

## Research Article

# Generalized Pareto Model: Properties and Applications in Neutrosophic Data Modeling

Zahid Khan <sup>1</sup>, Mohammed M. A. Almazah <sup>2</sup>, Omalsad Hamood Odhah <sup>3</sup>,  
and Huda M. Alshanbari <sup>3</sup>

<sup>1</sup>Department of Mathematics and Statistics, Hazara University Mansehra, Mansehra, Pakistan

<sup>2</sup>Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia

<sup>3</sup>Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Correspondence should be addressed to Zahid Khan; ktshazara@yahoo.com

Received 22 July 2022; Revised 14 August 2022; Accepted 20 August 2022; Published 12 September 2022

Academic Editor: Tahir Mehmood

Copyright © 2022 Zahid Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Pareto distribution is widely used to model industrial, biological, engineering, and other various types of data. A new generalized model, namely the neutrosophic Pareto distribution (NPD), is developed in this article. The proposed model is a neutrosophic variant of the classical Pareto distribution, potentially useful for analyzing vague, unclear, indeterminate, or imprecise data. The structure form of the proposed distribution is skewed to the right and determined to be unimodal. Several characteristics of the NPD are investigated under the neutrosophic framework. The expressions for basic properties such as mean, variance, raw moments, and shape coefficients are obtained. The maximum likelihood approach is presented for estimating the imprecise distributional parameters of the proposed model. The extended notions of the NPD are explained with various key functions in the domain of applied statistical methods. Finally, the practical benefits of NPD are proven by analyzing two real datasets.

## 1. Introduction

The distribution of extreme values for some natural phenomena (e.g., earthquakes, winds, floods, waves, and temperatures) is of relevance in a wide variety of practical situations. For instance, the distributions of high waves in the sea, the distribution of large floods in dams, and so on are important when designing these structures. Extreme value theory has exploded in popularity in recent years as a result of this interest [1]. By the end of the twenty-first century, there is a substantially increased interest in safety and reducing losses from man-induced and natural disasters [2]. The combination of highly felt social needs and the emergence of new theoretical methods has resulted in tremendous progress in this essential multidisciplinary field of research [3]. Note that the study of statistical characteristics for various natural catastrophes is

essentially required not only for understanding the physical nature of the underlying processes but also for risk assessment [4]. Edwards and Das [5] provide a comprehensive list of major domains where heavy-tailed distributions are shown to be useful. The Pareto distribution is one of the heavy-tailed distributions frequently encountered in physical systems to describe different natural disasters (e.g., volcanic eruptions, earthquakes, hurricanes, floods) [6]. The observed fluctuation in sea level, river flow discharges, asteroid craters, wind velocity, forest fires distributions, and some other natural disasters support the potential of a successful Pareto fitting model and obey the Pareto power law [7]. The Pareto distribution and its different variants are also especially well known in the literature for its ability to describe heavy-tailed data, which are commonly found in wealth distribution, actuarial science, life testing, finance, economics,

engineering, and survival analysis [8]. A large range of socioeconomic variables have heavy-tailed distributions that are reasonably well fitted by the Pareto model [9]. The shape of income distributions is governed by some underlying law [10]. Rootzen et al. [11] also listed a number of additional quantities that have been measured in diverse physical, biological, technical, and social systems and for which the Pareto rule has been proven to be a good fit. In short, several studies have asserted that the Pareto model is versatile in modeling many forms of data with large tails.

A traditional method for analyzing extreme values in a population is based on a precisely characterized extreme value Pareto distribution. The customary approach of Pareto distribution is appropriate to use when data consist of a set of exact values or distributional parameters are exactly defined values [12]. However, this strategy has been critiqued since employing exact data results in the loss of information contained in data. Measurements on quantitative variables always have a certain range of inaccuracy [13]. Apart from continuous measurements, there is abundance of situations where exact reporting is impossible due to the irregular nature of circumstances. For example, due to the fluctuating nature of water, the depth of an ocean cannot be exactly quantified but can only be approximated. This issue is remedied by using fuzzy and neutrosophic statistics rather than conventional statistical approaches [14]. The neutrosophic approach is a broader idea that combines a fuzzy concept set with the notion of a classical set [15]. The neutrosophic philosophy takes into account the presence of truth, falsity, and ambiguity [16]. The notion of neutrosophy is now being utilized in a variety of application areas [17–19]. The obtained data may be unclear in a number of real-world scenarios. Several researchers have turned to neutrosophical philosophy to solve the issues of incomplete data [20–22]. In the domain of neutrosophic statistics, conventional statistical approaches have been rigorous in their treatment of ambiguous data processing. New application areas for distribution theory are emerging and demanding further attention. The literature on statistical distributions is dense with several strategies for generalizing continuous distributions in order to improve their ability to describe a variety of datasets.

This study presents the NPD within the neutrosophic framework, thereby enhancing the model's flexibility in dealing with uncertain data sets from a variety of real-world circumstances. This work aims to investigate the usage and implementation of the NPD in healthcare data analysis and to demonstrate the practical advantages of the suggested model.

The remainder of this work is structured as follows: Section 2 contains a description of the proposed and other key characteristics. Simulation studies, including the quantile function of the NPD, are explained in Section 3. The estimation process under the neutrosophic logic is presented in Section 4. In Section 5, a concise explanation of significant theoretical findings is followed by some real-world examples. Section 6 summarizes the findings of the study.

## 2. Proposed Model with Some Useful Properties

This section gives an overview of the suggested distribution and presents it in a coherent framework. The following concepts provide a link between the proposed model and its uses in the applied statistical methods. If the random variable  $X$  with two parameters  $\lambda_n$  and  $\rho_n$  follows the Pareto model, then the density function ( $DF_n$ ) of the proposed distribution is defined as

$$\mathcal{G}_n(\mathcal{X}) = \frac{\lambda_n \rho_n^{\lambda_n}}{x^{\lambda_n+1}} \text{ for } x \geq \rho_n; \lambda_n, \rho_n > 0, \quad (1)$$

where  $\lambda_n = [\lambda_l, \lambda_u]$  and  $\rho_n = [\rho_l, \rho_u]$  are the neutrosophic shape and scale parameters, respectively, of the NPD. Note that the proposed model differs from the existing structure of the classical Pareto model, where shape and scale parameters are precisely determined. When the indeterminate part is considered zero in the proposed model, that is,  $\lambda_l = \lambda_u = \lambda$  and  $\rho_l = \rho_u = \rho$ , it becomes equivalent to the classical model. Various values of  $\lambda_n$  and  $\rho_n$  result in different density curves. A variety of density curves with different neutrosophic shape values and a fixed of scale are plotted in Figure 1.

