MultiCraft

International Journal of Engineering, Science and Technology
Vol. 14, No. 2, 2022, pp. 1-10

INTERNATIONAL JOURNAL OF ENGINEERING, SCIENCE AND TECHNOLOGY

www.ijest1-ng.com
www.ajol.info/index.php/ijest

A process capability index for zero-inflated processes

Surajit Pal [1], Susanta Kumar Gauri [2]*

[1] *SQC & OR Unit, Indian Statistical Institute, 110, N. Manickam Road, Chennai-600029, INDIA*
[2] *SQC & OR Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata-700108, INDIA*
Corresponding Author: e-mail: susantagauri@hotmail.com, Tel. No.: 091-033-2575-3350; Fax No.: 091-033-2577-6042
ORCID iDs: https://orcid.org/ 0000-0001-6066-5782 (Pal), https://orcid.org/ 0000-0003-4654-7696 (Gauri)

Abstract

The proportion of zero defect (ZD) outputs is as an integral characteristic of a zero-inflated (ZI) process or high quality process. Different ZI processes can almost equally satisfy the same USL of number of defects but they can produce substantially different proportions of ZD products. The application of conventional method for process capability evaluation fails to discriminate these processes because in the conventional method, the process capability is evaluated taking into consideration the USL of number of defects only. In this paper, a new measure of process capability for ZI processes is proposed that can truly discriminate different ZI processes taking into account the USL of number of defects as well as the proportion of ZD units produced in these processes. In the proposed approach, at first a measure of process capability index (PCI) with respect to the USL is computed, and then the overall PCI is obtained by multiplying it with an appropriately defined multiplying factor. A real-life application is presented.

Keywords: Process capability index; Zero-inflated process; Proportion of zero defect units; Zero-inflated Poisson process.

DOI: http://dx.doi.org/10.4314/ijest.v14i2.1

Cite this article as:
Pal S., Gauri S.K. 2022. A process capability index for zero-inflated processes. *International Journal of Engineering, Science and Technology*, Vol. 14, No. 2, pp. 1-10. doi: 10.4314/ijest.v14i2.1

Received: April 17, 2022; Accepted: July 13, 2022; Final acceptance in revised form: July 15, 2022

1. Introduction

Rapid technological advancement and implementation of automation and computerization have resulted in high quality processes in many manufacturing industries. In these processes, most of the time the process is in its perfect state and during this period, defects are rarely observed i.e., almost all the produced items are having zero number of defect. These zeros are known as structural zeros (Hu et al., 2011; Zamzuri, 2015). But when there is an equipment or process problem (alternatively called random shock), defects may occur. The number of defects in items are often modelled by a Poisson distribution or a negative binomial distribution. Such high quality manufacturing processes are often referred to as zero-inflated (ZI) processes with random shocks (Xie and Goh, 1993; Xie et al., 1995; Chang and Gan, 1999) or simply zero-inflated (ZI) processes (Lambert, 1992; Sim and Lim, 2008). Because of presence of the structural zeros, number of defects in a sample of size n collected from a ZI process always contains more number of zero defect (ZD) products than are expected under chance variation of the Poisson or negative binomial distribution. These extra zeros cause overdispersion (i.e. variance be larger than the mean). When the underlying defect distribution is Poisson, the process is called as ZI Poisson (ZIP) process and when the underlying defect distribution is negative binomial, the process is called as ZI negative binomial (ZINB) process. The ZIP model is generally used where the overdispersion is solely caused by the extra zeros. For count data where the overdispersion is caused by excess zeros and also by unobserved heterogeneity, the ZINB model is usually used (Phang and Loh, 2013; Chaney et al., 2013).

In literature, several propositions are reported on the control and monitoring of ZI processes (Chang and Gan, 1999; Sim and Lim, 2008; Zamzuri, 2015; Rakitzis et al., 2017; Tian et al., 2019; Alevizakos and Koukouvinos, 2021). A detailed review on the past and current trends for the models and monitoring of ZI processes is available in Mahmood and Xie (2019). Zhang and Yi (2022) have discussed about the Zero-inflated Poisson models with measurement error in the response. Tian et al. (2022) introduced the zero-inflated non-central negative binomial (ZINNB) distribution and presented the maximum likelihood estimation method for estimation of the parameters of the ZINNB distribution. In a manufacturing set up, process capability analysis of ZI process or high quality process has also an important role in the context of quality control. In order to survive in today's highly competitive markets, manufacturers need to produce items from a high quality process. Also, manufacturers place their orders only to those vendors who have high quality processes. Selection of the best vendor among the competing vendors having high quality processes is an important issue to these manufacturers. It is also of interest to know how well a high quality process will hold the specifications.

To the best of our knowledge, only Patil and Shirke (2012) and Pal and Gauri (2021) have attempted to measure capability of a ZI process. Number of defects is mainly smaller-the-better (STB) type of quality characteristic and the specified requirement for number of defects is usually defined in terms of upper specification limit (USL) of number of defects in a unit product. Patil and Shirke (2012) have modified the Perakis and Xekalaki (2005) proposed C_{pcu} index by incorporating the inflation of zero parameter into C_{pcu} index, and applied it for measuring the capability of a ZI Poisson (ZIP) process. But it fails to represent the true capabilities of ZI processes consistently. Particularly, for small value (≤ 0.5) of inflation of zero parameter, the value of the process capability index becomes unusually high, which gives a wrong impression about the capability of the concerned process. On the other hand, Pal and Gauri (2021) evaluated the capability of a ZIP process with respect to the USL by applying the concept of Borges and Ho (2001). Pal and Gauri's (2021) approach ensure that the computed capability index has a one-to-one correspondence with the expected nonconformance and it is not unreasonably high. However, they ignored the count of ZD products produced in a ZI process. Because of that, Pal and Gauri (2021) proposed approach fails to discriminate the ZIP processes which produces different proportions of ZD units but having the same proportion of nonconforming items with respect to the USL of number of defects.

