1. INTRODUCTION

Lego factory is manufacturing system that operates as a semi-automatic closed-loop production line, consisting of two parallel workstations performing identical tasks. parts undergo individual processing at either workstation then go to production line. To optimize the workflow, the company has a policy guiding the allocation of parts by designated percentage to the first workstation and the remaining percentage to the second. The company is contractually obligated to a daily production target, with penalties incurred for any shortfall proportional to the contractual agreement.

The company encountered financial situation due to rise of electricity and strained financial issues for two reasons: firstly, both parallel workstations require a substantial amount of power, and secondly, one workstation consumes more electricity than its counterpart. Simultaneously, if the workstations fail to operate at an adequate speed, achieving the daily production goal becomes uncertain, resulting in significant penalty costs, The aim of this project is to improve the performance of the manufacturing system by building a valid simulation model to capture the behaviour of the system and to improve the system.

The work of this project have being divided into several phase, first phase was to build valid simulation model to capture the system's behaviour as is. In this phase study on real system have being done, during the study, the following main activities was done, problem definition, conceptual modelling, data collection, preparation of input data, simulation model construction, simulation model verification and simulation model validation.

The second phase was performance evaluation of the system as is , using validated simulation model built during first phase, the main goal of this phase was to conduct investigation on the system performance using set of key performance indicators (KPIs) namely, production volume of the system with 90% probability, the bottleneck within the system, potential impact of altering quantity of parts circulating in the system on the overall performance of the system. In addition to these KPIs both daily energy consumption and daily operational costs was estimated. These assessments was done to understand the plant operational dynamics.

The third and fourth phases were system improvement and implementation respectively, the main goal of these phases was to introduce two allocation policies for the two parallel workstations and implement on the workstations. First policy focuses on short term improvement acting on routing probabilities of parts to the workstations, while second policy focuses on long term solution acting on the system's dynamics like machine states and buffer level.

2. MODELLING OF THE SYSTEM AS-IS SYSTEM.

This section discusses activities of project's first phase, during this phase important activities have being done, first activities was plant visit with a main goal of getting clear understanding of the system's logical operation and layout, useful information and data for simulation model building was collected. Starting with physical layout of the plant, there are 5 station in the plant organized in close network, two parallel workstations and all other stations which connected as series arrangement. The logical operation and sequences of workstations of the plant will be discussed in detail in the following sections.

2.1.Conceptual model

After collecting useful information from real system during the visit, the conceptual model build was built, by defining logical flow of parts in the system and interaction of different system components. The following diagram shows the conceptual model of the plant.

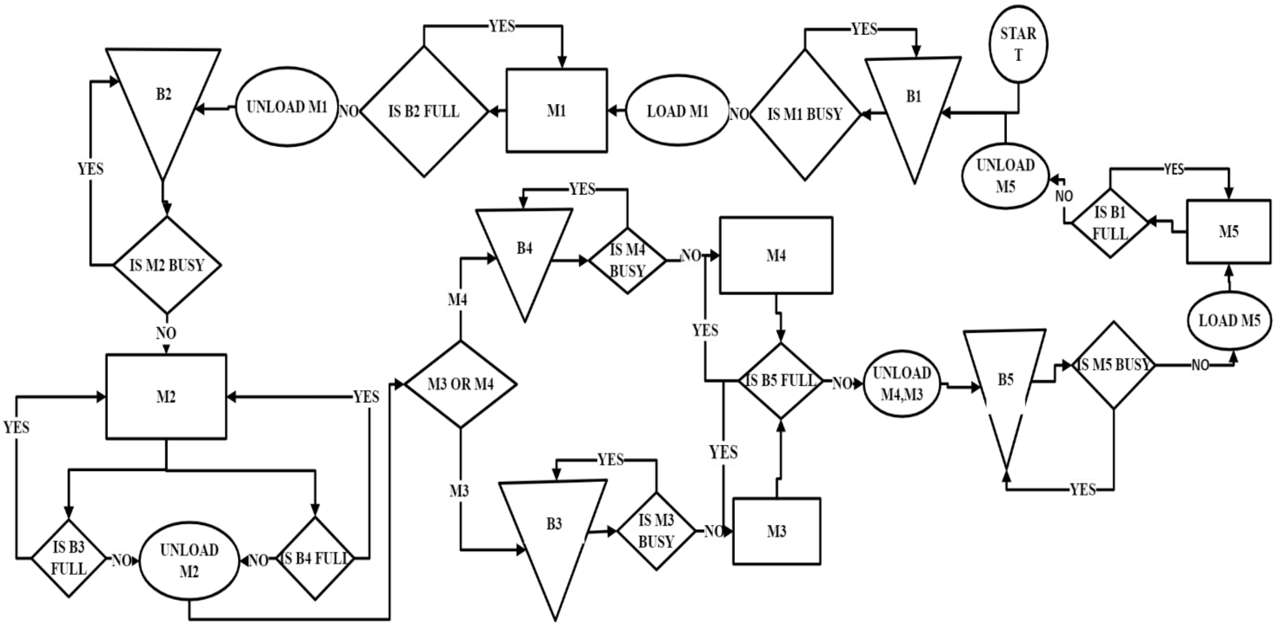

Figure 1: the conceptual model

Figure 1 shows how the real system works conceptually. The system operates as a closed system, with 12 pallets moving inside, and each pallet represents single part. These parts join the system after passing through station 5 ,(M5). They head to the buffer of station1 (B1), waiting to be loaded onto the station, if the machine (M1) is available the parts are loaded, and the processing begins.

The system follows a policy of blocking after service, meaning that stations can't unload a part if the following station's buffer is full after processing. If buffer of station2 (B2) isn't full, the parts are unloaded and taken to station 2 (M2). When they reach buffer of station2, they wait until the machine is available and then get loaded. After processing, the machine decides whether the part goes to station 3 or station 4 based on routing policy.

If the parts sent to station 3 (M3), the machine checks the buffer of station3 (B3) and unloads if it's not full. The same applies to station 4 (M4). A gate between station 2 and the parallel stations (3 & 4) executes the decision made by station 2. When the part arrives the buffer of station4 (B4), it is loaded onto the machine if it's available; otherwise, it is kept in the buffer, the same as at station 3.

After processing both parallel stations, they check the buffer of station5 (B5). If it's not full, the parts are unloaded. When the part reaches station 5, it is loaded onto the machine if the machine is available;

otherwise, it is kept in the buffer. After processing, the machine checks station1's buffer and unloads the part if it's not full and this continuous during course of operation of the plant.

2.2.Modal assumptions

This model is built on several assumptions which details as follows, there is 1 buffer in front of each station, Buffers have finite capacity, processing times of machines are stochastic, dispatching rule is first-in-first-out (FIFO), blocking after service policy (BAS), machines are perfectly reliable except machines in station1 and station5, failures always happen after processing (post-process failure), loading, and unloading are available when needed , machine processes one part at the time. The buffer capacity of each station can be see the table below.

B1	B2	B3	B4	B5
5	5	4	6	4

2.3.Execution logic

In this manufacturing system, parts initially enter the system through station1 and wait in the buffer of station1 to undergo processing. After the completion of this initial processing, the parts move on buffer of station2 through conveyor belt, where they await the second processing activity. Upon the completion of the second processing activity, **51%** of the parts proceed to station4, while the remaining **49%** are directed to station3 as deducted from the data.

After completion of processing activities at the parallel stations, the parts advance to station5. Which is where the final activity is carried out, and subsequently, the part exits the system. This sequential flow outlines the journey of the parts through the different stations in the manufacturing process.

