



---

*Research article*

## **On the implementation of a new version of the Weibull distribution and machine learning approach to model the COVID-19 data**

**Yinghui Zhou<sup>1</sup>, Zubair Ahmad<sup>2</sup>, Zahra Almaspoor<sup>2,\*</sup>, Faridoon Khan<sup>3</sup>, Elsayed tag-Eldin<sup>4</sup>, Zahoor Iqbal<sup>5</sup> and Mahmoud El-Morshedy<sup>6,7</sup>**

<sup>1</sup> School of Information and Communication Engineering, Communication University of China, Beijing, China

<sup>2</sup> Department of Statistics, Yazd University, P. O. Box 89175-741, Yazd, Iran

<sup>3</sup> PIDE School of Economics, PIDE Islamabad 44000, Pakistan

<sup>4</sup> Faculty of Engineering and Technology, Future University in Egypt New Cairo 11835, Egypt

<sup>5</sup> Department of Mathematics, Quaid-i-Azam University, Islamabad 44000, Pakistan

<sup>6</sup> Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

<sup>7</sup> Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

\* **Correspondence:** Email: zahra.ferry21@gmail.com.

**Abstract:** Statistical methodologies have broader applications in almost every sector of life including education, hydrology, reliability, management, and healthcare sciences. Among these sectors, statistical modeling and predicting data in the healthcare sector is very crucial. In this paper, we introduce a new method, namely, a new extended exponential family to update the distributional flexibility of the existing models. Based on this approach, a new version of the Weibull model, namely, a new extended exponential Weibull model is introduced. The applicability of the new extended exponential Weibull model is shown by considering two data sets taken from the health sciences. The first data set represents the mortality rate of the patients infected by the coronavirus disease 2019 (COVID-19) in Mexico. Whereas, the second set represents the mortality rate of COVID-19 patients in Holland. Utilizing the same data sets, we carry out forecasting using three machine learning (ML) methods including support vector regression (SVR), random forest (RF), and neural network autoregression (NNAR). To assess their forecasting performances, two statistical accuracy measures, namely, root mean square error (RMSE) and mean absolute error (MAE) are considered. Based on our findings, it is observed that the RF algorithm is very effective in predicting the death rate of the COVID-19 data in Mexico. Whereas, for the second data, the SVR performs better as compared to the other methods.

**Keywords:** family of distributions; healthcare sector; machine learning algorithms; mathematical properties; simulation; statistical modeling

## 1. Introduction

The first COVID-19 infected case was identified in late 2019 in China and then spread around the globe with extraordinary speed. By March 11, 2021, the COVID-19 confirmed cases were registered in 213 countries, and the World Health Organization (WHO) declared this disease a global pandemic; see Ngo et al. [1]. Due to this pandemic, every aspect of life has been disturbed and almost every region around the globe has faced unexpected situations. In every region, among other sectors affected by this pandemic, the health sector is the most affected area; see Pfefferbaum and North [2], Kim et al. [3], Campion et al. [4], Gloster et al. [5], Talevi et al. [6], and Wastnedge et al. [7]. As of December 13, 2021, 09:09 GMT (Greenwich Mean Time), totally confirmed cases (TCC) have touched a figure of 270488249, the total number of deaths (TND) has reached 5324113, and 243235043 infected persons have been recovered. For the latest updates and details about the COVID-19 events; see <https://www.worldometers.info/coronavirus/>.

The top fifteen countries with the higher TND, include (i) America with 817956 deaths, (ii) Brazil with 616941 deaths, (iii) India with 475636 deaths, (iv) Mexico with 296672 deaths, (v) Russia with 290604 deaths, (vi) Peru with 201,650 deaths, (vii) the United Kingdom with 146439 deaths, (viii) Indonesia with 143936 deaths, (ix) Italy with 134831 deaths, (x) Iran with 130722 deaths, (xi) Colombia with 129107 deaths, (xii) France with 120431 deaths, (xiii) Argentina with 116771 deaths, (ivx) Germany with 106331 deaths, and (xv) Ukraine with 91215 deaths. For a brief overview of country-wise statistics related to the COVID-19 pandemic, we refer to Bo et al. [8].

Due to the unprecedented situation of the COVID-19 pandemic, it is necessary to have the best description and efficient modeling of the COVID-19 events. Statistical methodologies are very useful in modeling and predicting lifetime events. Several statistical studies on this pandemic have appeared. For example, Moreau [9] predicted the COVID-19 phenomena in Brazil. Tuli et al. [10] predicted the growth trend of the COVID-19 pandemic. Rahman et al. [11] implemented the Weibull model for the COVID-19 data analysis. Almetwally [12] introduced a new inverted Topp-Leone (NITL) distribution and used it for modeling the COVID-19 mortality rate data.

These statistical studies are carried out either by implementing the existing models or by proposing new methodologies to update/modify the existing distributions. In the recent era of DT (distribution theory), the development of new methods to introduce new distributions is an important research topic. In this regard, numerous methods to update the distributional flexibility of the existing model have been introduced; see Alizadeh et al. [13] Chipepa et al. [14], Handique et al. [15], Tahir et al. [16], Zaidi et al. [17], Riad et al. [18], and Bakr et al. [19]. For more information about the applicability of statistical models in applied sectors, we refer to Xu et al. [20], and Luo et al. [21].

In this paper, we further contribute to the literature on DT by proposing a new approach, namely, a new extended exponential (NEExp) family. It can be used to obtain the updated versions of the classical/traditional (such as Weibull, beta, gamma, Gumble, Rayleigh, etc.) or other existing models. The NEExp family is proposed by incorporating the T-X distributions approach of Alzaatreh et al. [22]

with the exponential distribution with probability density function (PDF)  $e^{-t}$ , taken as a parent model.

The novelty and key motivations of the proposed method are the followings:

- The method introduced in this paper is new and has not been studied in the literature.
- The proposed method is a very simple and convenient approach of adding an extra parameter to obtain the updated versions of the existing models.
- The proposed approach helps to improve the flexibility and characteristics of the existing models.
- The proposed method provides a close fit to healthcare and other related data sets.

The remaining work carried out in this paper is organized as follows. In Section 2, we define the proposed family and discuss its special case. Certain mathematical properties of the NEEExp family are provided in Section 3. The estimation of the parameters and a simulation study are provided in Section 4. To illustrate the NEEExp family, two practical data sets are analyzed in Section 5. To forecast the COVID-19 data sets, the machine learning methods are discussed in Section 6. Finally, some concluding remarks, limitations of the proposed method, and future study plans are discussed in Section 7.

## 2. The proposed method and its special model

This section is divided into two subsections. In the very first subsection, we define the proposed NEEExp family of distributions. Whereas, the second subsection is devoted to studying a special case of the NEEExp family of distributions

### 2.1. The proposed method

Let  $V$  has a NEEExp family, if its distribution function (DF)  $F(v; \delta, \boldsymbol{\theta})$  is given by

$$F(v; \delta, \boldsymbol{\theta}) = 1 - \frac{\delta [1 - M(v; \boldsymbol{\theta})]}{\delta + M(v; \boldsymbol{\theta})}, \quad v \in \mathbb{R}, \quad (2.1)$$

where  $\delta > 0$  is an additional parameter and  $\boldsymbol{\theta}$  is a vector of parameters associated with the baseline DF  $M(v; \boldsymbol{\theta})$ .

Furthermore, in link to  $F(v; \delta, \boldsymbol{\theta})$ , the PDF  $f(v; \delta, \boldsymbol{\theta})$  of the NEEExp family is

$$f(v; \delta, \boldsymbol{\theta}) = \frac{\delta(\delta + 1)m(v; \boldsymbol{\theta})}{[\delta + M(v; \boldsymbol{\theta})]^2}, \quad v \in \mathbb{R}. \quad (2.2)$$

For  $v \in \mathbb{R}$ , the survival function (SF)  $S(v; \delta, \boldsymbol{\theta}) = 1 - F(v; \delta, \boldsymbol{\theta})$ , and hazard function (HF)  $h(v; \delta, \boldsymbol{\theta}) = \frac{f(v; \delta, \boldsymbol{\theta})}{1 - F(v; \delta, \boldsymbol{\theta})}$  of the NEEExp family are given by

$$S(v; \delta, \boldsymbol{\theta}) = \frac{\delta [1 - M(v; \boldsymbol{\theta})]}{\delta + M(v; \boldsymbol{\theta})},$$

and

$$h(v; \delta, \boldsymbol{\theta}) = \frac{(\delta + 1)m(v; \boldsymbol{\theta})}{[1 - M(v; \boldsymbol{\theta})][\delta + M(v; \boldsymbol{\theta})]},$$

respectively.

A special member of the NEExp family called, a new extended exponential Weibull (NEExp-Weibull) model is discussed in the next subsection. The NEExp-Weibull model is introduced by using the DF of the Weibull model in Eq (2.1). The DF  $M(v; \boldsymbol{\theta})$  of the Weibull model is given by

$$M(v; \boldsymbol{\theta}) = 1 - e^{-\varphi_2 v^{\varphi_1}}, \quad v \geq 0, \varphi_1 > 0, \varphi_2 > 0, \quad (2.3)$$

with PDF  $m(v; \boldsymbol{\theta})$  given by

$$m(v; \boldsymbol{\theta}) = \varphi_1 \varphi_2 v^{\varphi_1-1} e^{-\varphi_2 v^{\varphi_1}}, \quad v > 0, \varphi_1 > 0, \varphi_2 > 0,$$

where  $\boldsymbol{\theta} = (\varphi_1, \varphi_2)$ . By incorporating Eq (2.3) in Eq (2.1), we reach at the DF of the NEExp-Weibull model; see Subsection 2.2.

## 2.2. The NEExp-Weibull distribution: a special model

A random variable  $V$  has the NEExp-Weibull model with parameters  $\varphi_1 > 0, \varphi_2 > 0$ , and  $\delta > 0$ , if its DF  $F(v; \delta, \boldsymbol{\theta})$  and PDF  $f(v; \delta, \boldsymbol{\theta})$ , are given by

$$F(v; \delta, \boldsymbol{\theta}) = 1 - \frac{\delta e^{-\varphi_2 v^{\varphi_1}}}{\delta + 1 - e^{-\varphi_2 v^{\varphi_1}}}, \quad v \geq 0, \quad (2.4)$$

and

$$f(v; \delta, \boldsymbol{\theta}) = \frac{\delta(\delta + 1) \varphi_1 \varphi_2 v^{\varphi_1-1} e^{-\varphi_2 v^{\varphi_1}}}{[\delta + 1 - e^{-\varphi_2 v^{\varphi_1}}]^2}, \quad v > 0, \quad (2.5)$$

respectively.

