▩ 연구논문

# Sample Size Determination and Evaluation of Form Errors+

Sung Ho Chang
Dept. of Industrial Engineering, Kum Oh National University of Technology
Sunn Ho Kim
Dept. of Industrial Engineering, Myoung Ji University

## Abstract

In current coordinate measuring machine practice, there are no commonly accepted sample sizes for estimating form errors which have a statistical confidence. Practically, sample size planning is important for the geometrical tolerance inspection using a coordinate measuring machine. We determine and validate appropriate sample sizes for form error estimation. Also, we develop form error estimation methods with certain confidence levels based on the obtained sample sizes in various form errors: straightness, flatness, circularity, and cylindericity.

## 1. Introduction

The evaluation of form errors (e.g. straightness, flatness, circularity, and cylindericity) using a coordinate measuring machine(CMM) relies on discrete measurements. However, definitions of form errors in the current standards (ISO 1101, ANSI Y14.5) assume perfect (continuous) measurements, not discrete measurements. Therefore, there is no commonly accepted method for calculating form errors using discrete measurements; it is current practice to satisfy the definitions of the standards using discretely measured points. However, current practice does not consider the uncertainty of manufactured surfaces. As a result,

+ 본 연구는 한국과학재단의 핵심전문연구의 일부지원에 의해 수행되었음.

it is not possible to give statistical confidence to the estimated form errors or to suggest statistically reliable minimum sample points. At the same time, the number of measured points needed to be large enough to provide reliable results.

Theoretically, the minimum number of points to calculate form errors are straight forward. As an example, a minimum of three points are necessary to get a straightness error. Two points are used to estimate a straight line and one point is used to get the information about the uncertainty of the estimated straight line. If all three measured points lie on the perfect straight line, then there is no straightness error because the third point does not give any information. If those three points are not on a straight line, the third point gives information about the straightness error. However, there are no surfaces or curves whose uncertainty information (irregularity) can be explained by one point. Therefore, the theoretical minimum number of three points are not enough to obtain information about form errors. Additional measurements are needed to get statistically reliable information. By establishing a statistically reliable minimum, the manufacturer does not have to measure an inordinate number of points.

Several assessment methods have been developed to assess form errors using discrete measurements but they do not consider the sample size and statistical confidence. These assessment methods can be classified by their objective functions. The conventional least squares (LS) technique, which can be found in various introductory regression analysis books [Draper and Smith(1966), Neter et. al.(1985)], minimizes the sum of the squared linear deviations. The minimum deviation (MD) technique minimizes the sum of the absolute values of the arithmetic maximum and minimum linear deviations. The minimum average deviation (MAD) technique [Shunmugam, 1987] minimizes the sum of the absolute linear deviations. The minimum zone (MZ) technique [Murthy and Abdin, 1980] minimizes the sum of squared normal deviations. Recently Menq et. al(1990) suggested the sample size planning method. However, they did not consider the difference between concerned surfaces. Then, they gave the same sample size regardless of the concerned surface.

The different objective functions depend on the estimates of parameters. General expressions for the parameters of the least square (LS) principle are given by Murthy and Abdin(1980) and Kakino and Kitazawa(1978). For the minimum zone (MZ) principle, Murthy and Abdin(1980) and ElMaraghy et al. (1989) give the estimation of parameters. For the minimum average deviation principle, Shunmugam(1987) give the estimation of parameters. For the minimum deviation (MD) principle, Gota and Lizuka(1977) and Shimokohbe(1984) give the estimation of parameters.

Each approach obtains the best result satisfying its specific and limited objective function. These approaches estimate the form errors assuming the measured points include real maximum peak and valley points of the concerned surface. As a result, they give different information depending on which points are measured even on the same surface. Also these approaches do not consider the characteristics of manufactured surfaces which can have various geometrical shapes relative to the ideal shape specified on drawings [Weckenmann and Heinrichowski(1985), Bourdet, Clement and Weill(1984)]. In this paper we develop an alternate approach for determining the appropriate sample size and for estimating form errors based on the definitions proposed in the previous research [Chang et. al., 1990]