Different ZI processes can almost equally satisfy the same USL of number of defects but they can produce substantially different proportions of ZD products. Obviously, the most desirable process is one in which the proportion of ZD products is the maximum. Therefore, the proportion of ZD outputs should be considered as an integral characteristic of a ZI process, and a measure of process capability of a ZI process should reflect the same. This is possible only if a lower specification limit (LSL) for the proportion of ZD products is taken into consideration along with the USL of number of defects. In such a case, the two different specification limits are for two different types of attribute quality characteristics – LSL for proportion of ZD products and USL for number of defects in a unit. The number of ZD items follows a binomial distribution whereas the number of defects in items follows a Poisson or negative binomial distribution. Finding indices for these two different type of characteristics and combining them into a single index is a difficult task. No standard methodology can deal with such cases.

In this paper, a new approach for measuring capability of a ZI process is proposed. In this approach, at first a measure of PCI with respect to the USL for defects is computed, which is then multiplied by a factor to obtain the overall PCI. The proposed index can truly discriminate the performances of different ZI processes. This article is organized as follows. The literature review is presented briefly in section 2. The proposed approach for evaluation of process capability of a ZI process is described in section 3. The results of a comparative study are discussed in section 4. Application of the proposed method to a real life problem is described in section 5. The article is concluded in section 6.

2. Literature Review

Capabilities of processes are assessed in terms of different indices, e.g. C_p, C_{pk}, C_{pm}, and C_{pmk} (Kane, 1986; Kotz and Johnson, 2002; Chen et al., 2017; Polhemus, 2018). Historically, these indices are developed for a product characteristic that can be described as a continuous variable and follows normal distribution. The generalisation of these indices for continuous nonnormal quality characteristics are suggested by Clements (1989), Pearn and Chen (1995), Goswami and Dutta (2013) etc. However, in reality, many quality characteristics are neither continuous variable, nor they follow normal distribution. These data (e.g. defect, error, defective items etc.) are typically obtained by counting and are known as attribute data, which usually follow Poisson or binomial distribution. Therefore, standard formulas cannot be used for computation of capability indices of a process involving such characteristics. To alleviate the problem, some generalized indices, e.g. C-index (Borges and Ho, 2001), C_f index (Yeh and

Bhattacharya, 1998), C_{pc} index (Perakis and Xekalaki, 2005) and C_{py} index (Maiti et al., 2010) are proposed in literature. These indices are applicable to any process regardless of whether the quality characteristic is discrete or continuous and irrespective of its underlying probability distribution. The attribute characteristics are usually smaller-the-better (STB) type and the requirements are specified by an USL. Thus, the appropriate generalized indices for these characteristics are C_{fu}, C_u, C_{pcu} and C_{pyu}. Pal and Gauri (2020[a], 2020[b]) have compared the relative accuracies of these one-sided generalized indices for binomial as well as Poisson processes and they recommended for using C_u index for measuring capability of a Poisson or binomial process. It may be noted that in all these cases the process outputs are assumed to follow purely Poisson or purely binomial distribution, and it is further assumed that there is only USL for the process outputs.

The quality revolution caused by an increasingly competitive global market since 1990s coupled with the rapid advancement of technologies and automation in today's world has led to tremendous improvement in the quality of manufactured products. One assumption is that these processes are so good that, in general, most of the produced items are defect-free and only a few defective items are produced because of random shocks in the process (Chang and Gan, 1999; Xie and Goh, 1993). The random shocks cause occurrences of defective items each containing one or more number of defects. The number of defects follows a Poisson or negative binomial distribution. Such high-quality processes usually have more count of zeros than are expected under chance variation of its underlying Poisson or other count distribution (Sim and Lim, 2008). These processes are usually referred to as ZI processes (Lambert, 1992; Sim and Lim, 2008), and these processes are modelled as a mixture of a degenerate distribution at zero and a Poisson or negative binomial distribution.

There are many research articles on developing appropriate control charts for monitoring of ZI processes (Chang and Gan, 1999; Sim and Lim, 2008; Zamzuri, 2015; Rakitzis et al., 2017; Tian et al., 2019; Alevizakos and Koukouvinos, 2021). However, only Patil and Shirke (2012) and Pal and Gauri (2021) have attempted to measure the capability of a ZI process. Patil and Shirke (2012) have modified the Perakis and Xekalaki (2005) proposed C_{pcu} index by incorporating the inflation of zero parameter into C_{pcu} index. But it fails to represent the true capabilities of ZI processes consistently. Particularly, for small value (≤ 0.5) of inflation of zero parameter, the value of their process capability index becomes unusually high, which gives a wrong impression about the capability of the concerned process. On the other hand, Pal and Gauri (2021) have applied the concept of Borges and Ho (2001) for measuring the capability of a ZI Poisson process. Pal and Gauri's (2021) approach ensure that the computed capability index never be unreasonably high and it has one-to-one correspondence with the expected nonconformance from the process. However, Pal and Gauri (2021) proposed approach fails to discriminate the ZI processes which produce different proportions of ZD units but having the same proportion of nonconforming items with respect to the USL of number of defects.

3. The Proposed Approach for Evaluation of Capability of a Zero-inflated (ZI) Process

Suppose, a ZI process has an upper specification limit (USL) for the maximum number of permissible defects/nonconformities in a unit, denoted by c_{usl}. Then the outputs of a ZI process can be classified into three categories: zero defect (ZD) units (i.e. outputs which contain no defect), acceptable (AC) units (i.e. outputs which contain number of defects/nonconformities ranging from 1 to c_{usl}) and nonconforming (NC) units (i.e. outputs which contain more than c_{usl} number of defects/nonconformities). Suppose, the proportion of ZD units in the process is p_{ZD} and the LSL for proportion of ZD unit is specified as p_{ZD}^{lsl}. It may be noted that while number of ZD units in a sample follows binomial distribution, number of defects in a unit (c) follows Poisson distribution.

