applied sciences

MDPI

Article

Enhancing Control: Unveiling the Performance of Poisson EWMA Charts through Simulation with Poisson Mixture DATA

Nuşin Uncu [1,*] and Melik Koyuncu [2]

[1] Department of Industrial Engineering, Adana Alparslan Turkes Science and Technology University, Adana 01250, Turkey
[2] Department of Industrial Engineering, Cukurova University, Adana 01330, Turkey; mkoyuncu@cu.edu.tr
* Correspondence: nuncu@atu.edu.tr

Abstract: Poisson-Exponentially Weighted Moving Average (PEWMA) charts are one of the most frequently used control charts for monitoring count data. But as real-world data often shows overdispersion—prevalent in manufacturing, health care, economics, and marketing—the standard Poisson distribution falls short. One of the ways to tackle overdispersion is to use Poisson mixture distributions. Our study examines Average Run Length (ARL) performance in the presence of Poisson mixture distribution in the PEWMA control charts. Through meticulously designed experiments, we explore different control parameter combinations and employ simulation to evaluate the process. Our graphs illustrate the performance of the PEWMA control chart, offering desired in-control ARL across parameter combinations. Finally, the performance of the PEWMA control chart is presented for the real process data of fastener production.

Keywords: control charts; PEWMA; poisson mixture distribution; simulation

check for updates

Citation: Uncu, N.; Koyuncu, M. Enhancing Control: Unveiling the Performance of Poisson EWMA Charts through Simulation with Poisson Mixture DATA. *Appl. Sci.* **2023**, *13*, 11160. https://doi.org/10.3390/app132011160

Academic Editor: Jose Machado

Received: 30 August 2023
Revised: 27 September 2023
Accepted: 7 October 2023
Published: 11 October 2023

1. Introduction

Control charts are one of the most important Statistical Process Control (SPC) techniques used to diagnose and prevent problems in production and service processes such as production, maintenance, healthcare [1–5]. Control charts allow users to quickly identify assignable causes and implement necessary solutions. Various process control charts have been improved to effectively monitor process measurement data. Shewhart control charts are very common charts that are used in SPC. The scope of SPC is a two-phase application: Phase I (offline analysis), which is called the retrospective phase, and Phase II (online monitoring), which is called the prospective or monitoring phase. In the Phase I application, process data is analyzed, and the process is stabilized by detecting variability and taking actions, if any, for assignable causes. In this situation, where the process is under control, Phase I analysis is completed, and Phase II, where the process is monitored online, begins. While Shewhart control charts are effective during Phase I, they are less sensitive to small and medium-sized process shifts in Phase II. The cumulative sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts are superior alternatives to the Shewhart control chart for Phase II when process monitoring [6]. Although the performances of CUSUM and EWMA are closely aligned, EWMA stands out as the more user-friendly and straightforward option in terms of setup and usage. EWMA control chart was introduced by Roberts et al. [7] and is very effective in detecting small process shifts. There have been several theoretical studies of the Average Run Length (ARL) properties of the EWMA control charts. These studies provide ARL tables and graphs encompassing a variety of control chart parameter values. Borror et al. [8] examined the ARL performance of the EWMA control charts for the case of nonnormal distributions. The EWMA control limits were revised for Poisson counts [9]. Poisson EWMA (PEWMA) control charts have a wide range of applications as they can be used to monitor processes in both production and service sectors. A good parameter estimation also positively affects the performance of the control charts; otherwise, there may be problems such as the inability to detect variability promptly and an increase in the false alarm rate. Since historical data in PEWMA also control statistics, the points on the chart are not independent of each other. Therefore, the distribution of Run Length (RL) values is also non-geometric. For this reason, the ARL performance calculations of PEWMA control charts are more difficult than other control charts. In several studies, the RL distributions and ARL performances are calculated by simulation, integral equations, and Markov Chain Approach [10–17]. A new PEWMA chart is introduced by Sukparungsee [18] improved based on square transformation of Poisson data. In considering PEWMA as ARL-biased, Morais and Konth [19] used the Markov chain to determine ARL accurately and compared some competing results of previous studies. Although the Poisson distribution is frequently used to describe count data, it assumes that variance is equal to the mean (equi dispersion), which could lead to incorrect inferences in the case of overdispersion [20]. One of the most common methods to cope with different levels of overdispersion is to use Poisson mixture distributions [21]. Several various control charts have been proposed for different types of Poisson distributions in ref. [22] and Poisson mixture distributions in ref. [21].

This study involves a comparative analysis of the ARL performance of the PEWMA control chart. This analysis encompasses diverse parameter combinations within the context of Poisson mixture distributions. Simulation methodology is applied to generate data and facilitate the charting process.

The rest of the paper is organized as follows. In Section 2, we present the charting methodology. Section 3 delves into the examination of experimental design for different control chart parameters. The numerical results alongside the performance graphs are illustrated in Section 4. In Section 5, we provide a comprehensive discussion by comparing the findings of our study with the existing literature. Finally, the conclusion is given in Section 6.

2. Materials and Methods

The performance of control charts depends on their ability to rapidly and reliably detect an out-of-control situation as soon as it occurs. In this study, simulation is employed to illustrate the performance of PEWMA charts in the presence of a Poisson mixture distribution. First, we generated datasets that conform to the Poisson mixture distribution. Subsequently, these datasets were used as inputs to evaluate the ARL performance of the PEWMA control chart.

2.1. Ewma Control Charts

The EWMA control charts, first proposed by Roberts et al. [7], are especially preferred for single observations and Phase II applications, but they can also be used for rational subsets. The EWMA statistics (z_i) are defined in Equation (1).

$$z_i = \lambda x_i + (1-\lambda) z_{i-1} \tag{1}$$

where x_i is the value of ith observation for $i(i = 1, \ldots, n)$ and $0 < \lambda \leq 1$ is smoothing constant. To calculate value z_1, we need to provide the starting value z_0 which is equal to the process target value μ given in Equation (2).

$$z_0 = \mu. \tag{2}$$

Sometimes, the average of preliminary data is used as the starting value of the EWMA, so that $z_0 = \bar{x}$, where $\bar{x}$ is the sample mean. To illustrate that the z_i represents a weighted average of all previous sample means, replacing z_{i-1} by $[\lambda x_{i-1} + (1-\lambda) z_{i-2}]$ then Equation (3) is obtained.

$$z_i = \lambda x_i + \lambda(1-\lambda)[\lambda x_{i-1} + (1-\lambda) z_{i-2}] = \lambda x_i + \lambda(1-\lambda) x_{i-1} + (1-\lambda)^2 z_{i-2} \tag{3}$$

After performing a recursive substitution, z_i is obtained as shown in the following Equation (4).

$$z_i = \lambda \sum_{j=0}^{i=1} (1-\lambda)^j x_{i-j} + (1-\lambda)^i z_0 \tag{4}$$

The weights sum to unity since $\lambda(1-\lambda)^j$ decreases geometrically as j increases. Hence, the sum converges to one given in Equation (5).

$$\lambda \sum_{j=0}^{i=1} (1-\lambda)^j = \lambda \left(\frac{1-(1-\lambda)^i}{1-(1-\lambda)} \right) = 1-(1-\lambda)^i. \tag{5}$$

If the sample data consist of independent random variables x_i with a variance of σ^2, then the variance of z_i is given in Equation (6).

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2-\lambda} \right) [1-(1-\lambda)^{2i}]. \tag{6}$$

Hence, the control chart of EWMA is created by using the z_i values against the sample number i. The upper control limit (UCL) and the lower control limit (LCL) of the EWMA control chart are defined in Equations (7) and (8).

$$UCL = \mu + L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1-(1-\lambda)^{2i}]} \tag{7}$$

$$LCL = \mu - L\sigma \sqrt{\frac{\lambda}{(2-\lambda)} [1-(1-\lambda)^{2i}]} \tag{8}$$

where L is the width of the control limits, and the center line (CL) is μ.

