Process capability indices when two sources of variability present, a tolerance interval approach

Éva Pusztai | Sándor Kemény

Department of Environmental and Process Engineering, Budapest University of Technology and Economics, Budapest, Hungary

Correspondence
Éva Pusztai, Department of Environmental and Process Engineering, Budapest University of Technology and Economics, Műegyetem rkp 3, Budapest, Hungary.
Email: pusztai.e@mail.bme.hu

Abstract
The sound tolerance interval–based method and two $P_P$–based approximations are compared on the proportion of nonconforming parts. As output distribution of the process, one possible model is examined here: two sources of variations are in a one-way structure. It was found that the uncertainty of variance components estimates plays the major role in the goodness of the three calculation methods.

Summary: Statistical indices – like process capability ($C_P$) or process performance ($P_P$) index – make the relationship between the width of the specification interval and the extent of the process variability illustrative. The latter is characterized by the tolerance interval, which contains the major part of the population with high confidence. In the original concept this tolerance interval is calculated using oversimplified models. In practice this model is not conform with reality. As improvement two sources of variation is assumed in a one-way structure.

The proportion of non-conforming parts of the population is the quantity of interest. Non-conforming means that the characteristic is beyond the specification limits. According to this, the quantile of the distribution shall be determined that is equal to the specification limits. Thus, the task is to calculate the tolerance interval for the $N(\mu, \sigma_A^2 + \sigma_e^2)$ distribution.

In practical cases the variance components are unknown and are to be estimated. To estimate the ratio of non-conforming parts, two approximate calculation methods which are coherent with the definition of $P_P$ are investigated, as well.

The aim of this work is to compare the results of the two $P_P$ based approximate methods with the tolerance interval based (theoretically sound) calculation method. Two situations from the practice are considered – datasets from process validation and process monitoring environment are evaluated during which the effect of the number of experiments is given. Our study indicated that the main difference of the goodness of the approximations is from the uncertainty of parameter estimates.
1 | INTRODUCTION

Statistical methods are widespread in the field of quality management. This spread was effectively served by statistical indices like $C_p$, $C_{pk}$, $P_p$, and $P_{pk}$, which are included in the statistical software (Minitab, TIBCO Statistica) as well as the standards, guidelines, etc. They are also related to the Six Sigma approach.

The simplest (and still totally correct) concept is the potential capability. In this method, the 0.9973 probability range ($6\sigma$) is compared with the specification range ($USL-LSL$). Thus, if $C_p = 1$, the rate of nonconforming is 0.0027. The calculation requires knowledge of the $\sigma^2$ variance:

$$C_p = \frac{USL - LSL}{6\sigma}.$$ (1)

The variance is not known but is estimated in most cases. If the sample size is large enough, this estimate is fairly good, however. One may consider the uncertainty of $C_p$ (in the form of the confidence range) because of the imperfect knowledge of $\sigma^2$. Calculating a confidence interval for the $C_p$ is simple (Equation (2)):

$$P\left( \hat{C}_p \sqrt{\frac{\chi^2_{1-\frac{\alpha}{2},n-1}}{n-1}} \leq C_p \leq \hat{C}_p \sqrt{\frac{\chi^2_{1-\frac{\alpha}{2},n-1}}{n-1}} \right) = 1 - \alpha,$$ (2)

where $\hat{C}_p$ is the estimated value of the parameter, $\chi^2_{1-\frac{\alpha}{2},n-1}$ is the $1 - \frac{\alpha}{2}$ quantile of the chi-square distribution which has $n-1$ degrees of freedom, $n$ is the number of data points, and $\alpha$ is the significance level.

$C_p$ as the potential capability proved to be the most useful expression. It is easy to use (communicate) among those who are familiar with this parlour. Because of the simple usage, there is no way to consider the specific assumptions/conditions of the capability indices. The advantage properties are valid only if the model is sound. Namely, the following assumptions should be fulfilled:

- the process is in statistical control.
- The quality characteristic is a normally distributed random variable.
- There is only one source of variability.
- The expected value of the quality characteristic is equal to the target value (the process is centered).

