SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

A Novel Maxwell–EWMA Control Chart Integrated With Support Vector Regression For Biological Water Quality Monitoring

Ishah Maria Mathew and O.S. Deepa

Department of Mathematics, Amrita School of Physical Sciences, Coimbatore, Amrita Vishwa Vidyapeetham, India

ABSTRACT
Rapid industrialization in coastal regions has considerably enhanced the discharge of wastewater enduring several organic and inorganic pollutants into marine ecosystems. Such discharges can adversely distress seawater quality and generate serious environmental and social consequences. Therefore, drastic evaluation and monitoring of marine water quality are essential, particularly through oxygen-related indicators such as Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Dissolved Oxygen (DO), which offer imperative indication about the level of organic pollution and the global health of aquatic systems. In this study, a hybrid statistical monitoring approach is proposed for detecting changes in marine water quality with principal component analysis. The resulting principal component scores are subsequently incorporated into a neutrosophic based Exponentially Weighted Moving Average with Maxwell distributed quality characteristics (NVEWMA) for monitoring framework to improve the sensitivity of the control chart to variations in the underlying water quality process. The integrated PCA–NVEWMA framework is applied to a real-world marine water quality dataset to identifying abnormal variations. The findings demonstrate that the integration of PCA-based feature extraction with the NVEWMA control chart can serve as a valuable approach for continuous marine environmental monitoring and the early detection of undesirable changes in seawater quality.

Keywords: Neutrosophic VEWMA, Machine Learning, PCA, Random Forest, Gradient boosting, SVR.

1. INTRODUCTION

A popular Statistical Process Framework (SPC) framework introduced by [9], the EWMA control chart which is known for its ability to identify slight changes in the process parameters. Compared to the traditional control charts, EWMA provides improved sensitivity by giving more weight to current observations while keeping the past observations. The statistical properties and performance of EWMA charts were studied by [12], providing guidance for parameter selection. An adaptive EWMA chart proposed by [1] where the parameters are automatically adjusted to improve detection performance for various mean shifts, thereby increasing flexibility under varying process conditions.

Extended beyond the traditional methods, several studies have focused on developing control chart for non-normal process A V-type control chart introduced by [8] follows Maxwell distribution, which is designed for monitoring processes, further a Maxwell-CUSUM control chart by [7] was proposed to effectively detect changes in failure rates in manufacturing systems. These approaches often rely on accurate parameter estimation; in this regard, [11] investigated minimax estimation methods for the Maxwell distribution under different loss functions. In addition to all the prior works a statistical modeling based on the distribution of sum of independent gamma random variable was proposed by [13] which has broad application in process monitoring context.

Due to dimensionality and complexity of present-day industrial situations, Machine Learning approaches have gained popularity in SPC framework. These techniques provide enhanced capacity to handle complicated structures and non-linear interactions. To monitor non-linear process [10] proposed model using SVR. Similarly to demonstrate robustness and predictive ability to identify irregular behavior [6,15] have used ensemble learning techniques like Random Forest and Gradient Boosting. To enhance serially correlated data, [14] proposed a sequential learning framework.

Recent research has increasingly focused on extending SPC techniques to account for uncertainty through neutrosophic statistics. In the present-day indeterminacy [2] have proposed the neutrosophic geometric distribution to model data. To enhance monitoring performance [3] has developed neutrosophic EWMA control chart, similarly, to handle imprecise and uncertain data [4,5] has proposed attribute control chart. Together the studies illustrate a growing potential on neutrosophic approach for improving process monitoring while dealing with uncertain conditions.

These advancements demonstrate how crucial it is to combine machine learning methods with conventional SPC tools in order to improve monitoring efficiency and guarantee efficient process variation identification in complicated contexts.