For the NEExp-Weibull model with DF in Eq (2.4) and PDF in Eq (2.5), the SF  $\bar{F}(v; \delta, \boldsymbol{\theta}) = 1 - F(v; \delta, \boldsymbol{\theta})$ , HF  $h(v; \delta, \boldsymbol{\theta}) = \frac{f(v; \delta, \boldsymbol{\theta})}{\bar{F}(v; \delta, \boldsymbol{\theta})}$ , and CHF  $H(v; \delta, \boldsymbol{\theta}) = -\log(\bar{F}(v; \delta, \boldsymbol{\theta}))$  are given by

$$\bar{F}(v; \delta, \boldsymbol{\theta}) = \frac{\delta e^{-\varphi_2 v^{\varphi_1}}}{\delta + 1 - e^{-\varphi_2 v^{\varphi_1}}}, \quad v > 0,$$

$$h(v; \delta, \boldsymbol{\theta}) = \frac{(\delta + 1) \varphi_1 \varphi_2 v^{\varphi_1-1}}{[\delta + 1 - e^{-\varphi_2 v^{\varphi_1}}]}, \quad v > 0,$$

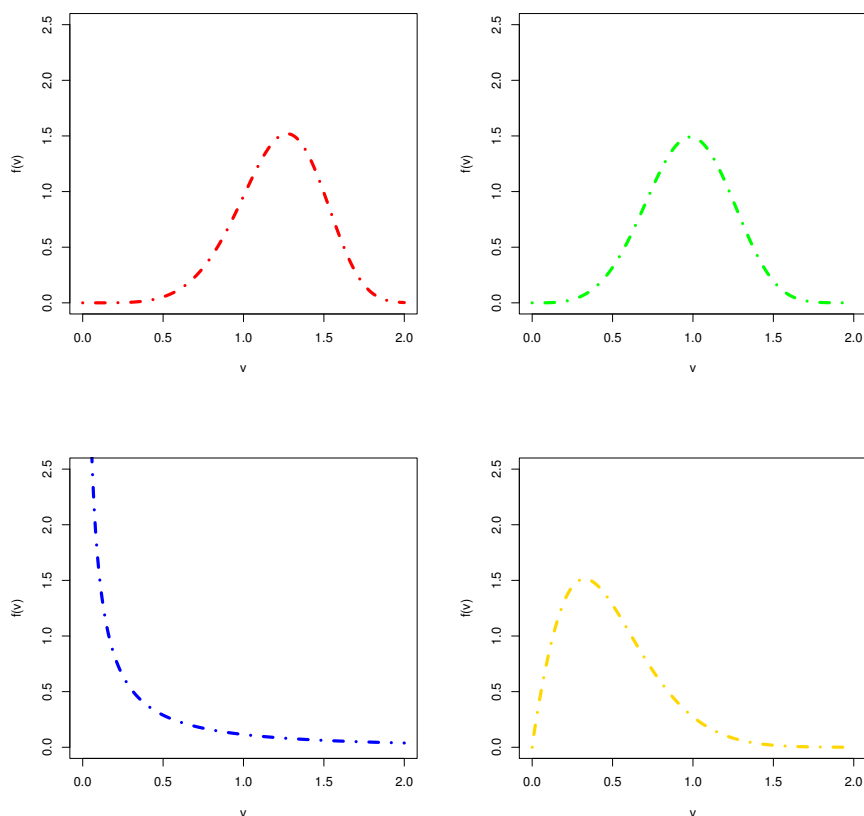
and

$$H(v; \delta, \boldsymbol{\theta}) = -\log\left(\frac{\delta e^{-\varphi_2 v^{\varphi_1}}}{\delta + 1 - e^{-\varphi_2 v^{\varphi_1}}}\right), \quad v > 0,$$

respectively.

A visual behavior of  $f(v; \delta, \boldsymbol{\theta})$  for (i)  $\varphi_1 = 5.5, \varphi_2 = 0.2, \delta = 9.5$  (red curve) (ii)  $\varphi_1 = 4.5, \varphi_2 = 0.6, \delta = 2.5$ , (green curve) (iii)  $\varphi_1 = 0.5, \varphi_2 = 1.6, \delta = 1.5$ , (blue curve) and (iv)  $\varphi_1 = 1.9, \varphi_2 = 2.6, \delta = 2.5$ , (gold curve) is provided in Figure 1.

From the visual illustration in Figure 1, we can see that  $f(v; \delta, \boldsymbol{\theta})$  possess different behaviors. For example, it takes (i) the left-skewed form (red curve), (ii) the symmetrical shape (green curve), (iii) the reverse-J shape (blue curve), and (iv) the right-skewed (gold curve).



**Figure 1.** A visual behavior of  $f(v; \delta, \boldsymbol{\theta})$  for different values of  $\varphi_1, \varphi_2$ , and  $\delta$ .

### 3. Mathematical properties

Here, we obtain different mathematical properties of the NEExp family with PDF  $f(v; \delta, \boldsymbol{\theta})$ . These properties include QF (quantile function) expressed by  $Q(u)$ ,  $r^{th}$  moment denoted by  $\mu_r'$ , MGF (moment generating function) represented by  $M_V(t)$ , RL (residual life), and RRL (reverse residual life) functions.

#### 3.1. The quantile function

The QF of the NEExp family can be obtained by inverting Eq (2.1). Let  $V$  have the NEExp family with DF  $F(v; \delta, \boldsymbol{\theta})$ , then, its QF is given by

$$v_q = Q(u) = F^{-1}\left(\frac{\delta u}{\delta + 1 - u}\right),$$

where  $u \in (0, 1)$ .

#### 3.2. The $r^{th}$ moment

Suppose  $V$  follows the NEExp family of distributions, then the  $r^{th}$  moment of the NEExp distributions is derived as

$$\mu'_r = \int_{\Omega} v^r \frac{\delta(\delta+1)m(v;\boldsymbol{\theta})}{[\delta + M(v;\boldsymbol{\theta})]^2} dv,$$

$$\mu'_r = \frac{(\delta+1)}{\delta} \int_{\Omega} v^r \frac{m(v;\boldsymbol{\theta})}{\left[1 + \frac{M(v;\boldsymbol{\theta})}{\delta}\right]^2} dv. \quad (3.1)$$

Consider the series

$$\frac{1}{(1+k)^2} = \sum_{a=1}^{\infty} (-1)^{a-1} a k^{a-1}. \quad (3.2)$$

By implementing Eq (3.2) with  $k = \frac{M(v;\boldsymbol{\theta})}{\delta}$ , we have

$$\frac{1}{\left(1 + \left[\frac{M(v;\boldsymbol{\theta})}{\delta}\right]\right)^2} = \sum_{a=1}^{\infty} (-1)^{a-1} a \left(\frac{M(v;\boldsymbol{\theta})}{\delta}\right)^{a-1}. \quad (3.3)$$

Using Eq (3.3) in Eq (3.1), we get

$$\mu'_r = \frac{(\delta+1)}{\delta^a} \sum_{a=1}^{\infty} (-1)^{a-1} a \int_{\Omega} v^r m(v;\boldsymbol{\theta}) M(v;\boldsymbol{\theta})^{a-1} dv,$$

$$\mu'_r = \frac{(\delta+1)}{\delta^a} \sum_{a=1}^{\infty} (-1)^{a-1} \int_{\Omega} v^r k_a(v;\boldsymbol{\theta}) dv, \quad (3.4)$$

where  $k_a(v;\boldsymbol{\theta}) = a m(v;\boldsymbol{\theta}) M(v;\boldsymbol{\theta})^{a-1}$  is the exponentiated PDF, and  $a$  is a power parameter. We can also write Eq (3.4), as follows

$$\mu'_r = \frac{(\delta+1)}{\delta^a} \sum_{a=1}^{\infty} (-1)^{a-1} \lambda_{a,r}, \quad (3.5)$$

where

$$\lambda_{a,r} = \int_{\Omega} v^r k_a(v;\boldsymbol{\theta}) dv. \quad (3.6)$$

Using the DF and PDF of the Weibull model in Eq (3.6), we get

$$\lambda_{a,r} = \int_0^{\infty} v^r \varphi_1 \varphi_2 v^{\varphi_1-1} e^{-\varphi_2 v^{\varphi_1}} \left(1 - e^{-\varphi_2 v^{\varphi_1}}\right)^{a-1} dv. \quad (3.7)$$

On solving Eq (3.7), we get

$$\lambda_{a,r} = \sum_{k=0}^{a-1} (-1)^k \binom{a-1}{k} \int_0^{\infty} v^r \varphi_1 \varphi_2 v^{\varphi_1-1} e^{-\varphi_2(k+1)v^{\varphi_1}} dv,$$

$$\lambda_{a,r} = \sum_{k=0}^{a-1} (-1)^k \binom{a-1}{k} \frac{\Gamma\left(\frac{r}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{r}{\varphi_1}} (k+1)^{\frac{r}{\varphi_1} + 1}}. \quad (3.8)$$

Using Eq (3.8) in Eq (3.5), we get the  $r^{th}$  moment of the NEEExp-Weibull model, given by

$$\mu'_r = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} (-1)^{a+k-1} \binom{a-1}{k} \frac{\Gamma\left(\frac{r}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{r}{\varphi_1}} (k+1)^{\frac{r}{\varphi_1}+1}}. \quad (3.9)$$

Using  $r = 1$ , in Eq (3.9), we get the first  $r^{th}$  moment of the NEEExp-Weibull model, given by

$$\mu'_1 = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} (-1)^{a+k-1} \binom{a-1}{k} \frac{\Gamma\left(\frac{1}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{1}{\varphi_1}} (k+1)^{\frac{1}{\varphi_1}+1}}.$$

Using  $r = 2$ , in Eq (3.9), we obtain the second  $r^{th}$  moment of the NEEExp-Weibull model, given by

$$\mu'_2 = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} (-1)^{a+k-1} \binom{a-1}{k} \frac{\Gamma\left(\frac{2}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{2}{\varphi_1}} (k+1)^{\frac{2}{\varphi_1}+1}}.$$

Using  $r = 3$ , in Eq (3.9), we obtain the third  $r^{th}$  moment of the NEEExp-Weibull model, given by

$$\mu'_3 = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} (-1)^{a+k-1} \binom{a-1}{k} \frac{\Gamma\left(\frac{3}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{3}{\varphi_1}} (k+1)^{\frac{3}{\varphi_1}+1}}.$$