Otherwise, having a $C_p$ value, one may only approximately know the rate of nonconforming. The potential capability is a limiting value. It is achieved if there is no shift in the process, thus $\mu = T$, where $T$ is the target value of the process. In real manufacturing processes, shifts exist ($\mu \neq T$), however. Thus, the chance for a part manufactured to occur beyond the $USL$-$LSL$ range is higher than 0.0027, even with $C_p = 1$. Assuming a certain $\frac{\mu - T}{\sigma}$ shift, one may calculate the rate of nonconforming. For example, $\frac{\mu - T}{\sigma} = 0.5$ is assumed in $6\sigma$ methodology. This has a good explanation if the idea of $C_p$ is used together with control charts.

The corrected process capability $C_{pk} = \max\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\}$ is still useful in giving a “gut feeling” of quality level. It is not that closely connected to the nonconforming ratio, however. The concept of $C_{pk}$ is useless unless certain (or maximum) extent of the $\frac{\mu - T}{\sigma}$ shift is assumed. This is still not far from reality in quality engineering.

The simple concept of $C_p$ (and partly that of $C_{pk}$) is based on a single source of variation, which may either be the process variation (part to part) or the measurement uncertainty.

In the manufacturing reality, it is typical that there are several sources of variation, eg, both part to part and because of measurement uncertainty. The frequently used approach (AIAG of automotive industry) neglects the measurement variation, when a gage R&R study is employed to ascertain that the measurement variation is small as compared with the part to part variation. This may work well, but a formal study is required whether the conditions are fulfilled.
The concept of process performance index is a further step away from the mathematical truth.\textsuperscript{11} Here, the so-called long-term variation is substituted for $\sigma^2$. Long-term variation\textsuperscript{16} is added (estimated) from the part to part component and the difference observed at different time points.

\[ P_P = \frac{USL - LSL}{6\sigma_{LT}} \]  

Measurement uncertainty is not mentioned, assuming it being negligible. This index differs from $CP$: 

\[ P_P \leq CP = \frac{USL - LSL}{6\sigma_{ST}} \]  

since $\sigma_{ST} \leq \sigma_{LT}$. Having the value of $P_P$, one may have no idea in the strict sense on the rate of non-conformity in spite of the original intention.

The use of $P_P$ is based on the following assumptions:

- the output distribution of the process is treated as if it would be a single normal distribution with a certain expected value and long-term variance.
- Virtually, there is a single normally distributed random variable.
- Multiple sources of variation are pretended as a single source of variation.
- The expected value of the quality characteristic is equal to the target value (the process is centered). If not, it can be handled with the $P_{PK}$ index.

It is usual in the automotive industry that suppliers are required to keep certain minimum values of $P_P$. Montgomery states,\textsuperscript{11} referring to the work of Kotz and Lovelace,\textsuperscript{17} that the usage of $P_P$ and $P_{PK}$ is practically statistical terrorism, which facilitates avoiding the understanding of the process.

Bothe states that performance parameters (AIAG) give very little (if any) information on a process, if it is unstable,\textsuperscript{18} as there is no defined output distribution.\textsuperscript{8} The authors of the present paper fully agree with this statement. The application of the process performance idea is illustrated in AIAG\textsuperscript{5} by processes where there is an additional source of variation (additional to the part to part variation).

$P_P$ index originally is intended to represent the goodness of the process that is the probability of occurring nonconformity. Thus, the practitioner may want to know the goodness of this approximation. Unfortunately, this answer is not simple. To judge the goodness of the approximations, one may need a theoretically sound model and algorithm of calculation. From among the several possible models, one special case will be examined here: when several sources of variation are in a one-way structure. If this model is applicable to an industrial manufacturing problem, it offers the possibility to use the in-control concept with a feasible alternative to the process capability calculation, utilizing the fact that the output distribution is modelled. This means that the results obtained here may not be valid in cases of other error structures, such as with a nested design or cross-design of multiple factors.

In this paper, our aim is to clarify how good are the models in the context of process capability, process performance, and ratio of nonconforming parts. Specific examples of failure are shown in Section 5. There are two somewhat different situations considered:

- Process validation (PPAP) with moderate size of well-defined and carefully executed experiments\textsuperscript{5};
- Process monitoring based on many but less carefully collected and less reliable data, from an operating process.

The “how good the models are” question is not simple by itself: on one side, the models come with approximations (calculation methods), and on the other side, the properties of parameter estimates have to be taken into account. The goodness depends on the size and structure of the experiment, and on the size of variance components.