## 2. Proposed Approach

We define form error as $6\sigma$ or the function of the surface shape parameter in the previous research [Chang et. al., 1990]. We usually do not know the real values of a standard deviation or a surface shape parameter. They are estimated from a estimated nominal surface. The nominal surface is estimated from sample measurements. Because reference measuring axes in a CMM are arbitrarily chosen, the nominal surface is estimated depending on the reference coordinates or axes (See Figure 1). Also we do not know the exact values of maximum and minimum deviations, then we predict those values. Therefore, there are two variations in estimating form errors: 1) variation in possible location of the nominal surface or feature, and 2) variation within the estimation of standard deviation which involves probability distribution. We use the linear regression method to estimate the nominal surface. To use the linear regression method, we make assumptions based on the fact that machining processes are always disturbed by various noises which are independent of the form of the surface. Hence the cumulative effect of these noises is subject to the central limit theorem and is governed by a Normal distribution [Greenwood and Williamson, 1966]. Under these assumptions, we can make basic assertions that involve probability distributions.

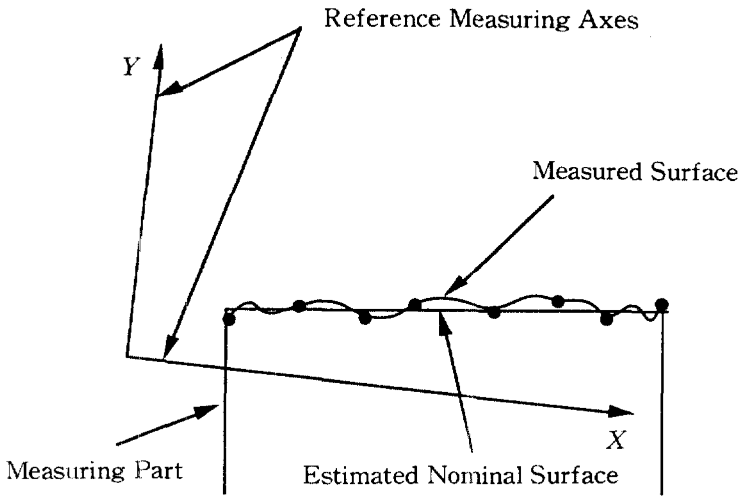
Let our manufactured surface be represented by the functional form

$$Z_i = f(X_i, Y_i) + \varepsilon_i$$

where $f(X_i, Y_i)$ : function of manufactured surface

$f(X_i, Y_i) = \beta_0 + \beta_1 x_i$ (for simple straightness case)

$\varepsilon_i$ : combined noise




〈Figure 1〉 Reference measuring axes in a CMM and estimated nominal surface

1. $\varepsilon_i$ is a normal random variable with mean zero and variance $\sigma^2$ (unknown), that is,

$$\varepsilon_i \sim N(0, \sigma^2),\ E(\varepsilon_i) = 0,\ V(\varepsilon_i) = \sigma^2.$$

2. $\varepsilon_i$ and $\varepsilon_j$ are uncorrelated, $i \neq j$, so that $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ and $Z_i$ and $Z_j$, $i \neq j$, are uncorrelated. Thus

$$E(Z_i) = f(X_i, Y_i), \qquad V(Z_i) = \sigma^2$$

Based on these assumptions, each observation comes from a normal distribution centered vertically at the level implied by the proposed model. The variance of each normal distribution is assumed to be homogeneous.

We can simply use estimated standard deviation for $6\sigma$ or surface shape parameter for special form of surface function to obtain form errors if there is no variations. However, we will not know much about the variations of the real nominal surface and the probability of parameter estimations. Therefore, we use a prediction interval length that considers the variation of real nominal surface and the variation of parameter estimation. The prediction interval length (PI) can be represented by the function of the sample size, a specified point and an estimated standard deviation in general linear regression analysis.