Using  $r = 4$ , in Eq (3.9), we obtain the second  $r^{th}$  moment of the NEEExp-Weibull model, given by

$$\mu'_4 = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} (-1)^{a+k-1} \binom{a-1}{k} \frac{\Gamma\left(\frac{4}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{4}{\varphi_1}} (k+1)^{\frac{4}{\varphi_1}+1}}.$$

Furthermore, the MGF of the NEEExp-Weibull model is given by

$$M_V(t) = \frac{(\delta + 1)}{\delta^a} \sum_{a=1}^{\infty} \sum_{k=0}^{a-1} \sum_{r=0}^{\infty} \frac{(-1)^{a+k-1} t^r}{r!} \binom{a-1}{k} \frac{\Gamma\left(\frac{r}{\varphi_1} + 1\right)}{(\varphi_2)^{\frac{r}{\varphi_1}} (k+1)^{\frac{r}{\varphi_1}+1}}.$$

### 3.3. The residual life

The RL of the NEEExp-Weibull model represented by  $R_t(v)$ , is given by

$$R_t(v) = \frac{\delta e^{-\varphi_2(v+t)\varphi_1}}{\delta + 1 - e^{-\varphi_2(v+t)\varphi_1}} \times \frac{\delta + 1 - e^{-\varphi_2 v \varphi_1}}{\delta e^{-\varphi_2 v \varphi_1}}.$$

Furthermore, the RRL of the NEEExp-Weibull model represented by  $\bar{R}_t(v)$ , is given by

$$\bar{R}_t(v) = \frac{\delta e^{-\varphi_2(v-t)\varphi_1}}{\delta + 1 - e^{-\varphi_2(v-t)\varphi_1}} \times \frac{\delta + 1 - e^{-\varphi_2 v \varphi_1}}{\delta e^{-\varphi_2 v \varphi_1}}.$$

## 4. Estimation and simulation study

In this section, we obtain the estimators  $(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\delta})$  of the parameters  $(\varphi_1, \varphi_2, \delta)$  by implementing the maximum likelihood estimation approach. Furthermore, for the evaluation of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$ , a simulation is also provided.

#### 4.1. Estimation

Let  $V_1, V_2, \dots, V_p$  be a sample of size  $p$  observed from the PDF  $f(v; \delta, \boldsymbol{\theta})$ . In link to  $f(v; \delta, \boldsymbol{\theta})$ , the likelihood function (LH)  $\lambda(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p)$  is given by

$$\lambda(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = \prod_{a=1}^p f(v_a; \delta, \boldsymbol{\theta}). \quad (4.1)$$

Using Eq (2.5) in Eq (4.1), we get

$$\lambda(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = \prod_{a=1}^p \frac{\delta(\delta+1)\varphi_1\varphi_2 v_a^{\varphi_1-1} e^{-\varphi_2 v_a^{\varphi_1}}}{\left[\delta+1-e^{-\varphi_2 v_a^{\varphi_1}}\right]^2}. \quad (4.2)$$

Corresponding to  $\lambda(\varphi_1, \varphi_2, \delta|v_1, v_2, \dots, v_p)$ , the log LH  $\pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p)$  is given by

$$\begin{aligned} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) &= p \log \delta + p \log(\delta+1) + p \log \varphi_1 + p \log \varphi_2 + (\varphi_1-1) \sum_{a=1}^p \log v_a \\ &\quad - \sum_{a=1}^p \varphi_2 v_a^{\varphi_1} - 2 \sum_{a=1}^p \log(\delta+1-e^{-\varphi_2 v_a^{\varphi_1}}). \end{aligned}$$

The partial derivatives of  $\pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p)$  are given by

$$\begin{aligned} \frac{\partial}{\partial \varphi_1} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) &= \frac{p}{\varphi_1} + \sum_{a=1}^p \log v_a - \varphi_2 \sum_{a=1}^p (\log v_a) v_a^{\varphi_1} \\ &\quad - 2\varphi_2 \sum_{a=1}^p \frac{(\log v_a) v_a^{\varphi_1} e^{-\varphi_2 v_a^{\varphi_1}}}{(\delta+1-e^{-\varphi_2 v_a^{\varphi_1}})}, \end{aligned}$$

$$\frac{\partial}{\partial \varphi_2} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = \frac{p}{\varphi_2} - \sum_{a=1}^p v_a^{\varphi_1} - 2 \sum_{a=1}^p \frac{v_a^{\varphi_1} e^{-\varphi_2 v_a^{\varphi_1}}}{(\delta+1-e^{-\varphi_2 v_a^{\varphi_1}})},$$

and

$$\frac{\partial}{\partial \delta} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = \frac{p}{\delta} + \frac{p}{(\delta+1)} - 2 \sum_{a=1}^p \frac{1}{(\delta+1-e^{-\varphi_2 v_a^{\varphi_1}})},$$

respectively.

On solving  $\frac{\partial}{\partial \varphi_1} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = 0$ ,  $\frac{\partial}{\partial \varphi_2} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p) = 0$ , and  $\frac{\partial}{\partial \delta} \pi(\delta, \boldsymbol{\theta}|v_1, v_2, \dots, v_p)$ , we get the estimators  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$ , respectively.



#### 4.2. Simulation study

Now, we evaluate the performances of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$  by conducting a SiSt (simulation study). The SiSt is performed by adopting the following steps

- To carry out the SiSt, the RNs (random numbers) from the NEEExp-Weibull model are generated using the inverse DF, given by

$$v_q = Q(u) = F^{-1}\left(\frac{\delta u}{\delta + 1 - u}\right).$$

- The SiSt is performed for two different combination sets of  $\varphi_1$ ,  $\varphi_2$ , and  $\delta$ , such as (a)  $\varphi_1 = 0.7$ ,  $\varphi_2 = 0.9$ ,  $\delta = 1.4$ , and (b)  $\varphi_1 = 1.5$ ,  $\varphi_2 = 1.4$ ,  $\delta = 1.25$ .
- For both two sets of  $\varphi_1$ ,  $\varphi_2$ , and  $\delta$ , RNs of sizes  $p = 25, 50, 75, \dots, 500$  are generated using the inverse DF method.
- The numerical values of the maximum likelihood estimators (MLEs)  $(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\delta})$  of the parameters  $(\varphi_1, \varphi_2, \delta)$  are obtained.
- Two statistical quantities/measures such as (i) mean square errors (MSEs) and (ii) Bias, are selected for assessing  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$ . The values of these quantities are given by

$$MSE(\hat{\phi}) = \frac{1}{500} \sum_{a=1}^{500} (\hat{\phi}_a - \phi)^2,$$

and

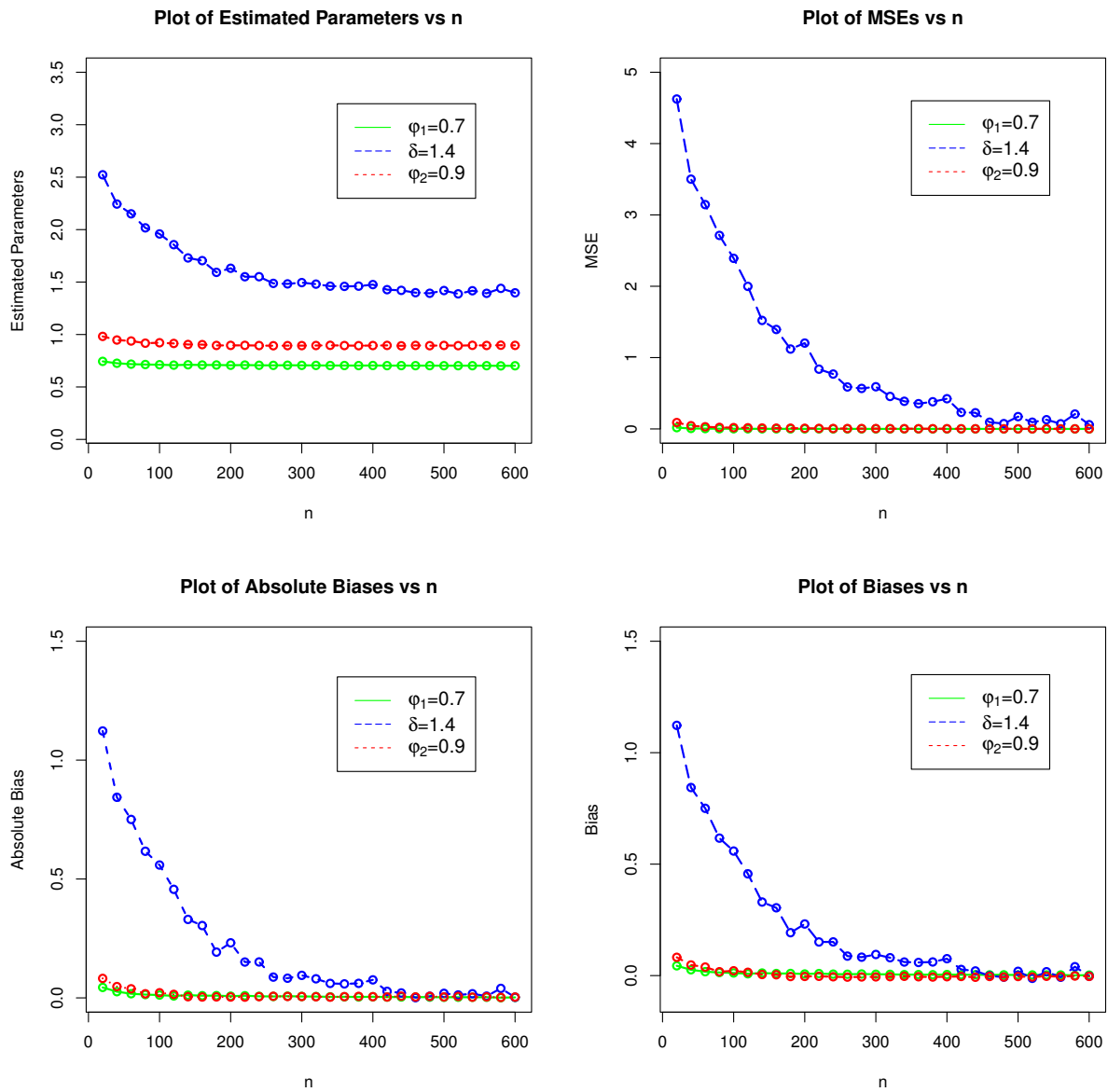
$$Bias(\hat{\phi}) = \frac{1}{500} \sum_{a=1}^{500} (\hat{\phi}_a - \phi),$$

respectively, where  $\phi = (\varphi_1, \varphi_2, \delta)$ .

