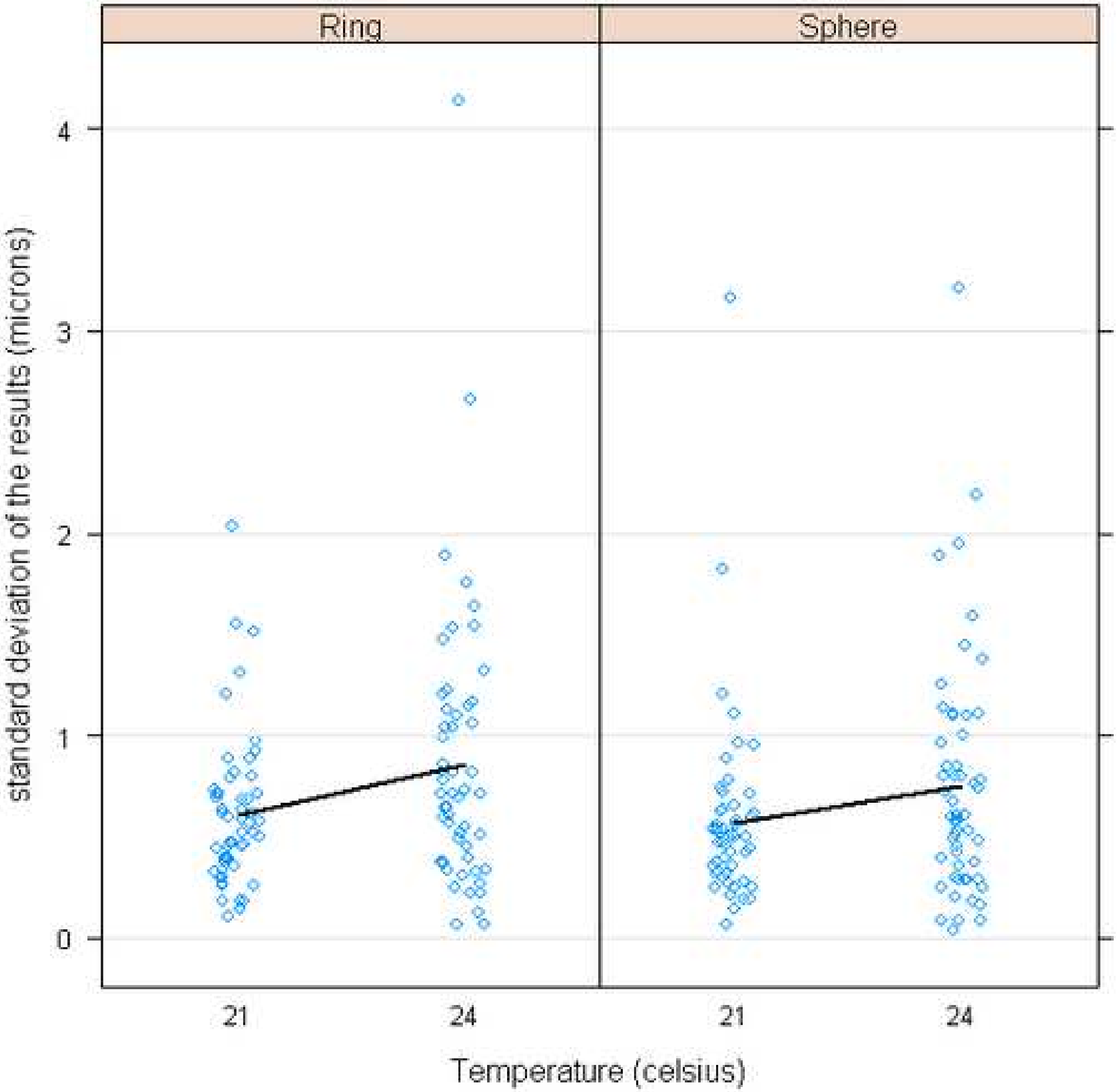

Figure 5
Click here to download high resolution image

