Reliable Computing 1 (3) (1995), pp. 275–284

Study on sampling techniques with CMMs

THOMAS J. MCLEAN and DAVID H. XU

Coordinate Measuring Machines (CMMs) coupled with computers have provided new, powerful capabilities in the field of manufacturing quality control Most CMMs in use today rely on point sampling to evaluate the specified dimensions and tolerances: we measure the coordinates in several selected points, and make conclusions about the entire surface Several different sampling patterns have been proposed In this paper, we describe theoretical and experimental results on choosing the best sample pattern for practical CMM applications

Исследования методик выборки с помощью КИМ

Т. Дж. Маклин, Д. Х. Ксу

Координатные измерительные машины (КИМ) в сочетании с компьютерами предоставляют новые широкие возможности в области контроля качества производства. Большинство используемых в настоящее время КИМ применяют точечную выборку для оценки заданных размеров и допусков, т е координаты измеряются в нескольких выбранных точках, и на основании этих измерений делаются выводы о всей поверхности. Для этих целей предложено несколько различных моделей выборки. В настоящей работе излагаются теоритические и экспериментальные выводы о выборе наилучшей модели для практических приложений КИМ.

1. Introduction: formulation of a real-life problem

The existing Coordinate Measuring Machines (CMMs) measure the precise values of 3–D coordinates x, y, and z of chosen points on a workpiece with an accuracy of 1 to 3 microns (see, e.g., [1]). The results of these measurements are used to evaluate the characteristics of the entire surface [1, 3–5]. For example, suppose that we have manufactured a surface that is supposed to be planar, and want to evaluate how planar the surface actually is. If we orient the surface in such a way that the desired plane coincides with $0xy$, then the ideal surface would be described by the equation $z = 0$. Since manufacturing is never perfect, the values of z for the points on the actual surface will be different from 0. The largest value M of $|z|$ for all the points (x, y, z) from this surface characterizes its planarity. To measure this characteristic, we select several points (x_i, y_i) on the plane, measure the corresponding values of z_i, and take the largest value $\max|z_i|$ as the estimate of M (see, e.g., [6]).

Similarly, if we want to check whether an edge of the workpiece is linear, we place it so that this edge is as close to the $0x$ axis as possible. Then, ideally, or all points (x, y, z) from this edge, we would get $y = z = 0$. For a real-life manufactured edge, we have $y \neq 0$ and $z \neq 0$ The distance r between the actual point (x, y, z) and the ideal edge $y = z = 0$ is equal to $r = \sqrt{y^2 + z^2}$ So, as a characteristic L of linearity of the edge, we can take the largest value

$L = \max r$ of this distance r for all points from the edge. To measure this characteristic, we choose several points x_i on the line, measure the corresponding values of y_i and z_i, compute $r_i = \sqrt{y_i^2 + z_i^2}$, and take $\tilde{L} = \max r_i$ as an estimate of L. The larger the number of points, the smaller is the error $L - \tilde{L}$ of this estimate.

The main problem with using CMM is that the larger the number of points for which we measure coordinates, the longer it takes to measure them. CMM is a million-dollar machine, and it is very expensive to use. In view of this, it is desirable to minimize the number of points for which we measure coordinates. Usually, the number of points which we can measure is limited by economic considerations, so, we have the following problem: the total number n of points is given; we must select the points in such a way that the resulting error will be the smallest possible. In this paper, we consider the problem of choosing the points for testing linearity.

For linearity, there are currently three main methods of selecting the points x_i on a given interval $[0, X]$:

- *Completely random sampling*, in which n values x_i are chosen randomly; all n choices are independent, and each variables x_i is uniformly distributed on the interval $[0, X]$. This method is currently the most recommended and the most thoroughly analyzed in the academic community.
- *Randomized block sampling*, in which we divide the interval $[0, X]$ into b blocks $[0, X/b]$, $[X/b, 2X/b], \ldots, [(b-1)X/b, X]$. Then, we select $x_1, \ldots, x_{(n/b)-1}$ to be independent random variables uniformly distributed in the first block; $x_{(n/b)}, \ldots, x_{2(n/b)-1}$ are independent random variables uniformly distributed in the second block, etc.
- *Equal distance sequential sampling*, in which the values x_i are equally spaced: $x_i = x_1 + (i-1)s$ for some step s. This method is the easiest to implement, and because of that, it is most widely used in practice.