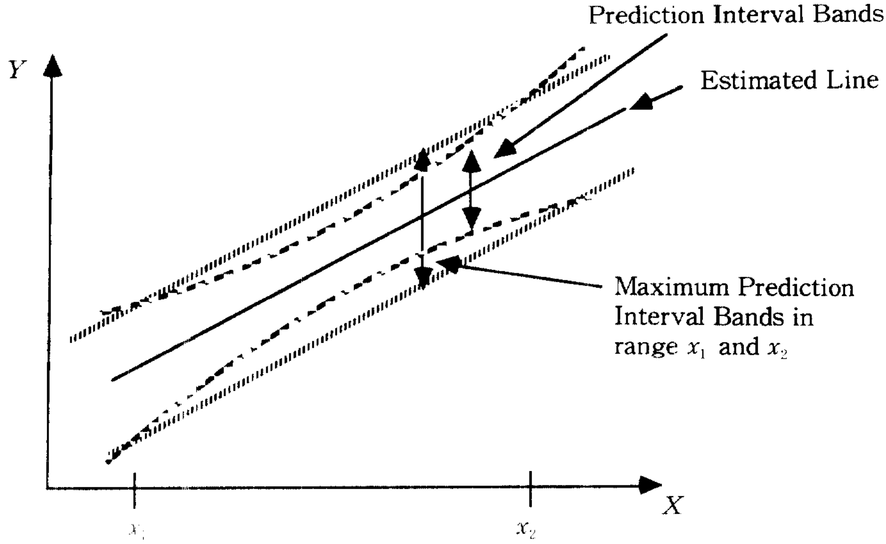
$$\mathrm{PI} = 2 * t(n-p, 1-\alpha/2) * h(n, \boldsymbol{P}_0) * \sqrt{\mathrm{MSE}}$$

where $t(n-p, 1-\alpha)$ : upper $(1-\alpha)$ percentage point of $t$-distribution with $(n-p)$ degrees of freedom

$h(n, \boldsymbol{P}_0)$ : function of sample size $n$ and a specified point $\boldsymbol{P}_0$

MSE : estimated variance

When a certain PI, with given sample size $n$, confidence level $(1-\alpha)$ and MSE, is approximately equal to $6\sigma$, we can say that it is an estimated form error. The bands of the PI, however, are curvilinear and our objective is to find linear bands which cover the maximum variations of the nominal surface and its estimation. The maximum PI is chosen at a given sample size ⟨Figure 2⟩.




⟨ Figure 2 ⟩ Illustration of Maximum Prediction Interval Bands

However, the interval estimate of the PI is a random variable because the sample standard deviation (or $\sqrt{MSE}$) is a random variable. The expected prediction interval is compared to $6\sigma$. We can say that PI at that sample size is an estimation of form error when the expected length of the maximum prediction interval, at a certain sample size with a certain confidence level, is approximately equal to $6\sigma$. The upper or lower confidence limits of the surface shape parameter at that sample size can be used to estimate form error.

The statements above can be represented in mathematical terms as follow;

$$\mathrm{PI}(\boldsymbol{P}_0) = 2 * t_{n-p,\,1-\alpha/2}\, h(n, \boldsymbol{P}_0)\sqrt{\mathrm{MSE}} \qquad (1)$$

where $\mathrm{PI}(\boldsymbol{P}_0)$ : length of prediction interval at $\boldsymbol{P}_0$

$h(n, \boldsymbol{P}_0) = \{1 + \boldsymbol{P}_0'(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}_0\}^{1/2}$

$\boldsymbol{P}_0$ : specified column vector of $\boldsymbol{P}$

$\boldsymbol{P}$ : observation design matrix
MSE : estimated variance of least square residuals
$n$ : # of measurements
$p$ : # of parameters estimated
$\alpha$ : confidence level.

Where $\boldsymbol{P}_0$ which gives the maximum value of PI is chosen among $\boldsymbol{P}$.

Since the residual error variance (MSE) follows a Chi-square distribution

$$\frac{(n-p)\,\mathrm{MSE}}{\sigma^2} \sim \chi^2_{n-p}$$

$$E[\sqrt{\mathrm{MSE}}] = \frac{\sigma}{\sqrt{n-p}} E[\chi_{n-p}]$$

the expected length of the prediction interval ($E$[PI]) can be represented as follows

$$E[\mathrm{PI}] = 2t_{n-p,1-\alpha/2}\, h(n, \boldsymbol{P}_0) \frac{\sigma}{\sqrt{n-p}} E[\chi_{n-p}] \qquad (2)$$

where $h(n, \boldsymbol{P}_0) = \{1 + \boldsymbol{P}_0'(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}_0\}^{1/2}$

$$E[\chi_{n-p}] = \frac{\Gamma[(n-p+1)/2]}{\Gamma[(n-p)/2]}\sqrt{2}$$

$\Gamma(n)$ = Gamma function

$$= \int_0^\infty x^{n-1} e^{-x}\, dx.$$

If we let $E[\mathrm{PI}] = 6\sigma$, we can determine the appropriate sample size needed to estimate the form error with the proposed form error definition with any degree of confidence. Alternatively we can determine appropriate confidence levels when the sample size is given which satisfies

$$t_{n-p,1-\alpha/2}\, h(n, \boldsymbol{P}_0) \frac{E[\chi_{n-p}]}{\sqrt{n-p}} = 3 \qquad (3)$$

In this paper we are only considering sample size determination when the confidence level is given for straightness, flatness, circularity, and cylindericity errors. The proposed approach for determining sample size and for estimating form errors not only satisfies the proposed definitions but also accounts for process variation in the estimating procedure.