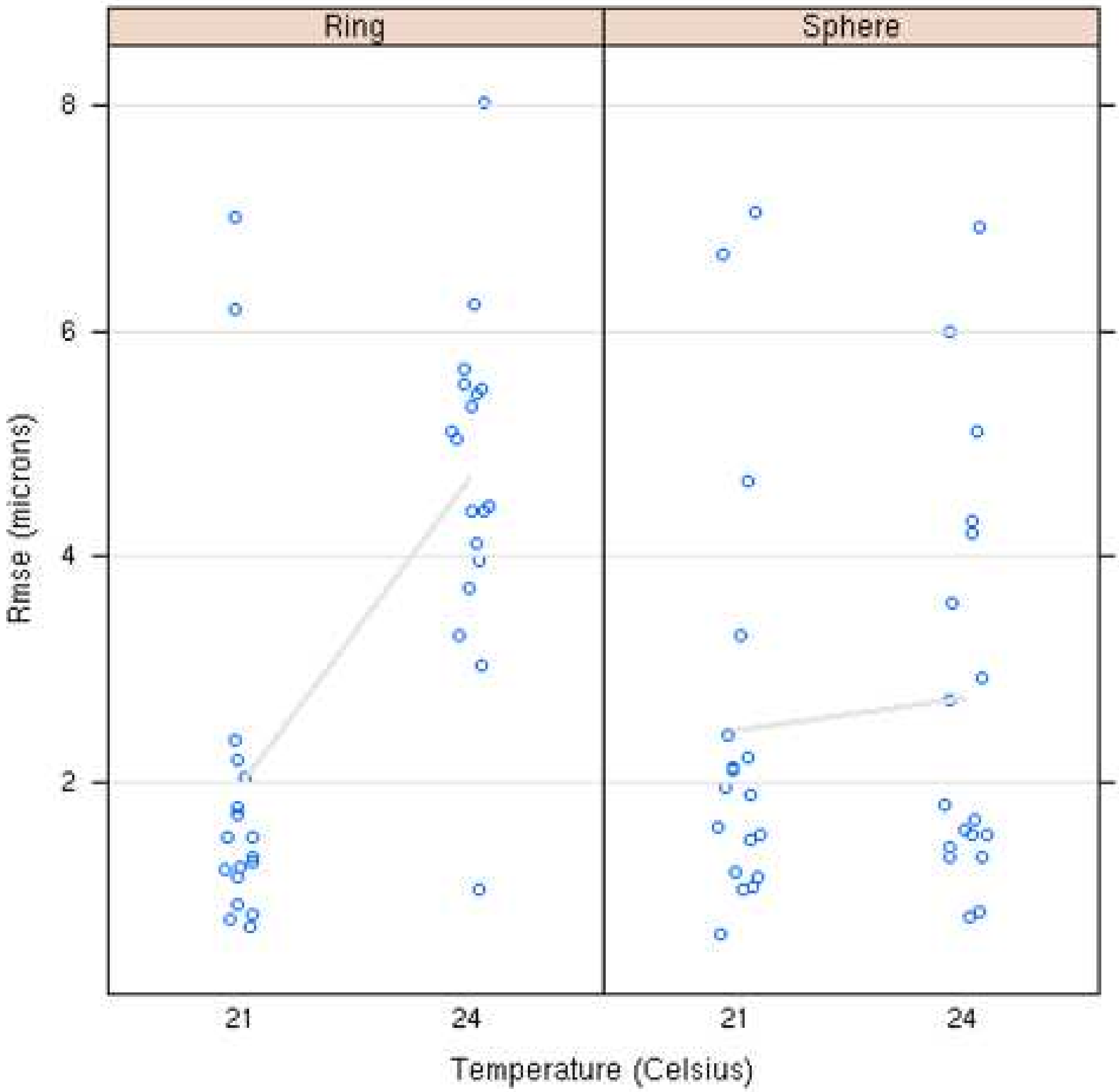

Figure 6
Click here to download high resolution image