The blocking mechanism, the system adheres to a policy known as "blocking after service." This policy dictates that stations cannot unload a part if the buffer of the next station is full after processing. The blocking scenarios at the stations differ because the buffer status depends on the production rate of the machine and its operational status while some stations are faster than other stations.

Upon analysis of collected data , The probabilities of stations being blocked were identified. For example, the blocking probability of station1 was 11.84%, while station2, station3, station4, and station5 exhibit blocking probabilities of **0.425%, 0.447%, 0.691%**, and **0%**, respectively. These probabilities indicate the likelihood of a station being temporarily unable to proceed due to a full buffer of next station.

2.4. Acquisition of input data

Acquisition of input data from the data collected in the second visit, during this visit some important data was collected from the system to have necessary information for building a simulation model. Analyse of collected data was done to see how the real system behaves specifically the processing, loading and unloading times of the stations, transportation times among the stations, time to failure and time to repair of the machines, routing probabilities of the parallel stations, blocking probabilities of the machines was determined, and buffer capacity of each station and transportation distance among the stations was deduced.

First step was data snooping by isolating data of each station, to examine them individually. Following this, a MATLAB code was developed to determine loading, processing, and unloading times for each station. For the loading times, specific calculation was executed that involves determining the difference between the load and process activities. When the activity is identified as process, the code subtracts the preceding activity, which is loading, yielding the loading time for each station, and saved as txt extension. After the loading step, the processing time for each station was calculated with help of MATLAB code. This is achieved by subtracting the process activity from the subsequent activity. If the identified activity is process, it subtracts the following activity, which could be unload, fail, or block. This process allows to obtain the processing time for every station and this processing time is stochastic.

After determining the processing time, the next step involves calculating the unloading time. Recognizing that loading occurs after unloading, carefully examination the activities of each station and the activities of the subsequent station was done. This analysis aims to prevent any instances of blocking and starvation times for stations. If the previous station was blocked, indicating a full buffer, difference of unloading and loading activities times of the same station gives unloading times. This strategy effectively eliminates the possibility of starvation and blocking times in such situations, flowing table shows the loading and unloading times of each station.

Stations	S1	S2	S3	S4	S5
Loading	3.69 seconds	3.04 seconds	3.57 seconds	2.95 seconds	2.94 seconds
Unloading	2.56 seconds	1.58 seconds	2.55 seconds	1.55 seconds	1.54 seconds

Additionally, the transportation time between stations was computed. Duration data collection visit single part was traced in the laboratory, collecting data for that specific part, To determine the time taken from station1 to station2, the time difference between when the part leaves station1 and when it arrives at station2 was determined, then the unloading time of station1was subtracted, this gives transportation time of station1 to station2. Considering the transportation length and sought to find the speed by using the transportation time and length. Once the speed was determined, the difference between buffer length and transportation conveyor length was computed and use that difference to obtain the transportation time without the buffer length. This process was repeated for all stations, the following table shows the transportation times among stations.

S1-S2	S2-S3	S2-S4	S3-S5	S4-S5	S5-S1
2.08s	10.43s	9.5s	0.397s	1.81s	2.93s

The time to fail and time to repair of both station1 and station5 was computed. Time to fail was determined by calculating the duration from the last repair to the next failure, while time to repair involved measuring the time between when the machine fail and its subsequent repair. It's crucial to note that these times are stochastic, meaning they exhibit randomness, and they require representation through distribution to model them as processing times for the stations.

2.4.1 Input data analyses

After determining acquiring input data, set of descriptive statistics was determined to gain insights of the behaviour of the data. This involved examining key measures such as the maximum, minimum, mean, median, variance, coefficient of variance, skewness, and kurtosis. These statistical metrics provided a comprehensive overview of the central tendencies, variability, and distribution characteristics of the dataset, enabling to better understand the nature and patterns within the data. after analysing these statistics the data. Input analysis was carried out to find out the distributions and their parameters to model the inputs of the simulation model, followed by fitting of candidate distribution to the processing times, time to failures and time to repair of each station.

First step was drawing the pdf of the processing time of station1. The aim was to see which candidate distributions could effectively represent the observed data. Fitting with various distributions such as

Lognormal, Gamma, Weibull, Exponential, and Uniform distributions. It became apparent that Uniform and Exponential distributions did not align well with the data as seen the picture (figure 2). However, Lognormal, Gamma, and Weibull distributions seems to fit, suggesting that they may accurately model the observed processing time distribution for station1.

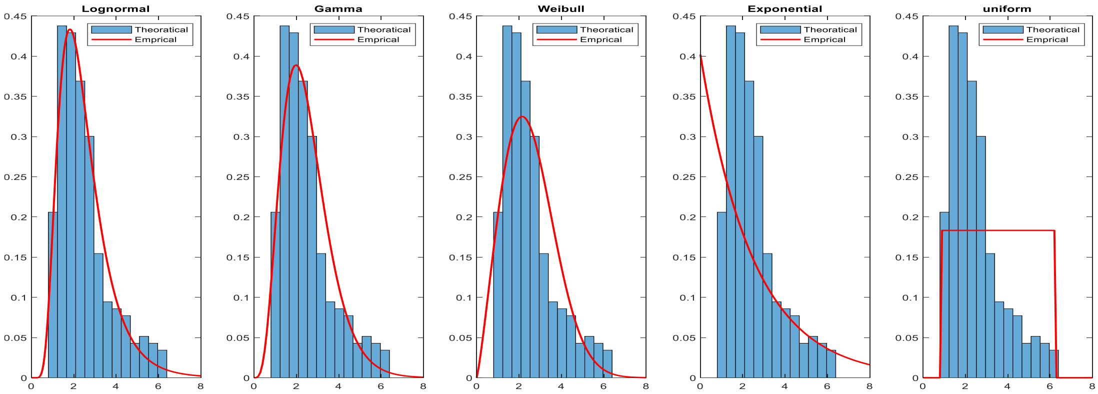

Figure 2: station1 processing times candidate distribution PDF

Additionally the Cumulative Distribution Function (CDF) of various distributions was applied to gain a clearer understanding of which distribution more effectively captures the processing time of station1. As seen the picture(figure 3) it became evident that Exponential, Uniform, and Normal distributions did not provide a better fit to represent this model. On the other hand, Gamma, Lognormal, and Weibull distributions appeared more suitable, with Lognormal being particularly promising. To validate and choose the best fit, a chi-square test was conducted as a further evaluation measure.

