A STATISTICAL FRAMEWORK FOR ENHANCING PROCESS CONTROL AND RELIABILITY USING AUTOENCODERS RANDOM SURVIVAL FORESTS AND NHPP MODELING

P. Sricharani[1], P. Satya Shekar Varma[2], M. V. Lavanya[3*]

•

[1]Professor, Department of Artificial Intelligence, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India
[2]Research Scholar, Computer Science and Engineering, National Institute of Technology, Warangal, Telangana, India
[3]Assistant Professor, Department of Mathematics and Statistics, GITAM School of Science, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India
Corresponding Author Email: *lavanya.mv9@gmail.com

Abstract

Statistical Process Control (SPC) plays a critical role in ensuring the reliability and quality of high-grade products. This study introduces two advanced machine learning methods to enhance SPC by integrating Auto encoders for anomaly detection and Random Survival Forests (RSF) for failure prediction. The Autoencoder model is employed to monitor real-time sensor data, learning the normal patterns of product quality and identifying deviations that indicate potential quality issues. By flagging anomalies when product performance metrics diverge from expected thresholds, the autoencoder helps to adjust SPC limits dynamically, improving responsiveness to emerging quality concerns. Additionally, RSF is used to predict the likelihood of product failure over time, based on historical failure data and process parameters. This predictive approach enables proactive interventions to prevent quality issues before they occur, enhancing long-term product dependability. Together, these machine learning methods create a comprehensive framework for real-time monitoring and failure prediction, providing a more adaptive and data driven approach to quality control. The integration of Auto encoders and RSF into the SPC methodology significantly advances the precision and effectiveness of product reliability assessment, offering a powerful tool for maintaining high-quality standards in manufacturing processes.

Keywords: Statistical Process Control (SPC), Autoencoders, Random Survival Forests (RSF), Anomaly Detection, Failure Prediction, Product Reliability, Quality Control, Machine Learning, Predictive Modeling

I. Introduction

In real-time analysis of lifespan data, especially concerning the failure times of high-quality products over a specified period, the integration of advanced statistical models with machine learning methods enhances Statistical Process Control (SPC) for quality monitoring and prediction. This study proposes a novel approach by incorporating the Freshet Distribution with the Non-

Homogeneous Poisson Process (NHPP) to design a new "time control chart." This chart, in combination with machine learning techniques—Autoencoders for anomaly detection and Random Survival Forests (RSF) for failure prediction—enables comprehensive monitoring and proactive quality interventions. The Frechet Distribution and NHPP model failure intensity (λ) across specific time intervals, providing a probabilistic foundation for real-time failure monitoring. The framework is represented by:

$$P(N(1,P) - N(1) = z) = \frac{e^{-\theta p}(\theta p)^z}{z!}, z = 0,1,2,\ldots \quad (1)$$

where $N(1)$ represents the random count of failures within [0, l], enabling a focus on instances significantly exceeding the Upper Control Limit (UCL) and prompting targeted quality investigations.

$$P(N(l) = z) = \frac{e^{-\mu(l)}\mu(l)^z}{z!}, z = 0,1,2,\ldots \quad (2)$$

In this equation, $\mu(t)$ represents the mean value function. Eq.1 characterizes a Non-Homogeneous Poisson Process (NHPP). The function $\mu(t)$ can be interpreted as the mean value function for an NHPP. The renowned Frechet distribution (FD) is selected as the distribution for $\mu(t)$ to provide a growth model based on the Non-Homogeneous Poisson Process (NHPP). Building on previous research by Roy, Pham, Srinivas Rao & Sricharani, Grabski, and others focused on mathematical models for SPC, this study extends the traditional approach by using

Autoencoders to detect anomalies in real-time sensor data. The Autoencoder model learns normal patterns in production data and identifies deviations, dynamically adjusting control limits to improve SPC responsiveness. Concurrently, RSF uses historical failure data and process parameters to predict the likelihood of product failure over time, allowing for timely interventions that prevent potential quality issues. Together, these tools form a robust framework for anomaly detection, real-time monitoring, and predictive failure analysis.

The structure of this paper is as follows: It begins by outlining the characteristics of the Frechet Distribution and NHPP, then introduces the construction of the time control chart. Comparisons with the Exponential model and the benefits of Autoencoder-driven anomaly detection and RSF-based failure prediction are discussed. Additionally, the application of order statistics for monitoring production is examined. The paper concludes with key insights into the enhanced reliability and adaptability of SPC through machine learning integration, underscoring the benefits of this approach for quality control in manufacturing environments.

II. Methodology

In this study, a combination of statistical modeling and machine learning techniques is used to create an enhanced control system for monitoring time between failures and identifying anomalies in product quality. This methodology integrates the Frechet distribution, Autoencoders for anomaly detection, and Random Survival Forests (RSF) for failure prediction, along with the Non-Homogeneous Poisson Process (NHPP) as the mean function to dynamically adjust control limits based on real-time data.

2.1. Frechet Distribution and its Characteristics

The Frechet distribution, known for its skewness and stability, is used in applications such as life testing and reliability analysis. The Frechet distribution finds extensive practical utility in diverse real-world scenarios, encompassing life testing, rainfall analysis, queue dynamics, wind speed modeling, earthquake prediction, sea current assessments, horse racing performance, flood analysis, and track race record evaluation The probability density function (PDF) associated with the Frechet distribution is expressed as follows

$$g(y) = \frac{k}{b}\left(\frac{y-b}{a}\right)^{1-k} e^{-\left(\frac{y-a}{b}\right)^{-k}}, k \epsilon (0, \infty) \tag{3}$$

Its cumulative distribution function (cdf) is

$$G(y) = e^{-\left(\frac{y-a}{b}\right)^{-k}} \; if \; y > a \tag{4}$$

The Frechet distribution is characterized as a skewed and maximally stable distribution. For the Frechet distribution, the mean, median, and variance are denoted as follows

$$Mean = a + b\Gamma\left(1 - \frac{1}{k}\right) \; for \; k > 1 \tag{5}$$

$$Median = a + \frac{b}{\sqrt[k]{\log_e 2}} \tag{6}$$

$$Variance = b^2\,\Gamma\left(1 - \frac{2}{k}\right) - \left(1 - \frac{1}{k}\right)^2, k > 0 \tag{7}$$

The NHPP with $\phi G(y)$ as the mean value function for our present study is

$$\mu(x) = \phi e^{-\left(\frac{y-a}{b}\right)^{-k}} \tag{8}$$

Frechet distribution contains 3 parameters namely k is a shape parameter, a is a location parameter b is a scale parameter where ϕ is number of damages in the manufactured goods.

2.2. Anomaly Detection with Autoencoders

Anomaly detection is critical for identifying unexpected deviations in product performance. Autoencoders are utilized to learn the typical behavior of the manufacturing process by encoding and reconstructing normal data patterns. During operation, the Autoencoder flags an anomaly if the reconstruction error exceeds a predefined threshold, indicating potential quality concerns or system instability. This enables real-time adjustments of SPC limits by detecting anomalies early in the process.