One possible approach for assessing the overall capability of a ZI process can be evaluating separately the capability indices with respect to p_{ZD}^{lsl} and c_{usl}, and then integrating the two indices into one index. But it seems to be quite a difficult task. Therefore, it is planned to obtain first a measure of process capability with respect to the USL (i.e. c_{usl}) and then to apply an appropriate multiplying factor taking into consideration the proportion of ZD units with respect to its specified limit. This will ensure that different ZI processes are discriminated properly even when they equally satisfy the given USL on the number of defects/nonconformities.

Borges and Ho (2001) suggested a measure of process capability, called c-index, which has one-to-one correspondence (mapping) between the proportion of NC items produced in a process and Z-value of the standard normal distribution. When only USL is specified, in this approach, the expected proportion of NC items in a process with respect to USL is mapped to the Z-score in the right side of standard normal distribution, and 1/3rd of this Z-score is considered as the measure of the process capability with respect to USL and it is denoted as c_u. Obviously, in this approach, c_u responds to changes in the NC region and not to changes in the distribution of the observed quality characteristic. Consequently, computation of c_u index is feasible in any process regardless of whether the quality characteristics are discrete or continuous and irrespective of their probability distributions. Another major advantage of c_u is its interpretation, which is similar to the interpretation of the index computed from a normally distributed process. Based on the value of c_u, one can easily estimate the expected proportion of conforming outputs in the process, which is $\Phi(3c_u)$ where $\Phi(.)$ denotes the standard normal cumulative distribution function.

Therefore, it is proposed to obtain the measure of process capability of a ZI process with respect to USL by using Borges and Ho's (2001) approach. Suppose, the proportion of NC units in the process is p_{NC}. Then we need to find out the Z-value in the right side of the standard normal distribution that results in probability area equal to p_{NC}. Let Z_U is the value of Z that results in probability area p_{NC} above it. The Z_U value can be obtained by using inverse cumulative probability of the standard normal distribution function as follows: $Z_U = \Phi^{-1}(1 - p_{NC})$. Then the process capability index of the ZI process with respect to USL will be obtained as $C_{pu}^{ZI} = \frac{1}{3} \times Z_U = \frac{1}{3} \times \Phi^{-1}(1 - p_{NC})$.

However, as mentioned earlier, count or proportion of ZD products produced is an important characteristics of a ZI process, and therefore, assessment of process capability ignoring the count of ZD products produced in a ZI process may be misleading. Different ZI processes can have the same proportion of nonconforming items (i.e. products having more than c_{usl} number of defects), but they can produce different proportions of ZD products. Obviously, the most desirable process is one in which the proportion of ZD products is the maximum and the most undesirable process one in which the proportion of ZD products is the minimum. The overall measure of process capability of a ZI process should reflect the same. Suppose, there is a lower specification limit (LSL) for the proportion of ZD products produced in a ZI process and let it is denoted as p_{ZD}^{lsl}. Now, if we consider $m = \frac{p_{ZD}}{p_{ZD}^{lsl}}$ as an multiplying factor to C_{pu}^{ZI}, then the above requirement will be satisfied. For example, suppose three ZI processes equally satisfy the USL of number of defects, but proportions of ZD products produced in these processes are less than p_{ZD}^{lsl} (the LSL), equal to p_{ZD}^{lsl} and more than p_{ZD}^{lsl}, respectively. Then the values of the multiplying factors for these processes will be less than one, equal to one and more than one. This will imply that the overall process capability index will be the minimum for the first process and the maximum for the third process, which is expected under consideration of proportion of ZD products produced in these processes. Thus, the overall process capability index of a ZI process (C_p^{ZI}) can be obtained as

$$\begin{aligned} C_p^{ZI} &= m \times C_{pu}^{ZI} \\ &= \frac{p_{ZD}}{p_{ZD}^{lsl}} \times \frac{1}{3}\Phi^{-1}(1 - p_{NC}) \end{aligned} \tag{1}$$

If $C_P^{ZI} \geq 1$, the concerned ZI process will be considered capable of producing products satisfying both the LSL of proportion of ZD units (p_{ZD}^{lsl}) and the USL of number of defects / nonconformities (c_{usl}). Otherwise, the ZI process will be considered not capable of satisfying the LSL of proportion of ZD units and/or the USL of number of nonconformities. It is worth to mention that if the value of $(1 - p_{NC})$ is less than 0.5, then the value of C_P^{ZI} will be negative. The value of $1 - p_{NC} < 0.5$ implies $p_{NC} > 0.5$. It gives sufficient indication that the process is producing plenty of nonconforming units with respect to the USL of number of nonconformities, and thus the process is not capable at all. So it is recommended to consider $C_P^{ZI} = 0$ if $1 - p_{NC} < 0.5$. This will ensure that the process capability index C_P^{ZI} is always greater than or equal to zero.

Again, when the value of proportion of NC units (p_{NC}) is zero, the value of $C_{pu}^{ZI} = \frac{1}{3}\Phi^{-1}(1) \cong \frac{4}{3} = 1.33$. In that case, depending on the proportion of ZD units (p_{ZD}) produced in the process and the LSL of proportion of ZD units (p_{ZD}^{lsl}), the maximum value of C_P^{ZI} can be $\hat{C}_P^{ZIP} = 1.33 \times \frac{\hat{p}_{ZD}}{p_{ZD}^{L}}$. For example, in a high quality ZI process with zero NC unit, if the LSL of the proportion of ZD units, i.e. p_{ZD}^{lsl} is specified as 0.80, the maximum value of C_P^{ZI} will not exceed $\hat{C}_P^{ZIP} = 1.33 \times \frac{1.0}{0.8} = 1.6625$.