1.1 Research Gap

The key research gaps can be summarized as follows:

1. **Inability to handle uncertain data:** Commonly used VEWMA control charts are generally designed for numerical observations and cannot effectively accommodate neutrosophic data.
2. **Limited focus on process variability:** Most existing neutrosophic control charts focus on monitoring process location, while less attention has been given to monitoring process variability under uncertainty.
3. **Limited integration of VEWMA and Neutrosophic concepts:** There is limited research on integrating the VEWMA control chart framework with neutrosophic theory to effectively monitor processes involving uncertain and indeterminate observations.
4. **Limited real-world applications:** Applications of neutrosophic VEWMA control charts to real-world datasets remain scarce, highlighting the need for practical validation and demonstrating the applicability of such approaches in uncertain real-world environments.

2.1 Neutrosophic Random Variable

Suppose that X_L be a random variable having mean μ and variance σ^2 , a neutrosophic random variable $X_N = X_L + X_L I_N; I_N \in [I_L, I_U]$ where $X_L I_N$ be the indeterminate part and $I_N \in [I_L, I_U]$ be the indeterminacy. When $X_L = 0$, the proposed neutrosophic random variable reverts to the classical random variable [2,3,4]. Neutrosophic logic is the generalization of fuzzy logic where an indeterminacy component is added additionally. Note that $I_{N^2} = I_N, \ldots, I_{N^n} = I_N, 0.I_N = 0; n \in N$

2.2 Machine Learning Techniques

2.2.1 Principal Component Analysis (PCA):

PCA is a linear transformation method that keeps much of the variability in multivariate data while lowering its dimensionality. Considering the data matrix X with a mean of zero, PCA is obtained by breaking down the covariance matrix $\Sigma = \frac{1}{n}X'X$ into its eigenvalues. The eigenvectors that come out of this process are the principal components, and the eigenvalues that go with them show how much variance each component explains. The leading principal components give a representation in a lower dimensional subspace:

$$Z = XW_k$$

where W_k has the eigenvectors that go with the k biggest eigenvalues.

Other machine learning techniques used are:

1. Support Vector Regression (SVR)
2. Boosting
3. Random Forest Regression

3 PROPOSED CONTROL CHART

3.1 Maxwell distribution:

Consider R as a continuous random variable which follows Maxwell distribution with scale parameter σ. The probability density function can be written as:

$$f(r,\sigma) = \sqrt{\frac{2}{\pi}\frac{r^2}{\sigma^3}}e^{-\frac{r^2}{2\sigma^2}}, \quad r > 0,\ \sigma > 0$$

The MLE of σ is defined as $\hat{\sigma} = \sqrt{\frac{1}{3n}\sum_{i=1}^{n} r_i^2}$ (8). Here $V = \hat{\sigma}^{\,2}$ is the square of the estimate following gamma distribution with parameters $\frac{3n}{2}$ and $\frac{2\sigma^2}{3n}$.

Hence

$$\boxed{\hat{\sigma}^{\,2} \sim \text{Gamma}\left(\frac{3n}{2}, \frac{2\sigma^2}{3n}\right)}$$

where $\hat{\sigma}^{\,2}$ = square of the MLE ,$\frac{3n}{2}$ = shape parameter of the Gamma distribution,$\frac{2\sigma^2}{3n}$ = scale parameter of the Gamma distribution,n = sample size and σ = original Maxwell scale parameter.

For a Gamma distribution with shape a and scale b,

$$E(Y) = ab, \qquad Var(Y) = ab^2.$$

Therefore, putting $a = \frac{3n}{2}$, $b = \frac{2\sigma^2}{3n}$, we have $E(\hat{\sigma}^{\,2}) = \frac{3n}{2}\left(\frac{2\sigma^2}{3n}\right) = \boxed{\sigma^2}$

and

$$Var(\hat{\sigma}^{\,2}) = \frac{3n}{2}\left(\frac{2\sigma^2}{3n}\right)^2 = \boxed{\frac{2\sigma^4}{3n}}.$$