2.3. Failure Prediction with Random Survival Forests (RSF)

RSF is applied to predict failure times and assess product reliability based on historical failure data and operational parameters. RSF provides survival probabilities over specified time frames, allowing for proactive interventions before potential failures. By incorporating survival analysis, RSF into quality control systems enhances the accuracy of failure predictions and helps prevent unexpected downtimes, supporting an adaptive and responsive quality control system.

2.4. Control chart monitoring the time between failures based on mean value function

Similar to Shewhart's theory of variable control charts, a control chart for such data would be produced using 0.9973 probability limits of the periods between failure random variable denoted as "t," if the random variable is thought to reflect the inter-failure time of a device. Taking into account equi-tailed probability, these limits and the center line are found as the solutions to the following equations.

$$G(I) = 0.00135 \tag{9}$$

$$G(I) = 0.5 \tag{10}$$

$$G(I) = 0.998655 \tag{11}$$

Let I_L , I_C, and I_U be respectively the solutions of Eq 9, Eq 10, and Eq 11 in the standard form

$$I_L = G^{-(0.00135)} \tag{12}$$

$$I_C = G^{-(0.5)} \tag{13}$$

$$I_U = G^{-(0.99865)} \tag{14}$$

In Fig 1, be respectively the solution of the equations Eq.12, Eq.13, Eq.14 for Frechat distribution with parameter values $k = 0.65, b = 1$ & $a = 0.1$. This parameter combination is taken in order to have close affinity with exponential distribution (0.03) as evidenced from the following probability density function curves.

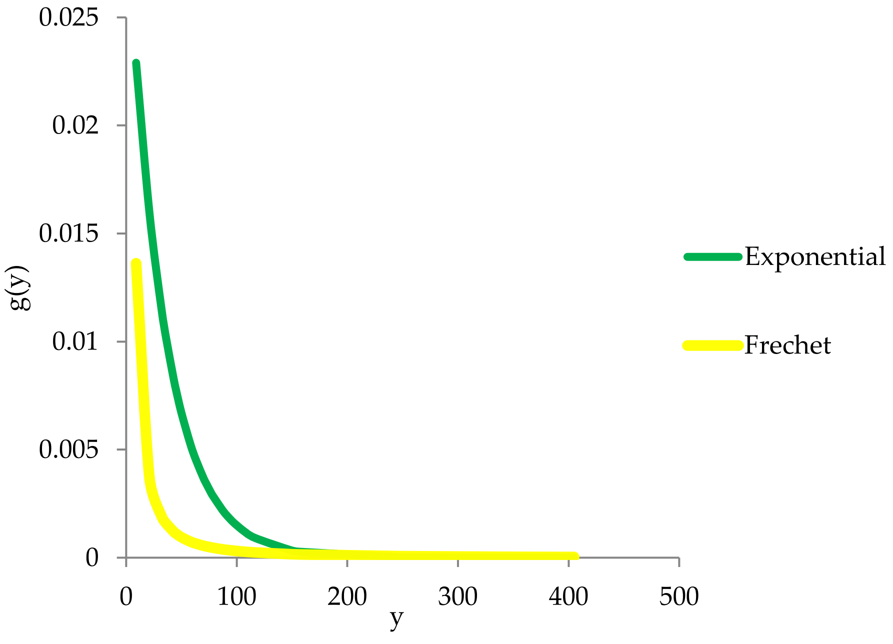


Figure 1: *Concordance of Exponential and Frehet distributions*

The NHPP based on $F(I)$ as the mean value function is the SRGM for our present study and is given by

$$\mu(I) = \phi(1 - e^{-}), I > 0, > 0 \tag{15}$$

Using the dataset from 11Kim, the time control chart is created using the mean value function corresponding to inter-failure time and is accompanied by three parallel lines that extend to the horizontal axis at points I_L, I_C, $and\ I_U$.

The illustration looks like this:

Table 1: *Real time data1*

Failure number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Failure time (hours)	9	21	36	43	45	50	58	63	70	70	71	77	78	87	91
Failure number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Failure time (hours)	92	95	98	104	105	116	140	156	247	249	250	337	384	396	405

In Fig- 2, The adequacy of the data's fit to the model is evaluated by calculating the correlation coefficient using a QQ-plot.

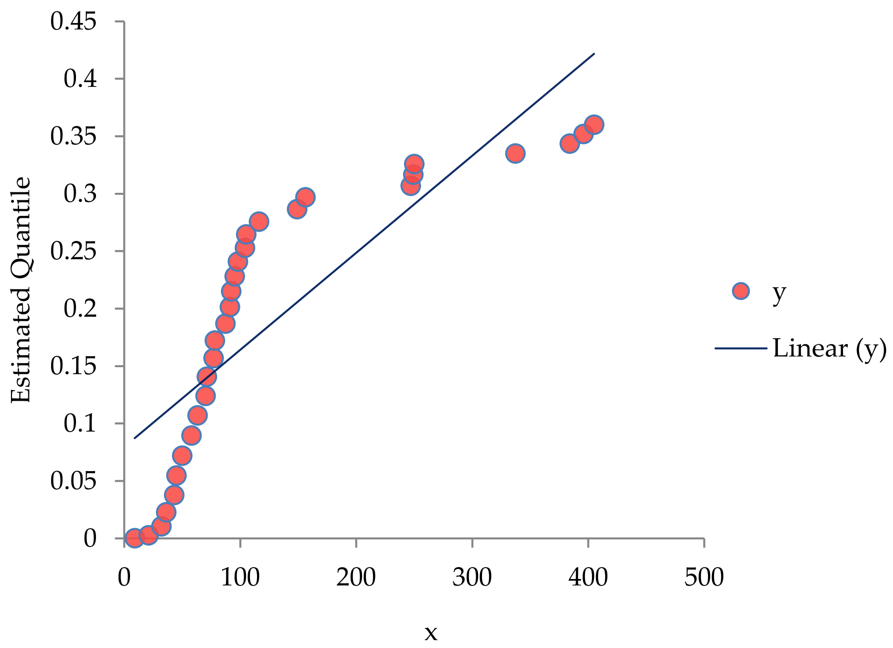


Figure 2: *QQ plot for the real time data1*

Using real-time data (Table 1), the adequacy of the model fit is verified by calculating the correlation coefficient and constructing a QQ-plot. The high correlation coefficient of 0.835 indicates a strong positive correlation, confirming the model's suitability for the data. The following table summarizes the estimation of parameters and mean value, upper, central and lower control limits using Eq.15.