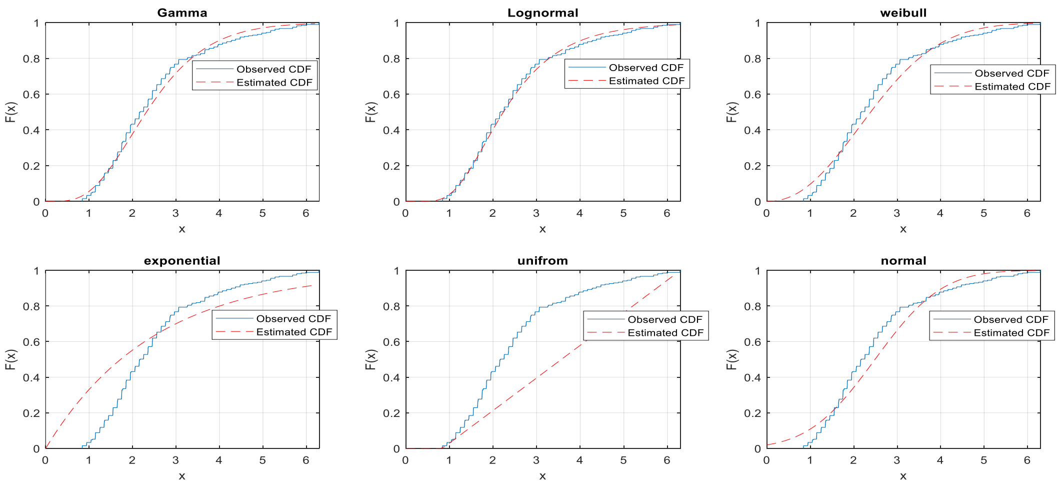

Figure 3:station1 processing times candidate distribution CDF

The Gamma, Lognormal, and Weibull distributions were selected to test for modelling the processing time of station1. Prior to conducting the tests, the parameters for each distribution was determined using Maximum Likelihood Estimators (MLEs). The obtained parameters are as follows:

Gamma distribution: Shape (G_k=4.8896) and Scale (G_θ=0.5088)

Lognormal distribution: Mean (μ=0.8056) and Sigma (σ=0.4563)

Weibull distribution: Scale (w_β=2.8215) and Shape (w_λ=2.2036)

Subsequently, these distributions was subjected to test, along with their respective parameters, to a chi-square test with a confidence interval of 95%. The null hypothesis (H0) that the processing time conforms to the estimated distribution was formulated, and the alternative hypothesis (H1) that the processing time does not conform to the estimated distribution.

For the Lognormal distribution with 16 cells, the degree of freedom (k-s-1) is 13, and the test statistic (x^2_0) is 18.0037. As $x^2_0 < x^2_{0.95,13}$(22.362), The null hypothesis at a significant level of α=0.05 cannot be rejected. Therefore, the processing time of station1 conforms to a Lognormal distribution (0.80,0.45).

In the case of the Gamma distribution with 16 cells, the degree of freedom (k-s-1) is 13, and the test statistic (x^2_0) is 35.8339. Since $x^2_0 > x^2_{0.95,13}$ (22.362), the null hypothesis at a significant level of α= 0.05, can be rejected. Therefore, the processing time of station1 does not conform to a Gamma distribution (4.88, 0.50).

Similarly, for the Weibull distribution with 16 cells, the degree of freedom (k-s-1) is 13, and the test statistic (x^2_0) is 56.4982. As $x^2_0 > x^2_{0.95,13}$ (22.362), the null hypothesis at a significant level of α= 0.05, can be rejected. Therefore, the processing time of station1 does not conform to a Weibull distribution (2.82, 2.20), In summary, the preference for representing the processing time of Station1 was lognormal distribution. The consistency and appropriateness of the lognormal distribution have been confirmed through statistical test, providing a reliable representation of Station1's processing time data.

In the case of Station2, the processing time of station2 was found almost deterministic. The mean processing time, as calculated from the data was 8.1031seconds, showcasing a remarkable consistency. The extent of variation, measured by the standard deviation, is notably low at 0.0060. and the confidence interval of [8.1023, 8.1038]. In simpler terms, this interval reflects our strong belief that the true processing time at Station2 lies within this specific range, underscoring the station's reliability and precision in its tasks, figure 4 shows CDF of processing time of station 2.

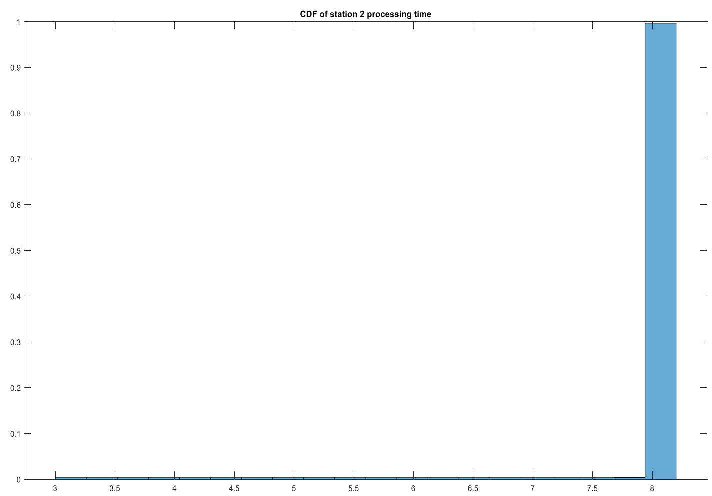

Figure 4:station2 processing time CDF

Upon examining Station3, employing a visual aid by creating a PDF of the data. These graphical representation allows to see patterns and potential distributions families that could accurately describe the processing time characteristics. Through our analysis, various distributions was tested, and the visual depiction indicated that a uniform distribution aligns well with representing the processing time of Station3 as shown the figure 5. In simpler terms, the histogram suggested that the observed data closely resembles the expected pattern of a uniform distribution processing time of station3.

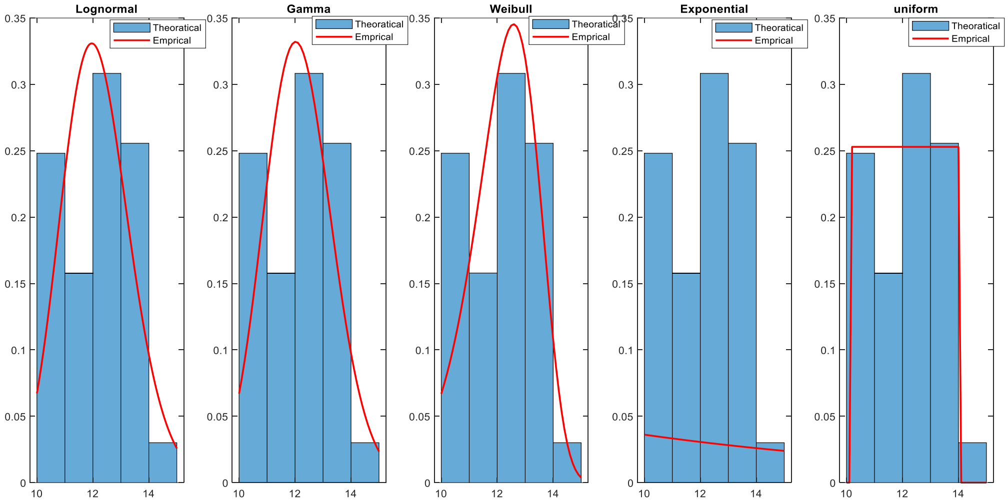

Figure 5:station3 processing times candidate distribution PDF

The cumulative distribution functions (CDFs) of different distributions was also applied to assess the representation of processing time for Station3. The visual analysis clearly indicated that the Uniform distribution is the most appropriate choice as shown figure 6. To validate and reinforce selection, chi-square test was conducted , emphasizing its role in confirming the accuracy of our conclusions.