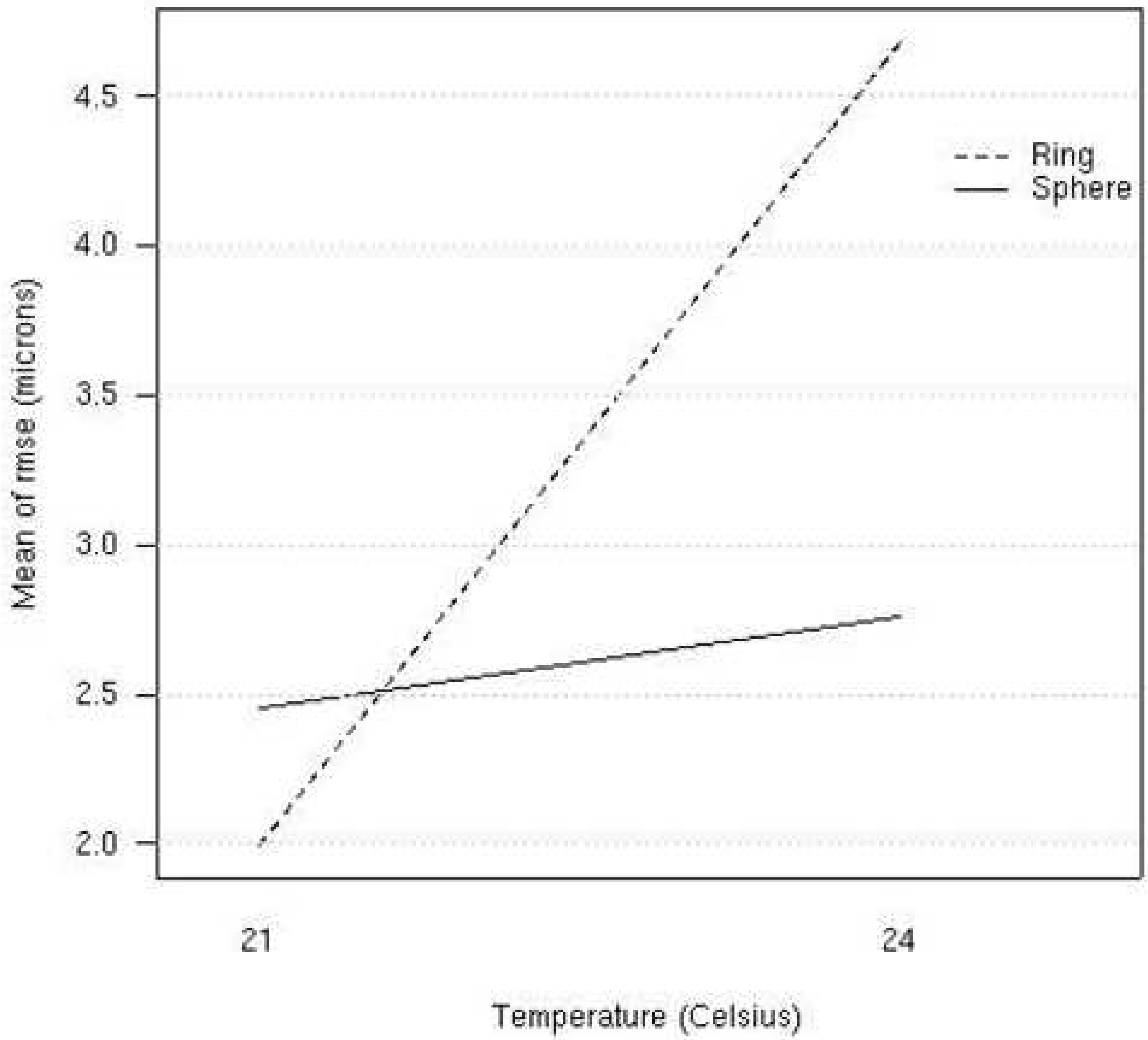

Figure 7
Click here to download high resolution image