2.2. Poisson EWMA Control Chart

Borror et al. [9] extended the EWMA control limits to PEWMA where observations are Poisson count, then the basic EWMA recursion remains unchanged given in Equations (1) and (2). Nevertheless, there are some issues that need to be addressed regarding the presence of the Poisson distribution. The random variable X, which represents the number of events occurring in a fixed interval, follows a Poisson distribution with a parameter $\alpha > 0$, and its probability mass function is given in Equation (9).

$$f(x) = \frac{\alpha^x e^{-\alpha}}{x!} \quad x = 0, 1, 2, \ldots \tag{9}$$

where α is the mean number of events. it is known that the standard deviation of Poisson distribution is $\sigma = \sqrt{\alpha}$, hence the control limits of the PEWMA chart are written in Equations (10) and (11).

$$UCL = \alpha + A_u \sqrt{\frac{\lambda\alpha}{(2-\lambda)} [1-(1-\lambda)^{2i}]} \tag{10}$$

$$LCL = \alpha - A_L \sqrt{\frac{\lambda\alpha}{(2-\lambda)} [1-(1-\lambda)^{2i}]} \tag{11}$$

where $L\sigma$ is represented by A and in symmetric cases $A_U = A_L = A$. The center line of the PEWMA chart is α. The chart gives an out-of-control signal when the point is not between the interval of UCL and LCL. The variance of the PEWMA statistic z_i is given in Equation (12).

$$var(z_i) \approx \lambda\alpha/(2-\lambda). \tag{12}$$

To demonstrate the PEWMA control chart, Figure 1 displays a plot of 100 randomly generated observations (n = 100) from a Poisson distribution with a mean of 5. The original data x_i is represented by the cyclic symbol, while the EWMA data z_i is shown with a solid line. The initial value, z_0, is set to 5, and λ is chosen to be 0.2 for the calculations.

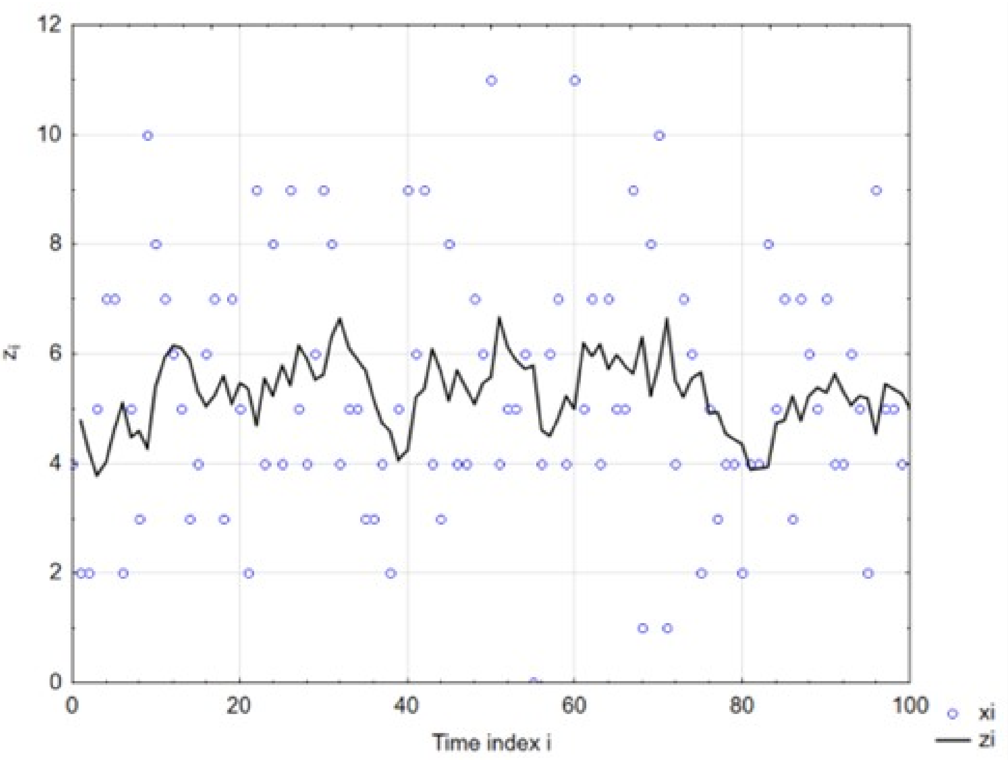

Figure 1. PEWMA of a single simulated realization.

2.3. Mixture Distributions

If a random variable X has a finite mixture distribution, the probability (density) function can be written as in Equation (13).

$$f(x) = \sum_{i=1}^{k} \pi_i f(x_i, \theta_i), \quad x \in R_x. \tag{13}$$

Here, $f_1, \ldots, f_k$ are probability (density) functions on the support set R_x where $x \in R_x$ and θ_i is the paramater vector of the function f_i for $i = 1, \ldots, k$. If we denote the proportion of the distribution f_i as π_i for $i = 1, \ldots, k$ within the finite mixture distribution, the parameter vector Φ of the finite mixture distribution can be expressed as follows [23].

$$\Phi = \begin{bmatrix} k \\ \theta \\ \pi \end{bmatrix}$$

If $f_1, \ldots, f_k$ comes from the same distribution family, the probability (density) function of the finite mixture distribution can be written as in Equation (14) where the sum of proportions is equal to 1 as in Equation (15).

$$f(x; \Phi) = \sum_{i=1}^{k} \pi_i f(x_i, \theta_i), \quad x \in R_x \tag{14}$$

$$\sum_{i=1}^{k} \pi_i = 1, \quad 0 < \pi_i < 1. \tag{15}$$

2.4. Poisson Mixture Distributions

The Poisson mixture distribution comprised of k components is represented similarly in Equation (14) where $\Phi = \begin{bmatrix} k \\ \theta \\ \pi \end{bmatrix}$ is the parameter vector [24]. The probability function of Poisson mixture distribution $f(x;\Phi)$ is given in Equation (16) and the sum of the proportions is equal to 1.

$$f(x;\Phi) = \sum_{i=1}^{k} \pi_i \frac{\alpha_i^x e^{\alpha_i}}{x!} \tag{16}$$

π_i: proportion of the ith Poisson distribution ($i = 1, \ldots, k$)
α_i: mean number of the ith Poisson distribution ($i = 1, \ldots, k$)

The mean of the Poisson mixture distribution (α_m) comprised of k components can be written as in Equation (17).

$$\alpha_m = \sum_{i=1}^{k} \pi_i \alpha_i, \;\; 0 < \pi_i < 1 \tag{17}$$

The variance of the Poisson mixture distribution (σ_m^2) can be calculated by using Equations (18)–(20).

$$var(X) = E[X^2] - (E[X])^2 \tag{18}$$

$$var(X) = E[X^2|\pi_i, \alpha_i] - (E[X|\pi_i, \alpha_i])^2 \tag{19}$$

$$\sigma_m^2 = \sum_{i=1}^{k} \pi_i(\sigma_i^2 + \alpha_i^2) - \alpha_m^2. \tag{20}$$

By capitalizing on the equality between the variance and mean of a Poisson distribution, we make a substitution of α_i instead of σ_i^2, leading to the derivation of the formula presented in Equation (21).

$$\sigma_m^2 = \sum_{i=1}^{k} \pi_i(\alpha_i + \alpha_i^2) - \alpha_m^2. \tag{21}$$

It is necessary to estimate the number of components k, and the parameters of the distribution, $\pi_1, \pi_2, \ldots \pi_k$ and $\alpha_1, \alpha_2, \ldots, \alpha_k$ to fit with a mixture distribution. k is determined using model selection criteria that are Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood (LL). The likelihood function and log-likelihood function are given in Equation (22) and in Equation (23), respectively.

$$L(\Psi) = \prod_{i=1}^{n} f(x_i, \Psi) = \prod_{i=1}^{n} [\sum_{j=1}^{k} \pi_j f(x_i, \theta_j)] \tag{22}$$

$$lnL(\Psi) = \sum_{i=1}^{n} lnf(x_i, \Psi) = \sum_{i=1}^{n} ln[\sum_{j=1}^{k} \pi_j f(x_i, \theta_j)] \tag{23}$$

Consider $L(\Psi)$ as the maximum value of the likelihood function for a mixed distribution model with the estimated number of components k. Subsequently, the AIC and BIC are computed using the formulas $2k - 2lnL(\Psi)$ and $-ln(n)k - 2lnL(\Psi)$, respectively. When deciding the number of components using model selection criteria, the value of k is selected based on the criterion of having the lowest AIC and BIC values, while simultaneously taking into account the highest LL value.