Completely random sampling can be viewed as a degenerate case of the randomized block sampling, for which there is only one block ($b = 1$).

In this paper, we show that both from the experimental and the theoretical viewpoint, equal distance sequential sampling is the best. The main experimental results of this paper have appeared in the thesis [7] of one of the authors (this thesis also contains *statistical estimates*, that we did not include into this paper). A brief description of our results was given in [2].

2. Theoretical analysis of the problem

2.1. The main idea

The main idea of of theoretical analysis is as follows: we are analyzing the shape of the manufactured objects. Manufacturing smoothes the discontinuities, and thus, the resulting function $r(x)$ is smooth. From the manufacturing considerations, we can estimate how smooth the function $r(x)$ is: namely, we can get an upper bound Δ for the derivative $r'(x)$: $|r'(x)| \le \Delta$. Therefore, the function $r(x)$ must satisfy the following inequality: $|r(x) - r(y)| \le \Delta \cdot |x - y|$ for all x and y.

Comment. The estimate Δ that we can get is reasonably crude. However, this crudeness does not bother us very much, because, as we will see later, the choice of the sampling does not depend on the value of Δ.

Due to this estimate, if we know the values $r(x_i)$ for n values $x_1 < x_2 < \cdots < x_n$, then, for every other x, we can find the *interval* $\mathbf{r}(x)$ of possible values of x. For example, if $x < x_1$, then, from $|r(x) - r(x_1)| \le |x - x_1|$, we can conclude that $r(x) \in \mathbf{r}(x) = [r(x_1) - \Delta \cdot (x_1 - x), r(x_1) + \Delta \cdot (x_1 - x)]$. If $x_i < x < x_{i+1}$, then we have two inequalities $|r(x) - r(x_i)| \le \Delta \cdot (x - x_i)$ and $|r(x) - r(x_{i+1})| \le \Delta \cdot (x_{i+1} - x)$, from which we conclude that $r(x) \in \mathbf{r}(x) = [r^-(x), r^+(x)]$, where

$$\begin{aligned} r^-(x) &= \max\left(r(x_i) - \Delta \cdot (x - x_i), r(x_{i+1}) - \Delta \cdot (x_{i+1} - x)\right) \quad \text{and} \\ r^+(x) &= \min\left(r(x_i) + \Delta \cdot (x - x_i), r(x_{i+1}) + \Delta \cdot (x_{i+1} - x)\right). \end{aligned}$$

The narrower the intervals, the better the sample. Let us describe this idea in mathematical terms. In this description, we will take into consideration that the measurement is never absolutely accurate, and therefore, the measured values $\tilde{r}_i$ may be slightly different from the actual values $r(x_i)$.

2.2. Definitions and the main result

Definition.

- *Let $X > 0$ and $\Delta > 0$ be positive real numbers. By $\mathcal{R}$, we will denote the set of all Δ–Lipschitz functions $r : [0, X] \to [0, \infty)$, i.e., the set of all non-negative functions for which $|r(x) - r(y)| \le \Delta \cdot |x - y|$ for all $x, y \in [0, X]$.*

- *Let a positive integer n be given. This integer is called a number of measurements. By a pattern $\vec{x}$, we mean an increasing sequence of n numbers from the interval $[0, X]$ (i.e., a sequence x_i for which $0 \le x_1 < x_2 < \cdots < x_n \le X$).*

- *Let $\delta > 0$ be a positive real number; this number will be called a measurement accuracy. By measurement results, we mean a sequence $\vec{r}$ of n non-negative real numbers $\tilde{r}_1, \ldots, \tilde{r}_n$. We say that a function $r \in \mathcal{F}$ is consistent with the measurement results $\vec{r}$ if $|r(x_i) - \tilde{r}_i| \le \delta$ for all $i = 1, 2, \ldots, n$. We say that the measurement results are consistent if there exists a function r that is consistent with them. If a function r is consistent with measurement results $\vec{r}$, then we define measurement error as the value $|L - \tilde{L}|$, where $L = \max r(x)$ and $\tilde{L} = \max(\tilde{r}_i)$.*

- *By a guaranteed error $E(\vec{x})$ of a pattern $\vec{x}$, we mean the largest possible measurement error for this pattern.*

Proposition. *The pattern $x_i = (i - 1/2) \cdot (X/n)$ has the smallest possible guaranteed error.*

Comment. So, in the sense of minimizing guaranteed error, the equal distance sequential sampling is the best choice.