Figure 1 shows that different indeterminate values of shape parameter resulted in different sturdy curves of the NPD. It is clear from Figure 1 that  $DF_n$  curves are not symmetric and distorted toward the right. The  $DF_n$  curve is portrayed as a thick layer instead of a single curve in the neutrosophic framework. The layer width (shaded area) indicates an imprecision region, and total area under the thick curve is one in view of completeness. Another intriguing aspect in probability theory applications is the neutrosophic cumulative function ( $CF_n$ ) of any density. The  $CF_n$  is a jointly coupled form of the  $DF_n$  is given by

$$\mathcal{G}_n(x) = 1 - \left(\frac{\rho_n}{x}\right)^{\lambda_n}. \quad (2)$$

The  $CF_n$  function estimates the probability that a random variable will have a value smaller than a given value. Figure 2 shows the  $CF_n$  curves for various interval values of the shape parameter of the proposed model.

Figure 2 depicts the cumulative densities of the proposed model for various interval values of shape and fixed value of scale parameter. In each panel of Figure 2, the  $CF_n$  curve is nondecreasing and ranges from 0 to 1. The nondecreasing nature of the  $DF_n$  implies that the DF cannot be negative and true for any distribution. Another useful function in the context of the applied statistical method is the possibility that an individual's life will outlive a certain period of time. This function is referred to as the survival function or simply the survival rate. In the neutrosophic framework, the survival function ( $SF_n$ ) of the proposed model may be represented as follows:

$$\mathcal{S}_n(x) = \left(\frac{\rho_n}{x}\right)^{\lambda_n}. \quad (3)$$

The graph of  $SF_n$  is referred to as a survival curve. Figure 3 depicts the survival curve for the proposed NPD.

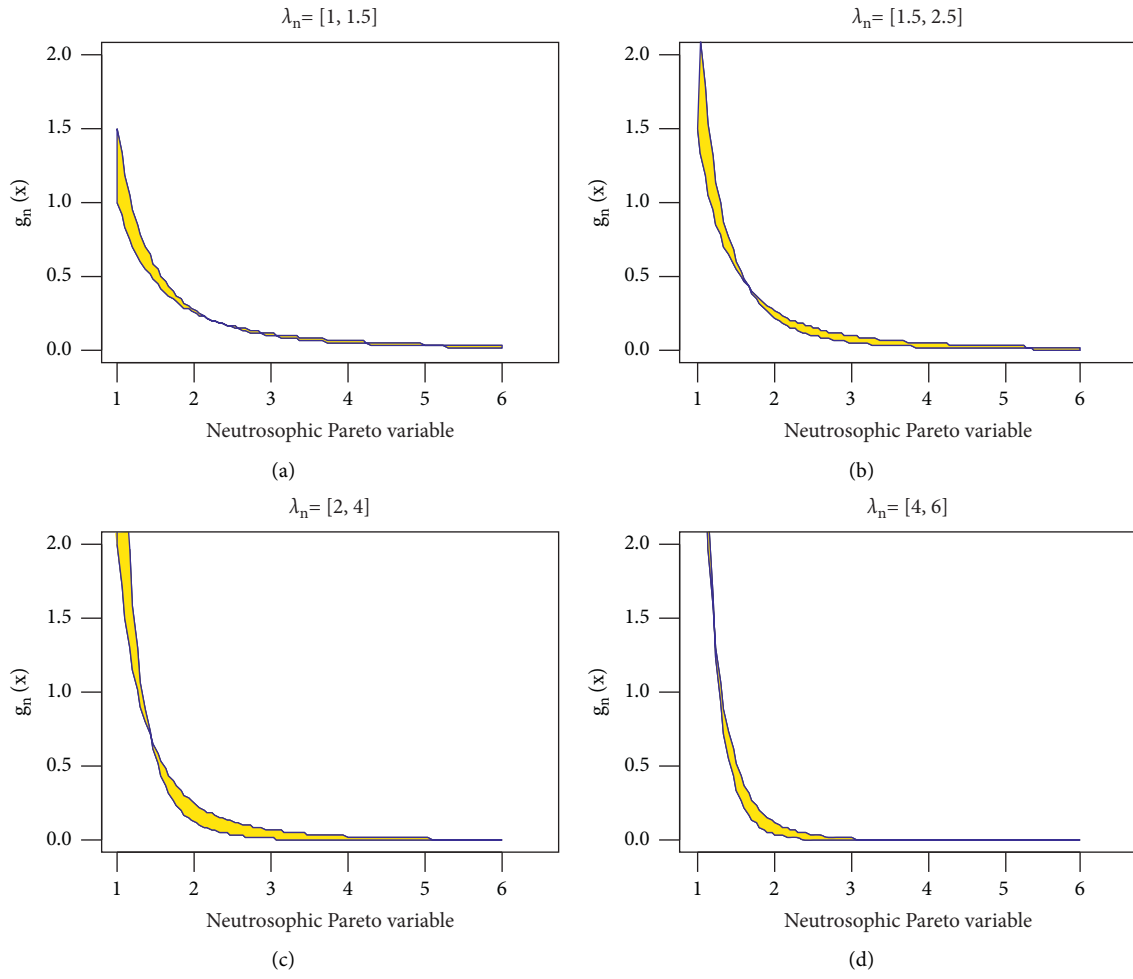


FIGURE 1: The  $DF_n$  curves of NPD at different values of  $\lambda_n$  and  $\rho_n = [1, 1]$ .

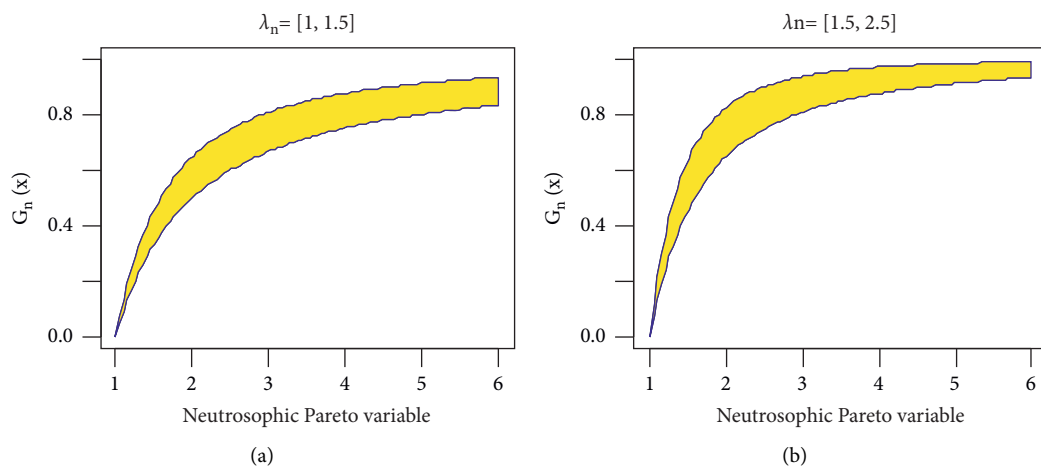


FIGURE 2: Continued.