2 | CALCULATING PROCESS CAPABILITY IN THE CASE OF A ONE-WAY BALANCED DESIGN, APPROXIMATIONS

Consider an experiment that includes a random factor $A$ with $a$ levels and each level has $n$ repeated measurements. The model of the observations from this experiment can be written as
\[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad (4) \]

where \( \mu \) is a reference value, \( Y_{ij} \) is the \( j \)th observation with the \( i \)th level of factor \( A \) (\( j = 1, \ldots, p \)), \( \alpha_i \) is the effect of the \( i \)th level of factor \( A \) (\( i = 1, \ldots, r \)), \( \epsilon_{ij} \) is the error term (including sampling error), which is an independently and identically distributed random variable \( N(0, \sigma_e^2) \). Whereas \( A \) is a random factor and \( \alpha_i \) is a normally distributed random variable with an expected value of \( 0 \) and \( \sigma_A^2 \) variance.\(^\text{19}\) Furthermore, the usual assumptions should be fulfilled\(^\text{20}\):

- the variance of the error term and that of \( \alpha_i \) are constant, ie, equal in each group (homoscedasticity);
- the \( \epsilon_{ij} \) errors are independent between groups and within groups;
- the \( \alpha_i \) effects are independent from each other and from the \( \epsilon_{ij} \) random error.

According to this model, the quality characteristic of interest is a normally distributed random variable with expected value \( \mu \) and \( \sigma_A^2 + \sigma_e^2 \) variance \( (\mu, \sigma_A^2 + \sigma_e^2) \).

For the purpose of evaluating the process capability, the output distribution is known if the structure and sources of variability are well-defined \( (\sigma_A^2, \sigma_e^2) \), detected, and the statistical assumptions are fulfilled. Only in this special case does it make sense to speak about the characteristics of the output distribution, namely, about the proportion of non-conforming parts.

There are two obvious choices to approximate the proportion of nonconforming parts, which is coherent with the definition of the \( P_P \).

The most convenient way would be to use the overall variance \( (s_{\text{overall}}^2) \) of the data to calculate the \( P_P \). In this course of calculation, the standard deviation is computed as if the data were obtained from a single population. This procedure is incorrect, because it does not consider the actual error structure of the dataset

\[
s_{\text{overall}}^2 = \frac{\sum \sum (y_{ij} - \bar{y})^2}{rp - 1}. \quad (5)\]

The expected value of the overall variance is

\[
E(s_{\text{overall}}^2) = \sigma_e^2 + \frac{rp - p}{rp - 1} \sigma_A^2 \quad (6)
\]

(with \( r \) levels of factor \( A \) and \( p \) repeated measurements).

Another approximation may be simply to add the variances, and this total variance is substituted in the formula of the \( P_P \). This method is called between/within variability in Minitab\(^\text{1}\)

\[
\sigma_{b/w}^2 = \sigma_{\text{total}}^2 = \sigma_A^2 + \sigma_e^2. \quad (7)
\]

Since the real variance components are unknown, they need to be estimated from the one-way ANOVA model.

\[
\hat{\sigma}_{\text{total}} = \sqrt{\hat{\sigma}_A^2 + \hat{\sigma}_e^2}. \quad (8)
\]

The variance components from the one-way ANOVA are estimated as

\[
\sigma_A^2 = \frac{s_A^2 - s_R^2}{n} \quad (9)
\]

\[
\sigma_e^2 = s_R^2. \quad (10)
\]

where \( \sigma_A^2, \sigma_e^2 \) are the estimated variance components, \( s_A^2 \) is the mean square which belongs to the effect of the factor \( A \), \( s_R^2 \) is the mean square of residuals.\(^\text{19}\)
It is clear that the deviation between the two approximations depends at least partly on the model, in this case, on the proportion of \( r \) and \( p \).

### 3 | TOLERANCE INTERVAL: THE SOUND CALCULATION METHOD

#### 3.1 | Tolerance interval

Although the tolerance interval is a rarely used statistical interval, it has several applications in the field of quality control and pharmaceutical industry.\(^{21-23}\) Such applications often require the calculation of tolerance intervals for different ANOVA models such as random/mixed effect models or nested designs.