# 3. Sample Size Determination

In this section, we explain the procedure for determining sample size using the maximum prediction interval approach for various functional forms. Observation matrix $\boldsymbol{P}$ is constructed with the assumption that each measurement is equidistance in every dimension (observations form a grid).

## 3.1 Simple Straight Line Function

The general simple straight line regression function from sample size $n$ is represented by

$$\hat{Y} = b_0 + b_1 X.$$

In the simple straight line case, the maximum value of $h(n, \boldsymbol{P}_0)$ can be obtained at one of two end points. For example, when $n = 3$:

$$\boldsymbol{P} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad (\boldsymbol{P}'\boldsymbol{P})^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \qquad \boldsymbol{P}_0 = (1, -1) \text{ or } (1, 1)$$

$$h(n, \boldsymbol{P}_0) = \{1 + \boldsymbol{P}_0' (\boldsymbol{P}'\boldsymbol{P})^{-1} \boldsymbol{P}_0\}^{1/2} = 1.354.$$

Likewise we obtain the maximum values of $h(n, \boldsymbol{P}_0)$ for different sample sizes. We find appropriate sample sizes which satisfy:

$$\{2t_{n-2, 1-\alpha/2} h(n, \boldsymbol{P}_0) \frac{E[\chi_{n-p}]}{\sqrt{n-2}}\} \approx 6 \qquad (4)$$

where $1-\alpha$ : confidence coefficient of the prediction interval.

The appropriate sample sizes for 95% ($\alpha = 0.05$) and 99% ($\alpha = 0.01$) confidence for the simple straight line function are 7 and 24. These sample sizes are found by increasing the sample size and searching the value of $E[\mathrm{PI}]/\sigma$ which is greater than or equal to 6.

## 3.2 Second Order Polynomial Curve Function

The general second order polynomial regression function from sample size $n$ is represented by

$$\hat{Y} = b_0 + b_1 X + b_2 X^2.$$

The maximum value of $h(n, \boldsymbol{P}_0)$ will be 1.396 when $n = 4$. In the same way we can get the maximum value of $h(n, \boldsymbol{P}_0)$ for different sample sizes. The appropriate sample sizes with 95% and 99% confidence for the second order polynomial curve function with the same procedure in the simple straight line function are 9 and 36.

### 3.3 Simple Plane Function

The general simple plane regression function from sample size $n$ is represented by

$$\hat{Z} = b_0 + b_1 X + b_2 Y.$$

As an example, the maximum value of $h(n, \boldsymbol{P}_0)$ will be 1.323 when $n = 4$. In the same way, we can get the maximum value of $h(n, \boldsymbol{P}_0)$ for different sample sizes. The result is similar to the simple straight line function except for the number of parameters to be estimated. The appropriate sample sizes with 95% and 99% confidence for the simple plane function are 8 and 25.

### 3.4 Second Order Surface Function

In this case we consider only the specific form of a surface

$$\hat{Z} = b_0 + b_1 X + b_2 X^2 + b_3 Y.$$

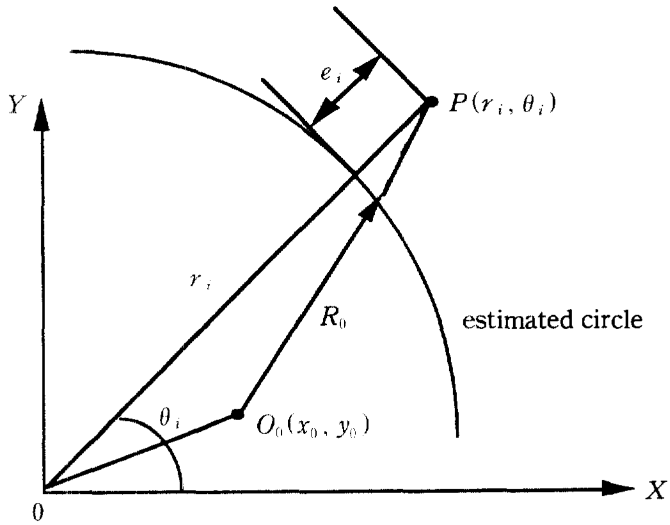
As an example, the maximum value of $h(n, \boldsymbol{P}_0)$ will be 1.414 when $n = 5$. In the same way, we can get the maximum value of $h(n, \boldsymbol{P}_0)$ for different sample sizes. The result is similar to the simple straight line function except for the number of parameters. The appropriate sample sizes with 95% and 99% confidence for the special second surface function are 9 and 36.