Table 2: *Frechet distribution estimators and its mean value control limits*

Frechet distributiom	
$\hat{k}$	0.65
$\widehat{k\phi}$	30.61188
$\mu(I_L)$	0.041327
$\mu(I_C)$	15.30594
$\mu(I_U)$	30.57056

Estimated values of $\mu(I)$ at the given failure times I1, I2, ... In along with the successive differences of these estimates are given in Table 3. The successive differences would indicate the estimated number of failures between consecutive failure times. The graph through $[I_i , \Delta\hat{\mu}(I_i)] , i = 1,2, \ldots , n - 1$ along with three parallel horizontal lines at $\mu(I_L) , \mu(I_C) , \mu(I_U)$ would be the required control chart and is given in Fig 3.

2.5. Anomaly Detection with Random Survival Forest (RSF)

To detect anomalies in the failure times, we can apply Random Survival Forest (RSF), a machine learning method designed for handling time-to-event data. In this context, RSF can help to identify outliers or abnormal failure times that deviate significantly from the expected distribution.

- Data Preprocessing: We start by cleaning the failure data, handling any missing values, and ensuring proper formatting for RSF input.
- RSF Model Training: RSF is trained on the failure times, using features such as the successive differences between consecutive failure times. The model learns the survival probabilities over time, which represent the likelihood of a failure occurring at a particular time.
- Anomaly Detection: The RSF model calculates survival probabilities for each failure time. Anomalies are flagged based on survival probabilities that fall below a specified threshold, indicating abnormal behavior.
- Control Chart with RSF: In conjunction with RSF, we plot a control chart for the successive differences of failure times. This chart helps visualize potential anomalies by comparing actual observations against predicted values from the RSF model.

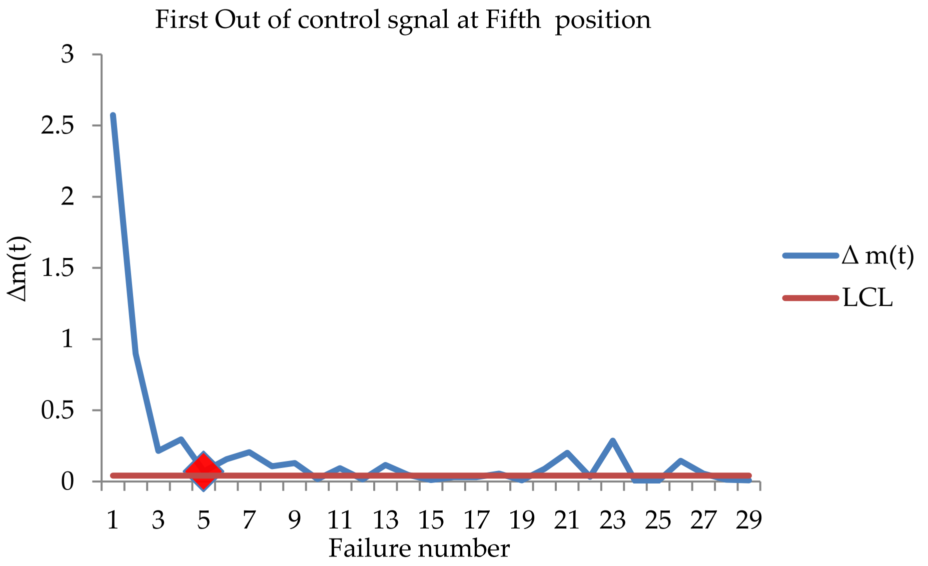
Table 3: *Consecutive differences placed on the $\mu(l)$ of FD*

Failure number	1	2	3	4	5	6	7	8	9	10
Failure time (hours)	9	21	36	43	45	50	58	63	70	70
$\mu(I)$	24.086	26.66	25.557	27.771	28.068	28.139	28.296	28.502	28.608	28.73
$\Delta\mu(I)$	2.5738	0.8973	0.2143	0.2965	0.0709	0.1573	0.2054	0.1067	0.1284	0.016
Failure number	11	12	13	14	15	16	17	18	19	20
Failure time (hours)	71	77	78	87	91	92	95	98	104	105
$\mu(I)$	28.753	28.846	28.86	28.977	29.023	29.034	29.066	29.096	29.152	29.161
$\Delta\mu(I)$	0.0926	0.0143	0.1167	0.0458	0.0109	0.0317	0.0301	0.056	0.0088	0.0889
Failure number	21	22	23	24	25	26	27	28	29	30
Failure time (hours)	116	140	156	247	249	250	337	384	396	405

$\mu(I)$	29.24	29.4505	29.484	29.7712	29.7755	29.7777	29.9232	29.9786	29.991	30
$\Delta\mu(I)$	0.2005	0.0335	0.2871	0.0043	0.0021	0.1454	0.0554	0.0124	0.0089	-

In Fig 3, The graph through $[I_i, \Delta\hat{\mu}(I_i)], i = 1, 2, \ldots, n-1$ along with three parallel horizontal lines at $\mu(I_L), \mu(I_C), \mu(I_U)$, would be the required control chart and is given in Figure 3. In Figure 3 the out of control is observed at the 5th observation below the LCL for Frechet Distribution.

3(a)



3(b)

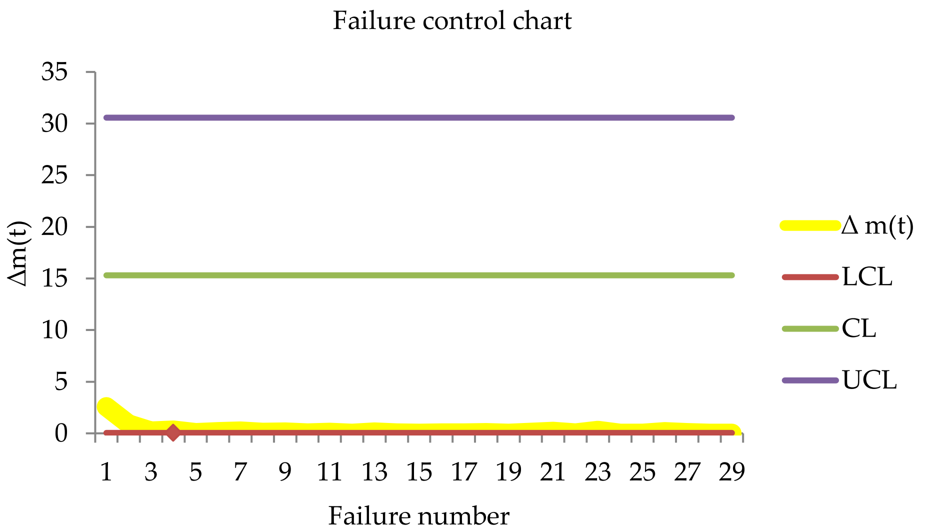


Figure 3: *Control chart based on successive differences of mean value function of FD*

2.6. Control Chart with RSF

The control chart incorporates three parallel lines representing:

- Upper Control Limit (UCL): The threshold above which observations are considered.
- Central Control Limit (CCL): The expected normal level of observations.
- Lower Control Limit (LCL): The threshold below which observations are considered

unusually low.