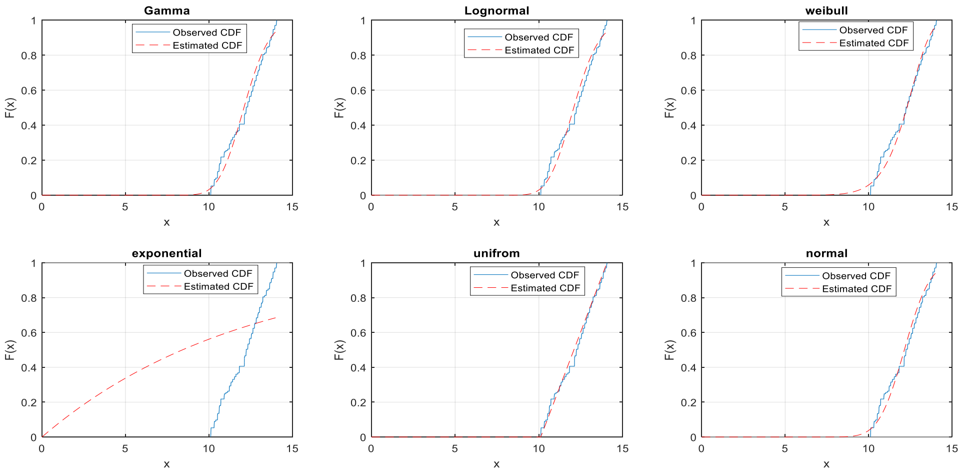

Figure 6:station3 processing times candidate distribution CDF

To confirm if the chosen distribution accurately represents the data, chi-square test was conducted. Before this, we determined the distribution parameters using the maximum likelihood estimator (MLEs), and found minimum 10.1135 and maximum 14.0672. then this distribution was subjected, along with its parameters, to a chi-square test with a 95% confidence interval. a null hypothesis (H0) was constructed proposing that the processing time conforms to the estimated distribution, while the alternative hypothesis (H1) suggested a deviation from the estimated distribution.

In examining the uniform distribution with 20 cells, and considered the degree of freedom (k-s-1) to be 17, so the resulting test statistic (x^2_0) was 20.5789. Given that x^2_0 falls below the critical value $x^2_{0.95,17}$ (27.5871), there is not statistical evidence to reject the null hypothesis at a significance level of α=0.05. Consequently, concluding that the processing time of Station3 adheres to a uniform distribution with parameters (10.11, 14.06).

Similarly, in the analysis of Station4's data, following the same approach to Station3. drawing a pdf and fitting various distributions to identify the most suitable representation for the processing time of Station4. Upon visual examination, it became evident that the uniform distribution aligns well with the observed processing time patterns of Station4.

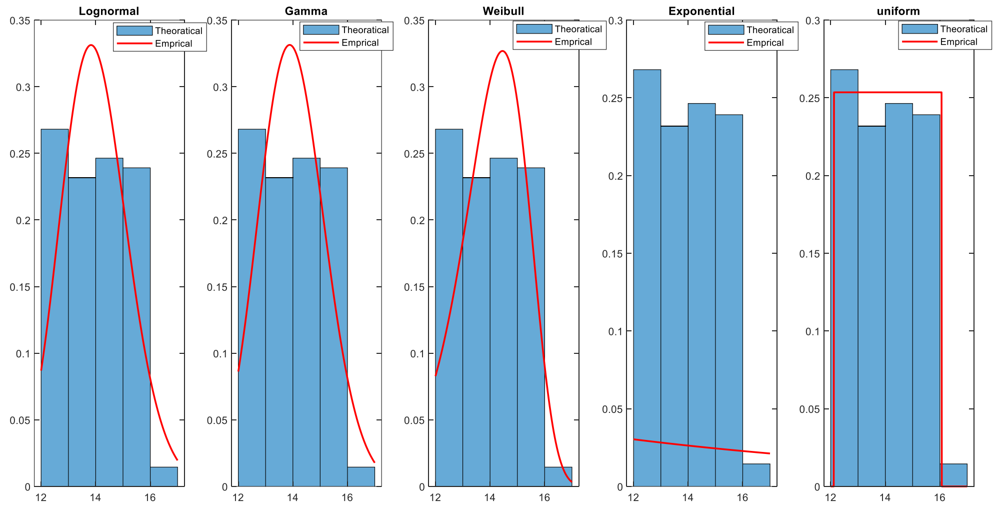

Figure 7:station4 processing times candidate distribution PDF

taking a similar steps as station3, exploring cumulative distribution functions (CDFs) of various distributions to evaluate how well they represent the processing time. A careful visual inspection pointed to the Uniform distribution as the most fitting choice, evident in the graphical representation. To affirm our selection, chi-square test was conducted , underscoring its significance in verifying the reliability of our findings.

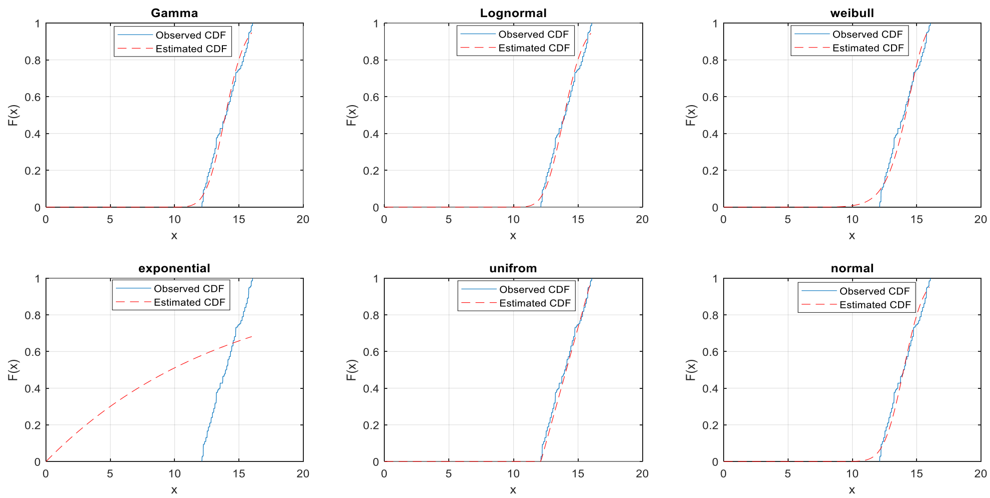

Figure 8:station4 processing times candidate distribution CDF

To confirm the accuracy of the chosen distribution in representing the data. the distribution parameters was determined using the maximum likelihood estimator (MLEs), finding values of minimum 12.1193 and maximum 16.0634. Then, this distribution, along with its parameters was subjected, to a chi-square test with a 95% confidence interval. Our null hypothesis (H0) suggested that the processing time conforms the estimated distribution, while the alternative assumption (H1) hinted at a deviation from the estimated distribution.

The chosen uniform distribution for Station4 was subjected to a chi-square test. With 20 cells and a degree of freedom (k-s-1) of 17, the resulting test statistic (x^2_0) was 17.942. Notably, as x^2_0 is below the critical value $x^2_{0.95,17}$ (27.5871), refrain from rejecting the null hypothesis at a significance level of α=0.05. Therefore, our conclusion stands that the processing time of Station4 aligns with a uniform distribution characterized by parameters (12.11, 16.06).

Ultimately, exploring the fitting of probability density functions (PDFs) to the processing time data for Station5. Upon visual inspection, Lognormal, Gamma, Normal, and triangular distributions appeared to offer suitable fits as figure 9 shows, while Weibull and uniform distributions seemed less fitting. To rigorously assess and select the most appropriate distribution, a chi-square test was conducted to determine which distribution aligns better with the observed data and successfully passes the test.

