Calculating Process Capability of Circular True Position Tolerances

Bert Uschold
Principal Design Engineer, Dexterity Engineering, LLC, Leominster, MA, USA

236 Pierce St.
Leominster, MA 01453
https://www.linkedin.com/in/bertuschold/
Bert.uschold@outlook.com 978-660-6791

Calculating Process Capability of Circular True Position Tolerances

ABSTRACT: There are many occasions where using a two dimensional, true position tolerance for a specifying the tolerance of a dimension is preferred to using a standard one-dimensional plus/minus tolerance. Unfortunately, it is difficult to demonstrate process capability for such dimensions. The key challenge is determining the standard deviation of the data. For over 30 years, many efforts have been made to determine process capability for positional tolerances. In some cases, the authors do not address calculating the standard deviation and in others they make errors in the process. This paper calculates a one-dimensional standard deviation in a non-typical but logical, methodical way, revealing a strategy to be easily determine the two-dimensional standard deviation. First find the center of the data and then the distance from each point to the center of the data. That distance can be used to calculate the standard deviation, but with a slight modification to the customary method. The end result is that the standard deviation for true position radius is the RSS of the standard deviation of the x and y components. That this technique is used in the field of geometric information systems supports its use for quality engineering.

Keywords: process capability; true position: circular true position

1. Main Text Introduction

In the medical device industry and others, it is frequently a requirement to establish process capability for critical dimensions. Proving a process can operate with a process capability index of 1.33 is a common, although not universal, requirement and the method to do so for standard plus/minus dimensions is well established. This is not the case for circular true position tolerances and numerous attempts have been made. When properly reported in a metrology report, the tolerance is given as a diametral value, although often the measured basic component dimensions locating the feature are also given. The primary challenge is finding the standard deviation. For various reasons which will not be discussed, neither the diametral value nor the individual component dimensions can be used. This paper will focus on properly calculating the standard deviation for a 2D dataset and then using that to calculate process capability.

a. Previous Attempts

Many authors have described methods of calculating process capability for true position tolerances. Krishnamoorthi (1990) describes a method where he uses the larger standard deviation of the two components. He uses that to calculate the area of a circular region of natural variability which is a circle whose radius is 3σ. He defines the Cp as a ratio of the area of the tolerance region to the area of natural variability. When the area of those two circles are equal, or when the diameter of the tolerance region equals 6σ, Cp is 1, as expected. However, if σ is halved, we would expect the Cp to be 2 but his formula calculates it as 4. Additionally, the method this paper will describe for finding standard deviation yields a value that is larger than the standard deviation of either of the x or y components. Therefore, his standard deviation is low, overstating process capability.

Davis, Kaminsky, and Saboo (1992) establish of method of determining non-conforming parts per billion (NCPPB). In the process, they create what they describe as a new process capability index which is used to calculate a failure percentage. If desired, a Cpk could be extrapolated from the NCPPB but they do not explicitly indicate how to calculate standard deviation and it appears they use the value for a single component. Several others, Rodriguez-Picon, et al (2019), consider the case where the data is not normally distributed and/or has a different variation for each axis. In each case, the method of calculating the standard deviation is not specified. Tahan and Levesque (2009) use Bothe's (2006) method of determining the standard deviation, which will be shown to be flawed. Phillips and Cho (1998) use Davis, Kaminsky, and Saboo (1992) et al's assumptions about σ. Grau (2009) addresses one sided tolerances. Although as reported, true position tolerances can be considered one-sided, when considering what they represent, they are not. Knowles, March, and Anthony (2002) and Diplaris and Sfantsikopoulis (2004) look at the more complicated problem of true position tolerances that include a maximum material condition modifier. This proposed method does not address that case.

Bothe (2006) proposes formulas for both Cp/Pp and Cpk/Ppk. To calculate short term performance capability, Bothe estimates standard deviation by first measuring the change in size of consecutively made holes. He then takes the average of those steps and divides by 1.128 to determine short term standard deviation. There are two reasons why this is not a good approximation. The first is that it depends on the order of the data. Because it measures the difference from point to point, if the data were analyzed in a different order, there would be a different result. However, since the data is random, this should not make a great difference. The more significant problem is that it weights a small movement the same as a large movement.