- In Figure 3, the out-of-control point occurs when the observation at the 5th failure time falls below the LCL, indicating an anomaly in the failure process.
- In figure 3, the control chart plots the successive differences, with three control limits, incorporating the RSF predictions. The out-of-control anomaly at the 5th observation is detected below the LCL, suggesting a failure that deviates from the expected pattern.
- The combination of Anomaly Detection and Random Survival Forest (RSF) allows for the effective monitoring of failure times and the detection of abnormal behavior. By leveraging RSF's survival probabilities, we can identify anomalies in the data, such as the out-of-control point observed in the control chart. This approach enhances the reliability of failure time analysis and provides a robust tool for predictive maintenance and system monitoring.

2.7. Comparative study

For comparison, Taking the Exponential model, the most frequently used model in reliability studies. The cumulative distribution function of Exponential distribution (ED) is

$$G(y) = 1 - e^{-\upsilon\delta}, \upsilon > 0, \ \delta > 0 \tag{16}$$

For our current investigation, the Nonhomogeneous Poisson Process (NHPP) is formulated with $G(y)$ serving as the mean value function

$$\mu(y) = \phi\left(1 - e^{-\delta y}\right), y > 0, \phi > 0, \delta > 0 \tag{17}$$

The following table summarizes the estimation of parameters and control limits for an Exponential distribution using Eq.17.

Table 4: *Parameter estimates of exponential model and their control limits*

Exponential model				
$\hat{\delta}$	$\hat{\phi}$	$\mu(I_L)$	$\mu(I_C)$	$\mu(I_U)$
0.03	30.00016	0.0405	15.0008	29.95966

Table 5: *Consecutive differences for the exponential distribution with RSF-based Anomaly detection*

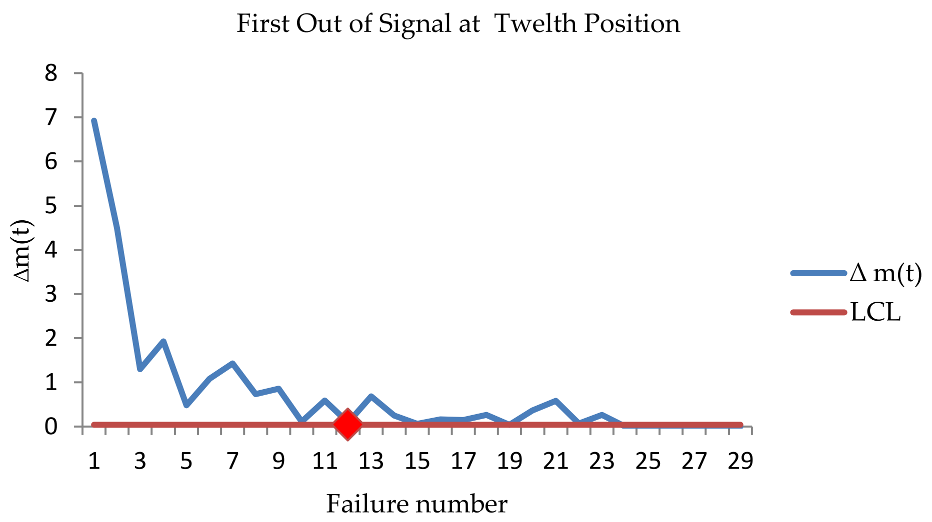
Failure number	1	2	3	4	5	6	7	8	9	10
Failure time (hours)	9	21	36	43	45	50	58	63	70	70
$m(t)$	7.0986	14.022	18.513	19.812	21.741	22.222	23.306	24.734	25.467	26.32
$\Delta m(t)$	6.9236	4.4909	1.2989	1.9297	0.4809	1.0833	1.4283	0.7334	0.8584	0.108
Failure number	11	12	13	14	15	16	17	18	19	20
Failure time (hours)	71	77	78	87	91	92	95	98	104	105
$m(t)$	28.753	28.846	28.86	28.977	29.023	29.034	29.066	29.096	29.152	29.161
$\Delta m(t)$	0.0926	0.0143	0.1167	0.0458	0.0109	0.0317	0.0301	0.056	0.0088	0.0889
Failure number	21	22	23	24	25	26	27	28	29	30
Failure time (hours)	116	140	156	247	249	250	337	384	396	405
$m(t)$	29.24	29.4505	29.484	29.7712	29.7755	29.7777	29.9232	29.9786	29.991	30

$\Delta m(t)$	0.2005	0.0335	0.2871	0.0043	0.0021	0.1454	0.0554	0.0124	0.0089	-

2.8. Anomaly Detection with Random Survival Forest (RSF)

Using Random Survival Forests (RSF), we apply a machine learning technique to detect anomalies in the failure times. RSF is particularly effective in handling censored data and can model the time-to-event data more robustly than traditional methods. RSF helps in identifying outliers by analyzing the successive differences of the failure times. The successive differences are computed between consecutive failure times, and any significant deviation from the expected range is flagged as an anomaly.

4(a)



4(b)

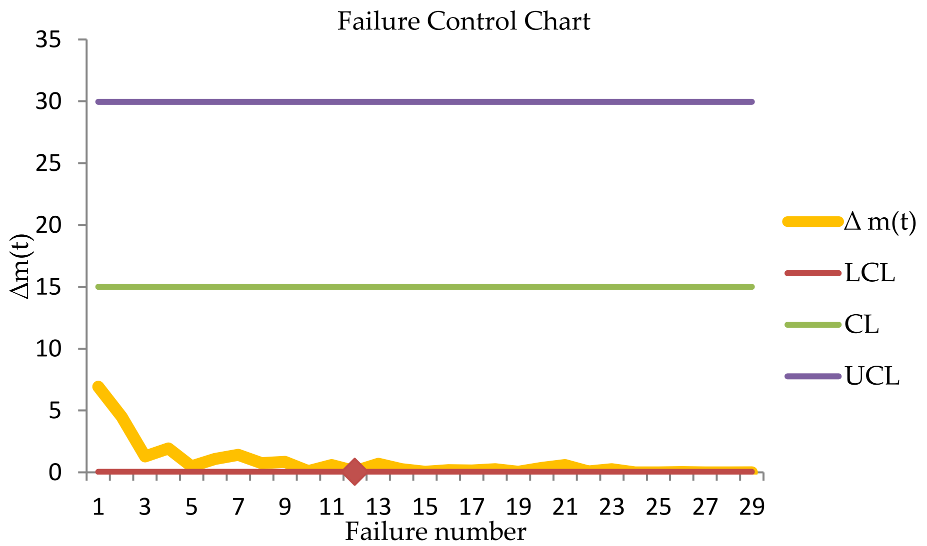


Figure 4: *Control Chart Based on Successive Differences of the Mean Value Function of ED Using RSF*

In Fig 4, The control chart in Figure 4 displays a graph for $i = 1, 2, \dots, n-1$, with three parallel horizontal lines. Notably, in Figure 4, an out-of-control scenario is evident at the 12th observation, falling below the Lower Control Limit (LCL) for the Exponential distribution.Upon examining Figures 3 and 4, both the Frechet distribution and Exponential distribution (using RSF) show out-of-control points below the Lower Control Limit (LCL), which suggests that the number of failures is lower than expected, implying a better-than-expected performance. The anomaly detected by RSF indicates that the 12th failure may have resulted from an improved process or a sudden change in system reliability. Such anomalies are indicative of improvements in quality and should be further investigated to determine the assignable causes.