2.5. Expectation Maximization Algorithm

The Expectation Maximization (EM) algorithm, initially introduced by Dempster et al. [25], was devised for the purpose of estimating parameters within mixture distribution models. The implementation of the EM algorithm relied on the complete-data likelihood function as depicted in Equation (24), where the log-likelihood $lnL(\Psi)$ is enhanced by the inclusion of randomly assigned 0–1 binary variables $z_{i,k}$, which identifies the ith observation as coming from the kth component.

$$lnL(\Psi, Z) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{i,k} \ln(\pi_k f(x_i, \theta_k)) \tag{24}$$

Starting from an initial $\Psi^{(0)}$, the EM algorithm iterates between an Expectation (E)-step and Maximization (M)-step.

- E-step
  The posterior probability of data point i belonging to component k at iteration r is denoted as $z_{i,k}^{(r)}$ is obtained by using Bayes' theorem and given in Equation (25).

$$z_{i,k}^{(r)} = \frac{\pi_{i,k}^{(r)} f(x_i, \theta_k^{(r)})}{\sum_{j=1}^{K} \pi_{i,j}^{(r)} f(x_i, \theta_j^{(r)})} \tag{25}$$

  where $\pi_{i,k}^{(r)}$ is the mixing proportion of component k at iteration r, and $f(x_i, \theta_k(r))$ is the probability density function of component k evaluated at data point x_i using parameters $\theta_k^{(r)}$.
- M-step
  Update the parameter estimates to maximize the expected $lnL(\Psi, Z)$ based on the computed posterior probabilities in the E-step. These steps are repeated until the expected log-likelihood converges to the maximum.

3. Performance Evaluation and Experimental Design

A dataset consisting of a mixture distribution is called a pure mixture distribution when it comes from the same distribution family, and it is referred to as a convolute mixture distribution when it comes from different distribution families. In evaluating the performance of the PEWMA control chart, we assumed that the data came from a Poisson mixture distribution within the same distribution family. We assumed that observations from this distribution are independent of each other and that the mixture proportions and distribution parameters remain constant over time.

There are several measures available to assess the performance of a chart, with one commonly used metric being the ARL to evaluate how well the control chart responds to a signal that is out of control. The formulation of the ARL is given in Equation (26)

$$ARL = \frac{\sum_{i=1}^{N} RL_n}{N}. \tag{26}$$

Various approaches exist for estimating the run-length characteristics, including Monte Carlo simulation, Markov chain analysis, and integral equations [26]. In this study, simulation is employed to estimate the run-length characteristics of the PEWMA control chart due to its flexibility and efficiency. The steps for generating simulation results can be explained as follows:

- Step 1: Set the simulation parameters, including the replication length (or the number of sample data) and the number of replications.
- Step 2: Select the parameters for the PEWMA control chart, including λ, σ, and A.

- Step 3: Define the Poisson mixture parameters (α values) to represent desired shifts in the mean, based on π values ranging from 0 to 1.
- Step 4: Run simulation.
- Step 5: Calculate the PEWMA statistics z_i and determine the control limits UCL and LCL.
- Step 6: Check if z_i falls within the control limits. If $z_i \leq$LCL or $z_i \geq$ UCL, record the sample index number as RL (Run Length).
- Step 7: Repeat Steps 4 and 6 for each replication.
- Step 8: Calculate the ARL by computing the average of all recorded RL values.
- Step 9: Return to Step 2 and run the simulation for different PEWMA control chart parameters.

In order to detect shifts in the process, the design parameters of the EWMA chart are set for multiple of σ in terms of control limits (L) and various λ values. There have been several theoretical studies of the ARL properties of the control chart such as [10,11,17]. These studies provide ARL of tables and graphs for a range of values of λ and L. In a PEWMA control chart, the ARL exhibits variability based on the sample mean and the control chart's parameters, namely λ and A. Furthermore, alongside these parameters, the ARL performance of the system is influenced by the proportions π_i of the Poisson mixture distributions. Different combinations of these parameters are designed before conducting experiments to ascertain the desired ARL performance. Table 1 displays the results of 44 experiments conducted across four main samples. These samples are generated based on a Poisson mixture distribution with the given α_i values, along with corresponding proportions π_i for each sample. In this study, each sample within the dataset consists of 1000 data points, generated using the software Arena version 14.00, and follows the characteristics of a Poisson mixture distribution. A simulation model was created to identify out-of-control points. This model utilizes the PEWMA control limits, as specified in Equations (10) and (11). The simulation is run for the combinations of A = 2.50, 2.55, 2.60, 2.65, 2.70, 2.75, 2.80, 2.85, 2.90, 2.95, 3.00, 3.05, 3.10, 3.15, 3.20, 3.25, 3.30, 3.35, 3.40, 3.45, 3.50 and λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.75. So, the total number of simulation experiments is 5544. The replication number for each simulation model is 100. For instance, consider the information provided in Table 1. In the case of Sample 2, a dataset of 1000 data points is generated from a Poisson mixture distribution characterized by means α_1 = 5 and α_2 = 5.5, along with proportions of 0.90 and 0.10, respectively. In the initial simulation iteration, ARL performance is evaluated by setting the parameters λ = 0.1 and A = 2.5. This process is then repeated for every combination of λ and A to conduct the experiments.

Table 1. Experimental design for samples with different mean and mixture proportions.

Sample	α_1 = 5 π_1	α_2 = 5.5 π_2	Sample	α_1 = 5 π_1	α_2 = 6 π_2	Sample	α_1 = 5 π_1	α_2 = 6.5 π_2	Sample	α_1 = 5 π_1	α_2 = 7 π_2
1	1.00	0.00	12	1.00	0.00	23	1.00	0.00	34	1.00	0.00
2	0.90	0.10	13	0.90	0.10	24	0.90	0.10	35	0.90	0.10
3	0.80	0.20	14	0.80	0.20	25	0.80	0.20	36	0.80	0.20
4	0.70	0.30	15	0.70	0.30	26	0.70	0.30	37	0.70	0.30
5	0.60	0.40	16	0.60	0.40	27	0.60	0.40	38	0.60	0.40
6	0.50	0.50	17	0.50	0.50	28	0.50	0.50	39	0.50	0.50
7	0.40	0.60	18	0.40	0.60	29	0.40	0.60	40	0.40	0.60
8	0.30	0.70	19	0.30	0.70	30	0.30	0.70	41	0.30	0.70
9	0.20	0.80	20	0.20	0.80	31	0.20	0.80	42	0.20	0.80
10	0.10	0.90	21	0.10	0.90	32	0.10	0.90	43	0.10	0.90
11	0.00	1.00	22	0.00	1.00	33	1.00	0.00	44	0.00	0.00

Analysis of Experimental Results

To observe the ARL performance of PEWMA control schemes, 1000 data are generated from a Poisson mixture distribution with parameter vector $\Phi = \begin{bmatrix} k = 2 \\ \theta = 5, 5.5 \\ \pi = (0.90, 0.10) \end{bmatrix}$. The ARL values are plotted against the parameter combinations of A = 2.50 to 3.50 and λ = 0.1 to 0.75 in Figure 2. The ARL is detected earlier with respect to large to small values of λ. For α_1 = 5, α_2 = 5.5 and a small shift in mean as 0.02σ, ARL performances are illustrated for λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.75. The ARL is detected earlier with respect to larger to smaller values of λ. However, for $A > 3.50$, the ARL performance of the PEWMA chart seems to become similar for different values of λ.