3.2 The proposed Neutrosophic VEWMA control chart:

The purpose of this study is to use an EWMA chart to track a Maxwell process in order to identify minute process fluctuations. We can use an EWMA chart to track V as it is a significant statistic for the Maxwell distribution. This charting scheme's plotting statistic is intended to be:

$$Z_{N,i} = \gamma_N V_{N,i} + (1-\gamma_N)Z_{N,i-1},$$

where $Z_{N,i-1}$ contains past information about the scale parameter and $V_{N,i}$ represents the current value of the maximum likelihood estimator of σ_N^2, (for i = 1,2, …). The above equation can be rewritten as

$$Z_{N,i} = \sum_{j=0}^{i-1} \gamma_N (1-\gamma_N)^j V_{N,i-j} + (1-\gamma_N)^i Z_{N,0}.$$

The weights $\gamma_N(1-\gamma_N)^j$ drop exponentially with the age of the sample observations. The in-control average run length will decide which γ_N should be used for process monitoring . $Z_{N,0}$, the starting value for the historical data, is selected to be equal to $\sigma_{N,0}^2$. The average preliminary data can be used to determine the target scale parameter $\sigma_{N,0}^2$ if no information is available. The mean and variance of the VEWMA statistic are σ_N^2 and $\frac{2\sigma_N^4}{3n}$ respectively as V_N has a gamma distribution with these values. Therefore, the mean and variance of VEWMA statistic are

$$E(Z_{N,i}) = \sigma_N^2(1+I_N) \quad and \quad Var(Z_{N,i}) = \frac{2\sigma_N^4}{3n}\left\{\frac{\gamma_N}{2-\gamma_N}\left(1-(1-\gamma_N)^{2i}\right)\right\}(1+I_N)^2.$$

Table 1: L_N factors to estimate $W_{N,i}$ coefficients.

Parameters		False Alarm rate (α)		
Smoothing constant (γ_N)	Sample size (n)	0.005	0.0027	0.002
[0.25, 0.255]	2	[3.070, 3.170]	[3.395, 3.495]	[3.556, 3.656]
	5	[2.893, 2.993]	[3.161, 3.261]	[3.288, 3.388]
	9	[2.830, 2.930]	[3.070, 3.170]	[3.181, 3.281]
[0.5, 0.55]	2	[3.341, 3.441]	[3.713, 3.813]	[3.900, 4.000]
	5	[3.043, 3.143]	[3.358, 3.458]	[3.510, 3.610]
	9	[2.910, 3.010]	[3.195, 3.295]	[3.328, 3.428]
[0.75, 0.75]	2	[3.500, 3.600]	[3.930, 4.030]	[4.126, 4.226]
	5	[3.160, 3.260]	[3.485, 3.585]	[3.652, 3.752]
	9	[3.001, 3.101]	[3.305, 3.405]	[3.447, 3.547]

Therefore, the control limits of a VEWMA control chart are given by,

$$\text{LCL} = W_{N,1}\sigma_N^2(1+I_N),$$

$$\text{CL} = \sigma_N^2(1 + I_N)$$
$$\text{UCL} = W_{N,2}\sigma_N^2(1 + I_N).$$

Here, CL represents the central line and

$$W_{N,1} = \left[1 - L_N\sqrt{\frac{2}{3n} \times \frac{\gamma_N}{2-\gamma_N}(1-(1-\gamma_N)^{2i})}\right],$$

$$W_{N,2} = \left[1 + L\sqrt{\frac{2}{3n} \times \frac{\gamma_N}{2-\gamma_N}(1-(1-\gamma_N)^{2i})}\right].$$

Since $V_{N,i}$ is a gamma random variable and $Z_{N,i}$ is the linear combination of $V_{N,i}$, the distribution of $Z_{N,i}$ is a gamma-series distribution, making it simple to determine the $W_{N,i}$ coefficients for a given L_N which are listed in Table 1.

4 REAL LIFE APPLICATION

The application of the proposed control charts is made on real time data. To show the effect of the proposed work the real time data was taken from https://rtwqmsdb1.cpcb.gov.in/#/ on a particular day for 15 districts . The parameters were BOD , DO, Nitrate, Chloride, Chemical oxygen demand, Total organic carbon in same units. The summary of the statistics and correlation values were noted and given in Table 2 and Table 3.