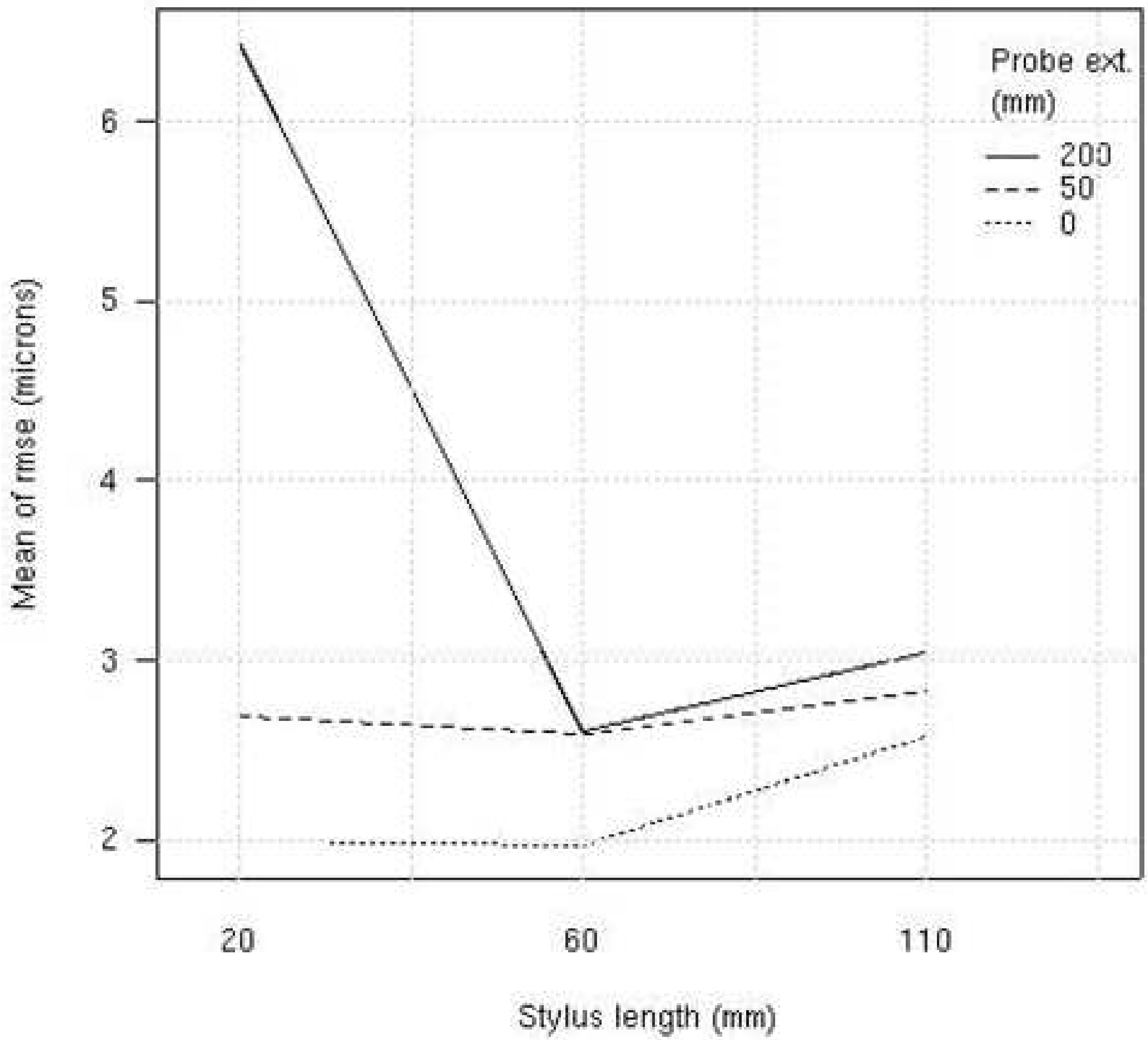

Figure 8
Click here to download high resolution image