### 3.5 Circular Function

The linearized deviation ⟨Figure 3⟩ is used [Shunmugam, 1986] to estimate the circle from $n$ observations which are represented by polar coordinates $(r_i, \theta_i)$. In ⟨Figure 3⟩, point 0 represents the $XY$ coordinate origin and also represents the origin of measured circle set by a CMM. Point $O_0$ represents the origin of estimated circle. Deviation, $e_i$, represents the distance from the measured point $P$ to the estimated circle in the direction to the origin set by a CMM.

$$e_i = r_i - (R_0 + x_0 \operatorname{Cos}\theta_i + y_0 \operatorname{Sin}\theta_i) \tag{5}$$

where $R_0$ = radius of the estimated circle
$x_0$, $y_0$ = coordinates of origin of the estimated circle.




〈 Figure 3 〉 Linearized Deviation from Circle

Then, the desired regression function can be written as follows:

$$r_i = R_0 + x_0 \mathrm{Cos}\,\theta_i + y_0 \mathrm{Sin}\,\theta_i \quad (6)$$

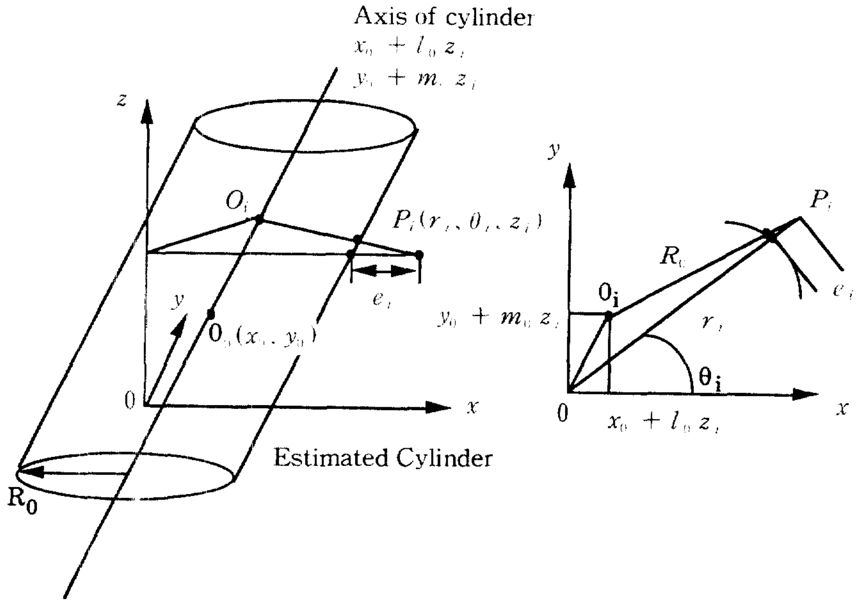
and if $\hat{Y} = r_i$, $b_0 = R_0$, $b_1 = x_0$, $b_2 = y_0$, $X_1 = \mathrm{Cos}\,\theta_i$ and $X_2 = \mathrm{Sin}\,\theta_i$, because circularity is the variation of the radius of the estimated circle, then

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 \quad (7)$$

Because $\mathrm{Cos}\,\theta$ cannot be represented by the linear combination of $\mathrm{Sin}\,\theta$ and there are three parameters to be estimated, Eq.(7) is exactly same as the simple plane function. However, sample points planning is different from the simple plane function, the appropriate sample sizes with 95% and 99% confidence for circular function are 7 and 22. Because we do not have to worry about balanced measurement in a circle, the number of measurements in a circular function is fewer than a simple plane function.