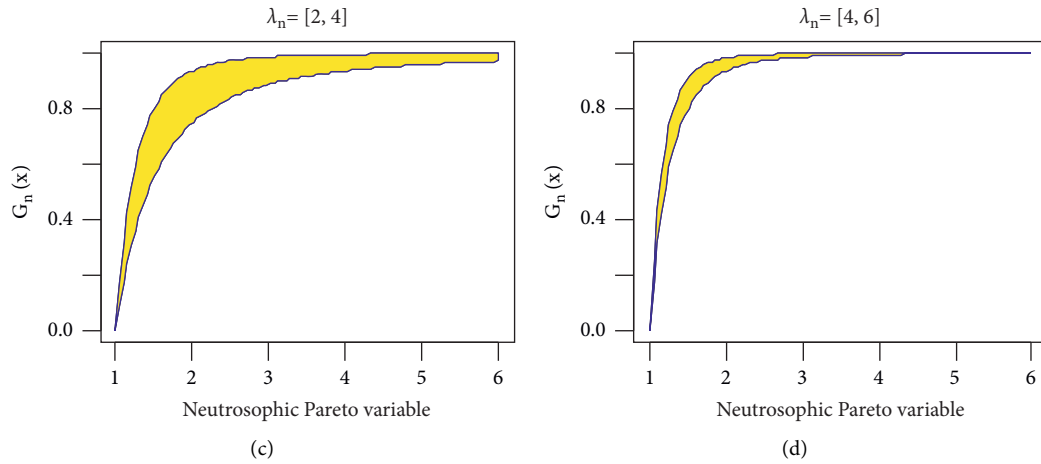


FIGURE 2: The  $CF_n$  curves of NPD at different values of  $\lambda_n$  and  $\rho_n = [1, 1]$ .

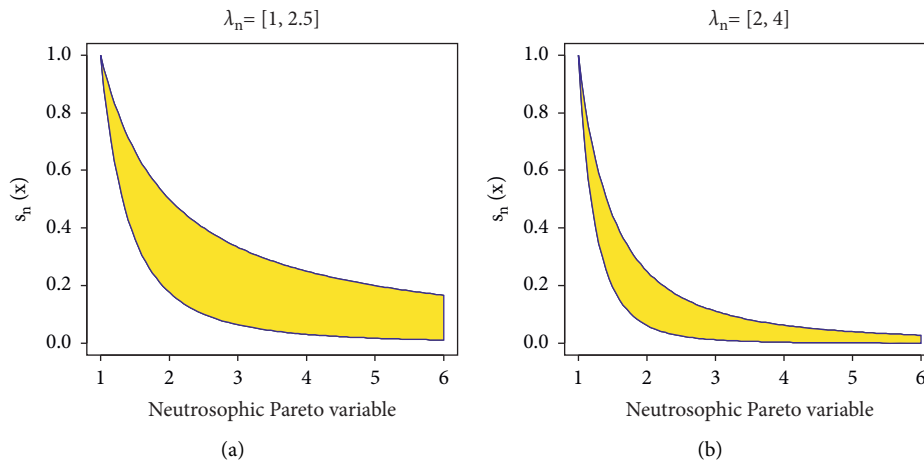


FIGURE 3: Survival curves of the NPD at different values of  $\lambda_n$  and  $\rho_n = [1, 1]$ .

The steep curve can demonstrate a short survival period, or a low survival rate can be shown by the steep curve, as seen in Figure 3(b). A flat or progressive survival curve indicates a longer survival rate, as seen in Figure 3(a).

Another critical function in reliability analysis is the neutrosophic hazard function ( $HF_n$ ), often known as the imminent failure rate. It is the ratio of the survival and density functions and may be calculated as follows for the suggested model:

$$\begin{aligned}
 h_n(x) &= \frac{g_n(x)}{\mathcal{S}_n(x)}, \\
 &= \frac{\lambda_n}{x}.
 \end{aligned}
 \tag{4}$$

The function  $h_n(x)$  provides the failure probability of an individual or item for a minimal time.  $HF_n$  may increase, decrease, stay constant, or reflect a more complex process. The graphical behavior of the hazard curve can be seen in Figure 4.

Figure 4 provides the hazard curves of NPD at the fixed value of the scale parameter and interval values of the shape parameter. Figure 4 indicates the decreasing trends of the hazard curves of the proposed model.

In this section, we have also further investigated the theoretical background and presented some key distributional properties of the proposed NPD in the context of neutrosophic logic. The distributional properties subject to parameterization as given in (1) are given as follows.

**Theorem 1.** *If  $X$  follows the NPD, then  $E(x) = (\lambda_n \rho_n / \lambda_n - 1)$ .*

*Proof.* By definition, the mean of the NPD is given by

$$\begin{aligned}
 E(x) &= \lambda_n \rho_n^{\lambda_n} \int_{\rho_n}^x \frac{x}{x^{\lambda_n+1}} dx, \\
 &= \left[ \lambda_l \rho_l^{\lambda_l} \int_{\rho_l}^x \frac{x}{x^{\lambda_l+1}} dx, \lambda_u \rho_u^{\lambda_u} \int_{\rho_u}^x \frac{x}{x^{\lambda_u+1}} dx \right].
 \end{aligned}
 \tag{5}$$

Equation (5) further yielded

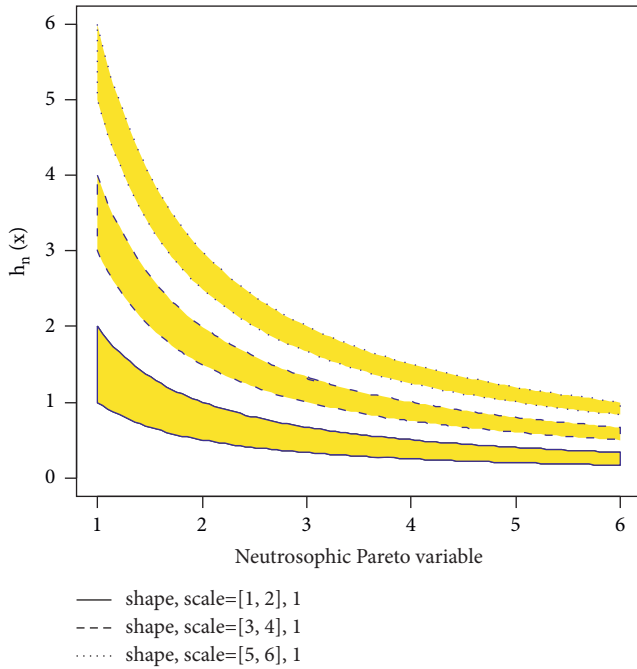


FIGURE 4: Hazard curves for the proposed NPD.

$$\lambda_l \rho_l^{\lambda_l} \int_1^x \frac{x}{\rho_l x^{\lambda_l+1}} dx = \frac{\lambda_l \rho_l}{\lambda_l - 1}, \tag{6}$$

$$\lambda_u \rho_u^{\lambda_u} \int_1^x \frac{x}{\rho_u x^{\lambda_u+1}} dx = \frac{\lambda_u \rho_u}{\lambda_u - 1}.$$

So,

$$\left[ \frac{\lambda_l \rho_l}{\lambda_l - 1}, \frac{\lambda_u \rho_u}{\lambda_u - 1} \right] = \frac{\lambda_n \rho_n}{\lambda_n - 1}, \tag{7}$$

hence proved.  $\square$

**Theorem 2.** If  $X$  follows the NPD, then  $\tilde{V}_n(x) = (\lambda_n \rho_n^2 / ((\lambda_n - 1)^2 (\lambda_n - 2)))$  is the variance of the proposed model.