In this study, the Frechet Distribution (FD) and Exponential Distribution (ED) are compared using Random Survival Forests (RSF) for anomaly detection in reliability analysis. The RSF method successfully detects anomalies in both distributions, with the Frechet Distribution showing an anomaly at the 5th observation and the Exponential Distribution showing an anomaly at the 12th observation. However, the Frechet Distribution results are more consistent and indicate fewer deviations from expected failure patterns compared to the Exponential Distribution, which shows more frequent outliers. Based on this analysis, the Frechet Distribution provides a more reliable and stable model for failure time analysis, making it the better choice for this study.

2.9. Monitoring the Prediction Process Using Order Statistics and Anomaly Detection with the Mean Value Function

The probability model of a continuous random variable Y describes the intervals between failures for a highquality product with a randomly chosen sample of size $'n'$, signifying $'n'$ inter-failure durations. $G(y)$ represents $Y's$ cumulative distribution function. Constructed for inter failure times, the time control chart displays benefits, warnings, and the stability of the failure process. In cases where $'r'$ is a natural number $(< n)$, the summations represent the intervals of time between each r^{th} error. Out-of-control signals relative to interfailure timeframes can be visualised using a control chart for the intervals between each r^{th} failure. The tr-controlcharts developed by 12Xie et al have certain shortcomings. Suggesting the utilization of the Failure Control Chart methodology to prevent such problems. If $(Y_1; Y_2; \dots; Y_r)$; $(Y_{r+1};\ Y_{r+2}; \dots;\ Y_{2r})$; $(Y_{2r+1};\ Y_{2r+2}; \dots;\ Y_{3r})$; etc are the i.i.d random variables having $G(y)$ as their common mode $Z_1 = Y_1, Z_2 = \sum_{i=1}^{2} Y_i, Z_3 = \sum_{i=1}^{3} Y_i, \dots,$ $Z_r = \sum_{i=1}^{r} Y_i$ becomes an ordered sample of size r representing the time to first failure, time to second failure, time to third failure, so on time to r^{th} failure respectively. The time to each r^{th} failure is represented by Z_r, the dots on the tr-chart, which is the control chart. The percentiles of the highest order statistics in a sample of size r would therefore function as control limits for the tr-chart when r is fixed. The following equations, which assume equi-tailed probability, provide solutions for the bottom, central, and upper limit lines, respectively.

$$|G(I)|^r = 0.00135 \tag{18}$$

$$|G(I)|^r = 0.5 \tag{19}$$

$$|G(I)|^r = 0.99865 \tag{20}$$

Let I_U, I_C, and I_L be respectively the solutions of Eq 18, Eq 19 and Eq 20 in the standard form

$$I_L = G^{-1}(0.00135)^{\frac{1}{r}} \tag{21}$$

$$I_C = G^{-1}(0.5)^{\frac{1}{r}} \tag{22}$$

$$I_U = G^{-1}(0.998655)^{\frac{1}{r}} \tag{23}$$

RSF (Random Survival Forests) and anomaly detection techniques into this model would enhance its ability to identify significant deviations or outliers in failure times. RSF can be applied to model the survival times and detect anomalies more accurately, further improving the monitoring process and ensuring the reliability of the product over time. In Fig 5, For exponential distribution, use the Frechet distribution. This combination of parameters $k = 0.02, a = 0, b = 1$, $\delta = 0.09$ is chosen to closely resemble the exponential distribution, as shown by the following probability curves of density functions.

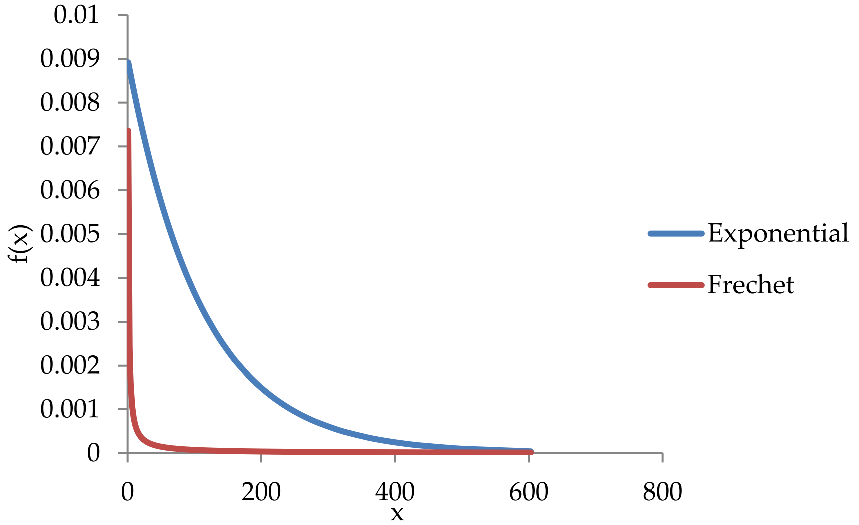


Figure 5: *Comparison of Frechet and Exponential distribution*

The NHPP with $G(x,y)$ as the mean value function as the FD for our present study is

$$\mu(x) = \phi(e^{-y})^{-k} \tag{24}$$

This model is demonstrated using real failure data, specifically 188 failures for the JET airplanes, as considered by Fatoki Olayode (2019). The following table presents the accumulated failures and the time to each 4th failure:

In this dataset, anomaly detection and Random Survival Forest (RSF) methods are applied to model and analyze the time to failure. The combination of RSF and anomaly detection allows for effective identification of outlier failures, which can help refine predictive maintenance strategies for aircraft systems. By using the RSF approach with anomaly detection, the model can highlight potential failures or abnormal behavior in the failure process. This allows for more accurate modeling of the failure times and provides a better understanding of the underlying processes.