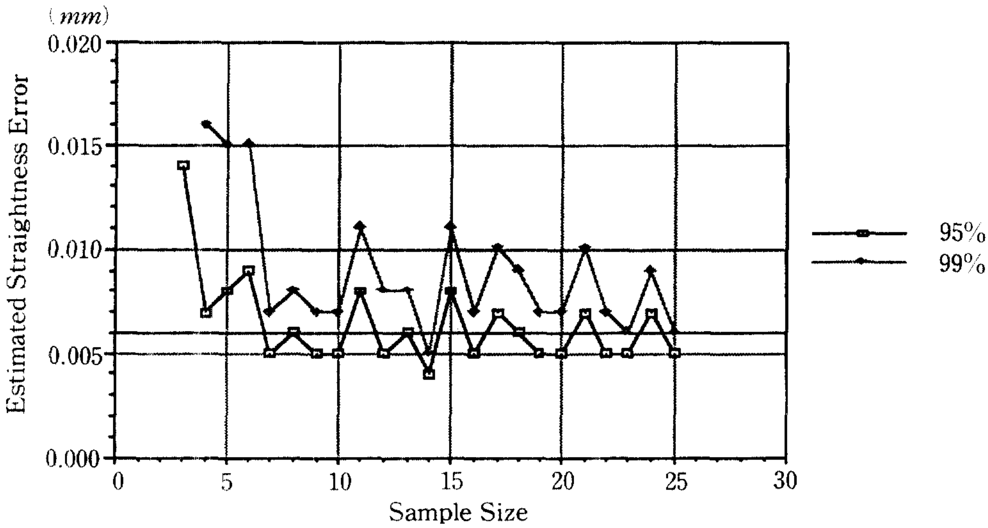
## 3.6 Cylindrical Function

The linearized deviation ⟨Figure 4⟩ is used [Shunmugam, 1986] to estimate the cylinder form $n$ observations which are represented by cylindrical coordinates ($r_i$, $\theta_i$, $z_i$). In ⟨Figure 4⟩, point 0 represents the origin of $XYZ$ coordinate and also represents the origin of the circular section of cylinder set by a CMM. Point $O$ represents the origin of the estimated cylinder. Point $O_i$ represents the origin of any circular section of the estimated cylinder. Deviation, $e_i$, represents the distance from the measured point $P_i$ to the estimated cylinder in the direction of the origin set by a CMM when the point $P_i$ is projected to $XY$ plane.