The basic idea of the tolerance interval is to calculate an interval that contains at least a specified proportion of the population with a given confidence.\(^{21,24}\) Suppose that there is a normally distributed random variable \( X \sim N(\mu, \sigma^2) \). In this case, the \( \beta \) content one-sided tolerance interval can be formulated as:

\[
P[P(X \leq U) \geq \beta] = 1 - \alpha.
\]

(11)

The opposite side:

\[
P[P(X \geq L) \geq \beta'] = 1 - \alpha,
\]

(12)

where \( U \) and \( L \) define the upper and lower limits, respectively, and \( 1 - \alpha \) means the degree of confidence. Note that \( \beta \) and \( \beta' \) are used for the two intervals so that the tail areas of the population of interest are not necessarily equal.

The limits of the intervals are based on the random sample from the population and can be estimated with \( \bar{x} + ks \) and \( \bar{x} - k's \) where \( \bar{x} \) is the mean, \( s \) is the standard deviation of the sample, \( k \) is the tolerance factor that belongs to the one-sided tolerance interval for the \( \beta \) proportion of the population, \( k' \) is the tolerance factor in the case of the one-sided tolerance interval for the \( \beta' \) proportion of the population.\(^{21}\) The value of the tolerance factor \( k \) (\( k' \)) is determined knowing the \( \beta \) (\( \beta' \)) proportion of the population, the level of confidence, and the degrees of freedom of the standard deviation. This calculation method assumes single source of variability in the process.

It is highly important that the choice of confidence level should be based on risk analysis, ie, it is necessary to weigh the risk of the quantile not being included in the interval (lower confidence level) against having a longer interval (higher confidence level). As Meeker et al mentioned, the analyst must determine the confidence level based upon what seems to be an acceptable degree of assurance for each application.\(^{21}\)

#### 3.2 | Calculating the proportion of nonconforming parts with tolerance interval, one-way balanced model

The real aim of the calculation of process capability metrics is to estimate the proportion of the population that is outside of the specification limits. In the case mentioned before, this is equivalent to the estimation of the values of the distribution function for the \( N(\mu, \sigma^2_A + \sigma^2_e) \) distribution at the specification limits. Since the parameters of the distribution are unknown (their estimate is based on the sample), the quantiles are uncertain. The confidence interval of the proportion of the nonconforming parts is equivalent with the confidence interval (eg, 95%) of a quantile of the distribution.\(^{21}\)

Thus, the task is to calculate the limit of the one-sided tolerance interval for the given proportion (\( \beta \)) of the population. The ratio of nonconforming parts is the same as the ratio where the limits of the interval coincide with the specification limits.

If the abovementioned (Section 2) assumptions of the random factor ANOVA are fulfilled, it is justified to calculate the tolerance interval and estimate the ratio of nonconforming parts. In the case of one-way ANOVA model, ie, \( N(\mu, \sigma^2_A + \sigma^2_e) \), the calculation of the tolerance interval is not simple. One complication is caused by that the estimate of the \( \sigma^2_A \) (ie, \( \hat{\sigma}^2_A \)) is calculated from the difference of two mean squares, which is not chi-squared distributed. An option to solve this problem is to use the Satterthwaite-approximation.\(^{24}\) Mee and Owen published a calculation method\(^{25}\) that is based on this approximation but, according to the simulation studies, it is proved to be highly conservative. This means...
that, the real confidence level is higher than the nominal level, thus the calculated interval is wider than the exact one would be.

Better performing (less conservative) methods were proposed by Vangel\textsuperscript{26} and later by Krishnamoorty and Matthew.\textsuperscript{27} The latter method is based on the concept of generalized confidence limits and applies Monte Carlo simulation to give an approximate one-sided tolerance limit. In addition, Krishnamoorthy and Peng\textsuperscript{28} have developed a calculation method that has a closed-form and therefore simulation is not required in the calculation. This is the so-called MOVER method (method of variance estimate recovery). According to their coverage studies, the MOVER tolerance limits are less conservative (more liberal) than the generalized tolerance limits. A conservative tolerance interval means that the real coverage probability of the interval is higher than the nominal one, ie, the interval is wider than that belonging to the declared probability. When the tolerance limits are used, the proportion of nonconforming parts is estimated by one minus the coverage probability. On the basis of the above, the use of a conservative calculation method leads to the underestimation of the proportion of the nonconforming parts, thus, a liberal interval is the better choice in this situation. On the basis of these favourable properties, the MOVER method will be applied in this paper.