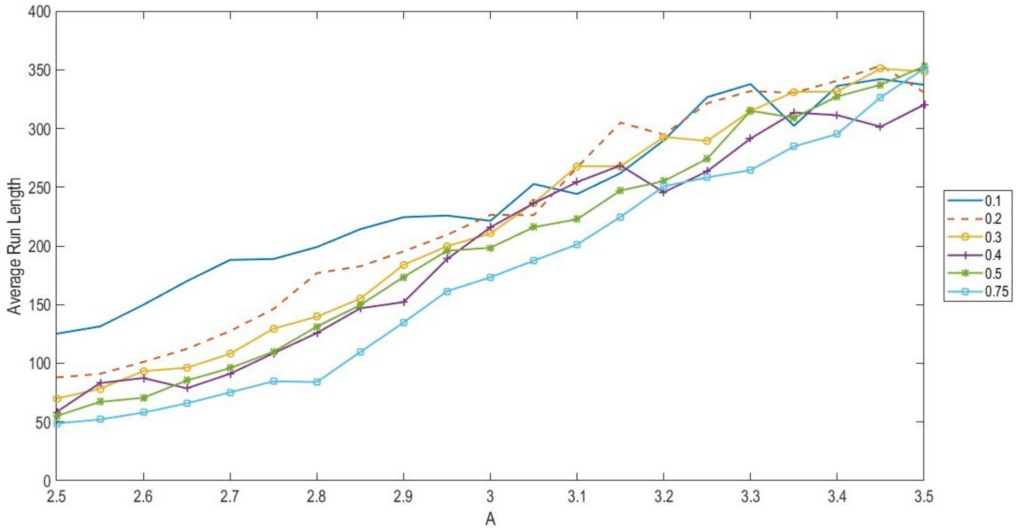

Figure 2. ARLs for various values of A and a series of λ with α_1 = 5, α_2 = 5.5, shift in mean = 0.02σ.

In Figure 3, for α_1 = 5, α_2 = 7 and a large shift in mean as 0.76σ, ARL performances are illustrated for λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.75. It is observed that out-of-control situations are quickly detected when the shift in the mean is large and λ is small.

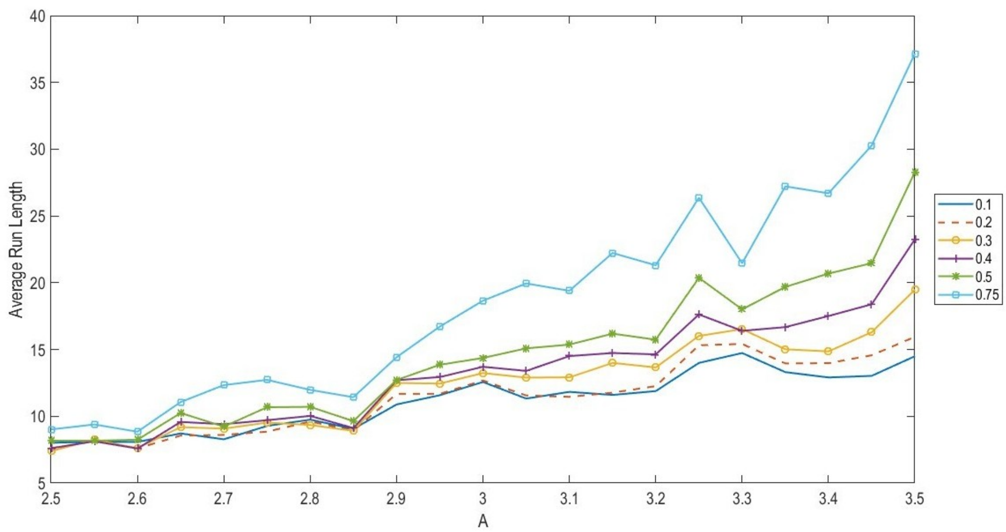

Figure 3. ARLs for various values of A and a series of λ with α_1 = 5, α_2 = 7, shift in mean = 0.76σ.

The performance of the ARL for a relatively large value such as 0.76σ of the mean shift, for different values of λ and for the A values determined between 2.70–3.20 is given in Figure 4. For given values of A, ARL is detected earlier for small values of $0.1 < \lambda < 0.3$, however for $\lambda > 0.3$ ARL performance declines generally for corresponding values of A.

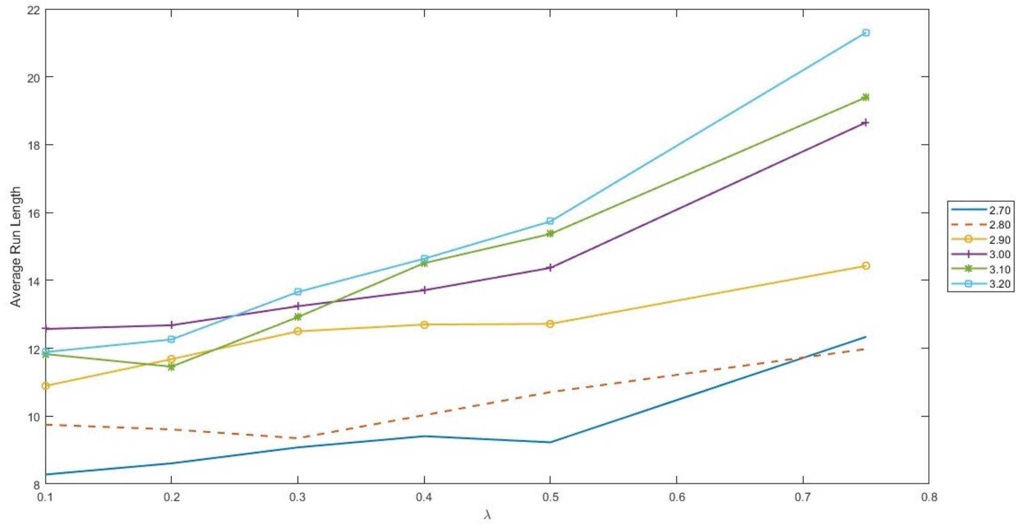

Figure 4. ARLs for various values of λ for a series of A with $\alpha_1 = 5$, $\alpha_2 = 7$, shift in mean = 0.76σ.

A 3D graphic is presented in Figure 5 to provide another view of how the ARL performance is influenced by the parameters λ, A and mean shifts. In Figure 5, the performance of ARL is given depending on λ and shift in mean for $A = 3.00$. It is seen that for larger shifts in mean greater than 0.10, λ does not affect the performance of ARL. However, the values of $\lambda > 0.5$ can be selected to detect out-of-control in the case of small shifts in the mean.

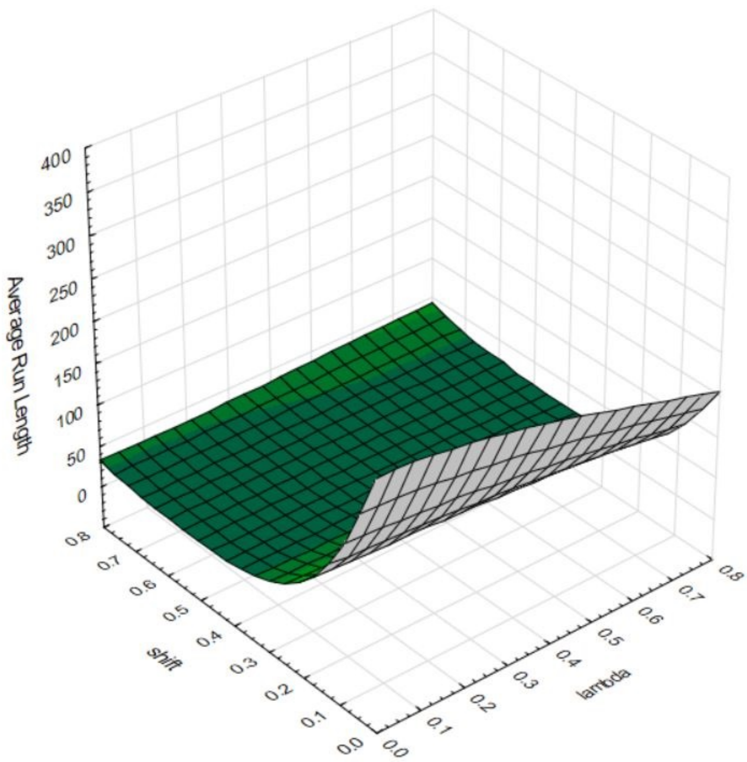

Figure 5. ARLs for various values of λ and shifts in mean (multiple of σ) with $A = 3.00$ and $\alpha_1 = 5$, $\alpha_2 = 7$.