Proof. This proof is reasonably simple. Namely, we will compute the guaranteed error of the pattern described in Proposition, and show that every other pattern has a larger guaranteed error.

For this pattern, the step s is equal to X/n, and $x_1 = s/2$. For every number $x \in [0, X]$, we can find x_i that is closest to this x by taking $i = \lceil x/s \rceil$. One can easily see that $|x - x_i| \le s/2$. Let $\vec{r}$ be any measurement results, and let $r \in \mathcal{R}$ be a function that is consistent with these results. Since $r \in \mathcal{R}$, we have $r(x) \le r(x_i) + \Delta \cdot (s/2)$. By definition of consistency, we have $r(x_i) \le \tilde{r}_i + \delta$. Therefore, $r(x) \le r(x_i) + \Delta \cdot (s/2) \le \tilde{r}_i + \Delta \cdot (s/2) + \delta$. Since $\tilde{r}_i \le \tilde{L}$, we have $r(x) \le \tilde{L} + \Delta \cdot (s/2) + \delta$. This is true for all x, and therefore, for $L = \max r(x)$, we have $L \le \tilde{L} + \Delta \cdot (s/2) + \delta$.

On the other hand, $\tilde{r}_i \le r(x_i) + \delta \le \max r(x) + \delta = L + \delta$. This inequality is true for all i and therefore, $\tilde{L} = \max \tilde{r}_i \le L + \delta$. Hence, $\tilde{L} \le L + \delta + \Delta \cdot (s/2)$.

Combining these two inequalities, we conclude that $|L - \tilde{L}| \le \delta + \Delta \cdot (s/2)$. So, for this pattern, $E(\vec{x}) \le \delta + \Delta \cdot (s/2)$.

Let us now show that for every other pattern $\vec{y}$, $E(\vec{y}) > \delta + \Delta \cdot (s/2)$. Let us first prove that if y_i is different from x_i, then either y_1 is largest than $s/2$, or the difference between $y_{i+1} - y_i$ is greater than s for some i, or $X - y_n$ is greater than $s/2$ (here $s = X/n$ is the step of our chosen pattern $\vec{x}$, that we are currently proving to be the best). Indeed, if none of these inequalities would be true, then we would have $y_1 \le s/2$, $y_{i+1} - y_i \le s$, and $X - y_n \le s/2$. Therefore, we would have $X = y_1 + (y_2 - y_1) + \cdots + (y_n - y_{n-1}) + (X - y_n) \le s/2 + s + \cdots + s + s/2 = X$, and the only possibility of equality is when all these inequalities are equalities, i.e., when $y_1 = s/2$ and $y_{i+1} - y_i = s$ for all i. But in this case, we would have $y_i = x_i$ for all i, and we assumed that $\vec{y} \ne \vec{x}$.

So, we have proven that if the pattern $\vec{y}$ is different from $\vec{x}$, then either $y_1 > s/2$, or $y_{i+1} - y_i > s$ for some i, or $X - y_n > s/2$. Let us show that in all three cases, $E(\vec{y}) > \delta + \Delta \cdot (s/2)$.

- First, let us consider the case when $y_1 > s/2$. Let the measurement results consist of identical numbers $\tilde{r}_i = -\delta$; then, $\tilde{L} = -\delta$. Let $r(x)$ be equal to 0 when $x \ge y_1$, and to $\Delta \cdot (y_1 - x)$ for $x < y_1$. Then, as one can easily check, r is consistent with $\vec{r}$. $L = \max r(x) = \Delta \cdot y_1$, so $L - \tilde{L} = \delta + \Delta \cdot y_1$. Since $y_1 > s/2$, we have $L - \tilde{L} \ge \delta + \Delta \cdot (s/2)$. By definition, $E(\vec{y}) \ge L - \tilde{L}$, and therefore, $E(\vec{y}) > \delta + \Delta \cdot (s/2)$.