4 | THE DATASET OF EXAMPLE 1

The performance of the following three calculation methods will be evaluated using a data set that is based on a real industrial problem:

- $P_p$ with the overall standard deviation,
- $P_p$ with the total standard deviation,
- tolerance interval.

However, because of confidentiality reasons, we are unable to completely report it as an example. The most important partial results will be given, thus the dataset can be reproduced using a simulation.

In this example, one wants to validate a production process, so the ratio of the population that is out of the specification limits (rate of nonconforming parts) has to be calculated. The model is a balanced one-way ANOVA. Factor $A$ has 10 levels and the overall mean is 8.6731. The most important results like the sums of squares (SS), the degrees of freedom (df), the mean squares, and the estimated variance components are summarized in Table 1.

The overall mean square is the estimated variance of data without considering the between-within structure of data, as if there was a single sample given. It is clearly different from the sum of the two mean squares (would be 0.0031).

5 | RESULTS AND DISCUSSION

Three calculation methods (detailed in Sections 2 and 3) will be compared:

- the first method is based on the overall standard deviation (without considering any error structure in data) this approach conforms to $P_p$,
- the second method is summing the variance components estimated by ANOVA,
- the third method is using the tolerance interval for the calculations.

According to the requirements, the product is considered conforming if the quality characteristic of interest falls between 8.6 ($LSL$) and 8.9 ($USL$).

| Effect       | Sum of Squares | Degrees of Freedom | Mean Squares | Estimated Variance Components |
|--------------|----------------|--------------------|--------------|-------------------------------|
| A (between)  | 0.0195         | 9                  | 0.0021       | 0.0001                        |
| Error (within)| 0.0903         | 90                 | 0.0010       | 0.0010                        |
| Overall      | 0.1098         | 99                 | 0.0011       |                                |
5.1 Estimation of the proportion of nonconforming parts using the overall standard deviation

The long-term variability ($LT$) is estimated with the overall standard deviation (Equation (4)), which is $0.0333$. In addition, according to the position of the overall mean, the process is shifted, thus $P_{PK}$ should be given:

$$P_{PU} = \frac{USL - \mu}{3\sigma_{LT}} = \frac{8.9000 - 8.6731}{3 \cdot 0.0333} = 2.2585$$ (13)

$$P_{PL} = \frac{\mu - LSL}{3\sigma_{LT}} = \frac{8.6731 - 8.6000}{3 \cdot 0.0333} = 0.7311$$ (14)

$$P_{PK} = \min\{P_{PL}; P_{PU}\} = 0.7311.$$ (15)

The proportion of nonconforming parts in the long term ($p_{TOTAL}^{LT}$) can be calculated using the values of the distribution function of the standard normal distribution. Obviously, as the true parameters are not known, their estimated values (and thus estimate of $Z$) are substituted. Generally, the manufacturers and their customers want to be sure about the minimum of $P_{P}$, and this is in the actual example of the lower limit.

$$p_{TOTAL}^{LT} = p_{LSL}^{LT} = \Phi(\hat{Z}_{LSL})$$ (16)

where $p_{LSL}^{LT}$ denotes the percentage of the parts below 8.6 ($LSL$), $\Phi(\hat{Z}_{LSL})$ is the probability distribution function at $\hat{Z}_{LSL}$. $\hat{Z}_{LSL}$ is the standard normal value calculated as:

$$\hat{Z}_{LSL} = \frac{\hat{\mu} - LSL}{\hat{\sigma}_{LT}}$$ (17)

In our example $\hat{Z}_{LSL}$ is 2.193, obtaining the $\Phi(2.193)$ value from the $Z$ table, it is 0.01414. Thus, the estimated proportion of the nonconforming parts is 0.01414.

5.2 Estimation of the proportion of nonconforming parts using the total variation

This approximate calculation method takes into consideration that two sources of variability are present in the process.

Firstly, the calculation of $\hat{\sigma}_{TOTAL}^{LT}$ is needed. According to Equation (7) and the last column of Table 1, this value is $0.0011$. From this point, the steps of the calculation are the same as in Section 5.1, but instead of $\hat{\sigma}_{LT}$, the $\hat{\sigma}_{TOTAL}^{LT}$ should be used. Thus, the $\hat{Z}_{LSL}$ is 2.183, $\Phi(2.183)$ based on the $Z$ table is 0.0145. The estimated proportion of nonconforming parts is 0.01451 in this case.