Table 2 : Summary of the Statistics of the water quality parameters

	DO	BOD	Nitrate	Chloride	Chemical_oxygen_demand	Total_organic_carbon
Min.	2.548	1.070	2.152	9.925	4.538	2.189
1st Qu.	4.262	2.357	2.152	21.855	9.383	4.821
Median	5.293	2.849	2.152	29.719	13.082	6.556
Mean	5.848	3.275	2.774	30.187	13.791	6.730
3rd Qu.	7.278	4.148	3.422	38.466	17.588	8.485
Max.	10.368	6.151	5.074	51.736	26.079	12.581

Thus, summary statistics provide an initial understanding of the distribution, central tendency, and variability of each water-quality parameter.

Table 3 : Correlation coefficient between pairs of water-quality parameters

	DO	BOD	Nitrate	Chloride	Chemical_oxygen_demand	Total_organic_carbon
DO	1.00000000	-0.07195799	-0.1524241989	0.0583579181	-0.1583279	-0.09832026
BOD	-0.07195799	1.00000000	0.1817833564	0.1404385057	0.9547784	0.99173603
Nitrate	-0.15242420	0.18178336	1.0000000000	-0.0009435878	0.2195136	0.17623263
Chloride	0.05835792	0.14043851	-0.0009435878	1.0000000000	0.2347664	0.11747083
Chemical_oxygen_demand	-0.15832789	0.95477843	0.2195136119	0.2347664237	1.0000000	0.96123609
Total_organic_carbon	-0.09832026	0.99173603	0.1762326289	0.1174708291	0.9612361	1.00000000

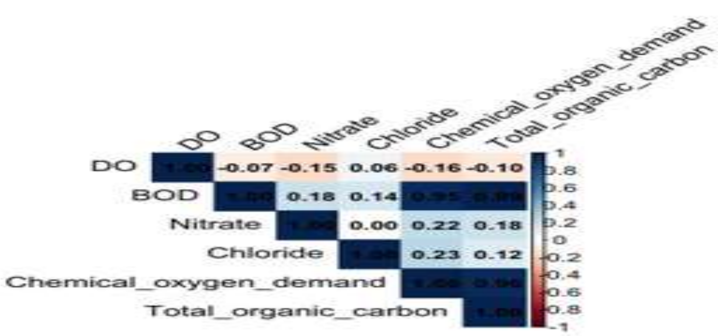

Figure 1 : Correlation coefficient between pairs of water-quality parameters

Correlation analysis is specifically useful because water-quality parameters may be interrelated. It can also aid in recognizing parameters that give analogous information that are important in multivariate analysis especially PCA.
Also, X bar control charts were plotted to check whether the data is out of control . Points outside the control limits or systematic patterns may indicate special-cause variation, such as pollution events, industrial discharge, seasonal effects, or changes in treatment processes.

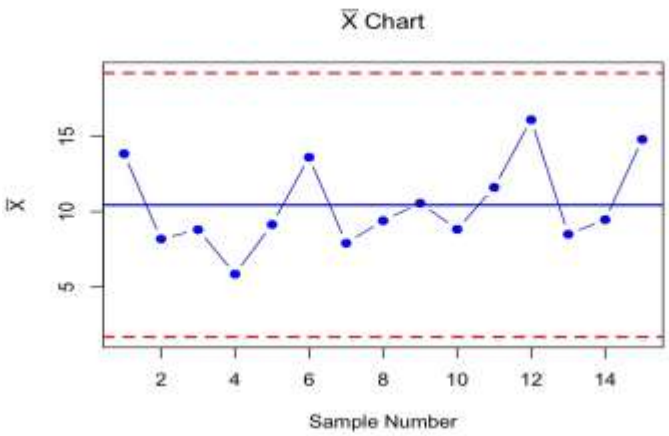

Figure 2 : X bar chart for water quality parameters

The figure shows all points are in control which indicates there is no seasonal variation. Hence by continuously monitoring various time interval, a small scaling factor was fixed so that the data was converted to neutrosophic series. We have considered γ_N as 0.5 and 0.55 with sample size 6 so the L_N will be closer to 3.3 and 3.4. The following steps were used in the proposed work.