When calculating standard deviation, the distance of each point to the mean is squared so that points farther from the mean have a greater impact to the standard deviation. This makes the calculated standard deviation to be low.

To determine long term performance capability, Bothe first calculates the distance from each point to the cluster's (data's) center, which this method agrees with. He then calculates the sample standard deviation and divides by the appropriate c_4 factor. It is in his method of calculating the sample standard deviation that there is an error resulting in a low value. Two key formulas are:

$$\sigma_{LT} = \frac{1}{c_4}\sqrt{\frac{\sum_{i=1}^{k}(r_i - \bar{r})^2}{k-1}} \quad (1)$$

and

$$\hat{\sigma}_{LT,C} = \sqrt{\frac{\sum_{i=1}^{k}(r_{C,i} - \overline{r_C})^2}{k-1}} \quad (2)$$

Formula (1) is for measuring performance capability (Bothe, 2006, 327) and formula (2) is for measuring potential capability (Bothe, 2006, 329). Above, r is the distance from a data point to nominal center and r_C is the distance from a data point to the center of the data. These are the standard formulas for calculating standard deviation but should not be used for true position tolerances. To find the standard deviation for a set of data, you need to find the center of the data and the distance of each point to that center of the data. For a 1D dataset, the mean does this. For a 2D dataset, which a true position tolerance is, you need to find the center of the 2D dataset. In the first formula, Bothe calculates the standard deviation of the radial values by using the mean of the values instead of the center of the data. An easy way to understand the difference is to imagine four points whose values are (1,0), (0,1), (-1,0), and (0,-1). See figure 1. For each point, the radius is 1. Using Bothe's method, one would report the mean as 1. Looking at the graph, we can see the center of that dataset is at (0,0).

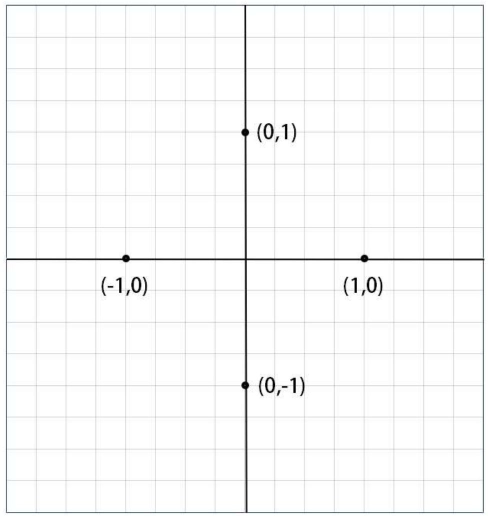

Figure 1.

In the second formula, he is closer because $r_{C,i}$ is the distance from an individual data point to the data center. Let's call it offset. Part of finding the standard deviation is to find that offset, which is what r_C is, and then square it. Bothe starts in that direction but before squaring the offset, he subtracts from it the mean of the offsets, necessarily reducing the calculated value of standard deviation to something below the actual standard deviation.

Finally, it must be noted that Krystek (2010) arrives at functionally the same formula for calculating C_{pk} as the following approach describes. Unfortunately, he assumes a standard deviation in his paper but does not describe how to calculate the standard deviation for a particular set of true position data. This paper will provide an appropriate method to calculate standard deviation. It also describes an alternative method to derive the C_p and C_{pk} formulas showing their similarity to the 1D case.

2. **Method**

Given a standard deviation and tolerance, it is well known how to calculate the process capability indices. The formulas are:

$$C_p = \frac{USL - LSL}{6\sigma}, \; C_{pk} = \frac{\min\left(USL-\mu\right),\left(\mu-LSL\right)}{3\sigma} \tag{3,4}$$

For convenience, this paper will use C_p and C_{pk} when referring to process capability in general, although the same method can be properly applied to P_p and P_{pk}. However, when seeking the standard deviation for a true position tolerance, because it is 2D data, the standard formula for calculating standard deviation cannot be used. To determine how to calculate the standard deviation for true position tolerances, this paper will describe the method of calculating process capability for standard dimensions but in an unorthodox and very methodical manner. In parallel, the method for calculating process capability of true position tolerances will be given to show that the two methods are in essence the same. In the process the method to calculate standard deviation is shown.