All the numerical and simulation results are obtained using `optim()` R-function with the argument `method = "L-BFGS-B"`. The results of the SiSt of the NEEExp-Weibull distribution are reported in Tables 1 and 2, and presented visually in Figures 2 and 3.

**Table 1.** The results of the SiSt of the NEExp-Weibull model for  $\varphi_1 = 0.7, \delta = 1.4, \varphi_2 = 0.9$ .

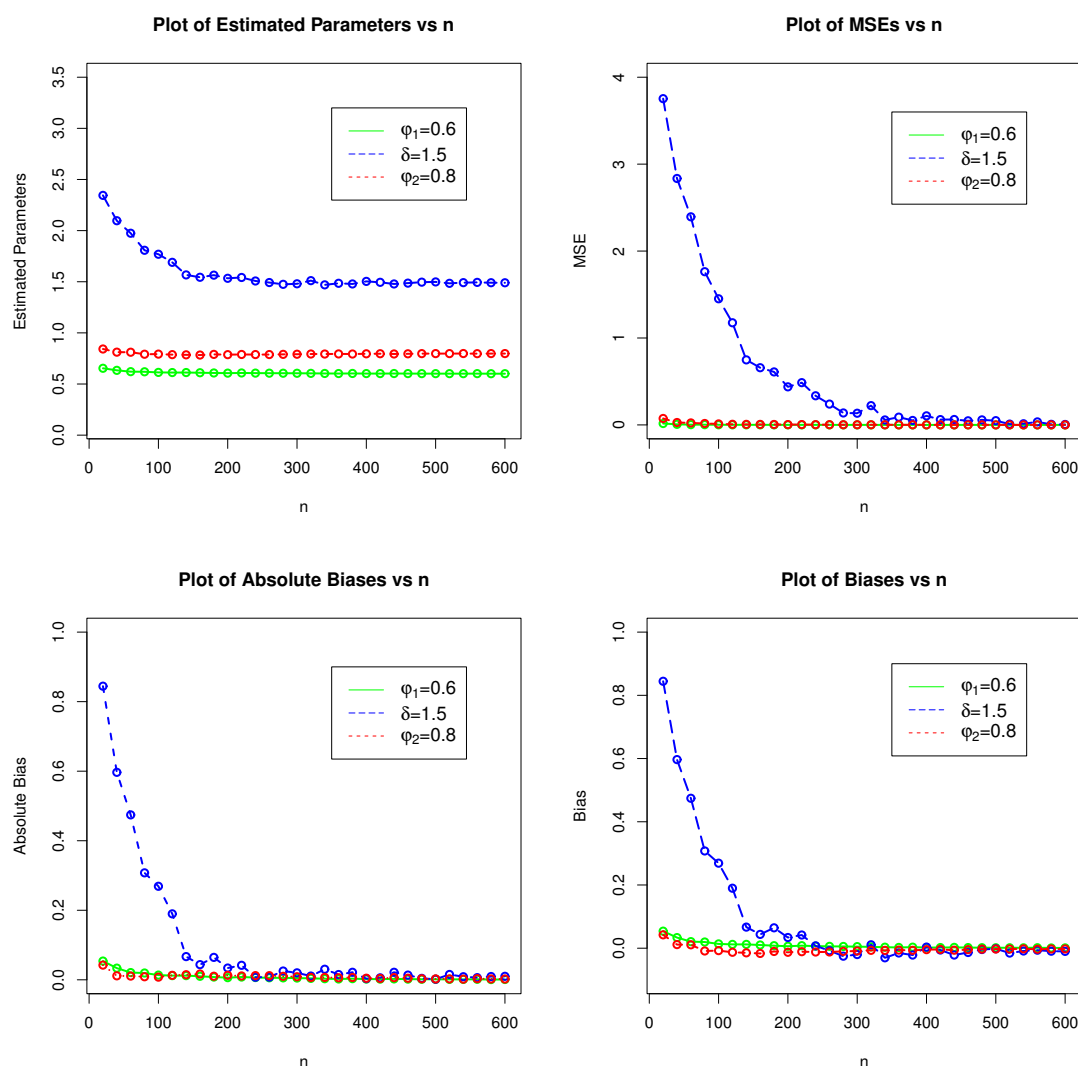
$n$	Parameters	MLEs	MSEs	Biases
20	$\varphi_1$	0.7439555	0.01816230	0.043955486
	$\delta$	2.5224710	4.62452592	1.122470958
	$\varphi_2$	0.9819897	0.08933255	0.081989688
40	$\varphi_1$	0.7261523	0.00675241	0.026152290
	$\delta$	2.2436960	3.50148162	0.843695947
	$\varphi_2$	0.9475541	0.04503070	0.047554140
60	$\varphi_1$	0.7174831	0.00469112	0.017483127
	$\delta$	2.1504600	3.14461417	0.750459996
	$\varphi_2$	0.9382455	0.03002095	0.038245459
80	$\varphi_1$	0.7139450	0.00303585	0.013945008
	$\delta$	2.0166510	2.71268772	0.616651319
	$\varphi_2$	0.9172897	0.02141037	0.017289700
100	$\varphi_1$	0.7118858	0.00250170	0.011885794
	$\delta$	1.9587070	2.39162374	0.558707085
	$\varphi_2$	0.9215668	0.01819720	0.021566823
200	$\varphi_1$	0.7071069	0.00095725	0.007106915
	$\delta$	1.6315320	1.20339825	0.231531527
	$\varphi_2$	0.8963361	0.00907772	-0.003663932
300	$\varphi_1$	0.7059804	0.00056999	0.005980353
	$\delta$	1.4947460	0.59116842	0.094745813
	$\varphi_2$	0.8942843	0.00510254	-0.005715678
400	$\varphi_1$	0.7046970	0.00042009	0.004696986
	$\delta$	1.4758080	0.42345175	0.075807616
	$\varphi_2$	0.8949554	0.00295727	-0.005044638
500	$\varphi_1$	0.7026197	0.00015114	0.002619658
	$\delta$	1.4191470	0.17256813	0.019147385
	$\varphi_2$	0.8958860	0.00155967	-0.004113986
600	$\varphi_1$	0.7021186	0.00011673	0.002118624
	$\delta$	1.3973690	0.05930217	-0.002630857
	$\varphi_2$	0.8964089	0.00084446	-0.003591110



**Figure 2.** A visual display of the SiSt for  $\varphi_1 = 0.7, \delta = 1.4, \varphi_2 = 0.9$ .

**Table 2.** The results of the SiSt of the NEExp-Weibull model for  $\varphi_1 = 0.6, \delta = 1.5, \varphi_2 = 0.8$ .

$n$	Parameters	MLEs	MSEs	Biases
20	$\varphi_1$	0.6538964	1.61446e-02	0.053896399
	$\delta$	2.3444560	3.753281849	0.844456065
	$\varphi_2$	0.8422871	0.074129355	0.042287133
40	$\varphi_1$	0.6335936	6.57586e-03	0.033593572
	$\delta$	2.0966990	2.835106550	0.596698837
	$\varphi_2$	0.8114326	0.027465620	0.011432631
60	$\varphi_1$	0.6207884	3.08374e-03	0.020788424
	$\delta$	1.9743140	2.393837599	0.474313588
	$\varphi_2$	0.8107312	0.022889902	0.010731190
80	$\varphi_1$	0.6195547	2.83132e-03	0.019554745
	$\delta$	1.8076620	1.762573786	0.307661602
	$\varphi_2$	0.7912179	0.016246953	-0.008782086
100	$\varphi_1$	0.6135883	1.52055e-03	0.013588293
	$\delta$	1.7689760	1.451201967	0.268976369
	$\varphi_2$	0.7923297	0.011392565	-0.007670316
200	$\varphi_1$	0.6062087	4.83673e-04	0.006208661
	$\delta$	1.5338360	0.437971185	0.033836372
	$\varphi_2$	0.7870588	0.003783507	-0.012941210
300	$\varphi_1$	0.6050252	2.87229e-04	0.005025202
	$\delta$	1.4799110	0.134430960	-0.020089413
	$\varphi_2$	0.7911718	0.001925689	-0.008828233
400	$\varphi_1$	0.6021233	1.08930e-04	0.002123257
	$\delta$	1.5034740	0.104238258	0.003473527
	$\varphi_2$	0.7954403	0.000832362	-0.004559674
500	$\varphi_1$	0.6014035	6.62443e-05	0.001403539
	$\delta$	1.4982590	0.048978064	-0.001741474
	$\varphi_2$	0.7974116	0.000411811	-0.002588439
600	$\varphi_1$	0.6009386	3.77904e-05	0.000938590
	$\delta$	1.4899270	0.005002637	-0.010072511
	$\varphi_2$	0.7976296	0.000242969	-0.002370384



**Figure 3.** A visual display of the SiSt for  $\varphi_1 = 0.6$ ,  $\delta = 1.5$ ,  $\varphi_2 = 0.8$ .

## 5. Analyzing the COVID-19 data taken from Mexico and Holland

The primary aim of the introduction of the proposed distribution is its implementation for data analysis in the health and other related sectors. This section illustrates the respective fact by analyzing two data sets. The first data set (Data 1) represents the mortality rates of the COVID-19 infected persons in Mexico. Whereas, the second illustration is based on taking another COVID-19 data from Holland.