Table 6: *Real-Time Data (188 Failures for JET Airplanes)*

Number	Failure times	Number	Failure times	Number	Failure times	Number	Failure times	Number	Failure times
1	194	39	14	77	32	115	46	153	14
2	413	40	14	78	9	116	230	154	111
3	90	41	29	79	438	117	26	155	97
4	74	42	37	80	43	118	59	156	62
5	55	43	186	81	134	119	153	157	39
6	23	44	29	82	184	120	104	158	30
7	97	45	104	83	20	121	20	159	7
8	50	46	7	84	386	122	206	160	44

9	359	47	4	85	182	123	5	161	11
10	50	48	72	86	71	124	66	162	63
11	130	49	270	87	80	125	34	163	23
12	487	50	283	88	188	126	29	164	22
13	57	51	7	89	230	127	26	165	23
14	102	52	61	90	152	128	35	166	14
15	15	53	100	91	5	129	5	167	18
16	14	54	61	92	36	130	82	168	13
17	10	55	502	93	79	131	31	169	34
18	57	56	220	94	59	132	118	170	16
19	320	57	120	95	33	133	326	171	18
20	261	58	141	96	246	134	12	172	130
21	51	59	22	97	1	135	54	173	90
22	44	60	603	98	79	136	36	174	163
23	9	61	35	99	3	137	34	175	308
24	254	62	98	100	27	138	18	176	1
25	493	63	54	101	201	139	25	177	24
26	33	64	100	102	84	140	120	178	70
27	18	65	11	103	27	141	31	179	16
28	209	66	181	104	156	142	22	180	101
29	41	67	65	105	21	143	18	181	52
30	58	68	49	106	16	144	216	182	208
31	60	69	12	107	88	145	139	183	95
32	48	70	239	108	130	146	67	184	62
33	56	71	14	109	14	147	310	185	11
34	87	72	18	110	118	148	3	186	191
35	11	73	39	111	44	149	46	187	14
36	102	74	3	112	15	150	210	188	71
37	12	75	12	113	42	151	57		
38	5	76	5	114	106	152	76		

In Fig 6, The goodness of fit of the data to the model is assess by calculating with the help of QQ-plot which resulted a reasonably good positive correlation coefficient of 0.69434.

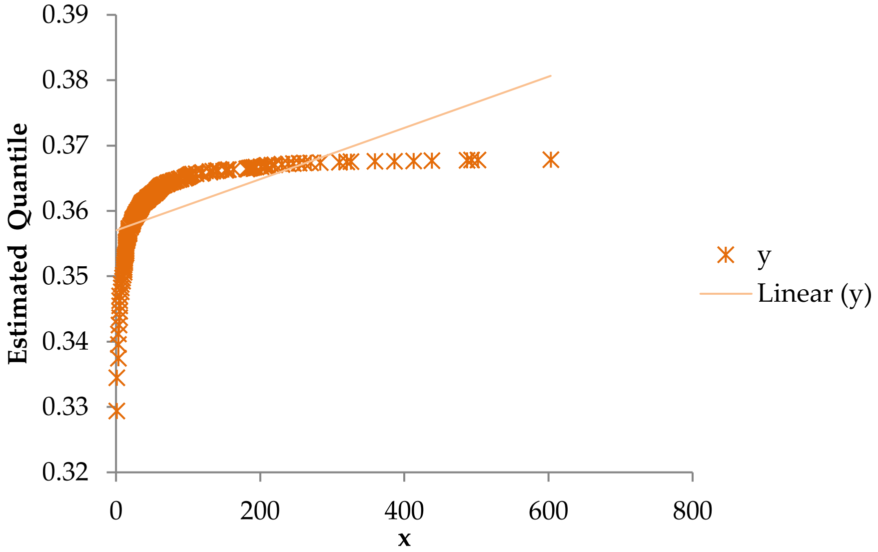


Figure 6: *QQ-Plot for real time data2*

The following Table 7 provides a thorough depiction of key metrics related to JET airplanes, including failure frequency, cumulative failure counts, and the temporal evolution associated with every 4th failure. The dataset could benefit from anomaly detection techniques, such as Isolation Forests, Autoencoders, or Z-Score methods, to highlight any outlying failure points or deviations from typical failure patterns that might signify system malfunctions or hidden patterns not immediately evident in the historical data.

Table 7: *Accumulated Failure Times*

Observation	Accumulated failures	4-order	Observation	Accumulated failures	4-order
1	771	8	25	110	227
2	225	17	26	468	236
3	1026	22	27	255	246
4	188	32	28	191	261
5	648	43	29	424	284
6	358	47	30	342	308
7	753	53	31	297	333
8	207	56	32	124	363
9	256	58	33	236	392
10	45	66	34	428	405
11	281	72	35	197	425
12	187	79	36	287	476
13	621	89	37	519	524
14	883	95	38	389	585
15	886	106	39	284	682
16	287	117	40	120	749
17	306	127	41	119	809
18	283	135	42	68	843
19	59	142	43	198	919
20	522	156	44	462	1031
21	724	173	45	211	1239
22	521	184	46	417	1596

23	423	200	47	287	2085
24	417	215			

Random Survival Forest (RSF) is a suitable method for analyzing and modeling the survival times of the JET airplanes, particularly when the failure time distribution is involved, as seen in the Frechet distribution analysis in Table 8. RSF can be used to predict the time-to-event or failure in such datasets, taking into account censored data (failures that were not observed within the study period). RSF models are robust for handling survival data, providing a powerful alternative to traditional methods like Cox proportional hazards models. Table 8 shows the parameter estimates for the Frechet distribution, along with the mean value and control limits, are meticulously presented. This section encapsulates the essential information needed to comprehend the distributional characteristics under consideration. The provided values, including the shape and scale parameters, offer insights into the Frechet distribution's behavior within the context of the study.

Table 8: *Frechet distribution parameter estimates and its mean value control limits*

Frechet distribution				
$\hat{k}$	$\hat{\phi}$	$\mu(t_L)$	$\mu(t_C)$	$\mu(t_U)$
0.02	114.7928	5.0219	39.9408	47.0046

The analysis of mean value function data, shown in Table 9, can also benefit from RSF, where survival probabilities can be calculated based on the accumulated failures (A.F) and other covariates. This would enhance the understanding of the dynamic relationship between accumulated failures and the corresponding mean failure times.