A. Find Cp and Cpk.

Step 1 - Define the Range of Acceptable Values, R

This is part of the process is extracting necessary data from the drawing specifications. For the standard method, the upper specification limit, USL, is the nominal value + tolerance. The lower specification limit, LSL is the nominal value – tolerance. Therefore, the range is:

$$R = USL - LSL \tag{5}$$

For the true position method, the limit is a circle centered at nominal whose diameter is the true position tolerance, TP.

R = TP

(6)

Step 2 - Find the mean/center of the data.

Finding the mean of the data set is the first step in calculating standard deviation. Thus, for the standard method, the mean is:

$$\mu = \frac{\sum x_i}{n} \tag{7}$$

where x_i is each measured value and n is the number of measurements. In the context of circular true position, we will call it the center instead of the mean. The center of the data is = ($\bar{x}$, $\bar{y}$). The values for $\bar{x}$ and $\bar{y}$ are:

$$\bar{x} = \frac{\sum x_i}{n}, \quad \bar{y} = \frac{\sum y_i}{n} \tag{8,9}$$

where x_i and y_i are the coordinates of each respective point.

Step 3 - Find the distance from each data point to the mean or center of the data.

With the mean or center defined, the next step in finding the standard deviation is to find the distance, $\underline{r_i}$, from each data point to the mean/center. For the standard method,

$$r_i = \text{ABS}|x_i - \mu| \tag{10}$$

and for the true position method,

$$r_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{11}$$

Step 4 - Find the standard deviation, σ.

This is the same for both cases. While is looks different from the typical standard deviation formula, given the definition of r_i, it is the same.

$$\sigma = \sqrt{\frac{\sum r_i^2}{n-1}} \tag{12}$$

Step 5 - Find the shortest distance from the mean/center to the defined limit, dcrit.

This is required for finding C_{pk} but is not required for C_p. For the standard method,

$$d_{crit} = \min[USL - \overline{x}, \overline{x} - LSL] \tag{13}$$

and for the true position method,

$$d_{crit} = \frac{TP}{2} - \sqrt{(x_{nom} - \overline{x})^2 + (y_{nom} - \overline{y})^2} \tag{14}$$

Step 6 - Calculate Pp and Ppk

We now have all of the necessary information to calculate Pp and Ppk for both methods, which after using slightly different terms in the first 5 steps, yields the same formulas.

$$C_p = \frac{R}{6\sigma}, \; C_{pk} = \frac{d_{crit}}{3\sigma} \tag{15}$$

When replacing R and d_{crit} with their defined values, more familiar and useful formulas are generated.

Standard method:

$$C_p = \frac{USL - LSL}{6\sigma}, \quad C_{pk} = \min\left[\frac{USL - \overline{x},}{3\sigma}, \frac{\overline{x} - LSL}{3\sigma}\right] \tag{16, 17}$$

True position method:

$$C_p = \frac{TP}{6\sigma}, \quad C_{pk} = \frac{\frac{TP}{2} - \sqrt{(x_{nom} - \overline{x})^2 + (y_{nom} - \overline{y})^2}}{3\sigma} \tag{18, 19}$$

3. Discussion

A. Explanation

The first calculation to find any process capability value is to find the mean of the data set. One might be tempted to say that is not necessary because potential capability is independent of the nominal value and thus the mean's relation to the nominal. However, the mean is required to calculate standard deviation.

The derivation of this approach was done using an electronic spreadsheet and as is often the case, the end result looks much simpler than all of the effort required to get there. What was also discovered and is not surprising is that the standard deviation data as calculated above can be found in another way. If σ_x and σ_y are treated as sides of a right triangle, then the hypotenuse yields the same result as (12) and can be re-written as:

$$\sigma = \sqrt{{\sigma_x}^2 + {\sigma_y}^2} \tag{20}$$

Equations (18) and (19) are unchanged. This observation makes programming a spreadsheet to calculate the true position simpler because, as mentioned, the x and y coordinates are also typically reported along with the true position value. See the application section for details.