By analyzing these two COVID-19 data sets, the numerical results of the proposed model is compared with the

- Weibull model with SF given by

$$S(v; \varphi_1, \varphi_2) = e^{-\varphi_2 v^{\varphi_1}}, \quad v, \varphi_1, \varphi_2 > 0,$$

- Logarithmic Weibull (L-Weibull) model with SF given by

$$S(v; \varphi_1, \varphi_2, \theta, \sigma) = \left( 1 - \frac{\sigma [1 - e^{-\varphi_2 v^{\varphi_1}}]}{\sigma - [\log(1 - e^{-\varphi_2 v^{\varphi_1}})]} \right)^\alpha, \quad v, \varphi_1, \varphi_2, \alpha, \sigma > 0,$$

- Novel exponent power-Weibull (NEP-Weibull) model with SF given by

$$S(v; \varphi_1, \varphi_2, \theta) = \left( 1 - \frac{1 - e^{-\varphi_2 v^{\varphi_1}}}{e^{e^{-\varphi_2 v^{\varphi_1}}}} \right)^\theta, \quad v, \varphi_1, \varphi_2, \theta > 0,$$

and

- New modified Weibull (NM-Weibull) model with SF given by

$$S(v; \varphi_1, \varphi_2, \sigma) = 1 - \frac{(1 - e^{-\varphi_2 v^{\varphi_1}})}{\sigma} [\sigma - e^{-\varphi_2 v^{\varphi_1}}], \quad v, \varphi_1, \varphi_2, \sigma \geq 1, \sigma \leq -1.$$

To figure out the best competitive model for the COVID-19 data, certain statistical tests such as the

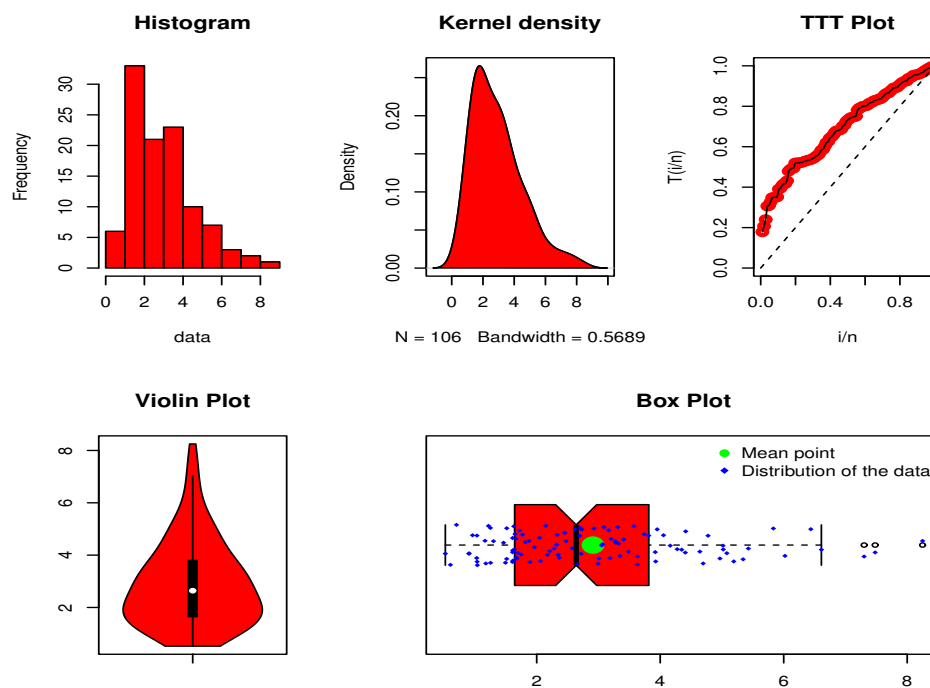
(i) AD (Anderson-Darling) test statistic given by  $AD = -p - \frac{1}{p} \sum_{a=1}^p (2a-1) [\log M(v_a) + \log \{1 - M(v_{p-a+1})\}]$ , (ii) CM (Cramer-von Mises) test statistic expressed by  $CM = \frac{1}{12p} + \sum_{a=1}^p \left[ \frac{2a-1}{2p} - M(v_a) \right]^2$ , and (iii) KS (Kolmogorov-Smirnov (KS) test statistic derived by  $KS = \sup_v [M_p(v) - M(v)]$ , are considered.

### 5.1. A data set considered from Mexico

Here, we analyze the mortality rates of the patients infected by the COVID-19 pandemic in Mexico; see <https://covid19.wh>. It is also studied by Almongy et al. [23] using a new extension of the Rayleigh distribution. This data set consists of 106 observations and is recorded from March 4, 2020, to July 20, 2020. The data set is given by: 4.4130, 3.0525, 4.6955, 7.4810, 5.1915, 3.6335, 6.6100, 8.2490, 5.8325, 3.0075, 5.4275, 3.0610, 3.3280, 1.7200, 2.9270, 5.3425, 5.0175, 2.6210, 2.1720, 2.5715, 3.8150, 7.3020, 3.9515, 3.1850, 1.7685, 3.1635, 2.3650, 1.6075, 4.6420, 6.4390, 4.4065, 5.0215, 3.6300, 2.9925, 3.2060, 1.6975, 2.2120, 4.9675, 3.9200, 4.7750, 1.7495, 1.8755, 3.4840, 1.6430, 5.0790, 4.0540, 3.3485, 3.5755, 3.2800, 1.0385, 1.8890, 1.4940, 1.6680, 3.4070, 4.1625, 3.9270, 4.2755, 1.6140, 3.7430, 3.3125, 3.0700, 2.4545, 2.3305, 2.6960, 6.0210, 4.3480, 0.9075, 1.6635, 2.7030, 3.0910, 0.5205, 0.9000, 2.4745, 2.0445, 1.6795, 1.0350, 1.6490, 2.6585, 2.7210, 2.2785, 2.1460, 1.2500, 3.2675, 2.3240, 2.3485, 2.7295, 2.0600, 1.9610, 1.6095, 0.7010, 1.2190, 1.6285, 1.8160, 1.6165, 1.5135, 1.1760, 0.6025, 1.6090, 1.4630, 1.3005, 1.0325, 1.5145, 1.0290, 1.1630, 1.2530, 0.9615.

Corresponding to the COVID-19 data (mortality rate of the COVID-19 infected persons) of Mexico, the summary measures are: minimum = 0.5205, 1<sup>st</sup> quartile = 1.6445, median = 2.6397, mean = 2.9112, 3<sup>rd</sup> quartile = 3.7970, maximum = 8.2490, variance = 2.640433, range = 7.7285, standard deviation = 1.624941, skewness = 0.9732453, and kurtosis = 3.666136.

Corresponding to the mortality rate of the COVID-19 infected persons in Mexico, some basic plots are presented in Figure 4. The plots in Figure 4, show that the first data set is right-skewed and possesses the increasing failure rate behavior.



**Figure 4.** Basic plots of the Mexico data.

Corresponding to the Mexico data, the values of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$  are presented in Table 3. The standard errors (SEs) (numerical values in the parentheses) of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$  are also presented in Table 3.

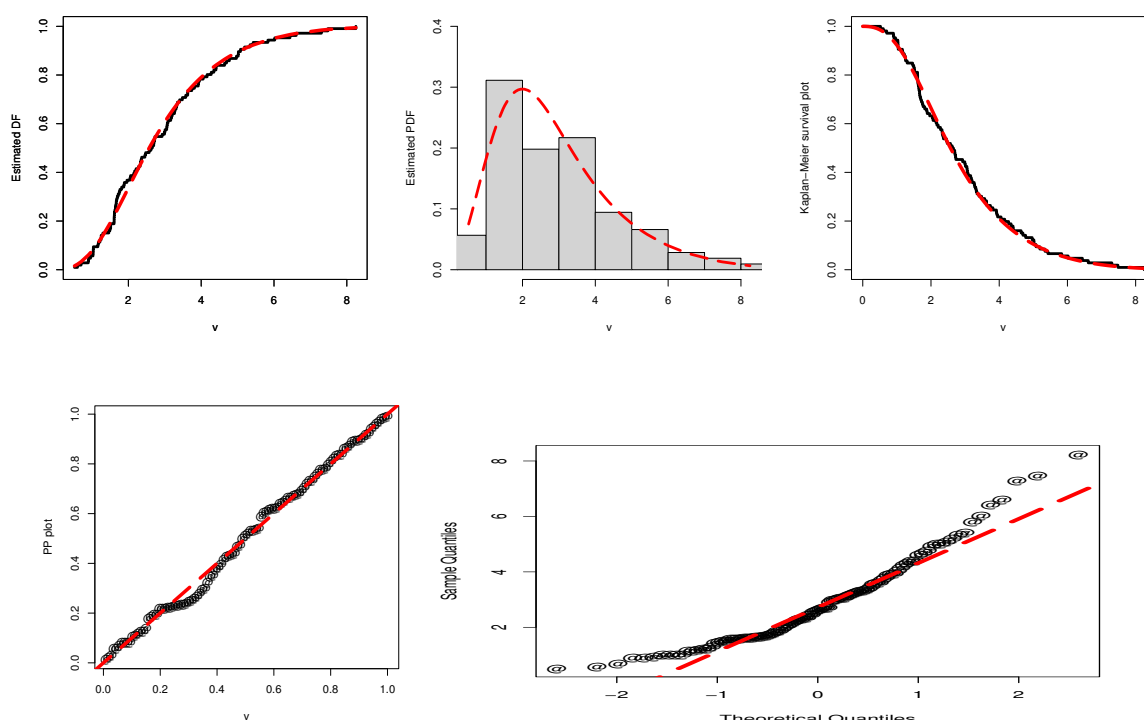
For the Mexico data, the values of the selected tests CM, AD, and KS of the fitted models are reported in Table 4. The associated p-value of fitted models is also provided in Table 4. From the numerical illustration in Table 4, we can see that the proposed model has the smallest values of CM, AD, and KS, and the largest p-value. These facts show that the proposed model is the best competitor. Besides the numerical illustration, a visual display of the performances of the proposed model is also provided in Figure 5.

**Table 3.** The numerical values of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ ,  $\hat{\delta}$ ,  $\hat{\alpha}$ ,  $\hat{\sigma}$ , and  $\hat{\theta}$  using the mortality rate data.