Table 9: *Mean Value Function for Accumulated Failure Times*

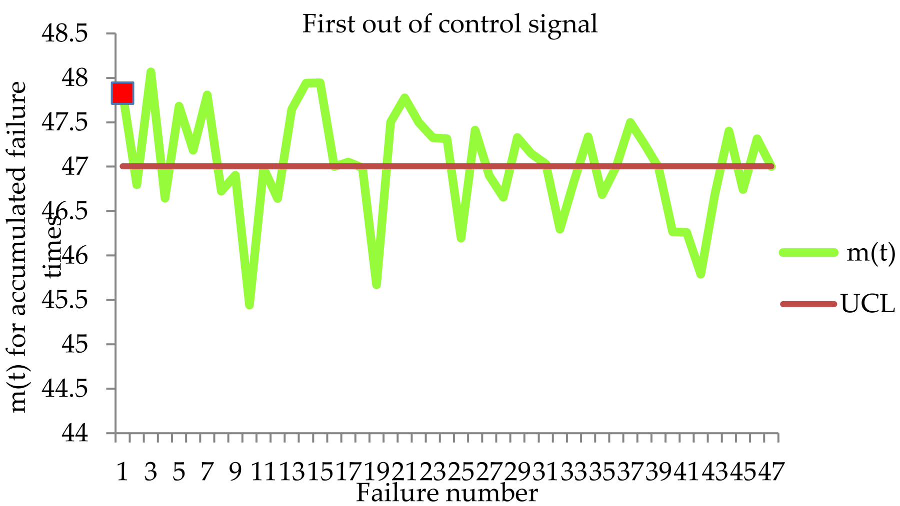
Observation	Accumulated failures	$m(t)$	Observation	Accumulated failures	$m(t)$
1	771	50.3916	25	110	31.9495
2	225	44.1301	26	468	50.0875
3	1026	50.8359	27	255	45.7180
4	188	41.4784	28	191	41.7278
5	648	50.6918	29	424	49.7215
6	358	48.8135	30	342	48.4995
7	753	50.7829	31	297	47.3305
8	207	42.9500	32	124	34.1860
9	256	45.7639	33	236	44.7627
10	45	16.9311	34	428	49.7611
11	281	46.7868	35	197	42.2069
12	187	41.3938	36	287	47
13	621	50.6507	37	519	50.3648
14	883	50.8228	38	389	49.3071
15	886	50.8233	39	284	46.8948
16	287	47	40	120	33.5755
17	306	47.6037	41	119	33.4194
18	283	46.8592	42	68	23.2716
19	59	20.9455	43	198	42.2843
20	522	50.3775	44	462	50.0457

21	724	50.7656	45	211	43.2290
22	521	50.3733	46	417	49.6488
23	423	49.7114	47	287	47
24	417	49.6488			

In Figure 7, control charts based on the mean value function depict the accumulated waiting time up to every 4th failure. The chart is crucial for identifying any potential anomalies in the failure data, such as out-of-control deviations. Anomaly detection techniques could help pinpoint abnormal data points, signaling unexpected failures or deviations from the expected system behavior.

The inclusion of Random Survival Forest (RSF) can also be integrated in this analysis to model the survival times and identify the factors contributing to failure. The RSF model would provide additional predictive power and insights into the reliability of JET airplanes, helping to pinpoint critical failure events earlier than traditional models.

7(a)



7(b)

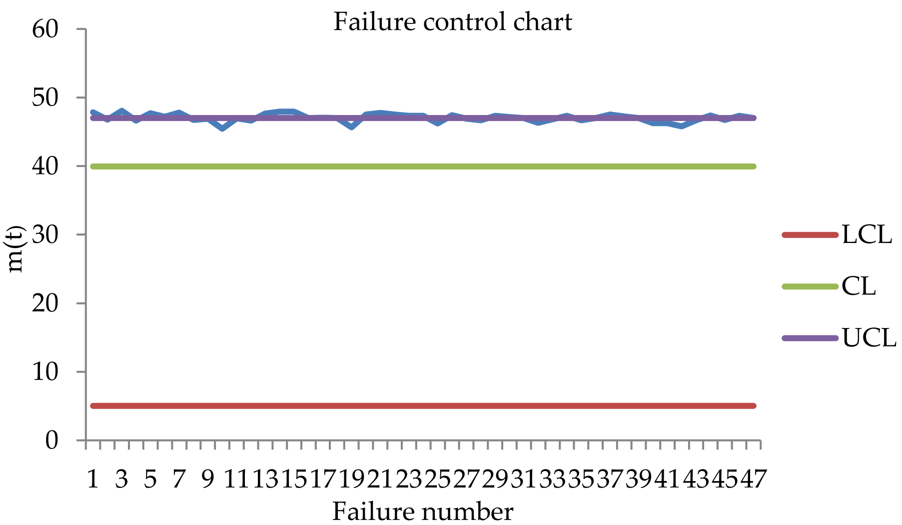


Figure 7: *Control Chart Based on Accumulated Failures of Mean Value Function of FD*

2.10. Comparative Study Based on Accumulated Failure Times

Comparing the model under study with the exponential model using the dataset provided in table 6, the results are as follows: Table 10 provides parameter estimates for the Exponential model along with their corresponding mean value function upper, central, lower control limits. This essential information includes estimates such as the scale parameter, location parameter, and shape parameter, crucial for characterizing the Exponential distribution. The presented values offer insights into the statistical properties of the model, facilitating a comprehensive understanding of its behavior within the context of the study.

Table 10: *Parameter estimates of Exponential model and their control limits*

Exponential model				
$\hat{\delta}$	$\hat{\phi}$	$\mu(t_L)$	$\mu(t_C)$	$\mu(t_U)$
0.009	50.84088	9.7453	42.7519	50.4237

Table 11 displays the Mean Value Function for accumulated failure times of ED. Each row provides detailed information, including observation number, accumulated failures, and the corresponding mean value. The progression of these values across observations facilitates a comprehensive analysis of the mean value function. The presented dataset contributes valuable insights into the dynamic relationship between accumulated failures and their associated mean values in the context of the ED model.

Table 11: *Mean Value Function for Accumulated failure times of ED*

Observation	Accumulated failures	$m(t)$	Observation	Accumulated failures	$m(t)$
1	771	47.8285	25	110	46.1942
2	225	46.7956	26	468	47.4102
3	1026	48.0677	27	255	46.9008
4	188	46.6444	28	191	46.658
5	648	47.6829	29	424	47.3274
6	358	47.1855	30	342	47.1471
7	753	47.8087	31	297	47.0287
8	207	46.7256	32	124	46.295
9	256	46.9040	33	236	46.8357
10	45	45.4419	34	428	47.3353
11	281	46.9822	35	197	46.684
12	187	46.6402	36	287	47
13	621	47.6473	37	519	47.497
14	883	47.9421	38	389	47.2552
15	886	47.9449	39	284	46.9912
16	287	47	40	120	46.2674
17	306	47.0538	41	119	46.2604
18	283	46.9882	42	68	45.7896
19	59	45.6700	43	198	46.6883
20	522	47.5018	44	462	47.3994
21	724	47.7759	45	211	46.7417
22	521	47.5002	46	417	47.3135
23	423	47.3255	47	287	47

24	417	47.3135

In Fig 8, control charts vividly portray the mean value function, capturing the cumulative waiting time up to every 4th failure. Noteworthy is the systematic plotting of mean values against each corresponding every 4th failure, producing a graphical representation comprising 47 data points extrapolated from the initial dataset of 188 observations. The manifestation of an out-of-control deviation in the Exponential distribution becomes apparent in the third observation, as the value exceeds the Upper Control Limit (UCL).