- Second, let us consider the case when $y_{i+1} - y_i > s$ for some i Let the measurement results consist of identical numbers $\tilde{r}_i = -\delta$; then, $\tilde{L} = -\delta$ Let $r(x)$ be equal to 0 when $x \notin [y_i, y_{i+1}]$, to $\Delta \cdot (x - y_i)$ for $y_i \le x \le (y_i + y_{i+1})/2$, and to $\Delta \cdot (y_{i+1} - x)$ for $(y_i + y_{i+1})/2 \le x \le y_{i+1}$. Then, as one can easily check, r is consistent with $\vec{r}$. $L = \max r(x) = \Delta \cdot (y_{i+1} - y_i)/2$, so $L - \tilde{L} = \delta + \Delta \cdot (y_{i+1} - y_i)/2$. Since $y_{i+1} - y_i > s$, we have $L - \tilde{L} \ge \delta + \Delta \cdot (s/2)$. By definition, $E(\vec{y}) \ge L - \tilde{L}$, and therefore, $E(\vec{y}) > \delta + \Delta \cdot (s/2)$.

- The third case is proven similarly to the first two; we have $r(x) = 0$ for $x \le y_n$, and $r(x) = \Delta \cdot (x - y_n)$ for $x \ge y_n$.

In all three cases, we have $E(\vec{y}) > \delta + \Delta \cdot (s/2) \ge E(\vec{x})$. So, the proposition is proven. □

3. Experimental results

We tested different sampling techniques on the experimental data from the National Institute of Standards and Technology (NIST) project [1]. These data include 12 measured lines on the following specimens:

- specimens A–H used by General Electric Co. as their standard planar specimens;

- specimen NAS: a NAS 979 standard artifact,
- a glass "optical flat" specimen REP. This specimen was measured using a "light box" to certify its surface flatness.

All the lines were measured using a Sheffield Cordax CMM available in the Automated Manufacturing Research Facility (AMRF) of NIST. For each line, a set of 400 readings over an approximately 2-inch distance was recorded. The largest deviation from linearity among these 400 measurements was taken as the actual value of L. Then, several (n) points were chosen according to the principles of equal distance, completely random, and randomized samplings, and the maximum of measured values over these chosen points was taken as $\tilde{L}$. We tested each method with $n = 3, 4, 5, \ldots, 40$. The plots that describe the dependency of the ratio $\tilde{L}/L$ (measured "straightness" $\tilde{L}$ to true straightness L) are given. For the majority of them, equal spacing does lead to a smaller error than the alternative two patterns.

This is especially clear when n is small. Indeed, in this case, $s = X/n$ is reasonably large so, $\Delta \cdot (s/2) \gg \delta$, and hence, the error component $\Delta \cdot (s/2)$ that is influences by the choice of the pattern is the major component of the total error $L - \tilde{L}$. When n increases, δ (the measurement error of the CMM) becomes the major component of the error $L - \tilde{L}$. In this case, the choice of the pattern becomes rather irrelevant. The CMM measurement error is random, so on several graphs, we see random fluctuations of the total error for large n.