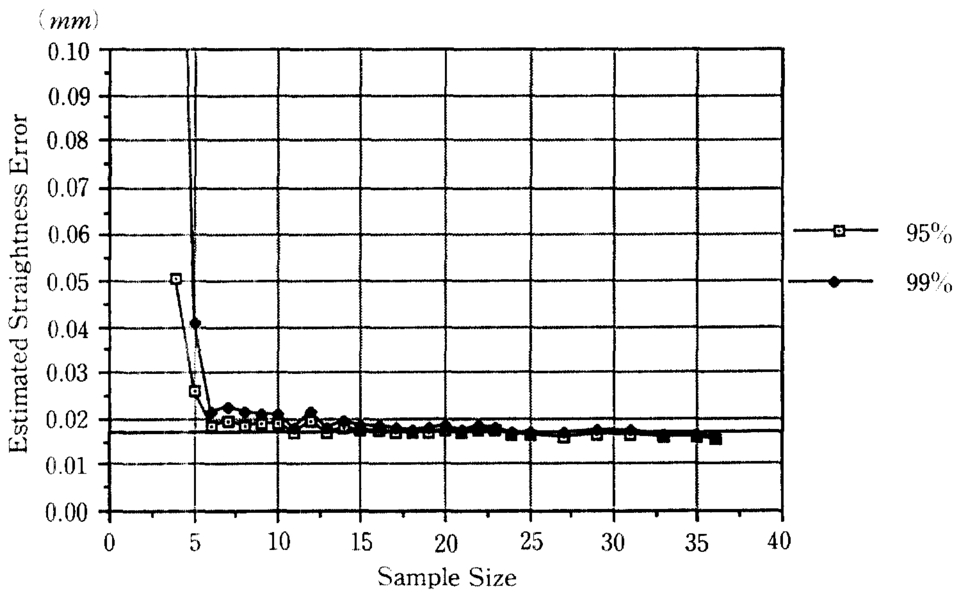



⟨ Figure 4 ⟩ Linearized Deviation from Cylinder

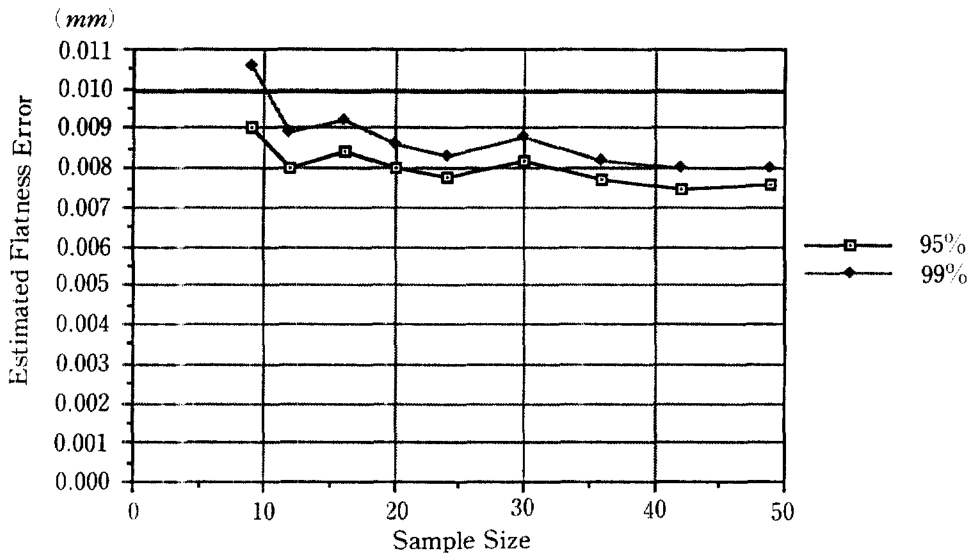
We conducted real part measurements using the CMM to assess the straightness and flatness errors. This experiment was performed using Sheffield Cordax RS-30 DCC CMM. A 165 $mm$ long rectangular bar was measured in increment of 1 $mm$. Based on these measurements we concluded that its straightness error is approximately 0.006 $mm$. We collected sample sizes 3 to 25 from these data with almost equi-interval. For each sample size, we estimate 95% and 99% confidence straightness errors ⟨Figure 5⟩. A 200 $mm$ long bar was measured in the same way and the estimated real straightness error was 0.017 $mm$. Sample sizes 4 to 36, with almost equi-interval, were collected. These estimated straightness errors are shown in ⟨Figure 6⟩. A 30 × 60 *mm* plate was measured in the same way and the estimated real flatness was 0.010 *mm*. Sample sizes of 9, 12, 16, 20, 24, 30, 36, 42, and 49 with almost equi-interval (grid type measurements), were collected. These estimated flatness errors are shown in ⟨Figure 7⟩.




⟨ Figure 5 ⟩ Variation of Estimated Straightness Error according to the sample size for 165 *mm* long bar




⟨ Figure 6 ⟩ Variation of Estimated Straightness Error according to the sample size for 200 *mm* long bar




〈 Figure 7 〉 Variation of Estimated Straightness Error according to the sample size for 60 × 30 *mm* plate

The estimated straightness and flatness errors have a tendency to decrease when the sample size is increased even though there are some fluctuations. The estimated flatness errors have a tendency to underestimate because of the large deviations in $x$ axis direction. We expected those fluctuations because our form error estimation approach uses $\sqrt{\text{MSE}}$. Even though we did not get the exact value of a straightness or flatness errors at the desired sample size, we got estimated values at the desired sample size especially with 99% confidence.

## 4. Conclusion

This paper has presented a procedure for determining the appropriate sample size and a formulation for evaluating form errors using a CMM. This new approach has the following characteristics.

1) It determines the sample size with a new criterion which is applied to the expectation of prediction interval with various confidence levels (95% and 99%).
2) It can be used to determine the confidence level when the sample size is given.

3) It uses the least squares criterion to estimate the desired feature in functional form.
4) It can be used to calculate Type I and II errors when the specification is given because it is statistically well defined.

The results of testing and verifying of this new approach are as follows.

1) It was carefully tested for determining the sample size for straightness, flatness, circularity and cylindericity.
2) The formulation was carefully tested for determining the straightness and flatness errors from simulated data and real measurement data.
3) Finally and most importantly, the results were tested and shown to be successful and satisfactory.