*Proof.* The variance of the NPD is given by

$$\tilde{V}_n(x) = E(x^2) - [E(x)]^2. \tag{8}$$

Now,

$$E(x^2) = \lambda_n \rho_n^{\lambda_n} \int_{\rho_n}^x \frac{x^2}{x^{\lambda_n+1}} dx, \tag{9}$$

$$= \left[ \lambda_l \rho_l^{\lambda_l} \int_{\rho_l}^x \frac{x^2}{x^{\lambda_l+1}} dx, \lambda_u \rho_u^{\lambda_u} \int_{\rho_u}^x \frac{x^2}{x^{\lambda_u+1}} dx \right].$$

Simplification of (9) provides

$$\left[ \frac{\lambda_l \rho_l}{(\lambda_l - 2)}, \frac{\lambda_u \rho_u}{(\lambda_u - 2)} \right] = \frac{\lambda_n \rho_n}{(\lambda_n - 2)}. \tag{10}$$

Equation (8) thus becomes

$$\tilde{V}_n(x) = \left[ \frac{\lambda_l \rho_l^2}{(\lambda_l - 1)^2 (\lambda_l - 2)}, \frac{\lambda_u \rho_u^2}{(\lambda_u - 1)^2 (\lambda_u - 2)} \right] \tag{11}$$

$$= \frac{\lambda_n \rho_n^2}{(\lambda_n - 1)^2 (\lambda_n - 2)} \text{ for } \lambda_n > 2. \quad \square$$

**Theorem 3.** If  $X$  follows the NPD, then  $2^{1/\lambda_n} \rho_n$  is the median value.

*Proof.* The median point can be derived from the distribution function as

$$\mathcal{G}_n(m_n) = 1 - \left( \frac{\rho_n}{m_n} \right)^{\lambda_n}, \tag{12}$$

$$= \frac{1}{2},$$

where  $m_n$  denotes the neutrosophic median value.

Furthermore, simplification of (12) for  $m_n$  yielded

$$m_n = 2^{1/\lambda_n} \rho_n = [2^{1/\lambda_l} \rho_l, 2^{1/\lambda_u} \rho_u]. \tag{13}$$

**Theorem 4.** The  $j$ th moment of the NPD is  $\lambda_n \rho_n^j / (\lambda_n - j)$

*Proof.* By definition, the  $j$ th moment of the NPD is given by

$$\mu_{jn} = \left[ \lambda_n \rho_n^{\lambda_n} \int_{\rho_n}^x \frac{x^j}{x^{\lambda_n+1}} dx \right], \tag{14}$$

$$= \left[ \lambda_l \rho_l^{\lambda_l} \int_{\rho_l}^x \frac{x^j}{x^{\lambda_l+1}} dx, \lambda_u \rho_u^{\lambda_u} \int_{\rho_u}^x \frac{x^j}{x^{\lambda_u+1}} dx \right].$$

From (14), we can write

$$\lambda_l \rho_l^{\lambda_l} \int_{\rho_l}^x \frac{x^j}{x^{\lambda_l+1}} dx = \frac{\lambda_l \rho_l^j}{(\lambda_l - j)}, \tag{15}$$

$$\lambda_u \rho_u^{\lambda_u} \int_{\rho_u}^x \frac{x^j}{x^{\lambda_u+1}} dx = \frac{\lambda_u \rho_u^j}{(\lambda_u - j)},$$

hence,  $\mu_{jn} = [\lambda_l \rho_l^j / (\lambda_l - j), \lambda_u \rho_u^j / (\lambda_u - j)] = (\lambda_n \rho_n^j / (\lambda_n - j))$  is required result, where  $j = 1, 2, 3, \dots$  is a general expression for the  $j$ th row moment about the origin of the NPD. By using the following relations, moments about the mean for NPD can be derived as

$$\begin{aligned}
\mu'_{1n} &= \mu_{1n} = \frac{\lambda_n \rho_n}{(\lambda_n - 1)}, \\
\mu'_{2n} &= \mu_{2n} - (\mu_{1n})^2, \\
&= \frac{\lambda_n \rho_n^2}{(\lambda_n - 1)^2 (\lambda_n - 2)}, \\
\mu'_{3n} &= \mu_{3n} - 3\mu_{2n}\mu_{1n} + 2(\mu_{1n})^3, \\
&= \frac{2\lambda_n (\lambda_n + 1)\rho_n^3}{(\lambda_n - 1)^3 (\lambda_n - 2)(\lambda_n - 3)}, \\
\mu'_{4n} &= \mu_{4n} - 4\mu_{3n}\mu_{1n} + 6\mu_{2n}\mu_{1n}^2 - 3\mu_{1n}^4, \\
&= \frac{3\lambda_n (3\lambda_n^3 + \lambda_n + 3)\rho_n^4}{(\lambda_n - 1)^4 (\lambda_n - 1)(\lambda_n - 3)(\lambda_n - 4)}. \quad \square
\end{aligned} \tag{16}$$

**Theorem 5.** *The coefficient of skewness of the NPD is  $(2(\lambda_n + 1)/(\lambda_n - 3))\sqrt{\lambda_n - 2/\lambda_n}$ .*

*Proof.* By definition, the coefficient of skewness for NPD is given by

$$\alpha_3 = \frac{\mu'_{3n}}{(\mu'_{2n})^{3/2}}, \tag{17}$$

where  $\mu'_{3n} = (2\lambda_n (\lambda_n + 1)\rho_n^3 / (\lambda_n - 1)^3 (\lambda_n - 2)(\lambda_n - 3))$  and  $\mu'_{2n} = (\lambda_n \rho_n^2 / (\lambda_n - 1)^2 (\lambda_n - 2))$ .