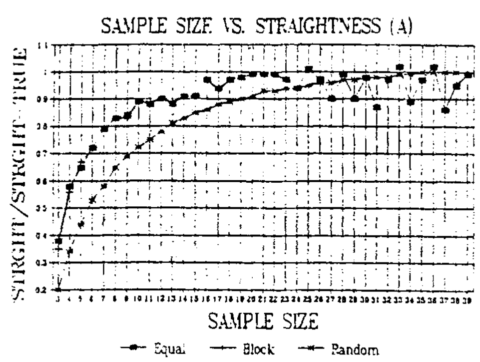

Figure 1

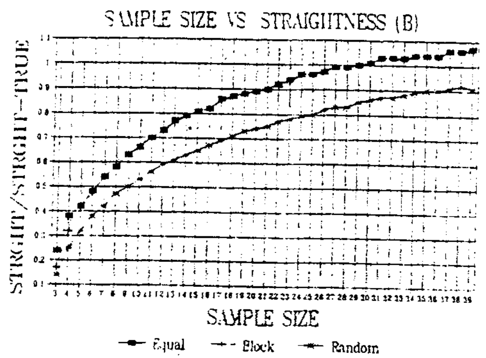

Figure 2

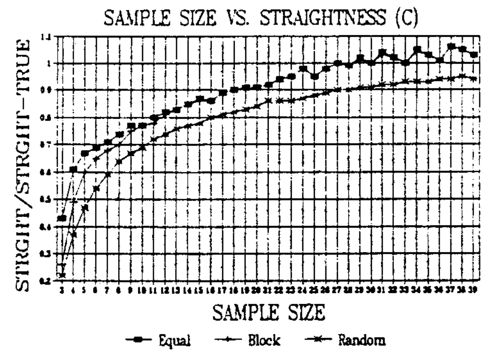

Figure 3

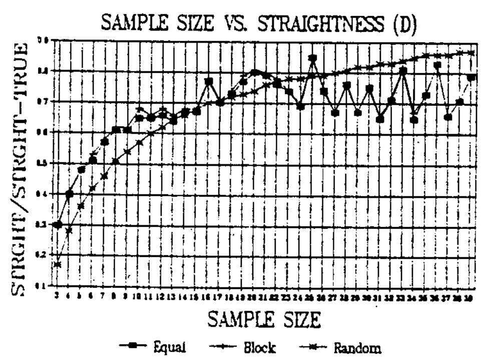

Figure 4

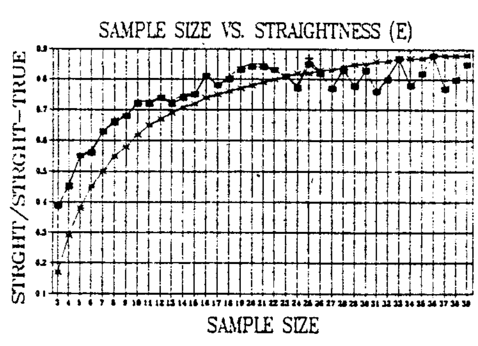

Figure 5

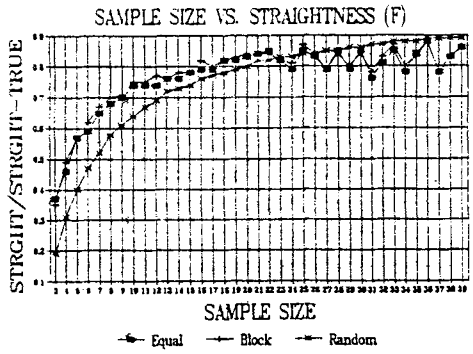

Figure 6

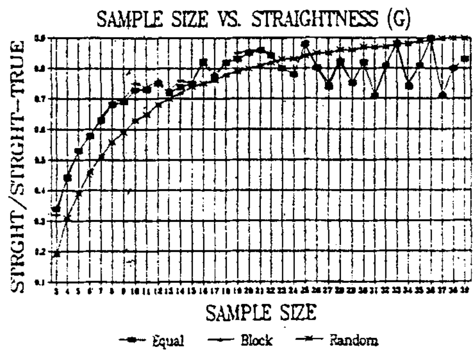

Figure 7

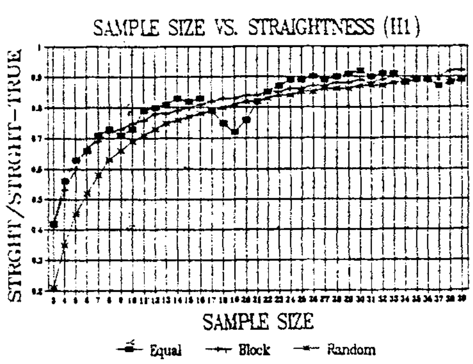

Figure 8

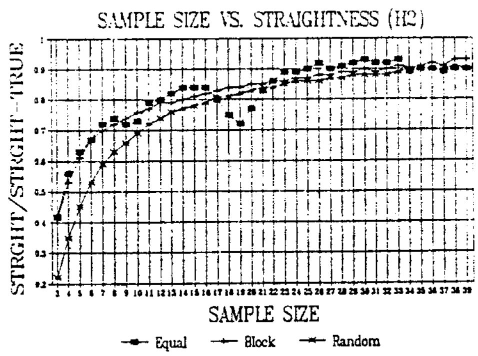

Figure 9

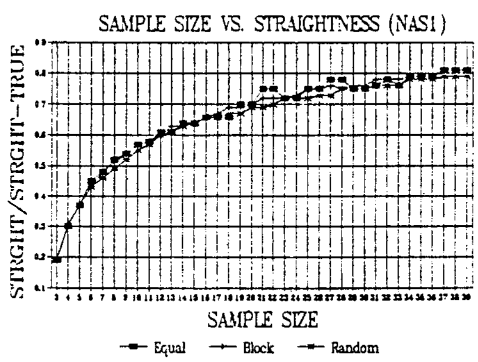

Figure 10

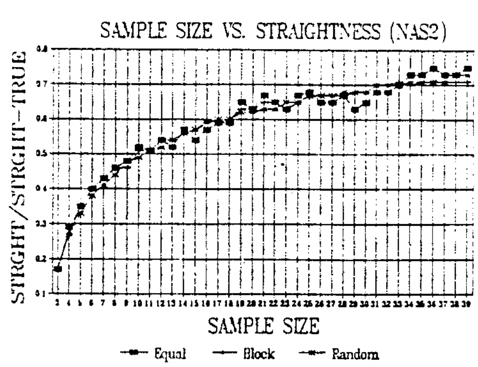

Figure 11

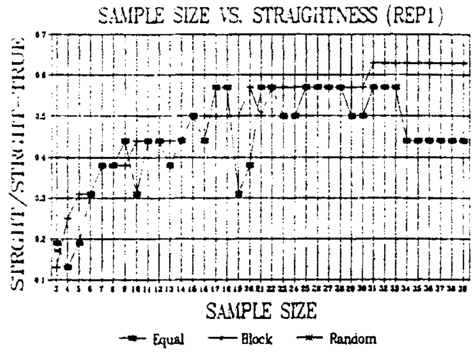

Figure 12

4. Conclusions

We have shown that when we check linearity of an edge of a workpiece by measuring coordinates of several sample points with a Coordinate Measuring Machine, the guaranteed error is the smallest when we use equal distance sampling.

References

[1] Hsu, J. P. and Hsu, T. W. *Status report on statistical inspection in coordinate measuring for ASA/NIST/NSF 1990–91 research fellowship program.* National Institute of Standards and Technology, 1991.