3.1 Procedure for obtaining estimate of C_p^{ZI}

An estimate of the overall process capability index ($\hat{C}_p^{ZI}$) of a ZI process can be obtained using the following steps:

1) ***Collect a sample of n units from the concerned zero-inflated (ZI) process and observe the numbers of nonconformities present in each of the sample items.***

   Let the observed number of nonconformities in the sample units ranges from d = 0, 1, 2,...to *m*. Suppose, number (frequency) of sample units each having exactly 'd' number of nonconformities is denoted by n_d (d = 0, 1, 2,...,*m*). Then, the total number of nonconformities can be computed as $D = \sum_{d=0}^{m} d \times n_d$.

2) ***Select an appropriate zero-inflated (ZI) probability distribution for describing the sample data.***

   Zhao et al. (2009) and Kumar and Ramachandran (2020) have provided test procedures for checking zero-inflation in the process data. Generally ZI Poisson (ZIP) model is used for modelling ZI count data where the overdispersion is solely caused by the extra zeros. Yang et al. (2011) proposed a method for outlier identification and robust parameter estimation in a ZIP process. For count data where the overdispersion is caused by excess zeros and also by unobserved heterogeneity, the most commonly recommended model is ZI negative binomial (ZINB). This is because it employs additional parameter that models additional variability (Chaney et al., 2013). Some other models that are used for such overdispersed count data are ZI double Poisson (ZIDP) model (Phang and Loh, 2013) and ZI generalized Poisson (ZIGP) model (Wagh and Kamalja, 2018). A few other models proposed in literature for modelling ZI data are Bayesian ZI regression model (Workie and Azene, 2021) and generalized linear mixed models (Favero et al., 2021).

Since, in real life, most commonly ZIP model is used for modelling ZI count data, here the procedure for obtaining the estimate of C_p^{ZI} are described considering that the concerned process data is well modelled by ZIP distribution. If the random variable Y represents number of nonconformities presents in a unit product and Y follows ZIP distribution, the probability mass function (pmf) for ZIP model can be written as

$$f(y;\Omega,\lambda)=\begin{cases}(1-\Omega)+\Omega e^{-\lambda} & \text{for } y=0\\ \Omega\frac{\lambda^y e^{-\lambda}}{y!} & \text{for } y=1,2,3,..\end{cases} \tag{2}$$

where Ω is the probability of occurrence of random shock and in this state of the process, nonconformities (defects) occur in the produced items according to Poisson distribution with parameter λ. The probability of occurrence of the other state is 1-Ω and in this state, only zero defect items are produced. The mean and variance of the ZIP distribution are $E(Y)=\Omega\lambda$ and $Var\ (Y)=\Omega\lambda[1+(1-\ \Omega)\lambda]$, respectively.

3) ***Estimate the parameters of the selected ZI distribution from the observed count data.***

The parameters Ω and λ can be estimated from the observed dataset by the method of maximum likelihood (Xie and Goh, 1993). The log-likelihood function of Ω and λ for the observed dataset can be written as

$$ln\ L(\Omega,\lambda)=\ln[\{P(Y=0)\}^{n_0}]+\ln\left[\prod_{d=1}^{m}\{P(Y=d)\}^{n_d}\right]$$

$$=\ n_0\ ln\left[(1-\Omega)+\Omega e^{-\lambda}\right]+\sum_{d=1}^{m}n_d ln\ (\Omega\frac{e^{-\lambda}\lambda^d}{d!}) \tag{3}$$

The partial derivatives of the log-likelihood function with respect to Ω and λ result in the following two likelihood equations:

$$\frac{n_0(-1+e^{-\lambda})}{(1-\Omega)+\Omega e^{-\lambda}}+\frac{n-n_0}{\Omega}=0 \tag{4}$$

$$\frac{-n_0\Omega e^{-\lambda}}{(1-\Omega)+\Omega e^{-\lambda}}+\frac{D}{\lambda}-(n-n_0)=0 \tag{5}$$

The maximum likelihood estimates (MLEs) of Ω and λ can be obtained by solving these two likelihood equations and that can be executed by using enumerative search procedures. The values of Ω and λ that maximizes the log-likelihood function are the MLEs of Ω and λ respectively. A Chi-square goodness-of-fit test must be performed for checking the adequacy of the fitted model for the sample data.

Suppose, the estimated parameters of the ZIP distributions are $\widehat{\Omega}$ and $\hat{\lambda}$.

4) ***Estimate the expected proportion of nonconforming units ($\hat{p}_{NC}$) in the concerned ZI process.***

A unit will be considered nonconforming if the number of nonconformities in it is more than c_{usl}. Then p_{NC} in the concerned ZIP process can be estimated as follows:

$$\begin{aligned}\hat{p}_{NC}&=P(Y>c_{usl})=1-[P(Y=0)+P(1\le Y\le c_{usl})]\\&=1-\left[\{(1-\widehat{\Omega})+\widehat{\Omega}e^{-\hat{\lambda}}\}+\sum_{y=1}^{c_{usl}}\widehat{\Omega}\frac{\hat{\lambda}^y e^{-\hat{\lambda}}}{y!}\right]\\&=\widehat{\Omega}(1-e^{-\hat{\lambda}})-\sum_{y=1}^{c_{usl}}\widehat{\Omega}\frac{\hat{\lambda}^y e^{-\hat{\lambda}}}{y!}\end{aligned} \tag{6}$$

5) ***Estimate the expected proportion of zero defect units ($\mathbf{p}_{ZD}$) in the concerned ZIP process.***