It must be noted that the method this paper uses to calculate the standard deviation of true position tolerances is also used to calculate geographic distributions. The company Esri uses it in their ArcGIS Pro software, where GIS stands for geographic information system. Instead of standard deviation, they call it standard distance and use it for 2D and 3D data fields. The formula they give on their website is:

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X}\,)^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \overline{Y}\,)^2}{n} + \frac{\sum_{i=1}^{n}(z - \overline{Z}\,)^2}{n}} \qquad (21)$$

where n is the total number of features. They do not use n-1 because they are looking at all of the features instead of a sample of them. It is functionally the same as (12) but accounts for a third axis and is stated in a different way.

B. Application

It is expected that when the method discussed here is used, often it will be calculated using an electronic spreadsheet. It is critical to recognize that in finding the standard deviation of the true position tolerance not to use the built-in standard deviation function. Here is one way to use a spreadsheet and the method described above to find standard deviation:

(1) Enter the x and y values for each data point into two columns.

(2) Find the mean of both the x values and the y values. This also defines the center of the data, $(\overline{x}, \overline{y})$.

(3) Find the standard deviation of the x values and y values, σ_x. and σ_y.

(4) The standard deviation, σ, is the square root of the sum of the squares of σ_x. and σ_y.

It is suggested to also calculate distance to the center point for each point to help identify which parts contribute the most to the standard deviation. The point that is farthest from the center might not be the same point that has the highest true position value.

C. Conclusion

This paper has shown that a methodical approach to calculating process capability and standard deviation for standard dimensions reveals, with slight modifications, a method to calculate process capability for true position tolerances. This method can easily be implemented on a spreadsheet using standard spreadsheet functions. It must be noted that this method has not been shown to apply when the tolerance includes a maximum material condition modifier, but it is expected that a modification can be found to incorporate that condition. As shown by usage in another industry, it also applies to three-dimensional data. Finally, it is interesting to observe that when determining process capability of the true position tolerance, the actual true position tolerance values are not used.

4. References

Bothe, D.R. 2006. Assessing capability for hole locations. *Quality Engineering* 18:3, 325-331.

Davis, R.D., F.C. Kaminsky, and S. Saboo. 1992. Process capability analysis for processes with either a circular or spherical specification region. *Quality Engineering* 5:1, 41-54.

Diplaris, S.C. and M.M. Sfantsikopoulis. 2006. Process capability requirement under maximum material condition. *The Journal of Engineering Manufacture* 220:10, 1629-1634.

Grau, Daniel. 2016. New Process Capability Indices for One-Sided Tolerances. *Quality Technology & Quantitative Management.* 6:2, 107-124.

Knowles, G., G. March, and J. Anthony. 2002. Evaluating Process Capability for Geometrically Toleranced Parts: A Practical Approach. *Quality Engineering* 14:3, 365-374.

Krishnamoorthi, K.S. 1990. Quality Engineering. *Capability Indices for Processes Subject to Unilateral and Positional Tolerances*. 2:4, 461-471.

Krystek, Michael P. 2010. Process capability indices and risk analysis for circular position tolerance zones. *Proceedings Volume 7544, International Symposium on Precision Engineering Measurement and Instrumentation.* 75446D.

Phillips, M.D., and B-R. Cho. 1998. Quality Improvement for Processes with Circular and Spherical Specification Regions. *Quality Engineering.* 11:2, 235-243.

Rodriguez-Picon, L.A., L.C Mendendez-Gonzalez, V.H. Flores-Ochoa, and M.A. Rodriguez-Medina. 2019. Capability indices for circular tolerance regions based on a Gaussian copula. *The International Journal of Advanced Manufacturing Technology*. 104:4143-4153.

Tahan, S.A. and S. Lévesque. 2009. Exploiting the Process Capability of Position Geometric Tolerance According to GD&T ASME Y14.5M. *2009 International Conference on Computers & Industrial Engineering.* 1267-1272.

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-standard-distance-spatial-statistic-works.htm