5.3 Estimation of the proportion of nonconforming parts based on the tolerance interval, one-way design

For the purpose of the calculation of one-sided tolerance limits, a function was written in R (version 3.6.1.), which is based on the MOVER method. To validate this, for an example problem—which is presented by Krishnamoorty and Peng— the tolerance limits were reproduced (not shown here). The results were the same as in the paper published, so the function is proved to be suitable for the further calculations.

The applied confidence level is 0.95. According to this calculation method, the proportion of nonconforming parts is 0.00663.
5.4 | Comparison of the results

The proportion of the nonconforming parts from the three calculation methods are summarized in Table 2.

The difference between the two \( P_P \)-based approximate methods is negligible, being 3.7E-4 (see Figure 1). However, the proportion of nonconforming parts with the tolerance interval–based (sound) calculation method is less than half of the results from the two approximate methods. It means that the \( P_P \)-based methods give incorrect approximations for the ratio of nonconforming parts in the case of this dataset.

The deviation between the three methods depends on the ratio of variances. According to the above, the question arises whether there are any combinations of variances when the deviations are negligible in the practical sense.

5.5 | Comparing the deviation between the three methods, effect of the size of variances

The difference between the results of the three calculation methods comes from two components. One of them is the different calculation methods; the other is the uncertainty of the estimations of the expected value and the variances. To separate these two effects, an ideal case of the first example will be investigated in Section 5.5.1. Here, we assume that the expected value and the variance components are known. In Section 5.5.2, the uncertainty caused by estimating the variance components will be investigated.

5.5.1 | Comparing the three calculation methods, known variance components

Suppose that the expected value and the variance components (between, within) of the output distribution are known. This way, the “true value” of the proportion of nonconforming parts can be calculated. In this part of the evaluation, the differences between the results (proportion of nonconforming parts) of the three calculation methods and the effects of the variances will be discussed.

The \( \sigma_A^2 \) component was taken as 0.0001, 0.0003, 0.0005 and the \( \sigma_c^2 \) component was changed in the range of 0.00072 to 0.0036 during the evaluation. The results from the two \( P_P \)-based approximations were calculated according to Equations (3) and (4), the specification limits in the beginning of Section 5 were given.

| Calculation Method                        | The Estimate of the Proportion of Nonconforming Parts |
|------------------------------------------|-------------------------------------------------------|
| \( P_P \) with overall std (Equation (3))| 0.01414                                               |
| \( P_P \) with total std (Equation (7))  | 0.01451                                               |
| Tolerance interval (MOVER method)        | 0.00663                                               |

\[
\sigma_A^2 = 0.0001, 0.0003, 0.0005 \quad \text{and} \quad \sigma_c^2 = 0.00072 \text{ to } 0.0036
\]

The estimated proportion of the nonconforming parts from the three calculation methods, Example 1

---

**FIGURE 1** The estimated proportion of the nonconforming parts from the three calculation methods, Example 1
The two limits of the tolerance interval according to $\mu \pm z_\beta \sqrt{\sigma_A^2 + \sigma_e^2}$ were calculated$^{28}$ where the known value of the $\mu$ expected value is 8.6731, $z_\beta$ is the critical value of the standard normal distribution which belongs to the $\beta$ (that is the given proportion of the population). On the basis of Section 3.2, the sum of the $\beta$ values—which belongs to the specification limits—gives the ratio of nonconforming parts.

Figure 2 shows the ratio between the proportion of the nonconforming parts calculated with the $P_P$ approximation using the total variance $P_{PL} = \frac{\mu - LSL}{\sigma_{total}}$ and the tolerance interval as a function of the size of the within variance component. It can be seen that the ratios of the results are near to 1.000. This means that the deviations are negligible in the case of all investigated within and between variance values. The $P_P$-based method using the total variance results in slightly higher proportion of nonconforming parts.

Comparing the three curves (the parameter of the curves is the between variance), the deviation in the results is higher with lower between variances, but still negligible in a practical sense.

The ratio of the results from the two $P_P$-based approximations are near 1.0000, but the highest deviation is almost 10%. It should be noted that the deviation between the two calculation methods depends on the relation of the number of groups and number of repetitions in the groups as well. In this case, this effect is not evaluated, so the conclusions are valid for the one-way balanced design where the number of groups and the number of repetitions are both 10.