1. Initially, mean and standard deviation of the original variables were used to standardize them. The normalized data was then subjected to PCA to get orthogonal principle components via the covariance matrix's eigen decomposition.
2. The main component score vector at time t was calculated as $\boldsymbol{P}_{N,t} = \boldsymbol{Z}_{\boldsymbol{N}_t}\boldsymbol{V}_{\boldsymbol{N}}$, where $\boldsymbol{V}_{\boldsymbol{N}}$ is the matrix of eigenvectors and $\boldsymbol{Z}_{N,t}$ is the standardized observation vector. $V_{N,t} = \frac{1}{3n}\sum_{j=1}^{n} P_{N,tj}^2$, was used as the input to the VEWMA statistics to describe the variability of the PC scores.
3. The VEWMA statistics were updated recursively using the formula $Z_{N,t} = \gamma_N V_{N,t} + (1-\gamma_N)Z_{N,t-1}$,where the smoothing parameter is denoted by γ_N. This integration increases the PCA-VEWMA chart's sensitivity to minor changes while allowing it to track the multivariate process's dominating variation structure.
4. Additionally, by modelling the PC scores and utilizing the generated residual components as inputs to the VEWMA statistic, machine learning models including Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting Machines (GBM) were integrated.

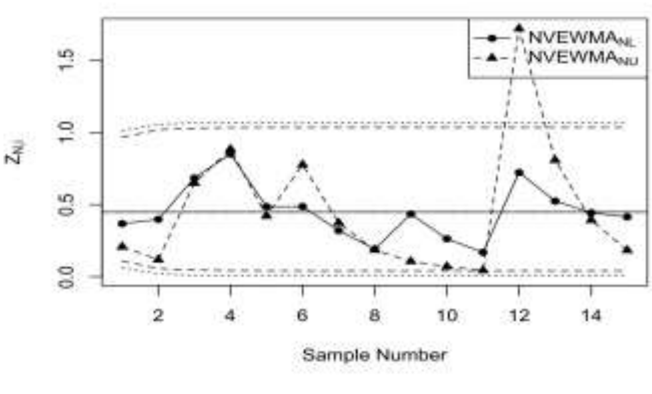

Fig 3. PCA Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2

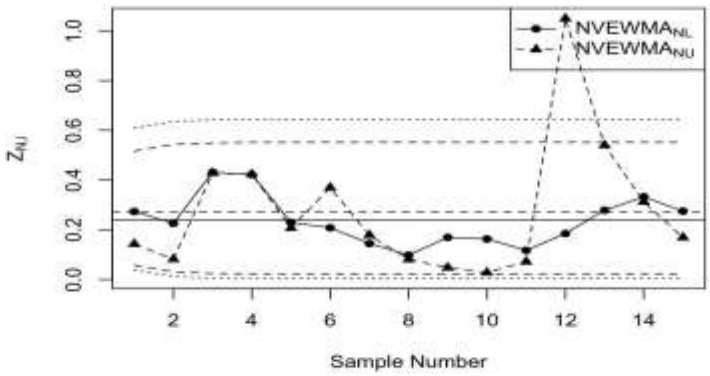

Fig 4. PCA-SVR Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2

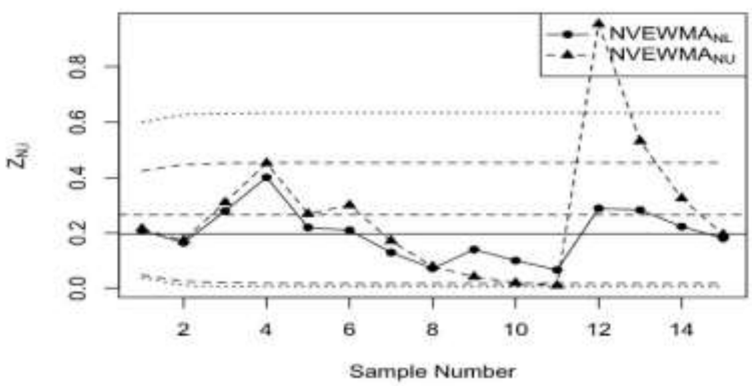

Fig 5. PCA-RF Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2

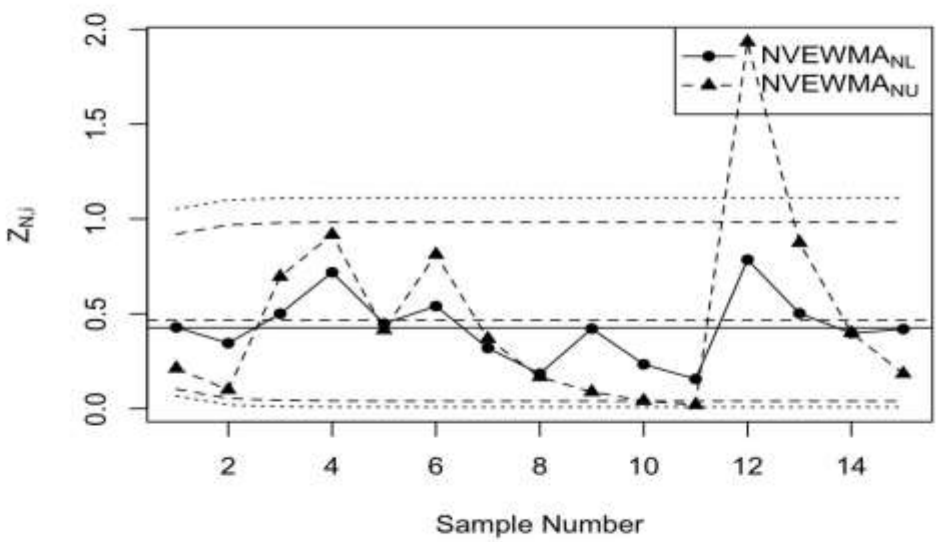

Fig 6 . PCA-GB Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2