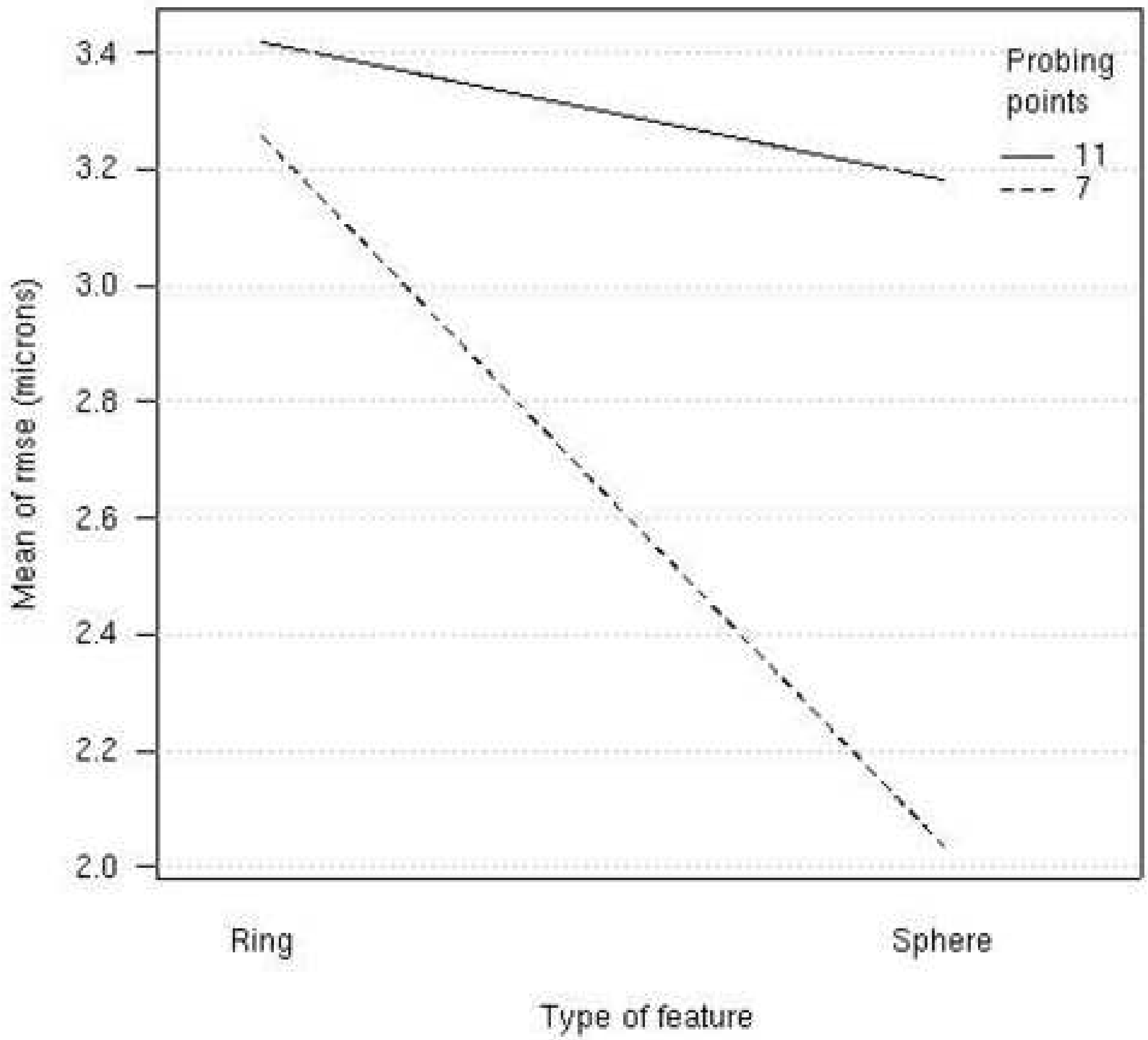

Figure 9
Click here to download high resolution image