The $P_P$-based method which does not consider the real structure of the dataset gives lower value for the ratio of nonconforming parts than the tolerance interval–based method (see Figure 3). The highest deviation is almost 10% (with $\sigma_A^2 = 0.0005$ and $\sigma_e^2 = 0.00072$).

The most significant conclusion is that the two $P_P$-based calculation methods (assuming known variance components, that is, without considering the uncertainty of the estimation) may give a reasonable approximation for the tolerance interval–based (correct) calculation method.

**FIGURE 2** The ratio between the proportion of the nonconforming parts calculated with the total variance used $P_P$ approximation and the tolerance interval as a function of the size of the within variance component

**FIGURE 3** The ratio between the proportion of the nonconforming parts calculated with the overall variance used $P_P$ approximation and the tolerance interval as a function of the size of the within variance component
5.5.2 Effect of the uncertainty of estimates

In practical situations, the distribution parameters are not known, only estimated from experimental data. The uncertainty of estimates depends heavily on the size of experiments. Example 1 (in Section 4) is a PPAP situation, the number of experiments is 10 in each group and the number of groups is 10, thus the total number is 100. On the basis of 1000 simulated datasets, the mean of the proportion of nonconforming parts (calculated with the sound, tolerance interval based method) was 0.00423 and the standard deviation was 0.00520. Since the estimates of parameters are poor—thanks to the small number of experiments—it is clear that the estimates of nonconforming parts are also poor.

Example 2 (modified version of the example of Section 4) is about an operating process, so the number of experiments is higher: there are 10 experiments in each group and the number of groups is 50. So, the total number of experiments is 500 (instead of 100). In this case, 1000 datasets were simulated, as well, the mean of the proportion of nonconforming parts was 0.07682 and its standard deviation was 0.01180. Because of the higher number of experiments, the estimates of nonconformity rates are much better than in the PPAP situation. The ratio of nonconforming parts calculated with the three methods (on the basis of evaluating a single dataset) is shown in Table 3.

In this case, the results are more different than in Example 1, containing smaller number of experiments. On the basis of Figure 4, the differences between the tolerance interval-based method and $P_R$-based approximate methods are significant (and much higher than in the first example). With the tolerance interval–based method, the proportion of nonconforming parts is 0.06943, which is more than twice as high as the results of the two $P_R$-based approximations. Considering the standard deviation of the estimate (which is 0.01180 according to 1000 simulations), it can be concluded that the deviation between the $P_R$-based methods and the sound (tolerance interval–based) calculation method is much higher than the standard deviation (the difference between the two $P_R$-based approximate methods is negligible).

6 CONCLUSIONS

Three different calculation methods for the proportion of nonconforming parts were investigated. A special case was considered, ie, the output distribution of the process is known and the variances are in one-way structure. In this case,
the theoretically sound calculation method is based on tolerance interval, and the results were compared with those from $P_R$-based calculations as approximations. On the basis of our results, it can be concluded that the main difference between the results of the three calculation methods comes from the uncertainty of the parameter estimations.

In addition, datasets from two different situations were evaluated. The first one is a process validation environment when the number of experiments is moderate and well-defined. The another one is process monitoring when the data are less carefully collected and less reliable. With 1000 simulated datasets, the standard deviations of the estimates from the tolerance interval-based method were calculated. In the process validation environment—because of the small number of experiments—the standard deviation of nonconformings is higher than the estimates. The result from the tolerance interval-based method is significantly different from the $P_R$-based approximations in the case of process monitoring.

It is important to note that the difference between the three methods greatly depends on the number of experiments (number of groups and repetitions), ratio of variances, and on the structure of the error term. To reveal the possible tendencies, more simulation studies are required.

ACKNOWLEDGEMENTS
The authors are grateful to Professor Jenő Reiczigel (University of Veterinary Medicine Budapest) and Mr. Péter Kunovszki for their advice.