Substituting in (17) yielded

$$\alpha_3 = \frac{2(\lambda_n + 1)}{(\lambda_n - 3)} \sqrt{\frac{\lambda_n - 2}{\lambda_n}}, \tag{18}$$

where  $\alpha_3 \in [\alpha_l, \alpha_u]$ .  $\square$

**Theorem 6.** *The coefficient of kurtosis for NPD is  $6(\lambda_n^3 + \lambda_n^2 - 6\lambda_n - 2)/\lambda_n (\lambda_n - 3)(\lambda_n - 4)$ .*

*Proof.* By definition, the coefficient of kurtosis is given by

$$\alpha_4 = \frac{\mu'_{4n}}{\mu'_{2n}{}^2}, \tag{19}$$

where  $\mu'_{4n} = (3\lambda_n (3\lambda_n^3 + \lambda_n + 3)\rho_n^4 / (\lambda_n - 1)^4 (\lambda_n - 1)(\lambda_n - 3)(\lambda_n - 4))$  and  $\mu'_{2n} = (\lambda_n \rho_n^2 / (\lambda_n - 1)^2 (\lambda_n - 2))$ .

Substituting in (19) yielded

$$\alpha_4 = \frac{6(\lambda_n^3 + \lambda_n^2 - 6\lambda_n - 2)}{\lambda_n (\lambda_n - 3)(\lambda_n - 4)}, \tag{20}$$

where  $\alpha_4 = [\alpha_l, \alpha_u]$ .  $\square$

### 3. Simulation Analysis of the Proposed Model

In this section, a Monte Carlo technique is employed to generate the random numbers that are expected to follow

TABLE 1: A summary of distribution properties of the proposed model based on simulated data.

Basic properties of NPD	Simulated results	Analytical results
Mean	[1.19, 1.33]	[1.20, 1.33]
Median	[1.12, 1.19]	[1.10, 1.19]
Standard deviation	[0.25, 0.47]	[0.25, 0.47]
Skewness	[3.95, 8.06]	[3.94, 8.02]
Kurtosis	[45.73, 350.05]	[45.62, 350.10]

NPD. In general, the Monte Carlo method refers to any technique for solving a problem that makes use of random outcomes. The objective of this study is to test the theoretical findings listed in Section 2 by simulating random samples from the NPD with known parameter values using the Monte Carlo approach. The inverse  $CF_n$  approach has been employed as the most straightforward technique to simulate random numbers from the proposed model. This approach enables us to make use of a computer built-in pseudo-random number generator for generating random numbers. The inverse  $CF_n$  of the proposed model is given by

$$Q_p = \rho_n [1 - U_i]^{-1/\lambda_n}, \tag{21}$$

where  $U_i$  randomly generated numbers from the uniform distribution, and  $Q_p$  is desired percentile value of the proposed NPD. Let  $10^5$  random samples are drawn according to the inverse  $CF_n$  method from the proposed model with  $\rho_n = [1, 1]$  and  $\lambda_n = [4, 6]$ . Analytical outcomes based on the analytical results given in Section 2 are calculated with baseline parameter values. Estimated values of different distribution properties along with exact results are provided in Table 1.

Table 1 displays the descriptive metrics of the proposed model for known distributional parameter values. The descriptive measures of the simulated data using the proposed model are in intervals due to assumed indeterminacies in defined parameters. The basic framework of the proposed model is validated by the strong agreement between simulated and analytical results.

### 4. Estimation of Neutrosophic Parameters

In this part, a well-known maximum likelihood (ML) technique is used to determine the neutrosophic parameters of the proposed NPD. The ML technique is defined by considering the parameters unknown and calculating the joint density of all observations in a dataset that are assumed to be identical and dispersed independently. Once the likelihood of the NPD is established, maxima of the function are determined. These ML estimators are essential in the statistical viewpoint because of minimal variance and asymptotic unbiasedness properties. Let  $x_1, x_2, \dots, x_m$  are identical, and independently observations from the  $m$  subjects follow the parametric model given in (1), and then, the joint density is given by

$$\begin{aligned}
\mathcal{L}(\eta_n, \rho_n | \mathbf{x}) &= \prod_{i=1}^m g_n(x_i | \lambda_n, \rho_n), \\
&= \prod_{i=1}^m \left[ \frac{\lambda_n \rho_n^{\lambda_n}}{x_i^{\lambda_n+1}} \right], \\
&= \lambda_n^m \rho_n^{m\lambda_n} \prod_{i=1}^m \frac{1}{x_i^{\lambda_n+1}}.
\end{aligned} \tag{22}$$

Taking the logarithm of (22) and symbolizing it by  $\omega_n(\mathcal{F}_i | \eta_n, \rho_n)$

$$\omega_n(\mathcal{F}_i | \eta_n, \rho_n) = \log \left[ \lambda_n^m \rho_n^{m\lambda_n} \prod_{i=1}^m \frac{1}{x_i^{\lambda_n+1}} \right]. \tag{23}$$

Simplification of (23) yielded

$$\begin{aligned}
\omega_n(\mathcal{F}_i | \eta_n, \rho_n) &= m \log(\lambda_n) + m \lambda_n \log(\rho_n) \\
&\quad - (\lambda_n + 1) \sum_{i=1}^m \log x_i.
\end{aligned} \tag{24}$$

Partially differentiating (24) by unknown values and equating to zero implies

$$\left[ \frac{\delta \omega_n(x_i, \eta_n)}{\delta \lambda_n}, \frac{\delta \omega_n(x_i, \rho_n)}{\delta \rho_n} \right] = [0, 0]. \tag{25}$$

Further solution of (25) provides the following estimates for unknown parameters of the NPD

$$\begin{aligned}
\hat{\rho}_n &= \text{Min} x_i, \\
\hat{\lambda}_n &= \frac{m}{\sum_{i=1}^m \log x_i - m \log(\rho_n)}.
\end{aligned} \tag{26}$$

Note that  $\hat{\rho}_n$  and  $\hat{\lambda}_n$  will be interval forms because of imprecise sample data. Additionally, we analyze the simulated dataset to demonstrate how the estimation procedure works in neutrosophic environment. Total  $10^4$  different random samples from the NPD are generated with values of  $\lambda_n$  is taken as [4, 6], whereas the value of  $\rho_n$  is fixed at [1, 1]. The behavior of ML estimator from unknown shape parameter and scale parameter is also investigated in terms of neutrosophic root mean square error (NRME). NRME is estimated according to the formula given as follows:

$$\text{NRME} = \sqrt{\sum_{j=1}^M \frac{(\hat{\delta}_j - \delta_j)^2}{M}}, \tag{27}$$

where  $\delta_j$  and  $\hat{\delta}_j$  are, respectively, actual and predicted value of the estimated parameter, and  $M$  is the total number of simulation runs. The R packages EnvStats and Metrics have been utilized to estimate the model's parameters and calculate the values of root mean square error. The estimated values of  $\lambda_n$  at a fixed value of scale parameter along with NRME values are reported in Table 2.

Table 2 shows that when the sample size increases, the value of the estimator tends to the benchmark value [4, 6], and NRME decreases to zero. This trending behavior reveals

TABLE 2: Estimation of the neutrosophic parametric with simulated data.