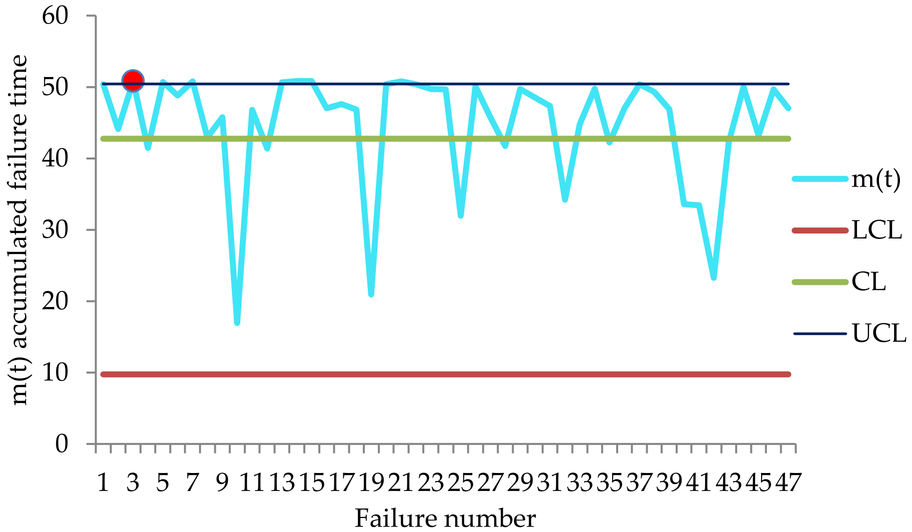


Figure 8: *Control chart based on accumulated failures of mean value function of ED*

Upon reviewing Fig 7 and Fig 8, it is evident that the initial out-of-control deviation is observed in the first observation for the Frechet Distribution (FD) and the third observation for the Exponential Distribution (ED). In both instances, the values surpass the Upper Control Limit (UCL), emphasizing a higher risk associated with quality concerns. The nature of these observations aligns with the principle that lower values indicate better quality. Remarkably, the Frechet Distribution exhibits an out-of-control signal at the UCL earlier than the Exponential Distribution, implying a more timely warning for potential quality issues. Consequently, in this comparative analysis, the Frechet model proves preferable to the exponential model, providing an earlier indication of out-of-control conditions based on the specified criteria.

Based on the comparative analysis of the Frechet distribution (FD) and Exponential distribution (ED), we find that the Frechet model provides an earlier indication of out-of-control conditions, making it a more effective choice for proactive quality control in JET airplanes. The use of anomaly detection and RSF models further strengthens this analysis by identifying outliers and predicting failure times with greater accuracy. These findings suggest that integrating these advanced techniques into the failure analysis process will improve the reliability and maintenance strategies for the aircraft systems, reducing risks and enhancing overall system safety.

III. Conclusion

This study focuses on the application of specialized control charts to assess product reliability, with a central line positioned at the median of varying quality and control limits derived from the quintailed most probable quantile at a coverage probability of 0.9973. The temporal occurrence of failures, defined as the duration between successive failures, serves as the primary variable in this analysis, with a software tool developed to monitor and visualize failure patterns. The quality metric used in this study is based on the average number of failures within a given time interval, with the unique feature of the time control chart emphasizing that points near the Lower Control Limit (LCL) indicate favorable product performance.

In comparing the Frechet Distribution (FD) and Exponential Distribution (ED), the study reveals that the Frechet model provides superior results for failure time analysis. The Frechet distribution showed earlier indications of out-of-control conditions, making it a more effective tool for proactive quality control in JET airplanes. Additionally, the use of Random Survival Forests (RSF) for anomaly detection further validated the superiority of the Frechet distribution, with RSF successfully identifying anomalies in both distributions. However, the Frechet distribution exhibited fewer outliers and more consistent results compared to the Exponential model, which showed more frequent deviations from expected failure patterns. This comparative analysis highlights that the Frechet distribution offers a more reliable and stable model for analyzing failure times, providing valuable insights into the product's performance and reliability. The integration of advanced techniques like anomaly detection and RSF models enhances the accuracy of failure time predictions, contributing to improved reliability and maintenance strategies. These findings suggest that the Frechet distribution, combined with anomaly detection methods, will significantly reduce risks and improve the safety and performance of aircraft systems, making it the better choice for failure analysis in this study.

References

[1] B. Srinivas Rao and P. Sricharani, "Time control charts through NHPP based on Dagum distribution," Int. J. Anal. Appl., vol. 16, no. 3, pp. 328339, Jan.2018, doi: 10.28924/2291-8639-16-2018-328.

[2] D. H. Lee et al., "Software reliability growth model with dependent failures and uncertain operating environments," Appl. Sci., vol. 12, no. 23, p.2383, Dec. 2022, doi: 10.3390/app122312383.

[3] F. Grabski, "Nonhomogeneous Poisson Process and Compound Poisson Process in the modelling of random processes related to road accidents," J. KONES, vol. 26, no. 1, pp. 39–46, Mar. 2019, doi: 10.2478/kones-2019-0005.

[4] H. C. Kim, "Assessing software reliability based on NHPP using SPC," Int. J. Softw. Eng. Appl., vol. 7, no. 6, pp. 61–70, Nov. 2013, doi: 10.14257/ijseia.2013.7.6.06.

[5] H. Pham, "A new software reliability model with Vtub-shaped fault detection rate and the uncertainty of operating environments," Optim., vol. 63, no.10, pp. 1481–1490, Oct. 2014, doi: 10.1080/02331934.2013.854787.

[6] H. Sabahno and A. Amiri, "New statistical and machine learning based control charts with variable parameters for monitoring generalized linearmodel profiles," Comput. Ind. Eng., vol. 184, p. 109562, 2023, doi: 10.1016/j.cie.2023.109562.

[7] M. Xie, T. N. Goh, and P. Ranjan, "Some effective control chart procedures for reliability monitoring," Reliab. Eng. Syst. Saf., vol. 77, pp. 143–150, Aug. 2002, doi: 10.1016/S0951-8320(02)00041-8.

[8] P. Roy, G. S. Mahapatra, and K. N. Dey, "An NHPPsoftware reliability growth model with imperfect debugging and error generation," Int. J. Reliab.Qual. Saf. Eng., vol. 21, no. 1–3, pp. 1–3, Apr. 2014, doi: 10.1142/S0218539314500089.

[9] R. De Blasis et al., "Weighted-indexed semi-Markov model: Calibration and application to financial modeling," Financ. Innov., vol. 9, no. 1, p. 35, Sep. 2023, doi: 10.1186/s40854-022-00418-6.

[10] R. P. de Oleveira et al., "Non-homogeneous Poisson and linear regression models as approaches to study time series with changepoints," Commun. Stat. Case Stud. Data Anal. Appl., vol. 8, no. 2, pp. 331–353, Mar. 2022, doi: 10.1080/23737484.2022.2056546.

[11] T. Salacinski, “Statistical process control using control charts with variable parameters,” Processes, vol. 11, no. 9, p. 2744, Sep. 2023, doi:10.3390/pr11092744.

[12] T. Smajdorova and D. Noskievicova, “Analysis and application of selected control charts suitable for smart manufacturing processes,” Appl. Sci., vol.12, no. 11, p. 5410, May 2022, doi: 10.3390/app12115410. igueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy - a cross-cultural study. Human Nature, 20:317–330.