The estimated probability of occurrences of random shock is $\widehat{\Omega}$. Then, the probability of absence of random shock is $1-\widehat{\Omega}$, and in the absence of random shock only ZD products are produced. On the other hand, when random shock takes place, the number of nonconformities in the produced items occurs according to Poisson distribution with parameter $\hat{\lambda}$. Therefore, the probability of ZD products due to occurrences of random shock is $P(Y=0)=\Omega e^{-\lambda}$. Thus, p_{ZD} in the concerned ZIP process can be estimated as follows:

$$\hat{p}_{ZD} = P(Y=0) = \widehat{\Omega} e^{-\lambda} + (1-\widehat{\Omega}) \tag{7}$$

6) Finally, obtain the estimate of the overall process capability index ($\hat{C}_p^{ZI}$)

The estimate of the overall process capability index of the concerned ZI process can be obtained as follows:

$$\hat{C}_p^{ZI} = \frac{\hat{p}_{ZD}}{p_{ZD}^{lsl}} \times \frac{1}{3}\Phi^{-1}(1-\hat{p}_{NC}) \tag{8}$$

7) **Obtain the confidence interval of the estimated overall process capability index ($\widehat{C}_p^{ZI}$)**

Since $\hat{C}_P^{ZI}$ is a point estimate obtained from sample data, it is necessary to construct confidence interval (CI) of the capability index C_P^{ZI} for inference purpose, especially when the sample size is relatively small.

It is known that the number of ZD units in a Poisson process follows a binomial distribution with parameters n (sample size) and proportion $p = e^{-\lambda}$, where λ is the parameter of the Poisson process. Accordingly, the mean value and 95% confidence intervals of number of ZD items in a Poisson process with parameter λ can be estimated (using normal approximation of the binomial distribution) as follows:

Mean number of ZD units = $n\hat{p} = ne^{-\hat{\lambda}}$ and

95% CI of ZD units = $\left[ne^{-\hat{\lambda}} - 1.96 \times \sqrt{ne^{-\hat{\lambda}}(1-e^{-\hat{\lambda}})},\ ne^{-\hat{\lambda}} + 1.96 \times \sqrt{ne^{-\hat{\lambda}}(1-e^{-\hat{\lambda}})}\right]$

Extending the same for a ZI Poisson process with parameters Ω and λ, the mean number of ZD units and 95% CI of ZD units in a ZI Poisson process can be estimated as follows:

Mean number of ZD units = $n\hat{p}_{ZD} = n\left[(1-\widehat{\Omega}) + \widehat{\Omega} e^{-\hat{\lambda}}\right]$ and

95% CI of ZD units = $\left[n\hat{p}_{ZD} - 1.96 \times \sqrt{ne^{-\hat{\lambda}}(1-e^{-\hat{\lambda}})},\ n\hat{p}_{ZD} + 1.96 \times \sqrt{ne^{-\hat{\lambda}}(1-e^{-\hat{\lambda}})}\right]$

It may be noted that the number of ZD products in a sample of size n collected from a ZI Poisson process cannot be less than $n(1-\widehat{\Omega})$.

Similarly, the mean number of NC units and 95% CI of NC units in a ZI Poisson process can be obtained as follows:

Mean number of NC units = $n\hat{p}_{NC} = n\left[\widehat{\Omega}(1-e^{-\hat{\lambda}}) - \sum_{y=1}^{CUSL} \widehat{\Omega} \frac{\hat{\lambda}^{y} e^{-\hat{\lambda}}}{y!}\right]$ and

95% CI of NC units = $\left[n\hat{p}_{NC} - 1.96 \times \sqrt{n\hat{p}_{NC}(1-\hat{p}_{NC})}, n\hat{p}_{NC} + 1.96 \times \sqrt{n\hat{p}_{NC}(1-\hat{p}_{NC})}\right]$

However, deriving the CI of C_P^{ZI} taking into account the CIs of ZD and NC units is quite difficult. Hence, it is proposed to use Nagata and Nagahata (1994) proposed generalized approximation formula for construction of two-sided CI of C_P^{ZI}. According to Nagata and Nagahata (1994),

$$(1-\alpha)\%\ \text{CI of } C_P^{ZI} = \left[\hat{C}_P^{ZI} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{9n} + \frac{\hat{C}_P^{ZI}}{2(n-1)}},\ \hat{C}_P^{ZIP} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{9n} + \frac{\hat{C}_P^{ZI}}{2(n-1)}}\right] \tag{9}$$

where, α is the level of significance and (1- α) is the confidence coefficient.

4. Comparative Study and Discussions

For the purpose of the comparative study, a ZI Poisson process is considered where the USL for number of defects is 3. So the outputs of the process can be classified into following three categories: zero defect (ZD) units (i.e. outputs which contain no defect), acceptable (AC) units (i.e. outputs which contain 1-3 number of defects) and nonconforming (NC) units (i.e. outputs which contain more than 3 defects).

In order to understand the possibilities of occurrence of number of units in these three categories, under different values of the parameters (Ω, λ) of ZI Poisson distribution, a few theoretical computations are carried out. At first, maintaining $\lambda = 2$ as constant, expected proportions of ZD, AC and NC units are computed under different values of Ω (probability of occurrence of random shock). The results of these computations are presented in Table 1. It can be observed from Table 1 that the proportion of ZD units varies substantially depending on the values of Ω, although the parameter (λ) remains constant (here λ = 2).

Table 1. Expected proportions of ZD, AC and NC units in ZI Poisson process when λ = 2 but Ω varies

Defect categories	Values of random shock probability (Ω)				
	Ω = 0.15	Ω = 0.12	Ω = 0.10	Ω = 0.08	Ω = 0.05
ZD (0 defect)	0.8703	0.8962	0.9135	0.9308	0.9568
AC (1-3 defects)	0.1083	0.0866	0.0722	0.0578	0.0361
NC (more than 3 defects)	0.0214	0.0172	0.0143	0.0114	0.0071

Now it is tried to find out various combinations of Ω and λ values for ZI Poisson processes where the proportion of NC units is the same. Although there are many such combinations, only four ZI Poisson processes are chosen, where proportion of NC units is the same 0.01. Table 2 shows the values of the parameters of the chosen four ZI Poisson distributions and the expected proportions of ZD, AC and NC units in these ZI Poisson processes.