Sample size	Estimate of $\lambda_n$	NRME
25	[4.338, 6.507]	[0.972, 1.458]
50	[4.162, 6.244]	[0.623, 0.935]
80	[4.109, 6.164]	[0.480, 0.720]
120	[4.070, 6.106]	[0.380, 0.570]
200	[4.042, 6.063]	[0.290, 0.435]

that ML neutrosophic estimators efficiently perform with a larger sample size. We can estimate and observe the performance of the scale parameter  $\rho_n$ , but results are not presented here due to a similar trend.

## 5. Applications of the Proposed Model

Two real datasets are utilized in this section to show how the proposed NPD may be implemented.

**5.1. The Dioxin Data.** Dioxins are a class of very poisonous chemical substances that are dangerous to humans [23]. In the environment, dioxins pose a threat. Dioxins are a matter of concern due to their extremely hazardous potential impact on human health. Experimental studies revealed that they could impact negatively on reproductive, developmental, and immunological systems and organs in the human body [24].

Additionally, they can affect hormones and result in cancer. Once dioxins enter cells, they remain there for an extended period of time due to their chemical persistence and ability to be absorbed by fatty tissue, where they are subsequently retained [25]. Dioxins are mostly produced as a byproduct of industrial operations; however, they can also be produced naturally. Dioxins are unintended byproducts of various manufacturing processes, including chlorine drying of paper pulp, smelting, and the production of some agricultural chemicals [26]. When it comes to dioxin discharge into the environment, unregulated waste incinerators are frequently the greatest offenders, because of incomplete combustion. The vast majority of dioxins in the food supply are found in dairy products, meat, shellfish, and fish. That is why securing our food supply is so important. Although dioxins are formed locally, their environmental dispersion is worldwide. Dioxins are prevalent in the environment on a global scale. Dioxins emission is therefore monitored by many countries on a regular basis. The total amount of dioxins emitted in Japan is monitored on a regular basis by the Ministry of Environment [27]. Because of good government policy, the quantity occurrence in the ecosystem or in food is now extremely small; routine levels of ingestion are extremely unlikely to cause acute toxicity. The current levels of dioxins in Japan indicate an extremely low risk of cancer. To assess the safety of dioxins exposure, the TDI (tolerable daily intake) is employed as an indication. It is the quantity of a chemical substance that may be safely absorbed into the body over a long period of time, per kg of body weight per day, known as TDI. Thus, TDI is a figure that is used as an indicator of how long it will take for daily intake to have an

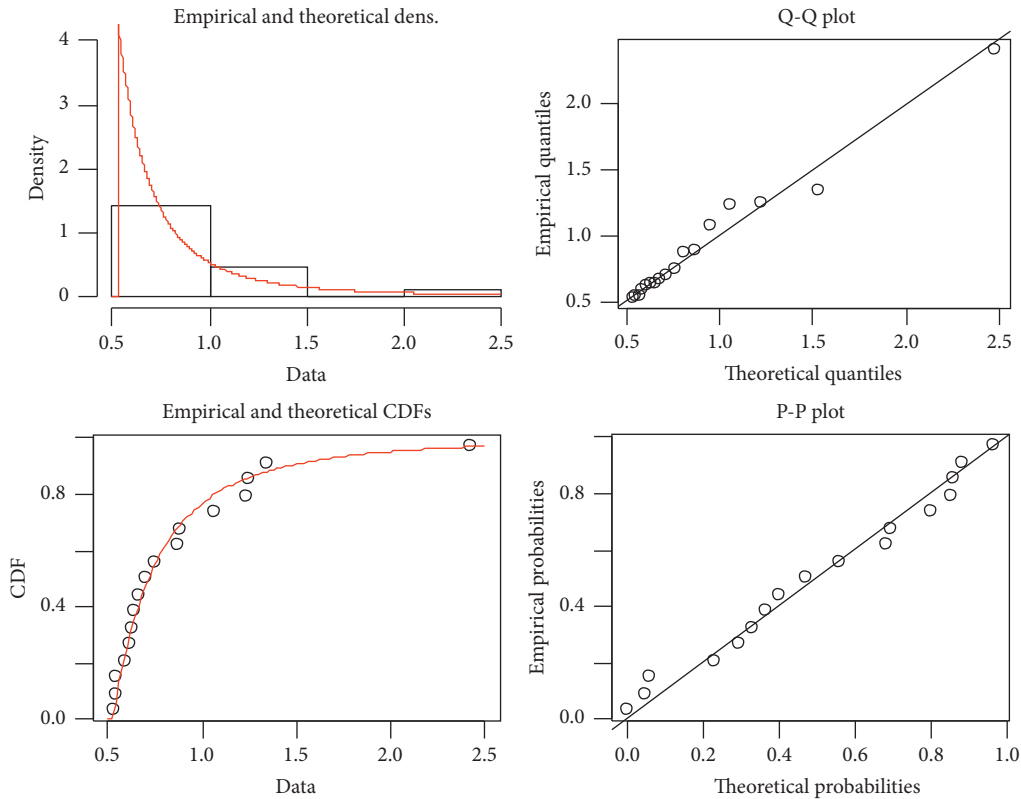


FIGURE 5: Fitting of Pareto distribution on average dioxins consumption from total food samples.

TABLE 3: Estimated dioxins consumption from total food samples collected with uncertainties.

Average daily ingestion of dioxins					
[0.80, 2.06]	[0.51, 1.26]	[0.12, 1.47]	[0.36, 1.47]	[1.04, 1.46]	[0.35, 1.19]
[0.42, 0.65]	[0.02, 0.91]	[0.04, 1.05]	[2.36, 2.44]	[1.029, 1.62]	[1.17, 2.26]
[0.05, 0.96]	[0.48, 1.35]	[0.18, 1.18]	[0.26, 1.24]	[0.14, 0.80]	

effect on health [28]. The amount of dioxins absorption from the average diet estimated annually for the time period 1998 to 2015 is published in the annual report 2017 on environmental statistics by the Ministry of Environment Japan [29]. First, the Pareto distribution on intake data for dioxins is evaluated using the distribution fitting package in *R* software. Figure 5 depicts basic probability plots and empirical density.

When the systematic deviations of the points from the straight line in each graph are considered, it is established that the Pareto distribution is an appropriate model for this dataset. Thus, the visualization plots in Figure 5 show that the process data are skewed. It is, therefore, possible to investigate the data in further depth by applying the previously proposed model. Although intakes of dioxins are initially precise quantities for demonstration purposes, we assume the uncertain sample values as shown in Table 3. The imprecise data are formed according to the strategy devised in [30].

Because of uncertain values, traditional Pareto analysis of these types of data is inappropriate. It is possible to use the suggested NPD to summarize the data containing indeterminacies. Table 4 provides a descriptive summary

TABLE 4: Analysis of dioxins intake dataset using the proposed model.