Model	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\theta}$
NEExp-Weibull	2.53401 (0.31876)	0.01723 (0.01562)	0.26201 (0.24135)	-	-	-
Weibull	1.92738 (0.14019)	0.09976 (0.02220)	-	-	-	-
L-Weibull	1.57117 (0.01001)	1.20372 (0.00879)	-	0.16022 (0.01580)	0.45036 (0.01773)	-
NEP-Weibull	1.71906 (0.03634)	0.95340 (0.03634)	-	-	-	0.15201 (0.01503)
NM-Weibull	1.89237 (0.15734)	0.11004 (0.03088)	-	-	12.06058 (20.5532)	-

**Table 4.** The analytical measures of the fitted models using Data 1.

Model	CM	AD	KS	p-value
NEExp-Weibull	0.05834	0.33674	0.06266	0.79940
Weibull	0.10264	0.66004	0.07147	0.65100
L-Weibull	0.06002	0.36130	0.07091	0.66060
NEP-Weibull	0.07017	0.43993	0.07862	0.52880
NM-Weibull	0.10723	0.68969	0.07024	0.67240

**Figure 5.** For the first data, the plots of the fitted DFs of the proposed and all the fitted models.

## 5.2. A data set considered from Holland

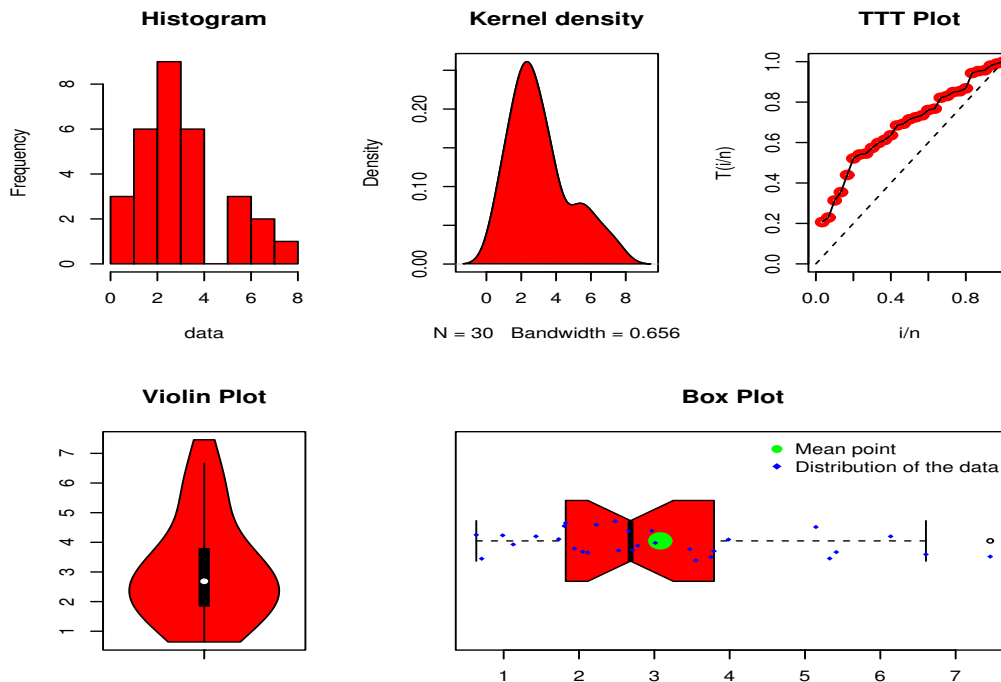
Here, we provide a second illustration of the proposed model by analyzing another COVID-19 data taken from Holland; see Almongy et al. [23]. This data set consists of 30 observations and is recorded between March 31, 2020, and April 30, 2020. The second data set is given by: 7.4590, 3.7490, 3.4700, 5.3280, 1.4285, 1.1270, 6.1370, 5.1445, 5.4160, 3.5495, 1.7305, 1.8235, 2.9640, 6.6055, 3.9840, 3.7920, 2.6535, 2.5240, 2.7155, 2.7775, 3.0135, 2.0485, 1.8055, 2.4800, 2.2310, 1.9415, 0.9870, 0.6365, 0.7080, 2.1175.

The summary measures of the mortality rate of the COVID-19 infected persons in Holland are: minimum = 0.6365, 1<sup>st</sup> quartile = 1.8530, median = 2.6845, mean = 3.0783, 3<sup>rd</sup> quartile = 3.7812, maximum = 7.4590, variance = 3.121073, range = 6.8225, standard deviation = 1.766656, skewness



$= 0.8339708$ , and kurtosis  $= 2.953478$ .

In link to the mortality rate of the COVID-19 infected persons in Holland, some basic plots are sketched in Figure 6. From the plots in Figure 6, it is obvious that the second data set, related to the mortality rate of the COVID-19 infected persons, is skewed to the right and has increasing failure rate behavior.



**Figure 6.** Basic plots of the Holland data.

Based on the Holland's COVID-19 data, the numerical values of estimators  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ , and  $\hat{\delta}$  are obtained in Table 5. Furthermore, the SEs of these estimators are also reported in Table 5.

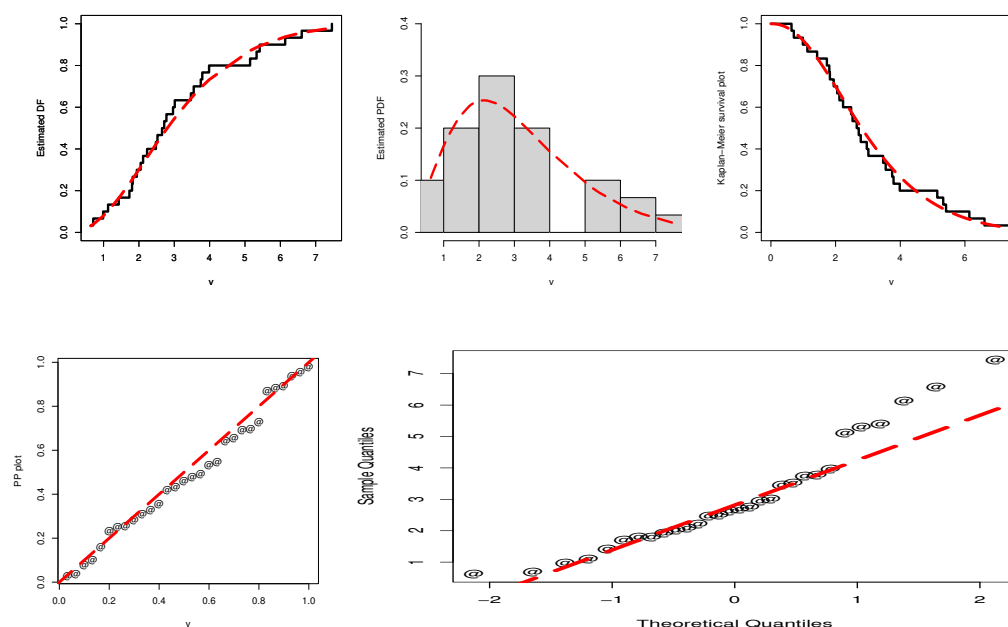
**Table 5.** The numerical values of  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ ,  $\hat{\delta}$ ,  $\hat{\alpha}$ ,  $\hat{\sigma}$ , and  $\hat{\theta}$  using the mortality rate data.

Model	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\theta}$
NEExp-Weibull	2.20197 (0.48765)	0.03704 (0.04876)	0.75896 (0.10983)	-	-	-
Weibull	1.92738 (0.14019)	0.09976 (0.02220)	-	-	-	-
L-Weibull	1.42252 (0.51193)	0.12239 (0.05844)	-	2.85548 (0.97654)	0.89686 (2.15028)	-
NEP-Weibull	1.77603 (0.07305)	1.03756 (0.07304)	-	-	-	0.11297 (0.02890)
NM-Weibull	2.00489 (0.28946)	0.06377 (0.04170)	-	-	-2.41345 (3.47713)	-

Corresponding to the Holland data, the p-value and values of CM, AD, and KS are reported in Table 6. From the numerical comparison of the fitted models in Table 6, it is obvious that the proposed model performs better than the other competitors as it has the largest p-value and smallest CM, AD, and KS values. In support of Table 6, the performances of the proposed model are also illustrated visually by plotting the estimated DF, PP, PDF, QQ, and SF; see Figure 7.

**Table 6.** The analytical measures of the fitted models using Data 2.

Model	CM	AD	KS	p-value
NEExp-Weibull	0.03088	0.20437	0.08557	0.96710
Weibull	0.04794	0.29434	0.10236	0.88030
L-Weibull	0.05165	0.31630	0.10144	0.88660
NEP-Weibull	0.03423	0.22143	0.10418	0.86740
NM-Weibull	0.03926	0.24960	0.08981	0.95100



**Figure 7.** For the first data, the plots of the fitted DFs of the proposed and all the fitted models.

## 6. Analyzing the COVID-19 data using machine learning algorithms

In this section, we implement three different machine learning algorithms, namely, SVR, NNAR, and RF to forecast the data set analyzed in Section 5. Before modeling, we split the data set into two parts; 80 percent as training data and 20 percent as testing data, followed by Qi and Zhang [24]. We apply all the models to training data and compare their forecasting performances using the testing data. To assess the out-of-sample (also known as post-sample) prediction accuracy, multistep ahead forecasts with RMSE and MAE are calculated.

### 6.1. Support vector regression

The SVR is a popular ML approach, which is used for regression as well as classification problems. The SVR was first developed by Cortes and Vapnik [25], and to date, it is one of the most widely used supervised learning methods that is based on structured risk minimization rule and statistical theory. The structured risk minimization rule maximizes prediction accuracy and mitigates the likelihood of over fitting.

In practice, it can effectively approximate linear and nonlinear problems and work well for numerous problems. The SVR uses various kernel functions to figure out the similarity between two data points to tackle the non-linear situations. The main advantage of SVR lies in its potential to capture the predictors' nonlinearity and then utilize it to enhance forecasting accuracy. In our case, the set of predictors contains the lagged values. The SVR assists to detect the margin of error which is bearable in the model; see Ribeiro et al. [26] and Bibi et al. [27]. The SVR equation with kernel function can be expressed as

$$F_t = \sum_{k=1}^h (\lambda_k - \lambda_k^*) N(c_k, c) + \mu, \quad (6.1)$$

where  $F_t$  is the outcome variable. The kernel function,  $N(c_k, c)$  denotes the inner product, while  $\mu$  is accommodated within the kernel function. In the literature, several kernel functions have been developed. Among them, the radial basis function (RBF) is the most popular, which can be illustrated as

$$N(c_k, c) = \exp \left\{ -\frac{\|c_k - c_N\|^2}{2\tau^2} \right\}, \quad (6.2)$$

where the Euclidean distance between the two predictors squared vectors is represented by  $\|c_k - c_N\|^2$  and  $\tau^2$  is basically the width of RBF; see Lu et al. [28]. Hence, in this study, we focus on the RBF kernel function for the SVR. Tuning the SVR model enable us to arrive at optimal parameters.

### 6.2. The random forest

Another ML approach is RF which is also known as the random decision forest. It falls within the supervised learning category. The RF is a very effective algorithm that is used for both regression and classification problems as proposed by Breiman [29]. Dietterich [30] argued that RF is considered the most efficient ensemble technique appearing in ML and fulfills good forecasting properties. The RF approach is employed in different areas, including stock trading, finance, e-commerce, and health care. It provides a forest out of a collection of decision trees that are usually estimated (trained) by employing the bagging approach.