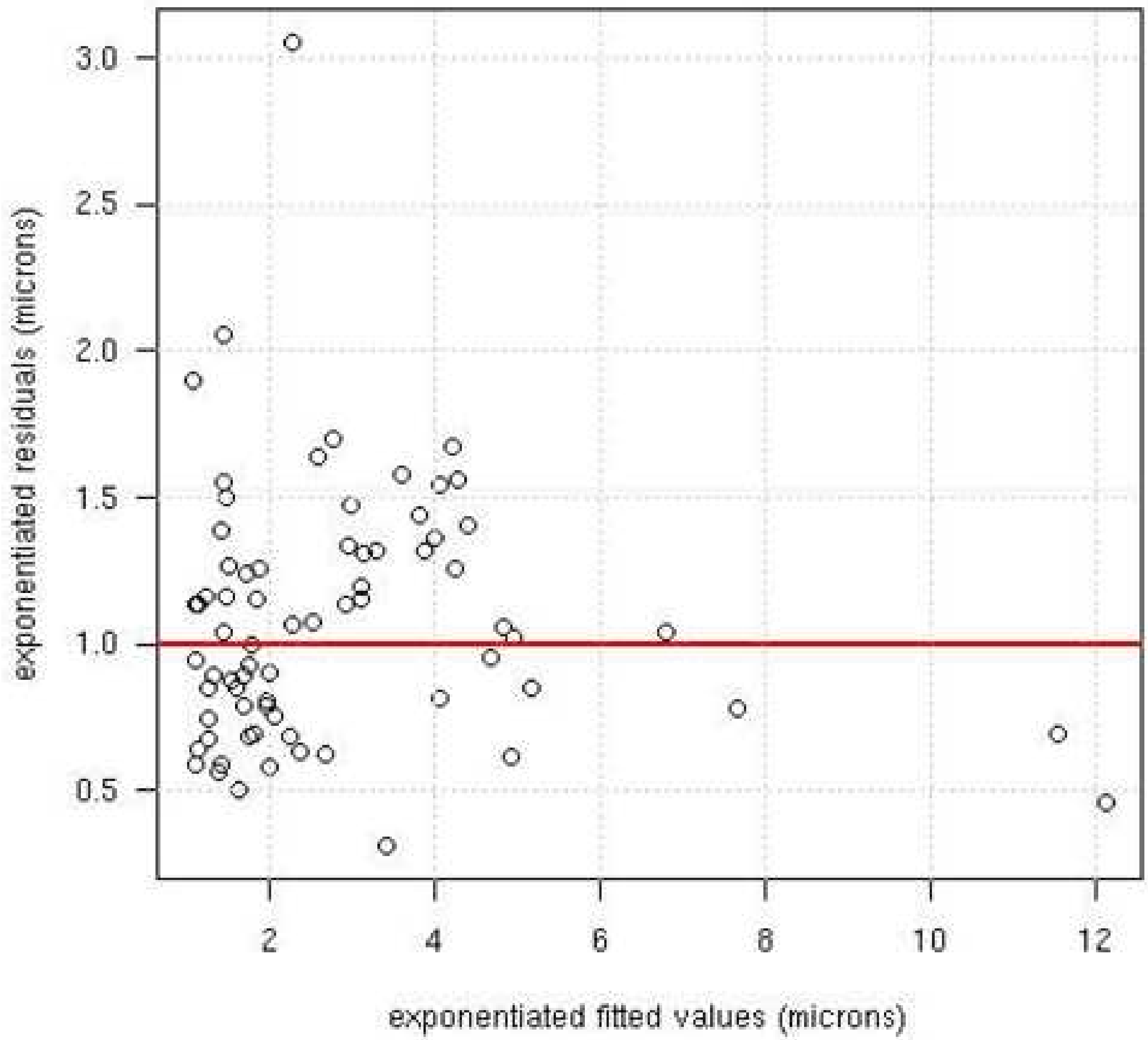

Figure 10
Click here to download high resolution image