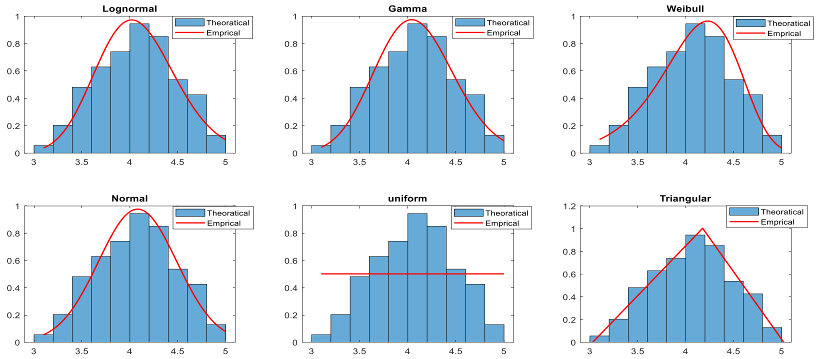

Figure 9:station5 processing times candidate distribution PDF

The Gamma, Lognormal, and Weibull distributions was selected to test for modelling the processing time of station5. Prior to conducting the tests, the parameters for each distribution was determined using Maximum Likelihood Estimators (MLEs). The obtained parameters are as follows:

Gamma distribution: Shape (G_k= 98.5374) and Scale (G_θ= 0.0414)

Lognormal distribution: Mean (μ= 1.4012) and Sigma (σ= 0.1015)

Normal distribution: Mean (μ= 4.0807) and Sigma (σ= 0.4085)

Triangular distribution: lower limit (a = 3.0293), upper limit (b = 5.0218) and mode (c = 4.1758)

To validate and assess the goodness-of-fit for the various distributions considered for Station5's processing time, a chi-square test with a 95% confidence interval was employed. With null hypothesis (H0) posits that the processing time adheres to the estimated distribution, while the alternative hypothesis (H1) suggests a deviation from the estimated distribution. This test serves as a critical step in confirming the appropriateness of the chosen distribution in representing the processing time data for Station5.

For the triangular distribution with 19 cells, the degree of freedom (k-s-1) is 16, and the test statistic (x^2_0) is 12.998258. As $x^2_0 < x^2_{0.95,13}$(24.9958), cannot be rejected the null hypothesis at a significant level of α=0.05. Therefore, the processing time of station5 conforms to a triangular distribution (3.02,5.02,4.17).

For Gamma distribution with 19 cells, the degree of freedom (k-s-1) is 16, and the test statistic (x^2_0) is 99.5852. Since $x^2_0 > x^2_{0.95,13}$ (26.2962), the null hypothesis can be rejected at a significant level of α= 0.05. Therefore, the processing time of station5 does not conform to a Gamma distribution (4.88, 0.50).

For the lognormal distribution with 19 cells, the degree of freedom (k-s-1) is 16, and the test statistic (x^2_0) is 118.3037. As $x^2_0 > x^2_{0.95,13}$ (26.2962), the null hypothesis can be rejected at a significant level of α= 0.05. Therefore, the processing time of station5 does not conform to a Weibull distribution (1.40, 0.10).

For the normal distribution with 19 cells, the degree of freedom (k-s-1) is 16, and the test statistic (x^2_0) is 101.5556. As $x^2_0 > x^2_{0.95,13}$ (26.2962), the null hypothesis can be rejected at a significant level of α= 0.05. Therefore, the processing time of station5 does not conform to a Weibull distribution (4.08, 0.40).

Consequently, analysis leads to conclude that the processing time of Station5 is best represented by a triangular distribution, with its parameters (3.0293, 5.0218, 4.1758). This decision is a result of comprehensive evaluations, including visual fits and a chi-square test, ensuring a robust validation process in selecting the most appropriate distribution.

In summary, the following table provides a summary of the processing times for each station along with the selected distributions to represent them and parameters.

Processing station 1	LOGN (0.80, 0.45)
Processing station 2	Deterministic (8.1031)
Processing station 3	UNIF (10.11, 14.06)
Processing station 4	UNIF (12.11, 16.06)
Processing station 5	TRI (3.02,5.02,4.17)

The TTR (time to repair) of station1 and station5 was also fitted the PDF and CDF to model the best distribution to represent, first the TTR of station1was fitted to different distributions, as shown in figure 10, the PDF and CDF of Gamma, Weibull, lognormal and exponential seems to fit the data while uniform doesn't seem to fit representing the time to repair of station1, then the parameters for each distribution was calculated using Maximum Likelihood Estimators (MLEs). The obtained parameters are as follows:

Gamma distribution: Shape (G_k= 1.2366) and Scale (G_θ= 8.9544)

Lognormal distribution: Mean (μ= 1.9485) and Sigma (σ= 1.0739)

Weibull distribution: Scale (w_β=11.5943) and Shape (w_λ=1.1325)

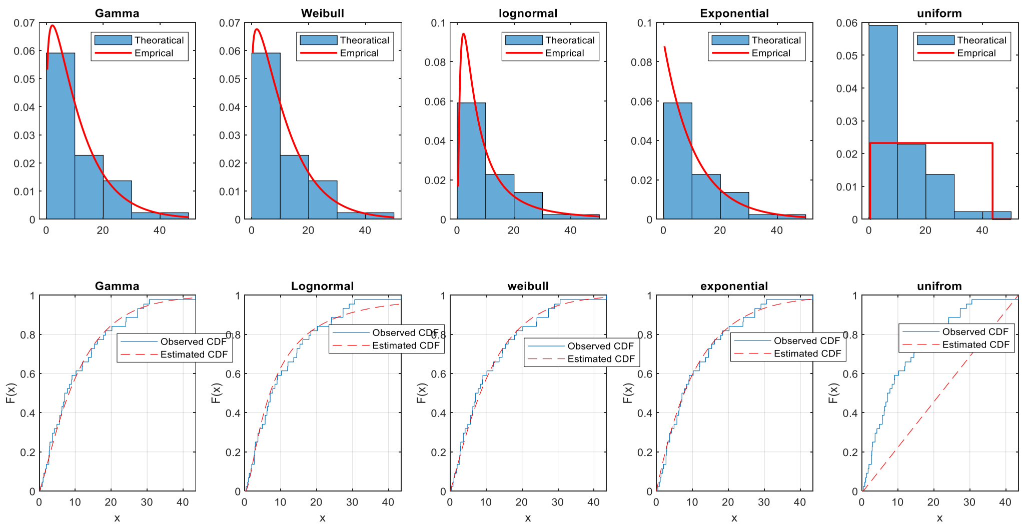

Figure 10: TTR of station 1 CDF and PDF

To validate these findings, the goodness-of-fit was assessed for various distributions considered for TTR (Time to Repair), initially a chi-square test was employed. However, this test consistently yielded passing results for all distributions. This was due to limited size of our dataset. Recognizing the need for an alternative approach, the Kolmogorov-Smirnov (K-S) test was used.

Conducting the K-S test on Weibull, Lognormal, and Gamma distributions, each with 44 samples and a significance level of 95%, critical values was discovered from the K-S table: $D_{0.95,44} = 0.19842$. Evaluating the K-S test statistics, D_0 values of 0.064451 was obtained for Weibull, 0.080911 for Lognormal, and 0.061149 for Gamma.

Upon analysis, the hypothesis H0 cannot be rejected with significance level 0.05, however it became evident that Gamma exhibited the lowest maximum absolute difference between the observed Cumulative Distribution Function (CDF) of the sample and the expected CDF. This conclusion is further supported by visual representation see the figure 11 below, and Data conforms Gamma (8.95,1.23).