The RF approach discovers the output based on the decision trees' forecasts. The forecasts are computed by averaging the output of several trees. The improvement is achieved in prediction by expanding the number of trees. In other words, as the number of trees in the forest grows, a more accurate forecast is obtained and circumvents the issue of over fitting as well. To estimate the RF, we use three hyper parameters i.e., the number of trees, number of nodes, and sample repetition. The number of nodes and trees are utilized as 3 and 500, respectively.

### 6.3. The neural network autoregression

In general, neural networks (NNs) are basically a network or circuit of neurons. The artificial neural network (ANNs) is composed of nodes or artificial neurons. NNs are highly flexible computing frameworks for analyzing a wide range of nonlinear problems. A key advantage of such networks is that they have not required prior information regarding the functional form in the model establishing process, rather highly determined by the characteristics of data; see Peng et al. [31].

A feedback NN is established with lagged realization as predictors and hidden layer(s) with dimension nodes. The neural network autoregression (NNAR) contains three layers such as (i) the input layer, (ii) the hidden layer, and (iii) the output layer. The NNAR model is fitted to forecast a time series by utilizing its past information as inputs  $F_t, F_{t-1}, \dots, F_{t-m}$ , the entire process refers to feedback delay, where  $t$  indicates the time delay parameter. The NNAR ( $m, n$ ) entails that the hidden layer consists of  $m$  delayed inputs and  $n$  nodes. The mathematical form of NNAR can be illustrated as

$$F_t = \omega_0 + \sum_{c=1}^m \theta \omega_c \left( \mathbf{s}_c + \sum_{n=1}^z \mathbf{s}_{cn} F_{t-i} \right) + \mu, \quad (6.3)$$

where  $\mathbf{s}_{cn}$  ( $c = 1, 2, 3, \dots, m, n = 1, 2, 3, \dots, z$ ) and  $\omega_c$  ( $c = 1, 2, 3, \dots, m$ ) are the weights of interconnection, and  $z$  is the length of input layers, and  $m$  is the length of hidden layers.

The prediction accuracy of all ML techniques is quantified by using two statistical accuracy criteria computed from a testing data set. Statistically, the forecast errors are more appropriate criteria to evaluate the predictive power and to finalize the best approach. In general, the most popular measures are MAE and RMSE. Hence, we compare the forecasting performance of ML techniques using these two measures in our study. Their mathematical expressions are given by

$$MAE = \text{mean}(|F_t - \hat{F}_t|),$$

and

$$RMSE = \sqrt{\text{mean}(F_t - \hat{F}_t)^2},$$

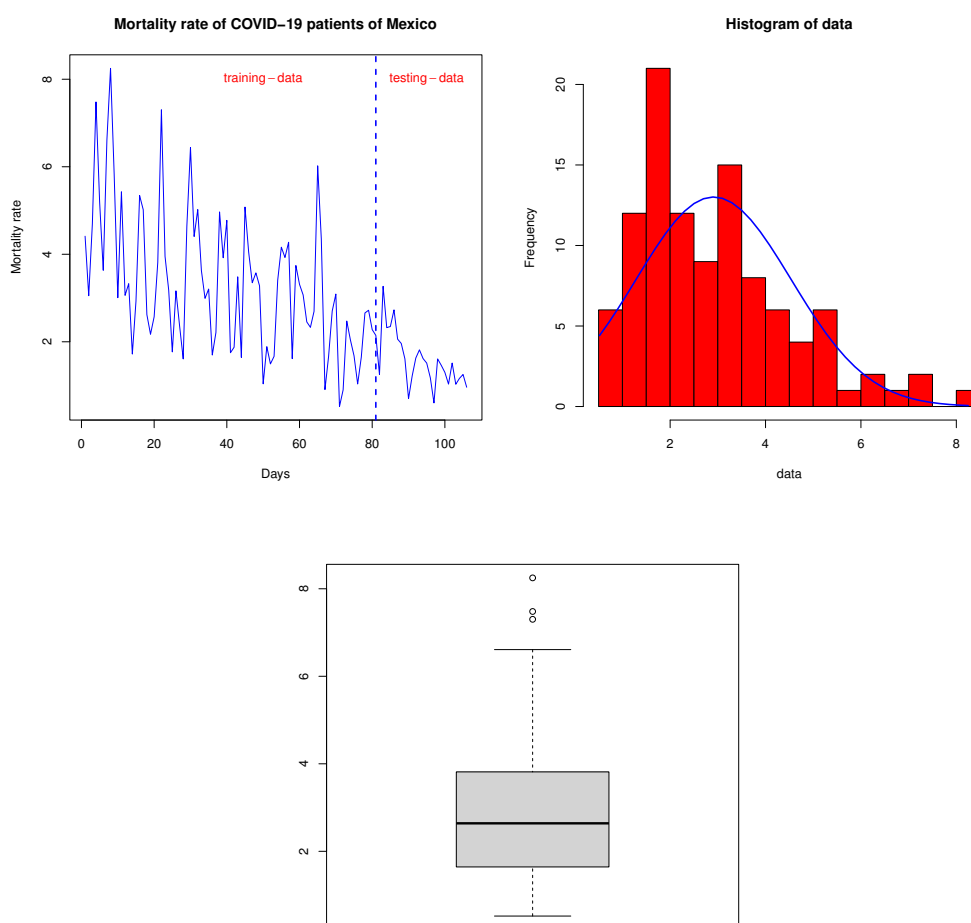
respectively.

### 6.4. The empirical results and discussion

This subsection is further divided into two parts. In the first part, we deal with the mortality rates of the patients infected by the COVID-19 pandemic in Mexico. In the second part, we deal with the death rates of the patients infected by the COVID-19 pandemic in Holland.

#### 6.4.1. Analyzing the Mexico data

From Figure 8, it can be observed that the entire trajectory of the mortality rate has experienced many ups and downs in the last few months, but the trend of data is decreasing over time. The mortality rate series is plotted in Figure 8, where the vertical blue dotted line splits the estimation and post-sample forecasting periods. The histogram plot and box plot are also provided in Figure 8, which reveal that the underlying time series is right skewed.



**Figure 8.** A virtual display of the mortality rate data in Mexico.

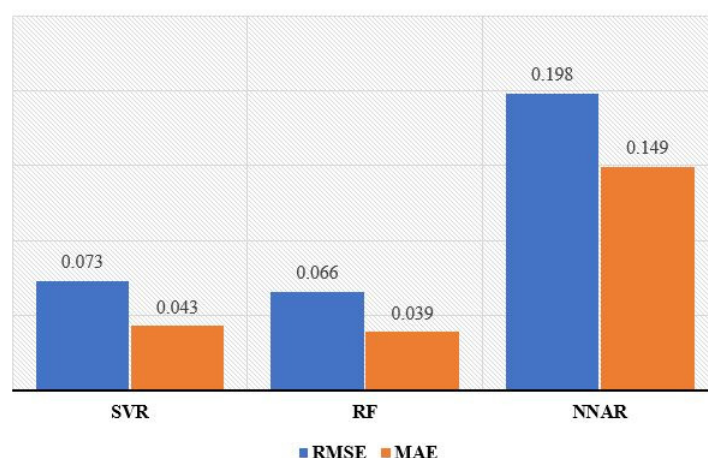
The two statistical accuracy measures for the Mexico data set are given in Table 7. The MAE and RMSE are calculated for ML algorithms such as SVR, RF, and NNAR. It is inferred that the RF outperformed the other competitor counterparts. For the RF method, the RMSE and MAE values are given by 0.066 and 0.039, respectively. Whereas, for the SVR, the values of the RMSE and MAE are given by 0.073 and 0.043, respectively. On the other hand, the RMSE and MAE values for the NNAR are, respectively, given by 0.198 and 0.149. From the values of the RMSE and MAE for the NNAR approach, it is clear that the values of NNAR are higher than the values of the RMSE and MAE computed for RF and SVR.

**Table 7.** The error metrics.

Criteria	SVR	RF	NNAR
RMSE	0.073	0.066	0.198
MAE	0.043	0.039	0.149

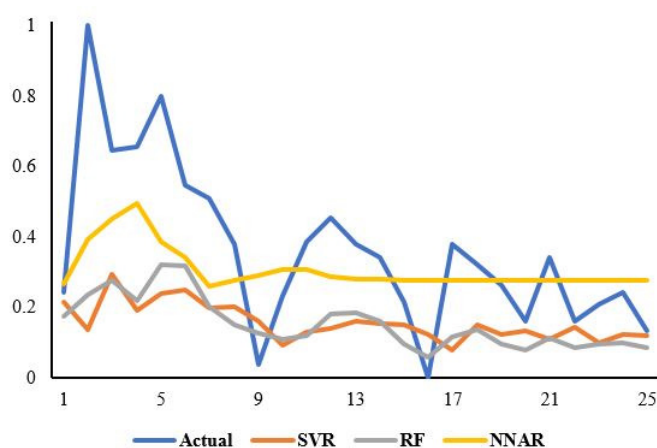
Furthermore, the visual illustration of forecast errors is provided in Figure 9. The plots in Figure 9

show that the ML algorithms, specifically, the RF remained an efficient tool in forecasting the mortality rates of COVID-19 patients.



**Figure 9.** The schematic representation of forecast errors for the mortality rate data in Mexico.

We also plot the prediction curves for COVID-19 deaths under the three machine learning algorithms to get a more intuitive picture of the prediction accuracy and results. In this regard, the line chart has been constructed, for the representation of predicted and actual values; see Figure 10. Some predicted values of the NNAR approach are very close, but few are substantially far away from the actual data. On the other hand, SVR prediction is highly stable over time and therefore beat the rival algorithms in predictive modeling.