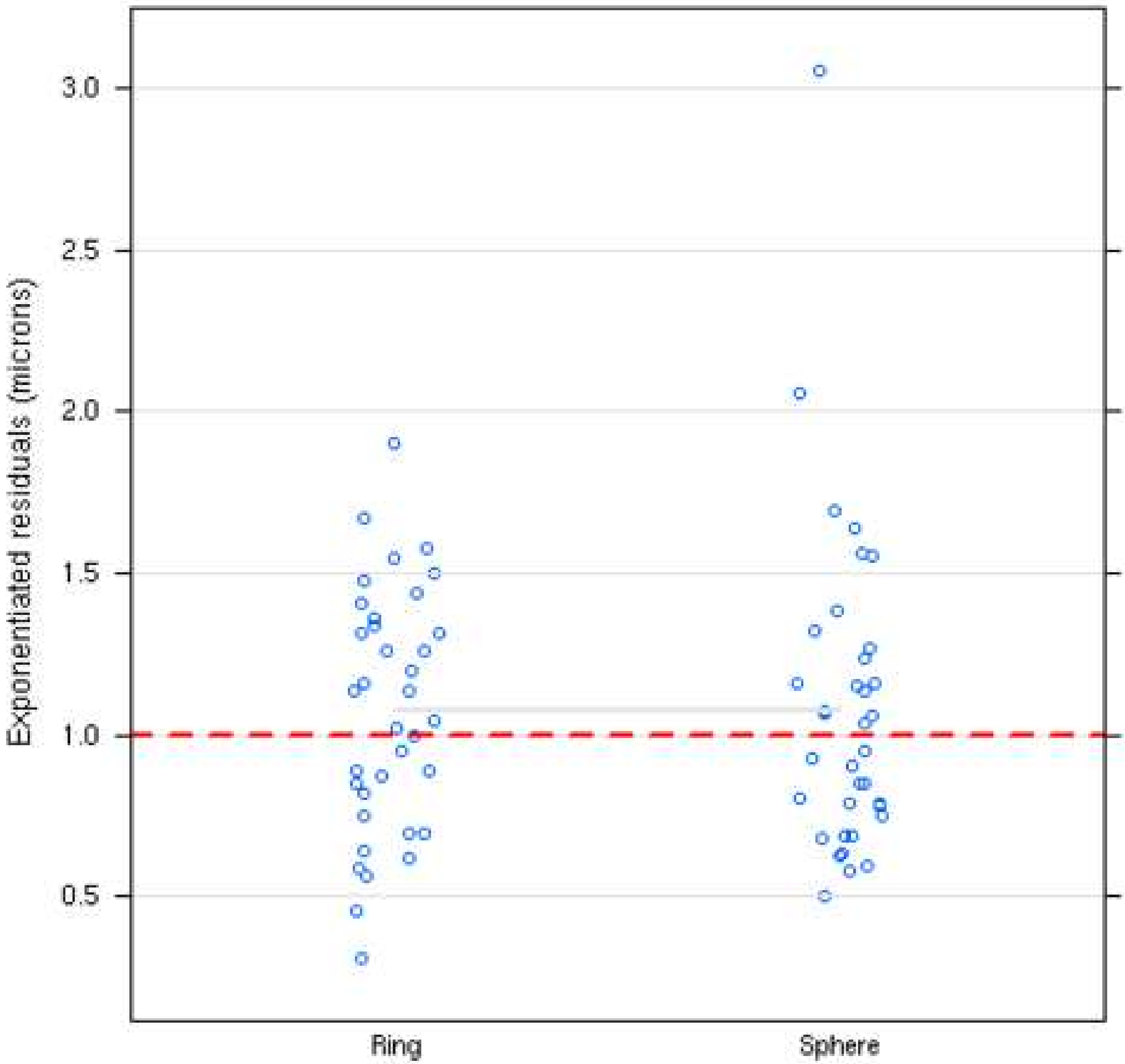

Figure 11
Click here to download high resolution image