The approach proposed in this paper can provide a useful basis for further research for estimating form errors using the CMM. The formulations developed for straightness and flatness errors can be extended to a higher order of dimensional geometric tolerances. Consequently, the formulation can be established to estimate true geometric errors using the CMM.

## References

[1] American National Standard Institutes (ANSI Y14.5M-1982), *Dimensioning and Tolerancing*.

[2] Bourdet, P., Clement, A., and Weill, R. (1984), "Methodology and Comparative Study of Optimal Identification Processes for Geometrically Defined Surfaces," *Proceedings of the International Symposium on Metrology for Quality Control in Production*, Tokyo.

[3] Chang, S. H., Herrin, G., and Wu, S. M. (1990), "Sample Size Planning and Estimation of Form Errors using a Coordinate Measuring Machine," *Technical Report 90-35*, Department of Industrial and Operations Engineering, The University of Michigan.

[4] Draper, N. and Smith, H. (1981), *Applied Regression Analysis*, 2nd. Ed., John Wiley & Sons. Inc.

[5] ElMaraghy, W. H., Wu, Z., and ElMaraghy, H. A. (1989), "Evaluation of Actual Geometric Tolerances Using Coordinate Measuring Machine Data," *ASME Design Technical Conferences*, Vol. 19-1.

[6] Fukuda, M. and Shimokohbe, A. (1984), "Algorithms for form error evaluation-methods of the minimum zone and least squares," *Proc. Int. Sym. Metrology*

Quality Control Production, Tokyo, Japan, pp. 197−202.

[7] Gota, M. and Lizuka, K. (1977), "An analysis of the relationship between minimum zone deviation in circularity and cylindericity," *Proc. Int. Conf. Prod. Eng.*, New Delhi, India, pp. $x$61−$x$70.

[8] Greenwood, J. A. and Williamson, J. B. P. (1966), "Contact of Nominally Flat Surfaces," *Proc. Roy. Soc.*, London, A295, pp. 300−319.

[9] International Organization for Standardization (ISO 1101-1983(E)), *Technical drawings: geometrical tolerancing-tolerancing of form, orientation, location and runout-generalities, definitions, symbols, indications on drawing.*

[10] International Organization for Standardization (ISO 1101/II-1974(E)), *Technical Drawings-Tolerances of form and of position-Part II: Maximum material principle.*

[11] Kakino, Y. and Kitazawa, J. (1978), "In situ measurement of cylindericity," *Ann. C.I.R.P.*, Vol. 27, No. 1, pp. 371−375.

[12] Menq, C. H., Yau, H. T., Lai, G. Y., and Miller, R. A. (1990), "Statistical Evaluation of Form Tolerances using Discrete Measurement Data," *ASME Winter Annual Meeting 1990.*

[13] Murthy, T. S. R. (1982), "A comparison of different algorithms for cylindericity evaluation," *Int. J. Mach. Tool Des.*, Vol. 22, No. 4, pp. 283−292.

[14] Murthy, T. S. R. and Abdin, S. Z. (1980), "Minimum zone evaluation of surfaces," *Int. J. Mach. Tool Des. Res.*, Vol. 20, No. 2, pp. 123−136.

[15] Nater, J., Wasserman, W., and Kutner, M. H. (1985), *Applied Linear Statistical Models*, 2nd. Ed., Richard D. Irwin, Inc.

[16] Shunmugam, M. S. (1986), "On assessment of geometric errors," *Int. J. Prod. Res.*, Vol. 24, No. 2, pp. 413−425.

[17] Shunmugam, M. S. (1987), "Comparison between linear and normal deviations of forms of engineering surfaces," *Precision Engineering*, Vol. 9, No. 2, pp. 96−102.

[18] Shunmugam, M. S. (1987), "New approach for evaluating form errors of engineering surfaces," *Computer Aided Design*, Vol. 19, No. 7, pp. 368−374.

[19] Shunmugam, R. C. and Whitehouse, D. J. (1971), "A unified approach to surface metrology," *Proc. Inst. Mech. Eng.*, London, UK, Vol. 185, pp. 697−707.

[20] Weckenmann, A. and Heinrichowski, M. (1985), "Problems with software for running coordinate measuring machines," *Precision Engineering*, Vol. 7, No. 2, pp. 87−91.