Summary statistics	Estimated measures
Scale parameter	[0.95, 1.02]
Shape parameter	[2.15, 2.36]
Mean	[1.76, 1.78]
Standard deviation	[1.92, 3.12]

of the consumption of dioxins from a typical diet using NPD.

Table 4 shows the estimated neutrosophic measures based on the suggested model. All the estimated values are expressed as intervals because of indeterminacies inherent in the analyzed dataset. Thus, the suggested model is more adaptable and capable of efficiently analyzing incomplete data or estimating the parameters with imprecision values.

**5.2. The Child Mortality Rate Data.** The second dataset used in this analysis provides the childhood mortality rates under the age of five, covering the period 1995 to 2020 for Saudi Arabia. The information has been gathered from a well-



TABLE 5: Interval estimates of infant death rates for the age less than five years.

Infant mortality rates					
[31.53, 31.81]	[29.33, 30.08]	[27.23, 28.67]	[25.09, 26.34]	[24.20, 24.88]	[22.00, 23.50]
[20.66, 22.09]	[19.74, 20.59]	[18.57, 20.03]	[18.04, 18.77]	[16.89, 17.89]	[15.92, 16.21]
[14.51, 15.92]	[13.92, 14.71]	[12.73, 14.32]	[12.20, 13.35]	[11.18, 12.68 ]	[10.21, 11.75]
[10.12, 11.03]	[9.12, 10.69]	[8.47, 9.42]	[8.59, 9.28]	[7.65, 9.03]	[7.77, 8.59]
[7.23, 7.98]	[6.81, 8.06]				

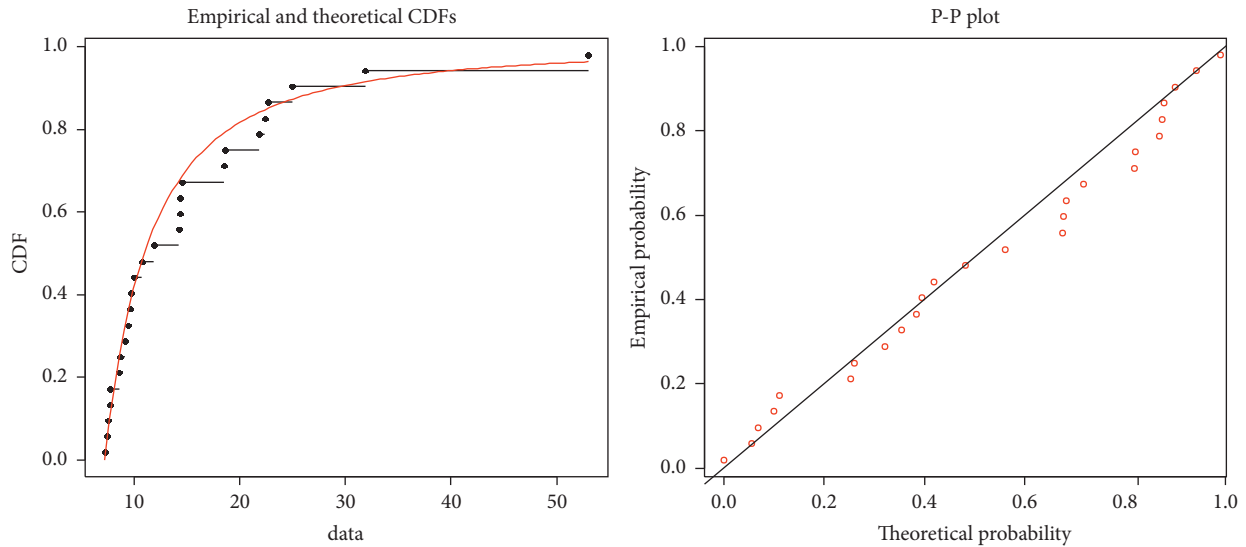


FIGURE 6: Probability plots of the average childhood mortality rates data using the Pareto model.

known source on the WHO’s global health indicators database, and it is usually expressed as a ratio per 1,000 live births [31]. Even though child fatality has significantly decreased at extremely low rates in many regions of the globe, it is still seen as a significant issue that requires great attention in the country’s policies. Significant worldwide progress has been achieved since 1990 in lowering child fatalities. There has been amazing development and remarkable progress in the health of children and adolescents in Saudi Arabia during the past two decades due to major factors such as malnutrition reduction, immunization of infectious illnesses, and diarrhea control [32]. The data from the source are crisp death rate values during the first five years of life. To aid comprehension of the previous notion of the suggested distribution, neutrosophic data are created using the approach provided in [30]. The interval childhood mortality rates for the period 1995–2020 are given in Table 5.

The noticeable uncertainty estimates in Table 5 are due to the fact that different estimation procedures typically used for reporting the mortality rates hinder the exact estimates. Depending on the number of census errors and the various estimating methodologies, there are likely to be fluctuations in estimates for any particular country. A distribution fitting *R* tool is used to depict the basic probability plots to test the applicability of the Pareto distribution on average child death rates, as shown in Figure 6.

In Figure 6, the subjective visual examination of the data suggests that the Pareto distribution is a reasonable model for the mortality data as observations are very close to the

TABLE 6: Descriptive statistics of the child death rates data using suggested distribution.

Descriptive measures	Estimated values
Mean	[1.90, 1.74]
Standard deviation	[0.33, 0.25]
Estimated shape parameter	[6.85, 8.02]
Estimated scale parameter	[1.52, 1.63]

straight line. As interval childhood mortality rates are utilized in this investigation, the conventional Pareto analysis is inapplicable. It is feasible to summarize data, including uncertainties using the proposed model. Using the suggested neutrosophic model, Table 6 displays a descriptive overview of the mortality statistics.

Table 6 provides the estimated uncertainty bounds of some essential statistics based on the proposed distribution. All estimated values are provided as intervals due to the intrinsic imprecision of the dataset being studied. Thus, the proposed model is more flexible and capable of evaluating an imprecise dataset more effectively.

## 6. Conclusions

The neutrosophic framework of the Pareto distribution and its applications in applied statistical methods are presented in this work. Statistical characteristics of the newly proposed model using the neutrosophic logic have been widely explored. The key expressions for the suggested model, such as

cumulative function, hazard function, reliability function, and survival function, have been derived and discussed in detail. The ML estimates for the unknown parameters of NPD have been developed. The theoretical characteristics of the proposed model have been evaluated using the Monte Carlo simulation approach. The effectiveness of the suggested NPD has been demonstrated by using a real dataset on average dioxins consumption data collected from food samples throughout the country of Japan.

A future study might concentrate on enhancing the capacity of the suggested distribution for various inference techniques and its utility for processing high-dimensional data.