These figures 3 -6 provide a clear illustration of how the proposed monitoring framework performs under increasing levels of uncertainty, along with its hybrid extensions that incorporate machine learning techniques such as SVR, RF, and GBM. It can be seen from graphs 3 to 6 that control charts with PCA shows better sensitivity to smaller process changes than control charts without PCA. Fig 3 shows PCA Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2. Here the upper and lower neutrosophic control limits are closer to each other . Therefore, the model shows an OOC situation which is consistent with the unfavorable environmental condition. As the indeterminacy parameter I_N grows the control limits are still wider. Thus the graph shows smooth behavior but with a minor delay in detecting smaller variations. Fig 4 shows PCA-SVR Neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2.The upper and lower neutrosophic control limits are much narrow compared to all other control charts thereby enhance monitoring performance after integrating SVR. The proposed method shows better sensitivity to smaller process changes while preserving robustness to noise. When compared to the baseline model, the SVR based method generates paths with less variability and better stability in the in control region. It is also seen that the OOC points are in the 12th sample for a particular indeterminacy values indicating that the SVR is able to capture non-linear relationships in data.

Fig 5 and 6 gives a clear view of PCA-RF and PCA-GB neutrosophic VEWMA control charts for Indeterminacy values I_N=0.2. The Random Forest model can handle complex interaction between variables and the GB based control chart shows smoother pattern than the Random Forest model but it's very sensitive to smaller shifts. The irregular signals are clearly identified for all levels of uncertainty with better separation on normal and abnormal process behavior. Moreover, the Gradient Boosting model yields a good balance between bias and variance which guarantees consistent and reliable monitoring performance. In both graphs the 12th point is found to be out of control.