Table 2. Parameters of ZI Poisson processes and expected proportions of different types of units

Process number	Parameters of ZI Poisson process		Expected mean number of defects	Expected proportions of		
	Ω	λ		ZD units	AC units	NC units
1	0.20161	1.36263	0.2747	0.850	0.140	0.010
2	0.16247	1.46699	0.2383	0.875	0.115	0.010
3	0.12501	1.60900	0.2011	0.900	0.090	0.010
4	0.08953	1.81849	0.1628	0.925	0.065	0.010

Now process capability indices for these ZI Poisson processes are evaluated using Patil and Shirke (2012), Pal and Gauri (2021) and the proposed approaches. The results of these analysis are presented in Table 3.

Table 3. Parameters of ZI Poisson processes and expected proportions of different types of units

Process number	Process parameters		Estimates of		
	Ω	λ	Patil and Shirke's C_{pcu}^{Z} index	Pal and Gauri's C_u index	Proposed C_P^{ZI} index
1	0.20161	1.36263	1.339	0.776	0.732
2	0.16247	1.46699	1.662	0.776	0.754
3	0.12501	1.60900	2.160	0.776	0.776
4	0.08953	1.81849	3.016	0.776	0.797

It can be noted that when Pal and Gauri's (2021) approach is used, the computed process capability indices are the same for all the four processes. This is because Pal and Gauri (2021) evaluated the capability of a ZI Poisson process considering only the proportion of NC units with respect to USL, and here, the proportion of NC units with respect to the USL is the same 0.01 for all the four processes. That is, the Pal and Gauri (2021) proposed approach fails to discriminate the ZI Poisson processes if the proportion of NC units is the same but proportion ZD units produced in these processes are different. On the other hand, the differences in the expected proportions of ZD units and/or AC units in different ZI Poisson processes are well reflected in the estimated process capability indices obtained by both Patil and Shirke's (2012) approach and the proposed approach.

However, the process capability indices obtained by Patil and Shirke's (2012) approach seems to have interpretation issues. Traditionally, a value of process capability index equal to 0.776 implies that the process is capable of producing 99% conforming products; a value of process capability index equal to 1 implies that the process is capable of producing 99.865% conforming products and the capability of the process is considered good; a value of process capability index equal to 1.33 implies that the process is capable of producing 99.995% conforming products and the capability of the process is considered very good. It may be noted that all the four ZI Poisson processes considered here are capable of producing 99% conforming products and therefore, the expected values of the process capability indices should be around 0.776. The process capability indices obtained by the proposed approach satisfy this requirement. In addition, the proposed approach can also discriminate the ZI processes based on proportions of ZD items produced. On the other hand, the values of process capability indices obtained by Patil and Shirke's (2012) approach are found to be substantially high, which gives a wrong impression about the capability of the concerned process.

5. The Real-life Application

The motivation for the current study was a real life problem of an Indian automobile industry. One of the critical components used in the automobile industry is supplied by three vendors. The vendors are coded as A, B and C. The vendors' processes are so well managed that most of the supplied components are usually free from defect barring a few items. Yet for ensuring adequate check on the quality of the component, a strict USL ($c_{usl} = 3$) is specified for the number of defects in a component. If any component contains more than three defects, it is considered nonconforming and gets rejected. The engineers of the company experienced that the performances of components differ considerably vendor wise when these components are used. So they were interested to assess the process capability of each vendor which will enable them for vendor selection objectively.

Most of the components produced by the vendors are of zero defect (ZD). The overdispersion (variance more than mean) in the sample data is mainly caused by the excess number of zeros in the sample data. Hence, the underlying distribution of the number of defects are assumed to be ZI Poisson (ZIP). However, the goodness-of-fit tests are performed to verify the same after estimating the parameters of ZIP distribution from the sample data using the maximum likelihood method. It was highlighted to the engineers that the efficiency of a ZIP process should be judged not only taking into account the proportion NC products produced but also taking into consideration the proportion of ZD products produced. Being enlightened by the above understanding, the engineers of the company decided to specify the LSL for the proportion of ZD units (p_{ZD}^{lsl}) as 0.95 in addition to the existing USL for the number of defects in a unit. So the process capabilities of the three vendors are now assessed taking into consideration both the specifications.

About 300 components are collected randomly from the grinding process under stable conditions for each of three vendors. The frequency of different number of defects in the sample components of each vendor is shown below in Table 4.

Table 4. Sample data on defects for all three vendors

Vendor	Sample Size (n)	Frequency of number of defects in the sample					
		0	1	2	3	4	5
Vendor A	304	296	7	4	3	0	0
Vendor B	296	287	5	3	0	1	0
Vendor C	305	284	8	6	4	2	1

A ZIP distribution is fitted to the sample data of each vendor using the maximum likelihood method. Let the estimated parameters of the fitted ZIP distribution for the i^{th} vendor are $\widehat{\Omega}_i$ and $\hat{\lambda}_i$ ($i = 1,2,3$). Putting these estimated parameter values in Equations (6) and (7), the expected proportion of NC components ($\hat{p}_{NC}$) and expected proportion of ZD components ($\hat{p}_{ZD}$) in the processes of all three vendors are obtained. The process capability indices for all three vendors are then obtained using Equation (8). The 95% confidence intervals of the capability indices are then computed using Equation (9). The estimated parameters of the fitted ZIP distributions, estimated proportions and capability indices for the three vendors are shown in Table 5.