The ARL performance of PEWMA chart is illustrated in Figure 6 where the parameters are set to $\lambda = 0.1$, $\alpha_1 = 5$, $\alpha_2 = 5.5$ with various shifts in mean and $2.50 < A < 3.50$. The performance of ARL varies almost linearly for different levels of shifts in mean. However, for values of $A > 3.00$, the performance exhibits non-stationary behavior.

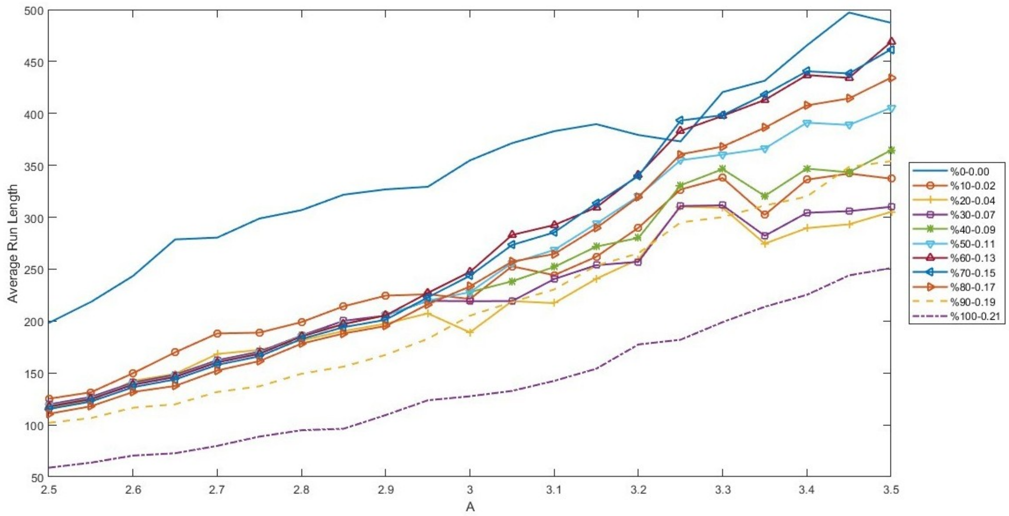

Figure 6. ARLs for various values of A and mixture proportions (π)-shift in mean with $\lambda = 0.1$, $\alpha_1 = 5$, $\alpha_2 = 5.5$.

Figure 7 demonstrates the changes in performance resulting from the shifts in mean where the parameters are set to $\lambda = 0.1$, $\alpha_1 = 5$, $\alpha_2 = 7$. In the case of larger shifts in mean, the same performance is observed for all values of A choosing $\lambda = 0.1$ when we compared the pure Poisson distribution that has no shifts in mean.

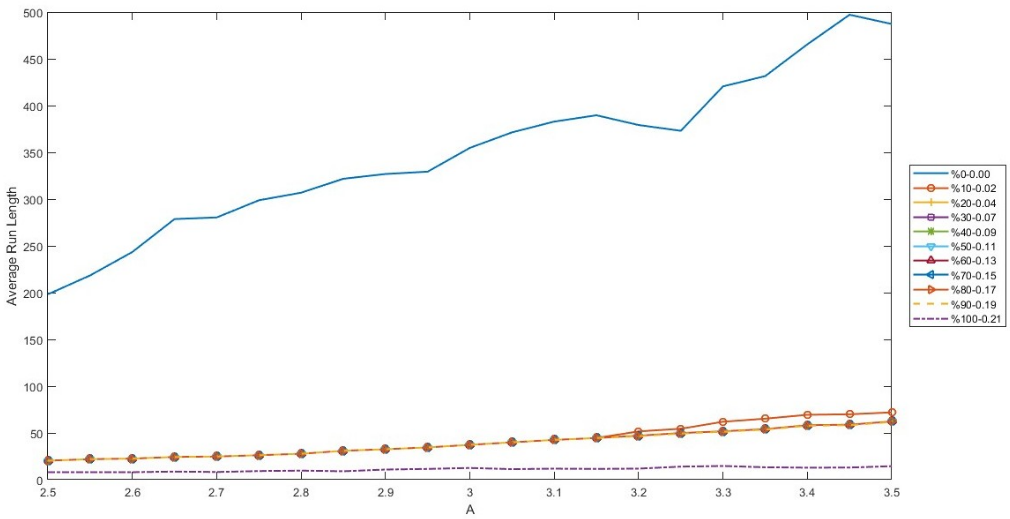

Figure 7. ARLs for various values of A and mixture proportions (π)-shift in mean with $\lambda = 0.1$, $\alpha_1 = 5$, $\alpha_2 = 7$.

4. Application in Fastener Quality Control

In order to demonstrate the applicability of the Poisson mixture distribution in the fields of statistical quality control and reliability, the quality control processes from the automotive supplier sector were taken into account. The company produces a diverse range of fastener types. Among these various fasteners, we specifically focused on the data related to large flange rivets and square nuts because our tests indicated that they fit the Poisson mixture distribution. A square nut is a nut with four sides. They are commonly used alongside a bolt to join two or more objects together. A large flange rivet is ideal for high-pressure applications that require a seal to prevent the ingress of liquids, solids, or moisture. The pictures of square nut and large flange rivet are given in Figure 8.

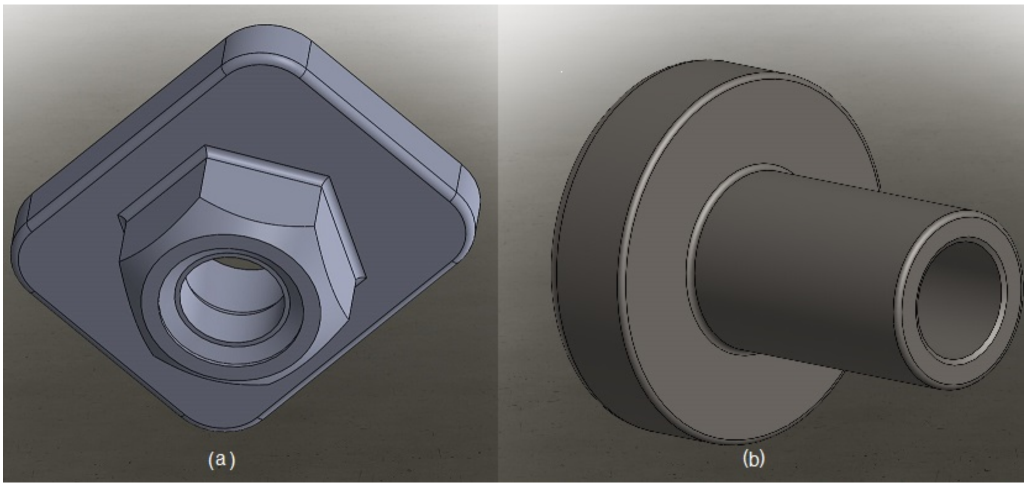

Figure 8. Products: (**a**) Square nut; (**b**) Large flange rivet.

As part of its quality control policy, the company conducts a count of defective products within 500 randomly sampled units from the production process. The defects recorded during quality control of the products include cracks, wrinkles, shear burst, imperfections, damage, and coating flaws. In Figure 9, some examples of the defects are shown for each product.

(a) (b)

Figure 9. Defects: (**a**) Shear on the square nut (**b**) Crack on the large flange rivet.

First, we applied Kolmogorov–Smirnov and chi-square tests to determine whether these data fit a discrete distribution such as Poisson, binomial and geometric distribution.