## Data Availability

The data that support the findings of this study are available within the manuscript.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups (project under grant no. RGP.2/34/43). This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, through the Research Funding Program (grant no. FRP-1443-21).

## References

- [1] T. G. Bali, "The generalized extreme value distribution," *Economics Letters*, vol. 79, no. 3, pp. 423–427, 2003.
- [2] E. Neumayer, T. Plümper, and F. Barthel, "The political economy of natural disaster damage," *Global Environmental Change*, vol. 24, no. 1, pp. 8–19, 2014.
- [3] S. Hochrainer and R. Mechler, "Natural disaster risk in Asian megacities: a case for risk pooling?" *Cities*, vol. 28, no. 1, pp. 53–61, 2011.
- [4] P. Hoeppe, "Trends in weather related disasters – consequences for insurers and society," *Weather and Climate Extremes*, vol. 11, pp. 70–79, 2016.
- [5] A. S. Edwards and K. Das, "Using statistical approaches to model natural disasters," *Am J Undergrad Res*, vol. 13, no. 2, pp. 87–104, 2016.
- [6] S. Hussain, S. H. Bhatti, T. Ahmad, and M. A. Shehzad, "Parameter estimation of the Pareto distribution using least squares approaches blended with different rank methods and its applications in modeling natural catastrophes," *Natural Hazards*, vol. 107, no. 2, pp. 1693–1708, 2021.
- [7] A. Chatterjee and B. K. Chakrabarti, "Fat tailed distributions for deaths in conflicts and disasters," *Reports in Advances of Physical Sciences*, vol. 1, no. 1, Article ID 1740007, 2017.
- [8] D. Hamed, F. Famoye, and C. Lee, "On families of generalized Pareto distributions: properties and applications," *Journal of Data Science*, vol. 16, no. 2, pp. 377–396, 2021.
- [9] M. Bee and S. Schiavo, "Powerless: gains from trade when firm productivity is not Pareto distributed," *Review of World Economics*, vol. 154, no. 1, pp. 15–45, 2018.
- [10] Y. M. Tripathi, C. Petropoulos, and M. Jha, "Estimation of the shape parameter of a Pareto distribution," *Communications in Statistics - Theory and Methods*, vol. 47, no. 18, pp. 4459–4468, 2017.
- [11] H. Rootzén, J. Segers, and J. L. Wadsworth, "Multivariate generalized Pareto distributions: parametrizations, representations, and properties," *Journal of Multivariate Analysis*, vol. 165, pp. 117–131, 2018.
- [12] Z. Khan, M. Gulistan, N. Kausar, and C. Park, "Neutrosophic Rayleigh model with some basic characteristics and engineering applications," *IEEE Access*, vol. 9, pp. 71277–71283, 2021.
- [13] F. Smarandache, *Introduction to Neutrosophic Measure, Integral, Probability*, Sitech Education publisher, Chennai, India, 2016.
- [14] F. Smarandache, "Neutrosophic set a generalization of the intuitionistic fuzzy sets," *Journal of Defense Resources Management*, vol. 1, no. 1, pp. 107–116, 2010.
- [15] Z. Khan, M. Gulistan, S. Kadry, Y. Chu, and K. Lane-Krebs, "On scale parameter monitoring of the Rayleigh distributed data using a new design," *IEEE Access*, vol. 8, pp. 188390–188400, 2020.
- [16] F. Smarandache, *Neutrosophical Statistics*, Sitech and Education Publishing, Chennai, India, 2014.
- [17] P. Majumdar, "Neutrosophic sets and its applications to decision making," in *Computational Intelligence for Big Data Analysis*, pp. 97–115, Springer, Berlin, Germany, 2015.
- [18] R. A. K. Sherwani, T. Arshad, M. Albassam, M. Aslam, and S. Abbas, "Neutrosophic entropy measures for the Weibull distribution: theory and applications," *Complex & Intelligent Systems*, vol. 7, no. 6, pp. 3067–3076, 2021.
- [19] J. Chen, J. Ye, S. Du, and R. Yong, "Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers," *Symmetry*, vol. 9, no. 7, p. 123, 2017.
- [20] M. A. Khan, A. Amin, S. A. Khan, and M. Gulistan, "Epidemiological studies on gastroenteritis in children in the Bannu district, Khyber Pakhtunkhwa, Pakistan," *Neutrosophic Sets and Systems*, vol. 47, pp. 1–8, 2021.
- [21] W. Q. Duan, Z. Khan, M. Gulistan, and A. Khurshid, "Neutrosophic exponential distribution: modeling and applications for complex data analysis," *Complexity*, vol. 2021, Article ID 5970613, 8 pages, 2021.
- [22] M. Aslam, "Analyzing wind power data using analysis of means under neutrosophic statistics," *Soft Computing*, vol. 25, no. 10, pp. 7087–7093, 2021.
- [23] A. Schecter, L. Birnbaum, J. J. Ryan, and J. D. Constable, "Dioxins: an overview," *Environmental Research*, vol. 101, no. 3, pp. 419–428, 2006.
- [24] R. A. Hites, "Dioxins: an overview and history," *Environmental Science & Technology*, vol. 45, no. 1, pp. 16–20, 2011.
- [25] S. Xu, S. Hansen, A. Rautio et al., "Monitoring temporal trends of dioxins, organochlorine pesticides and chlorinated paraffins in pooled serum samples collected from Northern Norwegian women: the MISA cohort study," *Environmental Research*, vol. 204, Article ID 111980, 2022.
- [26] S. Schröder, M. F. San-Román, and I. Ortiz, "Dioxins and furans toxicity during the photocatalytic remediation of emerging pollutants. Triclosan as case study," *Science of The Total Environment*, vol. 770, Article ID 144853, 2021.
- [27] B. A. Muzembo, M. Iwai-Shimada, T. Isobe et al., "Dioxins levels in human blood after implementation of measures

- against dioxin exposure in Japan,” *Environmental Health and Preventive Medicine*, vol. 24, no. 1, pp. 6–16, 2019.
- [28] N. González and J. L. Domingo, “Polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) in food and human dietary intake: an update of the scientific literature,” *Food and Chemical Toxicology*, vol. 157, Article ID 112585, 2021.
- [29] Annual Report on Environmental Statistics-2017, “Ministry of the environmental , government of Japan,” 2017, <https://www.env.go.jp/>.
- [30] Z. Khan, M. Gulistan, R. Hashim, N. Yaqoob, and W. Chamam, “Design of S-control chart for neutrosophic data: an application to manufacturing industry,” *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 4743–4751, 2020.
- [31] World health organization’s, “Data portal-health indicators,” 2020, <https://childmortality.org/>.
- [32] M. Almuneef, H. Saleheen, F. Albuhairan et al., “Child mortality in Saudi Arabia: time for action at all levels,” *International journal of pediatrics and adolescent medicine*, vol. 8, no. 3, pp. 165–171, 2021.