Table 5. Estimated parameters of the fitted ZIP distributions and capability indices

Vendor	Sample size (n)	Parameters of ZIP process		Estimated Proportions		$\hat{C}_p^{ZI}$	95% confidence limits of C_p^{ZI}
		$\widehat{\Omega}$	$\hat{\lambda}$	$\hat{p}_{ZD}$	$\hat{p}_{NC}$		
Vendor A	304	0.0660	1.1958	0.9539	0.00221	0.953	[0.868, 1.038]
Vendor B	296	0.0450	1.1263	0.9696	0.00125	1.029	[0.938, 1.120]
Vendor C	305	0.0828	1.7823	0.9311	0.00877	0.776	[0.704, 0.848]

It can be observed from Table 5 that the estimated $\hat{C}_P^{ZI}$ value for vendor B is above 1.0 and the estimated $\hat{C}_P^{ZI}$ value for vendor A is also reasonably good. But the estimated $\hat{C}_P^{ZI}$ value is quite poor for vendor C. The 95% confidence intervals of capability index C_p^{ZI} for vendor B and vendor A are overlapping and hence there may not be any significant difference between these two vendors' capability indices. However, the capability index of vendor C is significantly inferior compared to vendors B and A. Based on the estimated indices, the vendors are prioritized for placement of purchase order in the following sequence: vendor B, vendor A and vendor C.

It may be worth to highlight the followings: There is a single NC component in the sample of vendor B, but the proportion of ZD items is almost 97%. Whereas there is no NC item in the sample of vendor A, but the proportion of ZD items (95.4%) is just above the specified LSL. Although there is no NC item in the sample of vendor A, the estimated proportion of NC items in the lot is computed as 0.22%, which is slightly more than that of vendor B. This difference in proportion of NC items and the difference in the proportion of ZD items are reflected in the estimated process capability indices.

6. Conclusions

In today's manufacturing scenario, zero-inflated (ZI) processes are quite common, which is due to tremendous improvement in automation technologies. Often evaluation of capabilities of different high quality processes becomes necessary for their comparison and decision making. Keeping in mind that proportion of ZD products produced is an integral part of a ZI process, a measure of process capability index is proposed for ZI processes. In the proposed approach, at first a measure of process capability with respect to the USL of the number of defects is computed, and then the overall process capability index (PCI) is obtained by multiplying a factor defined based on the actual value of proportion of ZD units and the LSL of the proportion of ZD units. The proposed method is applied to tackle a real-life problem of an Indian automobile industry. In this article, ZI Poisson (ZIP) distribution is used for modelling the defect data. In many situations, especially when there is overdispersion in the ZI process, ZI negative binomial (ZINB) or ZI generalized Poisson (ZIGP) distributions are utilized for modelling the defect data. The proposed approach for computing the overall PCI can be used for those cases also. Existence of bivariate and multivariate ZI processes is also not uncommon in real-life manufacturing set up. Future studies may aim at developing appropriate methodologies for evaluating PCI for such processes.

Acknowledgment

Thanks to the referees for their comments that have enhanced the quality of this article.

References

Alevizakos, V., & Koukouvinos, C. (2021). Monitoring of zero-inflated binomial processes with a DEWMA control chart. *Journal of Applied Statistics*, Vol. 48, pp. 1319-1338. https://doi.org/10.1080/02664763.2020.1761950

Borges, W., & Ho, L. L. (2001). A fraction defective based capability index. *Quality and Reliability Engineering International*, Vol. 17, pp. 447-458. https://doi.org/10.1002/qre.438

Chaney, S. E. P., Charity, M., McDowall, D., & Aban, I. (2013). Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation*, Vol. 83, No. 9, pp. 1671-1683. https://doi.org/10.1080/00949655.2012.668550

Chang, T. C., & Gan, F. F. (1999). Charting techniques for monitoring a random shock process. *Quality and Reliability Engineering International*, Vol. 15, pp. 295–301. https://doi.org/10.1002/(SICI)1099-1638(199907/08)

Chen, K. S., Wang, K. J., & Chang, T. C. (2017). A novel approach to deriving the lower confidence limit of indices C_{pu}, C_{pl}, and C_{pk} in assessing process capability. *International Journal of Production Research*, Vol. 55, pp. 4963-4981. https://doi.org/10.1080/00207543.2017.1282644

Clements, J. A. (1989). Process capability calculations for non-normal distributions. *Quality Progress*, Vol. 22, pp. 95-100.

Favero, L. P., Hair, J. F. Jr., Souza, R. D. F., Albergaria, M., & Brugui, T. V. (2021). Zero-inflated generalized linear mixed models: a better way to understand data relationships. *Mathematics*, Vol. 9, No. 10, pp. 1-28. https://doi.org/10.3390/math9101100

Goswami, A., & Dutta, H. N. (2013). Some studies on normal and non-normal process capability indices. *International Journal of Mathematics and Statistics Invention*, Vol. 1, pp. 31–40.

Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *American Journal of Drug and Alcohol Abuse*. Vol. 37, pp. 367–375. https://doi.org/10.3109/00952990.2011.597280

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology*, Vol. 18, pp. 41-52.

Kotz, S., & Johnson, N. L. (2002). Process capability indices - a review, 1992-2000. *Journal of Quality Technology*, Vol. 34, pp. 2-19. https://www.jstor.org/stable/27919651

Kumar, C. S., Ramachandran, R. (2020). On some aspects of a zero-inflated overdispersed model and its applications. *Journal of Applied Statistics*, Vol. 47, pp. 506-523. https://doi.org/10.1080/02664763.2019.1645098

Lambert, D. (1992). Zero-Inflated poisson regression with an application to defects in manufacturing. *Technometrics*, Vol. 34, pp. 1-14. https://doi.org/10.2307/1269547

Mahmood, T., & Xie, M. (2019). Models and monitoring of zero-inflated processes: The past and current trends. *Quality and Reliability Engineering International*, Vol. 35, pp. 2540-2557. https://doi.org/10.1002/qre.2547

Maiti, S. S., Saha, M., & Nanda, A. K. (2010). On generalising process capability indices. *Journal of Quality Technology and Quantitative* Management, Vol. 7, pp. 279-300. https://doi.org/10.1080/16843703.2010.11673233

Nagata, Y., & Nagahata, H. (1994). Approximation formulas for the confidence intervals of process capability indices. *Okayama Economic Review*, Vol. 25, pp. 301-314.