[2] McLean, T. J. and Xu, D H. *Measuring surface with a coordinate measuring machine: intervals of precision.* In: "Abstracts for a Workshop on Interval Methods in Artificial Intelligence, International Conference on Numerical Analysis with Automatic Result Verification: Mathematics, Application and Software, February 25—March 1, 1993", Lafayette, LA, 1993, p 23.

[3] Murthy, T S. R. and Abdin, S. Z. *Minimum zone evaluation of surfaces.* Int. J. of Machines Tool Design and Research **20** (1980), pp. 123–136.

[4] Sweet, A. L., Noller, D., and Lee, S.-H. *Statistical design for the location of planes and circles when using a probe* Precision Engineering **7** (4) (1985), pp. 187–194.

[5] Traband, M. T., Joshi, S., Wysk, R. A., and Cavalier, T. M. *Evaluation of straightness and flatness tolerances using the minimum zone.* Manufacturing Review **2** (3) (1989), pp. 189–195.

[6] Weckermann, A. and Heinrichowski, M. *Problems with software for running Coordinate Measuring Machines.* Precision Engineering **7** (2) (1985), pp. 87–91.

[7] Xu, D. H. *Statistical inspection using coordinate measurement machines* University of Texas at El Paso, Mechanical & Industrial Engineering Department, Master Thesis, 1992.

Received: December 15, 1992
Revised version: April 20, 1995

Department of Mechanical & Industrial Engineering
The University of Texas at El Paso
El Paso
TX 79968
USA