Overall, from figures 2–6 ,it is clearly seen that the control limits become wider with the increase of indeterminacy parameter $\mathrm{I_N}$ and greater tolerance to variability, which reflects the capability of the neutrosophic framework in dealing with uncertainty. This reduces the risk of false alarms but may slow down detection of minor modifications. Among the hybrid techniques, PCA-SVR leads to smoother and stable monitoring, PCA-RF produces sharper identification and shifts, and PCA-GB has a balance tradeoff between sensitivity and robustness.

5 Conclusion

The proposed study develops neutrosophic VEWMA control chart for monitoring multivariate marine water quality data in the presence of uncertainty by incooperating PCA and ML models. The developed method combines VEWMA for improving shift detection and neutrosophic concept to handle indeterminacy along with PCA for dimensionality reduction. The results show that PCA-NVEWMA control chart with SVR model ensures smooth and stable monitoring ,whereas RF model sharpens detection and GB model achieves a balance between sensitivity and robustness. Hence , we can conclude that the proposed approach provides an efficient tool for monitoring complex environmental conditions particularly under uncertain and non-linear conditions.

REFERENCES

1. Aly, A.A., Hamed, R.M., Mahmoud, M.A.: Optimal design of the adaptive exponentially weighted moving average control chart over a range of mean shifts. Communications in Statistics – Simulation and Computation **46**(2), 890–902 (2015)
2. Aslam, M., Albassam, M.: Neutrosophic geometric distribution: Data generation under uncertainty and practical applications. AIMS Mathematics **9**(6), 16436–16452 (2024)
3. Aslam, M., Al-Marshadi, A.H., Khan, N.: A new X-bar control chart for using neutrosophic exponentially weighted moving average. Mathematics **7**, 957 (2019)
4. Aslam, M., Bantan, R.A.R., Khan, N.: Design of a new attribute control chart under neutrosophic statistics. International Journal of Fuzzy Systems **21**, 433–440 (2019)
5. Aslam, M., Bantan, R.A., Khan, N.: Design of a new attribute control chart under neutrosophic statistics. International Journal of Fuzzy Systems **21**, 433–440 (2019)
6. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Annals of Statistics **29**(5), 1189–1232 (2001)

7. Hossain, M.P., Sanusi, R.A., Omar, M.H., Riaz, M.: On designing Maxwell CUSUM control chart: An efficient way to monitor failure rates in boring processes. The International Journal of Advanced Manufacturing Technology (2019)
8. Hossain, M.P., Omar, M.H., Riaz, M.: New V control chart for the Maxwell distribution. Journal of Statistical Computation and Simulation **87**(3), 594–606 (2017).
9. Hunter, J.S.: The exponentially weighted moving average. Journal of Quality Technology **18**(4), 203–210 (1986)
10. Khediri, I., Limam, M., Weihs, C.: Support vector regression control chart for monitoring nonlinear profiles. Computational Statistics & Data Analysis **54**(11), 2769–2781 (2010)
11. Li, L.: Minimax estimation of the parameter of Maxwell distribution under different loss functions. American Journal of Theoretical and Applied Statistics **5**(4), 202 (2016).
12. Lucas, J.M., Saccucci, M.S.: Exponentially weighted moving average control schemes: Properties and enhancements. Technometrics **32**(1), 1–12 (1990)
13. Moschopoulos, P.G.: The distribution of the sum of independent gamma random variables. Annals of the Institute of Statistical Mathematics **37**(3), 541–544 (1985)
14. Qiu, P., Xie, X.: Transparent sequential learning for statistical process control of serially correlated data. Technometrics **63**(2), 172–184 (2021)
15. Sabahno, H., Amiri, A.: Machine learning based control charts using random forest for monitoring process parameters. Mathematics **11**(16), 3566 (2023)
16. Mathew, I. M., & Deepa, O. S. (2025). Neutrosophic EWMA and DEWMA control chart on exponential and transformed exponential distributions. International Journal of Neutrosophic Science, 26(4).