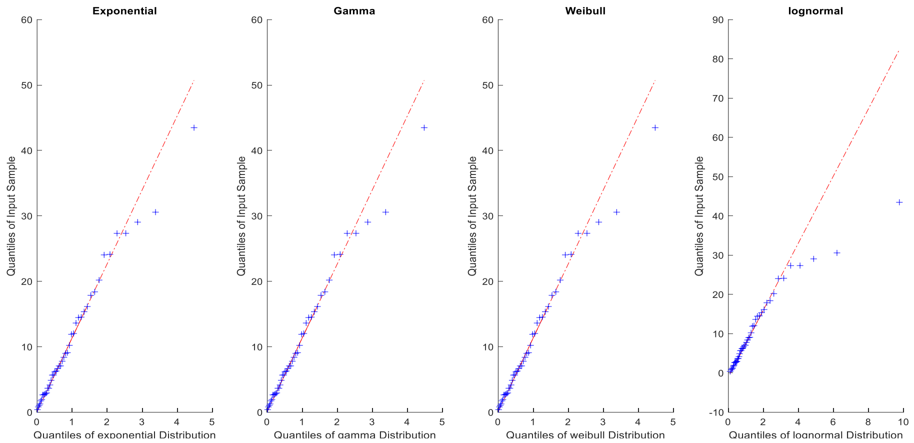

Figure 11:TTR of station 1 QQ plot

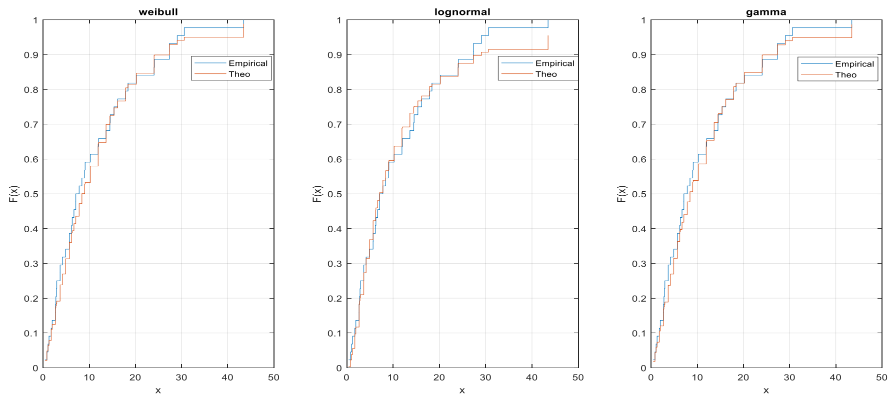

Figure 12:TTR of station 1 CDFs of theoretical and empirical

The TTR of station5 was fitted to different distributions, as shown figure 13, the PDF and CDF, Gamma, Weibull and lognormal seems to fit the data while exponential and uniform doesn't seem to fit the time to repair of station5, the parameters for each distribution was calculated using Maximum Likelihood Estimators (MLEs). The obtained parameters are as follows:

Gamma distribution: Shape ($G\_k = 3.2870$) and Scale ($G\_\theta = 3.6551$)

Lognormal distribution: Mean ($\mu = 2.3263$) and Sigma ($\sigma = 0.6305$)

Weibull distribution: Scale ($w\_\beta = 13.5670$) and Shape ($w\_\lambda = 2.1184$)

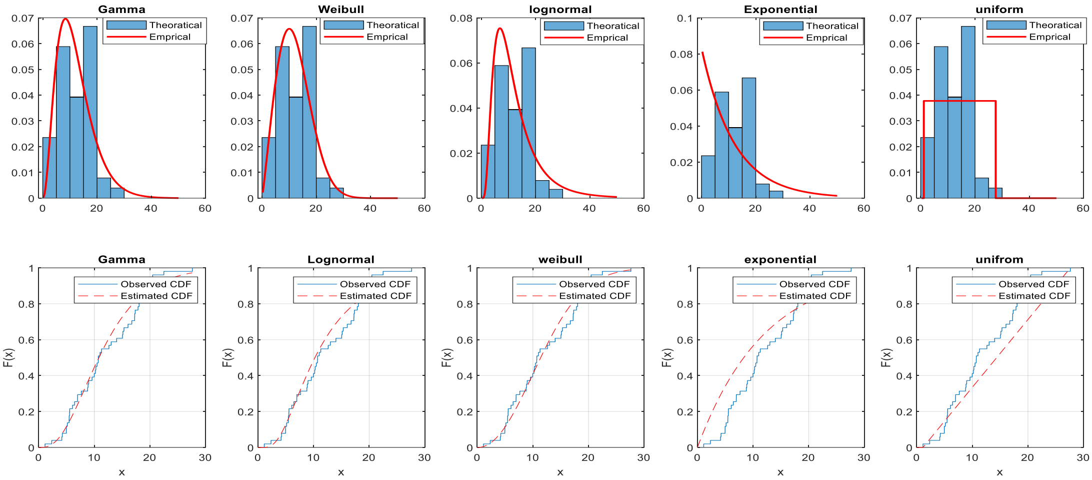

Figure 13:TTR of station 5 CDF and PDF

For Gamma distribution with 6 cells, the degree of freedom (k-s-1) is 3, and the test statistic (x^2_0) is 5.8235. Since x^2_0 > $x^2_{0.95,13}$ (7.8147), the null hypothesis cannot be rejected at a significant level of α= 0.05. For the lognormal distribution with 6 cells, the degree of freedom (k-s-1) is 3, and the test statistic (x^2_0) is 7.4706. As x^2_0 < $x^2_{0.95,13}$ (7.8147), the null hypothesis cannot be rejected at a significant level of α= 0.05. For the Weibull distribution with 6 cells, the degree of freedom (k-s-1) is 3, and the test statistic (x^2_0) is 6.5294. As x^2_0 < $x^2_{0.95,13}$ (7.8147), the null hypothesis cannot be rejected at a significant level of α= 0.05.

In summary, after the chi-square test, gamma has the highest p-value (0.12052), indicating the TTR of station5 follows a Gamma distribution (3.6551, 3.2870). The table below summarizes the TTR for station1 and station5 using selected distributions.

TTR1	GAMMA (8.95, 1.23,)
TTR5	GAMMA(3.65, 3.29).

The failures of two machines, M1(station1) and M5(station5), was assessed using data analysis. For Station 1, the total failure time was calculated by summing all instances when the machine failed, resulting in a cumulative downtime of 764.383 seconds over the data duration of 1.5 hours. This downtime accounts for 14% of the total operational duration. Similarly, for Station 5, the failure duration was determined using the same method as Station 1, resulting in a failure time of 1105.123 seconds. This constitutes 20% of the total duration of 1.5 hours. It is evident that the failure rate of Station 1 is lower than that of Station 5. Consequently, the repair time for failures at Station 1 should be shorter than that of Station 5, which aligns with the data presented in the table above. In summary, the table below shows the failure probabilities for the two stations.

Failure probabilities of M1	14%
Failure probabilities of M5	20%

2.5.Simulation model building and validation

In the process of building and validation of the simulation model, Arena software was used to construct the model with estimated input parameters representing the processing time of stations and other important parameters. To imitate the real system, the model was developed based on a conceptual model. Initially, 12 parts was created and assigned routes to station1, as the system being modelled is a closed system. And each station was equipped with a finite buffer capacity, managed using queue blocks. The model included seize blocks to check machine availability, delay blocks for loading and processing, scan blocks to monitor buffer capacity of the following station, and additional delay blocks for unloading.

To incorporate routing probabilities for parts to station3 and station4, a brunch block was implemented. Additionally, the failure of station1 and station5 were simulated using brunch blocks, assuming failures occur after processing. These brunch blocks were set with probabilities of failure and corresponding repair times and these blocks were put after the processing delay blocks to execute logical flow of process and failure. Furthermore, the transportation conveyors were modelled using a route block, accounting for transportation time within the system.