ORCID
Éva Pusztai https://orcid.org/0000-0002-1997-2630
Sándor Kemény https://orcid.org/0000-0001-5610-3664

REFERENCES
1. Minitab. Minitab 19.1. PA: Minitab, State College; 2019.
2. TIBCO Software Inc. 2018. Statistica 13.4. http://tibco.com.
3. ISO 22514-4 Statistical methods in process management—capability and performance—part 4: process capability estimates and performance measures; 2016.
4. ASTM E2281-15. Standard Practice for Process Capability and Performance Measurement. In: ASTM International. PA: West Conshohocken; 2015.
5. AUTOMOTIVE INDUSTRY ACTION GROUP. Statistical process control (SPC) – Reference material. ALEG; 2005.
6. FDA, Guidance for Industry—Process Validation: General Principles and Practices Rockville, MD, January 2011.
7. Henderson GR. Six Sigma Quality Improvement with Minitab. 2nd ed. New York: John Wiley and Sons, Inc.; 2011.
8. Bothe DR. Measuring Process Capability. 1st ed. New York: McGraw-Hill; 1997.
9. Kushe RH, Hurley P. Confidence bounds for capability indices. Journal of Quality Technology. 1999;24(4):188-195. https://doi.org/10.1080/00224065.1999.11979400
10. Wu CW, Pearn WL, Kotz S. An overview of theory and practice on process capability indices for quality assurance. Int. J. Production Economics. 2009;117:338-359. https://doi.org/10.1016/j.ijpe.2008.11.008
11. Montgomery DC. Introduction to Statistical Quality Control. 7th ed. New York: Wiley; 2013.
12. Ryan TP. Statistical Methods for Quality Improvement. 3rd ed. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2011.
13. Kotz S, Johnson NL. Process Capability Indices - A Review, 1992-2000. Journal of Quality Technology. 2002;34(1):2-19. https://doi.org/10.1080/00224065.2002.11980119
14. Mahshid R, Mansourvar Z, Hansen HN. Tolerance analysis in manufacturing using process capability ratio with measurement uncertainty. Precision Engineering. 2018;52:201-210. https://doi.org/10.1016/j.precisioneng.2017.12.008
15. Pearn WL, Liao MY. Measuring process capability based on $C_{PK}$ with gauge measurement errors. Microelectronics Reliability. 2005;45(3): 739-751. https://doi.org/10.1016/j.microel.2004.09.005
16. Steiner SH, MacKay RJ. Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes. 1st ed. Milwaukee: ASQ Quality Press; 2005.
17. Kotz S. Lovelace CR. Arnold: Process Capability Indices in Theory and Practice London; 1998.
18. Wheeler DJ, Chambers DS. Understanding Statistical Process Control. 1st ed. Knoxville: SPC Press; 1992.
19. Lorenzen TJ, Anderson VL. Design of Experiments, A No-Name Approach. 1st ed. New York: CRC Press; 1993.
20. Milliken GA, Johnson DE. Analysis of Messy Data, Volume 1: Designed Experiments.2nd ed. London: Chapman & Hall; 2009.
21. Meeker WQ, Hahn GJ, Escobar LA. Statistical Intervals: A Guide for Practitioners and Researchers. 2nd ed. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2017.
22. Komka K, Kemény S, Bánfai B. Novel tolerance interval model for the estimation of the shelf life of pharmaceutical products. Journal of Chemometrics. 2010;24(3-4):131-139. https://doi.org/10.1002/jchem.1294
23. Bánfai B, Ganzler K, Kemény S. Content uniformity and assay requirements in current regulations. Journal of Chromatography A. 2007; 1156(1-2):206-212. https://doi.org/10.1016/j.chroma.2006.10.067
24. Krishnamoorty K, Mathew T. Statistical Tolerance Regions: Theory, Applications and Computation. 1st ed. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2009.
25. Mee RW, Owen DB. Improved factors for one-sided tolerance limits for balanced one-way ANOVA random model. Journal of the American Statistical Association. 1983;78(384):901-905.
26. Vangel MG. New methods for one-sided tolerance limits for a one-way balanced random-effects ANOVA model. Technometrics. 1992;34(2):176-185.
27. Krishnamoorthy K, Mathew T. Generalized confidence limits and one-sided tolerance limits in balanced and unbalanced one-way random models. Technometrics. 2004;46(1):44-52. https://doi.org/10.1198/004017004000000103
28. Krishnamoorthy K, Peng J. Approximate one-sided tolerance limits in random effects model and in some mixed models and comparisons. Journal of Statistical Computation and Simulation. 2014;85(8):1651-1666. https://doi.org/10.1080/00949655.2014.887082

How to cite this article: Pusztai É, Kemény S. Process capability indices when two sources of variability present, a tolerance interval approach. Journal of Chemometrics. 2020;34:e3213. https://doi.org/10.1002/cem.3213