Pal, S., & Gauri, S. K. (2020[a]). Measuring capability of a binomial process. *International Journal of Engineering, Science and Technology*, Vol. 12, pp. 25-37. https://doi.org/10.4314/ijest.v12i1.3

Pal, S., & Gauri, S. K. (2020[b]). Measuring capability of a Poisson process: Relative goodness of the estimates by different approaches. *International Journal of Engineering, Science and Technology*, Vol. 12, pp. 1-13. https://doi.org/10.4314/ijest.v12i4.1

Pal, S., & Gauri, S. K. (2021). Measuring capabilities of zero-inflated processes. *International Journal of Engineering, Science and Technology*, Vol. 13, No. 3, pp. 37-48. https://doi.org/10.4314/ijest.v13i3.4

Patil, M. K., & Shirke, D. T. (2012). Process capability index for zero-inflated poisson process. *International Journal of Mathematical Sciences & Engineering Applications*, Vol. 6, pp. 87-97.

Pearn, W. L., & Chen, K. S. (1995). Estimating process capability indices for non-normal Pearsonian populations. *Quality and Reliability Engineering International*, Vol. 11, pp. 389-391. https://doi.org/10.1002/qre.4680110510

Perakis, M., & Xekalaki, E. (2005). A process capability index for discrete processes. *Journal of Statistical Computation and Simulation*, Vol. 75, pp. 175-187. https://doi.org/10.1080/00949650410001687244

Phang, Y. N., & Loh, E. F. (2013). Zero inflated models for overdispersed count data. *International Journal of Health and Medical Engineering,* Vol. 7, pp. 1331-1333.

Polhemus N. W. (2018). *Process Capability Analysis: Estimating Quality*. Boca Ration, Florida: CRC press, Taylor & Francis Group.

Rakitzis, A. C., Maravelakis, P. E., & Castagliola, P. (2017). CUSUM control charts for the monitoring of zero-inflated binomial processes. *Quality and Reliability Engineering International,* Vol. 32, pp. 465-483. https://doi.org/10.1002/qre.1764

Sim, C. H, & Lim, M. H. (2008). Attribute charts for zero-inflated processes. *Communications in Statistics – Simulations and Computation,* Vol. 37, pp. 1440-1452. https://doi.org/10.1080/03610910801983145

Tian, Wz., Liu, Tt. & Yang, Yt. (2022). Zero-inflated non-central negative binomial distribution. *Applied Mathematics-A Journal of Chinese Universities*, Vol. 37, pp. 187–198. https://doi.org/10.1007/s11766-022-4070-0

Tian, W., You, H., Zhang, C., Kang, S., Jia, X. & Chien, W.K. (2019). Statistical process control for monitoring the particles with excess zero counts in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, Vol. 32, pp. 93-103. 10.1109/TSM.2018.2882862

Wagh, Y. S., & Kamalja, K. K. (2018). Zero-inflated models and estimation in zero-inflated Poisson distribution. *Communications in Statistics-Simulations and Computation*, Vol. 47, pp. 2248-2265. https://doi.org/10.1080/03610918.2017.1341526

Workie, M. S., & Azene, A. G. (2021). Bayesian zero-inflated regression model with application to under-five child mortality. *Journal of Big Data*, Vol. 8, pp. 1-24. https://doi.org/10.1186/s40537-020-00389-4

Xie, M., & Goh, T. (1993). Spc of a near zero-defect process subject to random shocks. *Quality and Reliability Engineering International*, Vol. 9, pp. 89-93. https://doi.org/10.1002/qre.4680090205

Xie, W., Xie, M., & Goh, T. N. (1995). Control charts for processes subject to random shocks. *Quality and Reliability Engineering International*, Vol. 11, pp. 355-360. https://doi.org/10.1002/qre.4680110506

Yang, J., Xie, M., & Goh, T. N. (2011). Outlier identification and robust parameter estimation in a zero-inflated Poisson model. *Journal of Applied Statistics*, Vol. 38, pp. 421-430. https://doi.org/10.1080/02664760903456426

Yeh, F. B., & Bhattacharya, S. (1998). A robust process capability index. *Communications in Statistics - Simulation and Computation*, Vol. 27, pp. 565-589. https://doi.org/10.1080/03610919808813495

Zamzuri, Z. (2015). An alternative method for fitting a zero inflated negative binomial distribution. *Global Journal of Pure and Applied Mathematics*, Vol. 11, pp. 2419-2425.

Zhang, Q., & Yi, Gy. (2022). Zero-inflated Poisson models with measurement error in the response. *Biometrics*, Mar 9. https://doi.org/10.1111/biom.13657.

Zhao, Y., Lee, A. H., Burke, V., & Yau, K. W. (2009). Testing for zero-inflation in count series: application to occupational health. *Journal of Applied Statistics*, Vol. 36, No. 12, pp. 1353-1359. https://doi.org/10.1080/02664760802653586

Biographical notes

Dr. Surajit Pal is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Chennai Centre, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization. He has published about forty five papers in different National and International journals.

Dr. Susanta Kumar Gauri is a Faculty Member in the Statistical Quality Control and Operations Research Unit of the Indian Statistical Institute, Kolkata, India. His fields of interest are quality engineering, process optimization, statistical quality control and multiple response optimization. He has published about sixty papers in different National and International journals.