This result led us to consider that these data might come from a mixture distribution. We used the package Flexmix in R version 4.3.0 to determine whether these data fit the Poisson mixture distribution. This package calculates the AIC, BIC and LL values using the Equations (22) and (23) to determine the number of components of the Poisson mixture distribution. To determine the number of components, run parameters are set at 10 replications and 1000 iterations per replication. In the replication where AIC and BIC are the smallest and LL is the largest, the k value is determined as the optimal number of components. Accordingly, the k values for the square nut and large flange rivet are selected as 4 and 3, respectively. The results are given in Tables 2 and 3 for the square nut and large flange rivet, respectively.

Table 2. Computed values for selecting the number of components for square nut data.

Replication	Iter	k	AIC	BIC	LL
1	2	1	8059.960	8063.398	−4028.980
2	20	2	4236.763	4247.077	−2115.382
3	43	3	3123.799	3140.990	−1556.900
4	64	3	3123.799	3140.990	−1556.900
5	135	4	2905.724	2929.791	−1445.862
6	111	5	2905.724	2940.667	−1445.862
7	152	4	2905.724	2929.791	−1445.862
8	166	4	2905.724	2929.791	−1445.862
9	149	4	2905.724	2929.791	−1445.862
10	175	4	2905.724	2929.791	−1445.862

Table 3. Computed values for selecting the number of components for large flange rivet data.

Replication	Iter	k	AIC	BIC	LL
1	2	1	10,346.030	10,349.747	−5172.015
2	12	2	4435.795	4446.946	−2214.897
3	7	3	1790.279	1808.865	−890.139
4	10,000	4	1793.368	1819.405	−889.693
5	10,000	4	1793.368	1819.405	−889.693
6	10,000	4	1793.368	1819.405	−889.693
7	10,000	4	1793.368	1819.405	−889.693
8	10,000	5	1793.368	1830.836	−889.693
9	10,000	5	1793.368	1830.839	−889.693
10	10,000	5	1797.383	1830.836	−889.691

Once the number of Poisson mixture components was determined for each set of product sample data, we used the Flexmix package in the R programming environment to compute the proportions and parameters for each Poisson mixture distribution. The proportions and parameters corresponding to each product sample data are given in Table 4.

Table 4. Poisson mixture distribution parameters for each product data.

Product	Proportions				Parameters			
	π_1	π_2	π_3	π_4	α_1	α_2	α_3	α_4
Square Nut	0.14	0.39	0.09	0.38	11.165	60.348	163.485	29.883
Large Flange Rivet	0.06	0.53	0.41	-	115.11	0.697	21.782	-

To verify whether the fitted Poisson mixture distributions represent the actual data distribution visually, the histograms of real data and generated data are illustrated in Figures 10 and 11. Next, we employed the Mann–Whitney U test to determine whether the generated data accurately represents the real data. The test hypotheses are formulated as follows:

Hypothesis 0 (H0): *The real data and generated data come from the same distribution*.

Hypothesis 1 (H1): *The real data and generated data do not come from the same distribution*.

The p-values are 0.51 and 0.15 for square nut and large flange rivet, respectively. Therefore, p-values greater than 0.10 indicate that the data generated from Poisson mixture distributions significantly represent the real data.

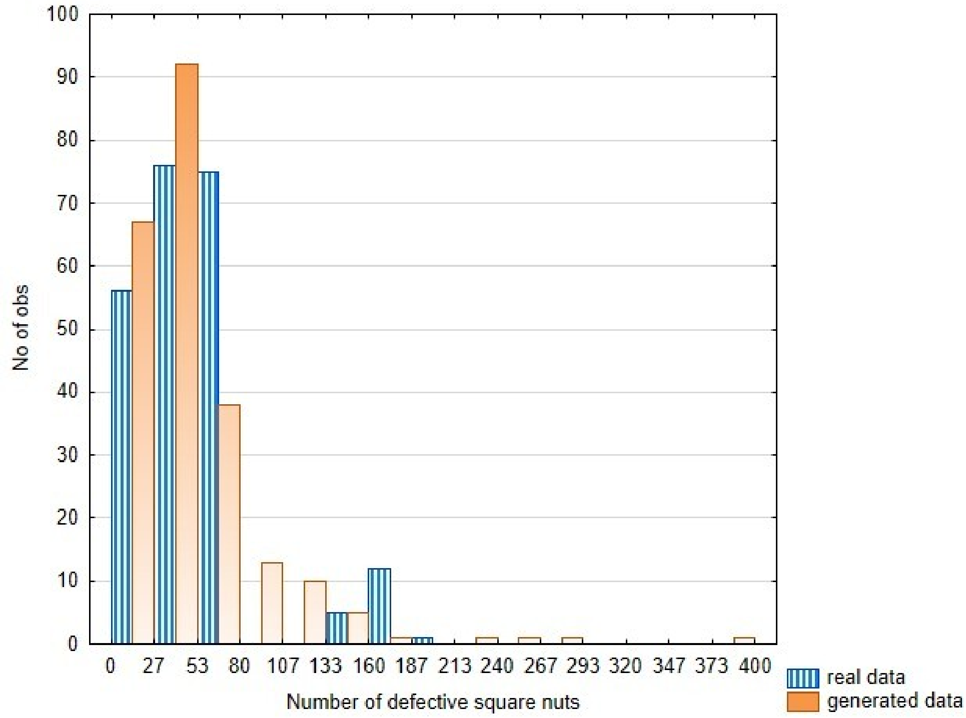

Figure 10. Comparison of the real and generated data distribution for square nut.

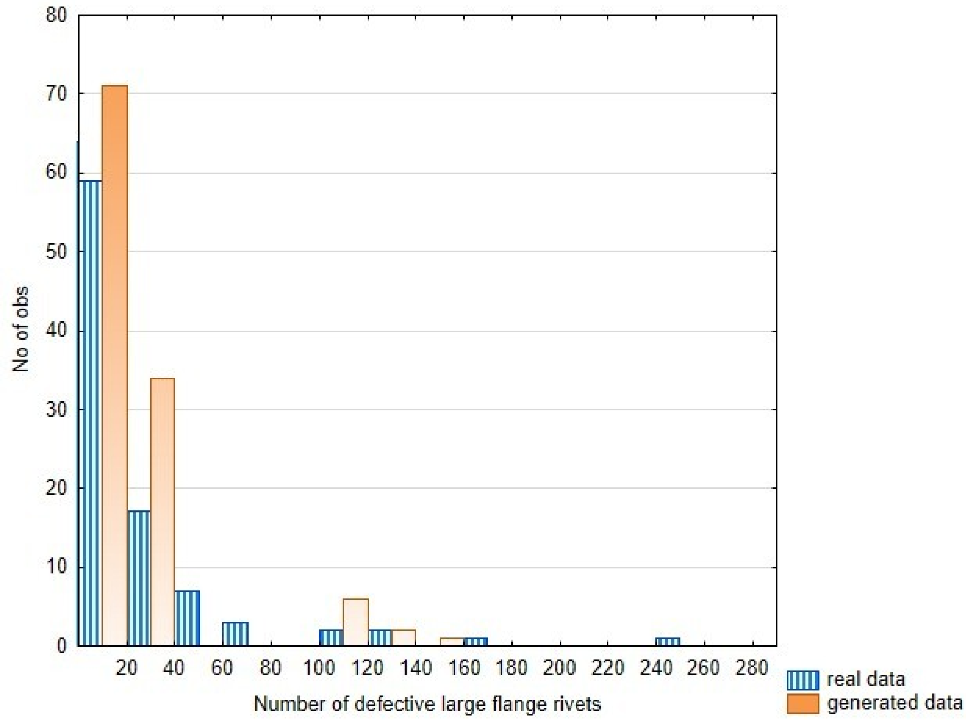

Figure 11. Comparison of the real and generated data distribution for large flange rivet.

It is seen that the PEWMA control charts, in which we have drawn square nut and large flange rivet data shown in Figures 12 and 13, respectively, are out of control very quickly while the parameters $L = 2.7$ and $\lambda = 0.2$. The charts seem very sensitive to Poisson mixture data that have more than two proportions with highly shifted process means. The reason for this is the violation of the assumption that the data in the PEWMA control chart come from a Poisson distribution, in addition to the presence of overdispersion in the data.