In the model validation phase, the simulation model was run with 10 replications, each lasting 3 hours. Recording important key performance indicators namely, production volume and system time, then then key performance indicators (KPIs) was compared with the same KPIs from the real system. focusing on the production volume, the model was validated using the 3-hour dataset provided my professors, where the real system's production volume was 858 parts, while simulation model output on average was 859 parts with a standard deviation of 2.0976. Utilizing the Confidence Interval approach, the real system's production volume falls within the range of [857, 860] with difference value of 2 parts, according to our approach(CI), since the worst-case error is below the difference value, the simulation model was accept as valid model.

Furthermore, the system time was analysed. By analysing a 3-hour dataset given, it was revealed that real system's system time was **150.6 seconds** (the time it take one part to from entry of station to the exist of station five). Running the simulation model yielded an average system time of **149.8** seconds, with a standard deviation of 0.3137. Employing the Confidence Interval approach again, this result was established an interval of [149.64, 150.09]. In this case, the worst-case error is below the difference level of 1-second, leading to acceptance of the model's validity. Consequently, the simulation model was considered valid for both production volume and system time.

3. Analysis of the AS-IS System

To analyse the system's performance as is , first warm up period should be removed in order to analyse its behaviour when it reaches steady state. To remove warm period, Welch method was applied, initially the simulation was ran for an extended period of 3 days, conducting 10 replications. By analysing the output of the simulation and applying welch method with different window sizes the warm period was identified. As seen in the figure below the 250 window size was selected and subsequently the warm period was revealed as 40 minutes.

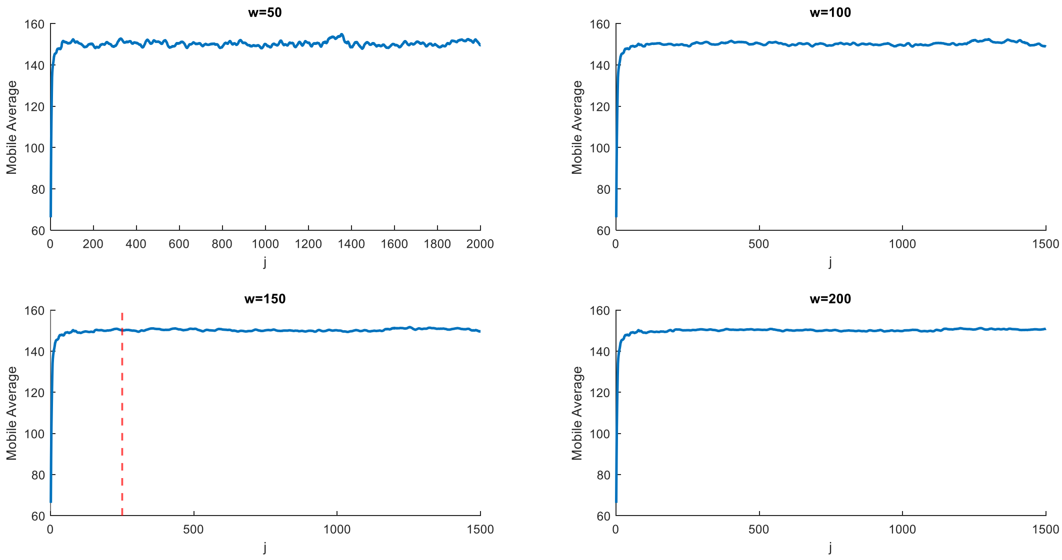

Figure 14: Welch moving average window

3.1 Daily production volume with 90% probabilities

After determining the warm-up time, the system's production volume with a 90% probability, was estimated. the system can reach with **90%** probability is **6094** parts. Additionally, then interval estimation of the quantile was performed using formula below, which yields a confidence interval of [6042.83, 6139.57], with a lower bound of **6042.83** and an upper bound of **6139.5.**

$$\ell = \left\lfloor nq - z_{1-\alpha/2}\sqrt{nq(1-q)} \right\rfloor \qquad u = \left\lceil nq + z_{1-\alpha/2}\sqrt{nq(1-q)} \right\rceil$$

3.2 Determination of system's bottleneck

To identify bottlenecks in the system, the performance of each station was analysed, The saturation levels of machine1, machine2, machine3, machine4, and machine5 are **6.31%, 63.46%, 46.17%, 56.45%,** and **31.78%** respectively. Moreover, the average buffer levels for station1, station2, station3, station4, and station5 are 0.9143, 4.1254, 0.3375, 0.3598, and 0.7691 parts respectively, these values clearly show station2 is the bottleneck of the system. The average daily production volume of the plant is **6772** parts per day.

3.3 Impact of varying quantity of parts in the system

Examining the impact of varying the number of circulating parts, and observing changes in performance parameters. With a 25% degrease number of parts circulating around the system which is 9 parts, the saturation levels remained relatively stable, and the bottleneck persisted at station2. The saturation levels for machine1, machine2, machine3, machine4, and machine5 became 6.19%**,** 62.55%**, 45.89%, 55.37%,** and **31.44%** respectively. Average buffer levels for station1, station2, station3, station4, and station5 decreased to 0.2700, 2.2514, 0.2923, 0.3514, and 0.6557 parts. The daily production volume was degreased to **6668.4** parts per day.

When the number of circulating parts increased to 15 parts, the saturation levels changed slightly. Machine1 reached **6.29%,** while machine2, machine3, machine4, and machine5 recorded **63.56%, 46.80%, 55.99%**, and **31.97%** respectively. Average buffer levels increased, with station1 at 3.3024 parts and station2, station3, station4, and station5 at 4.4128, 0.3838, 0.3610, and 0.9122 parts respectively. And the average daily production volume was increased to **6781** parts.

In conclusion increase number of parts circulating in the system, has positive impact on the daily production volume while the saturation and average work in progress (WIP) was not effected that much, so increasing number of parts circulating in the system should be considered to make sure reaching daily production target.

3.4 Daily energy consumption

The daily energy consumption of the system was calculated by running the simulation for a 24-hours excluding warm up period. Subsequently, the time during which the stations were busy was determined, considering the busy periods of the machines, loading, and unloading times of machines. For the machine busy times the saturation of each machine was determined from the simulation and then multiplied the total opening time to calculate the busy time of the machine, to include the loading and unloading busy times, since loading and unloading times are constant therefore number of worked parts in each station was counted and multiplied the loading and unloading times of each station and sum up the busy time of each machine and loading and unloading busy time of each station. Also the failure times of machine 1 and machine 5 was identified by considering the periods of machine failures. To calculate the idle time of the stations, the busy time and the failure time was subtracted from the total working time of each station, which amounts to 24 hours.

Consequently, the busy and idle times for each station was obtained and multiplied them by the power consumption of the stations when busy and idle. This led us to the daily energy consumption of the system, totalling **3693535.9 KJ,** the table below summarizes the daily energy consumption of the plant**.**

Parameter	S1	S2	S3	S4	S5	Total
Busy time	47776.5 s	86150s	60406.4s	64200.2s	57910.5s	316443.8s
Idle time	28281.4s	249.9s	25993.5s	22199.8s	11879.2s	88603.9s
Idle energy	113125.8KJ	749.8KJ	51987KJ	88799.2KJ	29697.9KJ	284359.9KJ
Busy energy	477765.2 KJ	775350.5KJ	1208129.3KJ	513601.5KJ	434329.3KJ	3409175.9KJ
Toal energy	**590891 KJ**	**776100.3KJ**	**1260116.4 KJ**	**602400.7KJ**	**464027.3 KJ**	**3693535.9 KJ**

Furthermore, the daily cost of the system was computed by multiplying the total energy of the system which is 3693535.9 KJ and unit energy cost of 0.0005 €/kJ, and the total is **1846.76 €**/day on average, there is no penalty cost because our production volume was higher than the daily target of **6750** parts. Since our model consistently produces an average of **6772** parts in a steady state, the confidence interval of mean cost with standard deviation of **8.46** € and 95% confidence interval so the marginal error was 6.055 and therefore the confidence interval was [1840.7, 1852.815].