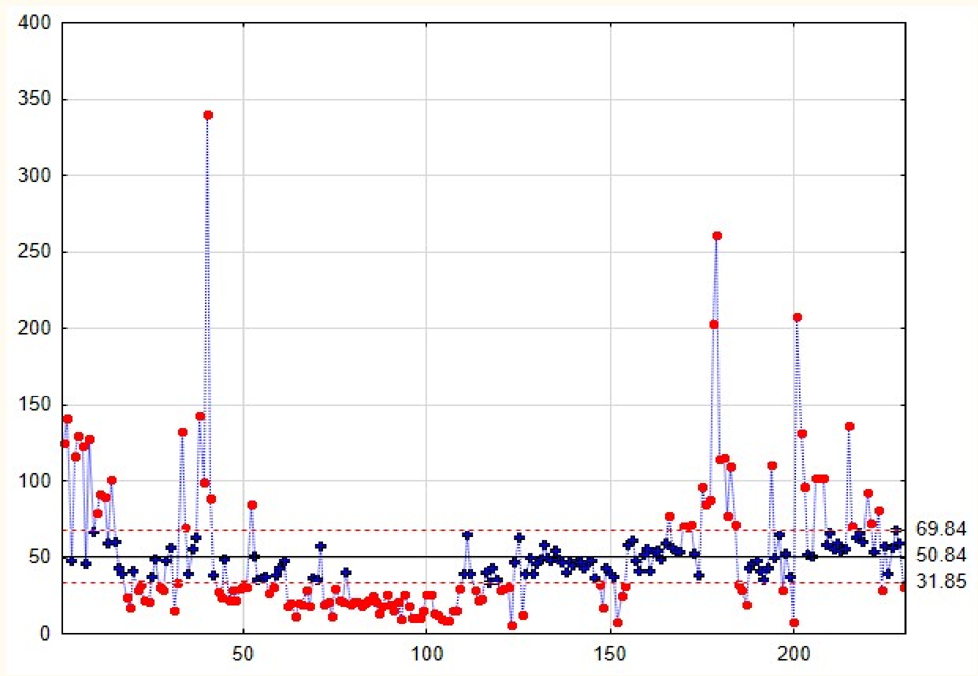

Figure 12. PEWMA control chart for square nut. Red and blue symbols represent out of control and in control points, respectively. Red dashed lines are control limits and bold line is center line.

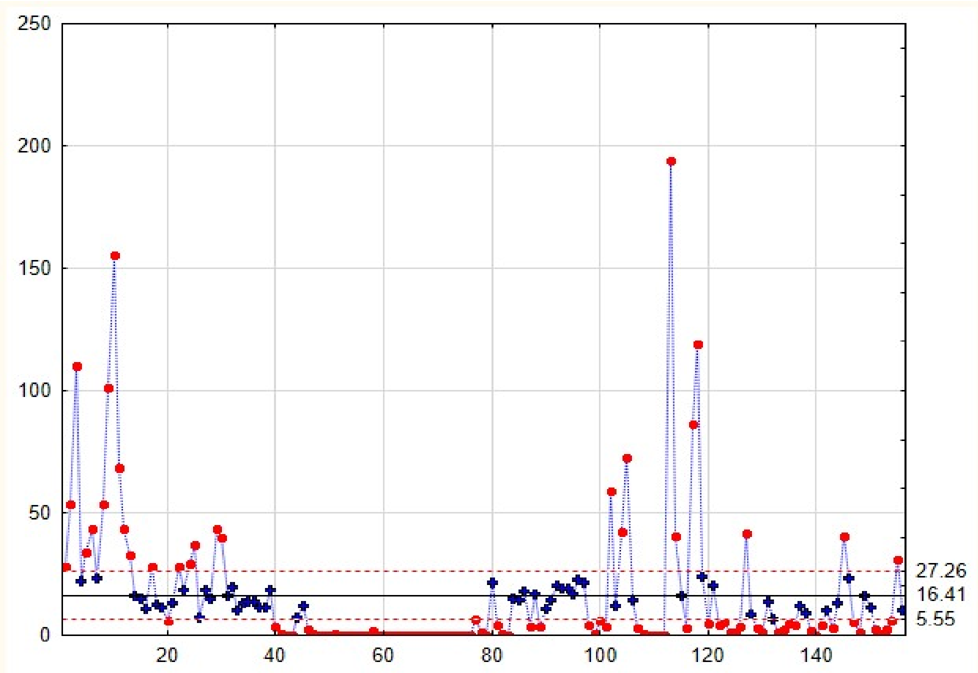

Figure 13. PEWMA control chart for large flange rivet. Red and blue symbols represent out of control and in control points, respectively. Red dashed lines are control limits and bold line is center line.

Reliability

Since reliability focuses on the overall performance and defective probabilities of products or systems, a strong quality control process contributes to achieving the desired reliability of a product. For square nuts and large flange rivets with batch sizes of 500, we can define the defective product numbers as in Equation (16) by using the determined number of components (k), proportions (π) and parameters (α) of each Poisson mixture distribution. By utilizing these probability density functions, we can calculate the probabilities of having the number of defective products greater or less than a certain value for both products within a batch.

5. Discussion

In the literature, there are notable instances of studies delving into the enhancement and progression of the PEWMA control chart, which was originally introduced by Borror et al. [9]. One of the studies offering suggestions for obtaining a more effective control chart was found in ref. [18]. PEWMA has argued that it is more effective by applying a square root transformation to control chart data.

Overdispersion in process data has been handled by categorizing a certain data rate as "noisy data" in SPC. This approach was adopted as a precautionary measure to mitigate the effects of overdispersion. In view of this concept, Jesus et al. [21] suggested in their study that handling overdispersion with the Poisson mixture distribution would eliminate the problem and increase the ARL performance. However, the charts they deal with are not EWMA charts as similar to Boaventura et al. [27], but for the c and u charts developed for attribute data. As far as we have seen in the literature, our work is a novel approach for PEWMA control charts, taking into account the Poisson mixture distribution to deal with overdispersed data.

The outcomes of our simulation experiments, conducted using the generated Poisson mixture data and a range of design parameter combinations, illuminated several significant findings. Notably, we observed that the selection of suitable values for the parameters played a pivotal role in shaping the performance of the PEWMA control charts.

We embarked on an exploration of the performance of PEWMA control charts when confronted with process data fitted to a Poisson mixture distribution. The key objective was to gain insights into how the PEWMA control charts respond under varying conditions of process mean shifts induced by the incorporation of nuisance data. Our investigation aimed to shed light on the optimal selection of design parameters to ensure efficient and effective detection of ARLs across different shifts in mean. To achieve this, we employed a method that involved the generation of nuisance data by mixing data from a Poisson distribution characterized by diverse parameters into a dataset, subsequently resembling another Poisson distribution. This approach endowed the new dataset with a Poisson mixture distribution while introducing noticeable shifts in the process mean. This allowed us to simultaneously investigate the influence of distribution characteristics and mean shifts on the PEWMA control chart performance.

In the simulation experiments conducted with this Poisson mixture data and different combinations of design parameters, we observed that selecting a large value for λ resulted in better performance for almost all values of A when nuisance data caused small shifts in the process mean. Conversely, when nuisance data caused large shifts, selecting a small value for λ shows better performance for almost all values of A. Informative graphical representations and tables have been meticulously crafted to facilitate users in the optimal parameter selection. These images serve as valuable tools to quickly identify ARLs, enabling fast and effective decision making based on average shift magnitudes.

As a result of this study, Table 5 representing the parameter combinations A and λ for some shift levels (multiple of σ) in the mean, taking into account the Poisson mixture distribution, to determine the ARLs in different PEWMA control schemes. This template will be useful for selecting parameters before monitoring the process. For example, λ should be chosen small to detect smaller shifts. However, for larger shifts, it is observed that the ARL values do not differ much from each other when $A < 2.90$ and $\lambda < 0.4$.

Table 5. ARLs for different PEWMA control schemes with Poisson mixture data.

Shift in Mean Multiple of σ	A = 3.50 λ = 0.75	A = 3.20 λ = 0.5	A = 2.90 λ = 0.4	A = 2.85 λ = 0.3	A = 2.70 λ = 0.2	A = 2.50 λ = 0.1
0.00	419.33	360.49	245.72	237.14	223.07	125.27
0.25	247.24	147.58	88.77	77.22	71.30	61.33
0.52	72.04	49.99	32.80	31.29	26.65	20.39
0.76	37.16	15.47	12.7	8.91	8.61	8.02

6. Conclusions

The structure of control charts is based on the assumption that process observations come from a specific distribution, such as the normal distribution. One of the criteria for evaluating their performance is the ARL when the fundamental distributional assumption is violated. In the literature, many studies have assessed the performance of control charts by using different distributions instead of the fundamental one. In the case of the PEWMA control chart, it is assumed that observations follow a Poisson distribution.

In this study, we assessed the performance of the PEWMA control chart in the presence of Poisson mixture data instead of Poisson data. The Poisson mixture data were generated using simulation, with parameters such as sample mean and the proportion of data for each component. Subsequently, the simulation was executed to obtain the ARL while calculating the PEWMA statistics and control limits. The results were analyzed for various parameters of PEWMA control charts, and we provided necessary interpretations to enhance their performance. Additionally, we presented a real-world application for process control in fastener production. We analyzed the data using the EM algorithm to determine the number of components and distribution parameters of the Poisson mixture distribution. Finally, we constructed PEWMA charts with the desired parameters.

The fact that there is more variability in process data, especially in real cases, makes the use of mixture distributions important. In future studies, the exploring efficacy of existing control charts for different mixture distributions will shed light on how to deal with over-dispersed data. This examination will also provide a foundation for the creation of more robust control charts specifically designed to address data characterized by various mixture distributions.

Author Contributions: Conceptualization, M.K. and N.U.; methodology, M.K. and N.U.; software, M.K. and N.U.; validation, M.K. and N.U.; formal analysis, M.K. and N.U.; investigation, M.K. and N.U.; writing—original draft preparation, M.K. and N.U.; writing—review and editing, M.K. and N.U.; visualization, N.U.; supervision, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All datasets are publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EWMA	Exponentially Weighted Moving Average
PEWMA	Poisson-Exponentially Weighted Moving Average
CUSUM	Cumulative Sum
SPC	Statistical Process Control
RL	Run Length
ARL	Average Run Length

References

1. Petrović, S.; Milosavljević, P.; Lozanović Šajić, J. Rapid evaluation of maintenance process using statistical process control and simulation. *Int. J. Simul. Model.* **2018**, *17*, 119–132. [CrossRef] [PubMed]
2. Wan, Q.; Chen, L.; Zhu, M. A reliability-oriented integration model of production control, adaptive quality control policy and maintenance planning for continuous flow processes. *Comput. Ind. Eng.* **2023**, *176*, 108985. [CrossRef]
3. Hajej, Z.; Nyoungue, A.C.; Abubakar, A.S.; Mohamed Ali, K. An integrated model of production, maintenance, and quality control with statistical process control chart of a supply chain. *Appl. Sci.* **2021**, *11*, 4192. [CrossRef]
4. Koetsier, A.; van der Veer, S.N.; Jager, K.J.; Peek, N.; de Keizer, N.F. Control charts in healthcare quality improvement. *Methods Inf. Med.* **2012**, *51*, 189–198.
5. Flegner, P.; Kačur, J.; Durdán, M.; Laciak, M. Statistical Process Control Charts Applied to Rock Disintegration Quality Improvement. *Appl. Sci.* **2020**, *10*, 8343. [CrossRef]
6. Montgomery, D.C. *Introduction to Statistical Quality Control*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
7. Roberts, S. Control chart tests based on geometric moving averages. *Technometrics* **2000**, *42*, 97–101. [CrossRef]
8. Borror, C.M.; Montgomery, D.C.; Runger, G.C. Confidence intervals for variance components from gauge capability studies. *Qual. Reliab. Eng. Int.* **1997**, *13*, 361–369. [CrossRef]
9. Borror, C.M.; Champ, C.W.; Rigdon, S.E. Poisson EWMA control charts. *J. Qual. Technol.* **1998**, *30*, 352–361. [CrossRef]
10. Crowder, S.V. A simple method for studying run–length distributions of exponentially weighted moving average charts. *Technometrics* **1987**, *29*, 401–407.
11. Crowder, S.V. Design of exponentially weighted moving average schemes. *J. Qual. Technol.* **1989**, *21*, 155–162. [CrossRef]
12. Maravelakis, P.E.; Castagliola, P.; Khoo, M.B. Run length properties of run rules EWMA chart using integral equations. *Qual. Technol. Quant. Manag.* **2019**, *16*, 129–139. [CrossRef]
13. Brook, D.; Evans, D. An approach to the probability distribution of CUSUM run length. *Biometrika* **1972**, *59*, 539–549. [CrossRef]
14. Testik, M.C.; McCullough, B.; Borrar, C.M. The effect of estimated parameters on Poisson EWMA control charts. *Qual. Technol. Quant. Manag.* **2006**, *3*, 513–527. [CrossRef]
15. Champ, C.W.; Rigdon, S.E. A comparison of the Markov chain and the integral equation approaches for evaluating the run length distribution of quality control charts. *Qual. Control. Appl. Stat.* **1992**, *37*, 407–408. [CrossRef]
16. Maravelakis, P.E.; Castagliola, P. An EWMA chart for monitoring the process standard deviation when parameters are estimated. *Comput. Stat. Data Anal.* **2009**, *53*, 2653–2664. [CrossRef]
17. Lucas, J.M.; Saccucci, M.S. Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics* **1990**, *32*, 1–12. [CrossRef]
18. Sukparungsee, S.; Mititelu, G. A Modified Poisson Exponentially Weighted Moving Average Chart Based on Improved Square Root Transformation. *Thail. Stat.* **2016**, *14*, 197–202.
19. Cabral Morais, M.; Knoth, S. Improving the ARL profile and the accuracy of its calculation for Poisson EWMA charts. *Qual. Reliab. Eng. Int.* **2020**, *36*, 876–889. [CrossRef]
20. Avci, E.; Alturk, S.; Soylu, E.N. Comparison count regression models for overdispersed alga data. *Int. J. Res. Rev. Appl. Sci.* **2015**, *25*, 1.
21. Jesus, B.D.; Ferreira, P.H.; Boaventura, L.L.; Fiaccone, R.L.; Bertoli, W.; Ramos, P.L.; Louzada, F. Statistical process control of overdispersed count data based on one-parameter Poisson mixture models. *Qual. Reliab. Eng. Int.* **2022**, *38*, 2324–2344. [CrossRef]
22. Mustafa, F.; Sherwani, R.A.K.; Raza, M.A. A new exponentially weighted moving average control chart to monitor count data with applications in healthcare and manufacturing. *J. Stat. Comput. Simul.* **2023**, 1–21. [CrossRef]
23. McLachlan, G.J.; Lee, S.X.; Rathnayake, S.I. Finite mixture models. *Annu. Rev. Stat. Its Appl.* **2019**, *6*, 355–378. [CrossRef]
24. Karlis, D.; Xekalaki, E. Mixed poisson distributions. *Int. Stat. Rev. Int. Stat.* **2005**, *73*, 35–58. [CrossRef]
25. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22. [CrossRef]
26. Qiu, P. *Introduction to Statistical Process Control*; CRC Press: Boca Raton, FL, USA, 2013.
27. Boaventura, L.L.; Ferreira, P.H.; Fiaccone, R.L.; Ramos, P.L.; Louzada, F. New statistical process control charts for overdispersed count data based on the Bell distribution. *An. Acad. Bras. Ciências* **2023**, *95*, e20200246. [CrossRef] [PubMed]