4 System Improvement

In this section, improvement of the system was done applying by two allocation policies to the parallel stations. The objective is to reduce the overall daily cost of the system, focusing on lowering both the energy cost and penalty cost. Several constraints needs consideration. One essential constraint is related to meeting the daily production target; failure to do so results in penalties for each part not produced. Another constraint involves the production capacity of the machines, ensuring that each machine operates within its assigned capacity. Additionally, there is a constraint tied to the buffer capacity of the stations. These constraints collectively guide efforts to optimize the system's efficiency while minimizing associated costs.

4.1 The short term improvement of the system

In addressing the challenge of determining the appropriate percentage allocation of parts to parallel stations within the context of our first allocation policy, two distinctive alternatives have being explored. The first alternative takes on a more calculation perspective, employing an analytical model. Here, the focus is on precisely balancing the power of parallel stations to determine the percentage that results in the lowest energy consumption, the busy and idle time of the parallel stations was calculated and then multiplied the busy power and idle power of each station by comparing these energy the percentage of parts leading to lowest cost was calculated and percentage was 62% to station4.

The second alternative was running the model AS-IS to determine the highest output of station2 can be achieved since it is the bottleneck of the system, without being constrained by the capacities of station3 and station4. This can be achieved by adjusting the buffer sizes of station3 and station4 to infinite values. This allows isolation and assess the maximum production capacity of station2 without any bottlenecks caused by downstream stations.

Following this, the model was reran this time directing all parts exclusively to station4 and adjusting its buffer to nominal size . This allows to assessing the comparative production abilities of station4 in relation to the outputs from station2, through this comparison, the production capacity of the station2 was discovered which is 6592 parts, while station4 can produce up to 4523 parts. Calculating the ratio of station4's output to station2's, the maximum production volume that by fully utilizing station4 can produce is with 68.6%, illustrating its potential contribution when operated at full capacity in the production line.

The selection procedure for determining the most cost-effective short-term solution involves selection of alternatives the two alternatives, alongside an evaluation of the system in its current configuration was compared. This methodical approach enables us to accurately assess which option yields the lowest daily operating costs for the system, for the two alternatives we execute 10 independent replications for each alternative and we calculated the average daily cost for each alternative the table below shows the daily cost in €.

Rep	AS-IS	Alternative1(62%)	Alternative2 (68.6)
1	1832.87387	1755.724624	1757.78177
2	1846.849745	1752.675436	1751.096372
3	1847.05916	1765.178216	1754.546814
4	1856.587668	1757.319886	1763.242983
5	1834.433247	1738.415324	1752.906531
6	1853.437432	1734.744868	1749.3885
7	1851.221251	1751.301526	1749.045932
8	1856.339336	1766.840297	1758.299227
9	1840.172208	1741.23909	1743.993058
10	1848.705749	1755.153292	1762.422754
Mean	1846.767967	1751.859256	1754.272394
Variance	71.65476665	116.0936923	38.25466857

The selection procedure was done to determine the alternative with the lowest cost, comparing the costs calculated when the system was in AS-IS, to the cost calculated from alterntive1 and CI is **[8.425799e+01,1.055594e+02].** and again comparing the cost calculated when was system AS-IS, to the cost calculated from alterntive2 and CI **[8.429734e+01,1.006938e+02]**. both alternatives are better than AS-IS, therefore alternative1 has lower average cost than alternative2, so that the most cost-effective option is Alternative 1, which involves routing 62% of the parts to station4, thereby achieving the lowest daily cost.

4.1.1 The implementation of short term improvement on the system

The short term solution above was implemented on the real system, these routing probabilities was applied on the real system and it was ran by 45 min to collect data validate the effectiveness of the policy. The energy cost and penalty cost form the real system for 45 min was calculated. then the simulation model was ran 10 replication for 45min and calculated the energy cost and penalty cost, the cost determined from the real system was 54.68€ and the cost calculated for 10 replication in the modal is shown this table below.

Rep	Cost in €
1	54.58741
2	56.20922
3	54.97596
4	52.96449
5	52.83313
6	53.76104
7	55.0935
8	57.2526
9	56.95965
10	55.98054

the mean cost of the modal was 55.06€ with a variance of 2.43€, confidence interval of mean cost with 95% confidence level and marginal error of 1.073976 was therefore calculated as [53.9, 56.13].

4.2 The long term improvement of the system

For the long-term solution, with a goal of minimizing the system's overall daily costs by reducing both energy and penalty costs, while taking the system's constraints into account. One such constraint was the necessity to meet daily production targets; not meeting these targets incurs penalties for each shortfall part. Another constraint was each machine's production capacity, which must not be exceeded to ensure optimal operation. Furthermore, the system is constrained by the buffer capacities at each station. Keeping in mind these constraints three long-term solution alternatives was developed.

The first alternative focuses on balancing the saturation between station2 and station4 to maintain the current system dynamics. The aim was to keep the system's bottleneck at station2, thus ensuring equilibrium between the saturation levels of station2 and station4. Parts are routed to station4 as long as its saturation falls below that of station2 and parts are routed to station 3 when saturation of station 4 reach that of station2 this is done prevent any shift of the bottleneck to station4.

The second alternative introduces a buffer threshold strategy, where the buffer capacity of station4 is limited at a 90% threshold. Parts are routed from station2 to station4 whenever station4's buffer is below this threshold, aiming to optimize flow without overwhelming station4.

The third alternative was time balancing, equates the times needed to process the parts that are in buffer and in the machine of station2 and those of station4. If the time required to process parts that are in station4's buffer and machine is less than that of station2, parts are directed to station4; otherwise, they are sent to station3. For each of these alternatives, 10 independent replications was conducted to determine the average daily cost, aiding in the identification of the most cost-efficient solution, the table below shows the cost of three alternative and the cost of AS-IS system in €.

Rep	AS-IS	Alternative1	Alternative2	Alternative3
1	1832.87387	1853.995637	1758.821431	1702.280968
2	1846.849745	1860.178379	1756.46106	1702.57993
3	1847.05916	1859.50247	1760.519481	1700.429951
4	1856.587668	1859.733851	1746.164622	1701.342907
5	1834.433247	1852.281793	1748.759443	1701.742468
6	1853.437432	1859.366467	1741.335739	1698.228457
7	1851.221251	1857.053455	1753.763675	1697.765448
8	1856.339336	1860.60932	1760.06002	1701.871933
9	1840.172208	1860.117045	1747.694006	1701.333962
10	1848.705749	1859.800162	1759.529827	1696.910853
Mean	1846.767967	1858.263858	1753.31093	1700.448688
Variance	71.65476665	8.371087555	47.04697869	4.20312259

Through the application of ranking and selection procedure to determine the alternative with the lowest cost with indifference zone of 0.5 we found that alternative3 which is the time balancing is the best alternative, alternative2 is the second best alternative and alternative1 is worse, then AS-IS system.

Alternatives	Mean	Ranking
AS-IS	1846.77	3
Alternative1	1858.26	4
Alternative2	1753.31	2
Alternative3	1700.5	1