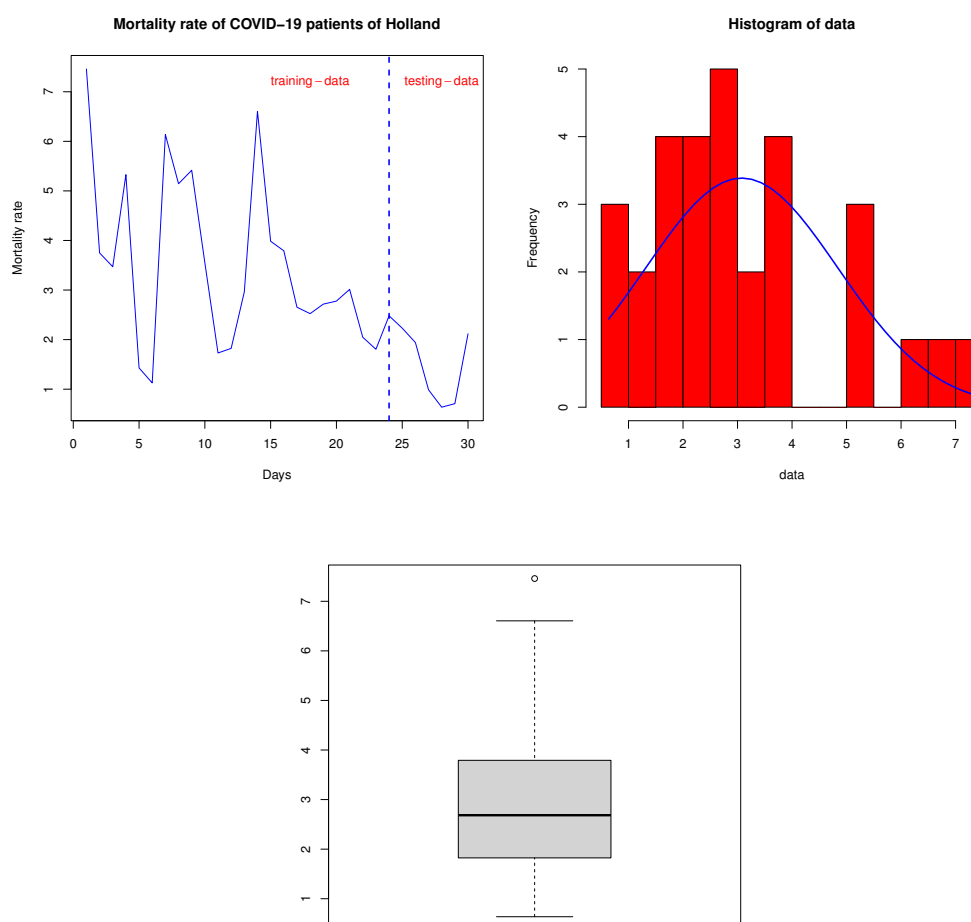


**Figure 10.** Forecasting performance of ML algorithms for the mortality rate data in Mexico.

#### 6.4.2. Analyzing the Holland data

Here, we estimate and predict the daily death data of COVID-19 for Holland. It can be observed from Figure 11 that the entire trajectory of mortality rate has experienced numerous episodes in several

months, but the times series is declining over time. The mortality rate series is plotted in Figure 11, where the vertical blue dotted line represents the estimation and post-sample forecasting periods. The histogram plot and box plot are also presented in Figure 10, which demonstrates that the underlying series is right skewed.



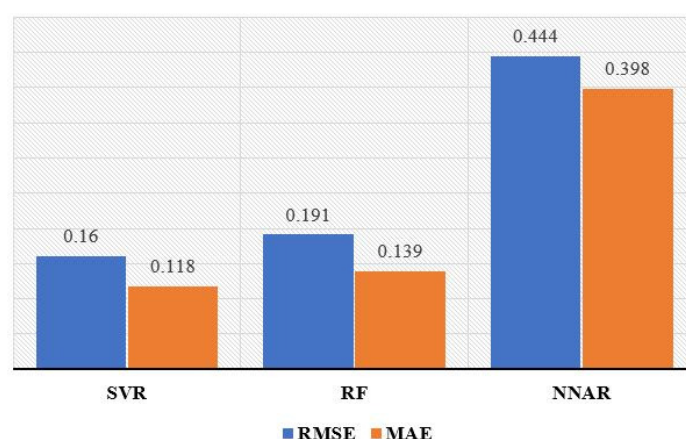
**Figure 11.** A virtual display of the mortality rate data in Holland.

The two considered statistical accuracy measures for the Holland data set are reported in Table 8. The MAE and RMSE are computed for ML algorithms, namely SVR, RF, and NNAR. From Table 8, we can observe that the SVR showed superior forecast performance as compared to the rival counterparts. For the SVR method, the RMSE and MAE values are 0.16 and 0.118, respectively. On the other hand, for the RF approach, these values are given by 0.191 and 0.139, respectively. Whereas, the forecast errors (RMSE and MAE) of NNAR are given by 0.444 and 0.398, respectively. From the above discussion, it is obvious that the forecast errors of the NNAR approach are substantially higher than the forecast errors of RF and SVR.

**Table 8.** The error metrics.

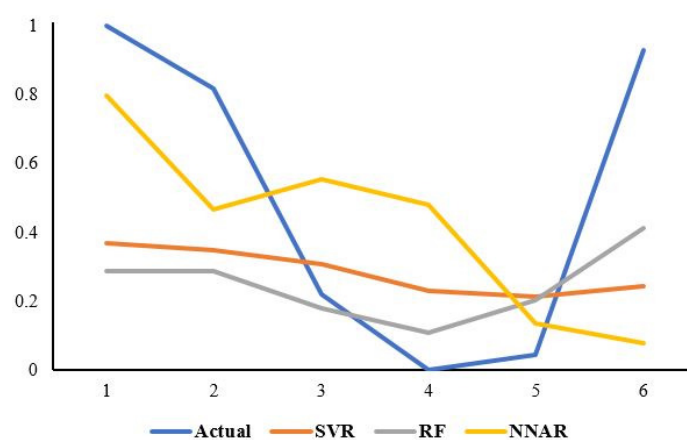
Criteria	SVR	RF	NNAR
RMSE	0.160	0.191	0.444
MAE	0.118	0.139	0.398

The graphical comparison of forecast errors is also depicted in Figure 12. It manifests that the ML algorithms, specifically SVR remained an effective tool in forecasting the post-sample trajectory of the mortality rate of COVID-19 patients.



**Figure 12.** The schematic representation of forecast errors for the mortality rate data in Holland.

We portray the prediction curves for COVID-19 deaths using these three ML algorithms to get a more clear picture of the models performances. In this regard, we use the line chart for the representation of forecasted and observed data; see Figure 13. From the plots in Figure 13, it can be seen that the observed test series is highly volatile and noisy. Despite this volatility, SVR and RF have shown good results, particularly the RF performance is highly satisfactory.



**Figure 13.** Forecasting Performance of ML algorithms for the mortality rate data in Holland.



## 7. Conclusions, limitations, and future research plan

The efforts in this paper added another useful approach to the literature on statistical methodologies by introducing a new family of distributions. The new family was named a NEEExp family of distributions. Based on a NEEExp family, an updated version of the Weibull model called, a NEEExp-Weibull model was studied. The MLEs of the NEEExp-Weibull model were obtained. The evaluation of the MLEs of the NEEExp-Weibull distribution was carried out through a brief SiSt. The usefulness of the NEEExp-Weibull model was shown by analyzing two data sets taken from the healthcare sector. The first data set was representing the mortality rate of COVID-19 patients in Mexico. Whereas, the second data was also related to the COVID-19 events taken from Holland. Using these two COVID-19 data sets, the NEEExp-Weibull distribution was compared with the Weibull and three other well-known models such as the L-Weibull, NEP-Weibull, and NM-Weibull distributions. Based on four well-known comparative tools, it is observed that the NEEExp-Weibull distribution was the best competitive model as compared to the Weibull and other well-known modified forms of the Weibull distribution. Therefore, based on the numerical results and findings of this study, it is observed that the NEEExp-Weibull distribution may be the best suitable choice to use for analyzing the medical and other related data sets. Besides the statistical modeling, for prediction purposes, we further implemented three ML methods including SVR, RF, and NNAR using the same data sets. To compare their forecasting performances, two well-known statistical accuracy quantities such as the RMSE and MAE were computed. We found that the RF algorithm was very efficient in forecasting using the first data. However, for the second data set, the SVR showed superior performance in contrast to other methods.

Besides the certain advantages of the NEEExp-Weibull distribution over the Weibull and other competitive distributions, the NEEExp-Weibull distribution has also some certain limitations, for example

- The proposed NEEExp-Weibull model is a continuous probability distribution, and it is employed to analyze the mortality rates of COVID-19 patients. Therefore, the NEEExp-Weibull model could not be implemented to analyze other forms of the COVID-19 data that are discrete in nature, for example, (i) the number of deaths, (ii) the number of confirmed cases, or (iii) the number of recovered cases, etc.
- Due to the introduction of the additional parameter, the NEEExp-Weibull distribution has a complicated form of its PDF. Therefore, the expressions for the estimators of its parameters are not in explicit forms. Therefore, computer software must be used to obtain the estimated values of the parameters.
- Since, the PDF of the NEEExp family has a complicated form, therefore, more computational efforts are needed to derive the key mathematical properties.

In the future, we are motivated to obtain the discrete version of the proposed NEEExp-Weibull distribution to counter the discrete data sets. In this work, we only used the maximum likelihood estimation method to estimate the parameters of the NEEExp-Weibull distribution. In the future, we are intended to estimate the parameters of the NEEExp-Weibull distribution using other classical methods such as ordinary least square, weighted ordinary least square, maximum product spacing methods, etc. Neutrosophic statistics is a generalization of classical statistics and is implemented when the data

sets are generated from a complex process. In the future, we are also motivated to study the neutrosophic extension of the NEEExp-Weibull distribution.

## References

1. B. T. Ngo, P. Marik, P. Kory, L. Shapiro, R. Thomadsen, J. Iglesias, et al., The time to offer treatments for COVID-19, *Expert Opin. Invest. Drugs*, **30** (2021), 505–518. <https://doi.org/10.1080/13543784.2021.1901883>
2. B. Pfefferbaum, C. S. North, Mental health and the COVID-19 pandemic, *N. Engl. J. Med.*, **383** (2020), 510–512. <https://doi.org/10.1056/NEJMp2008017>
3. E. J. Kim, L. Marrast, J. Conigliaro, COVID-19: magnifying the effect of health disparities, *J. Gen. Intern. Med.*, **35** (2020), 2441–2442. <https://doi.org/10.1007/s11606-020-05881-4>
4. J. Champion, A. Javed, N. Sartorius, M. Marmot, Addressing the public mental health challenge of COVID-19, *Lancet Psychiatry*, **7** (2020), 657–659. [https://doi.org/10.1016/S2215-0366\(20\)30240-6](https://doi.org/10.1016/S2215-0366(20)30240-6)
5. A. T. Gloster, D. Lamnisos, J. Lubenko, G. Presti, V. Squatrito, M. Constantinou, et al., Impact of COVID-19 pandemic on mental health: an international study, *PloS One*, **15** (2020), e0244809. <https://doi.org/10.1371/journal.pone.0244809>
6. D. Talevi, V. Socci, M. Carai, G. Carnaghi, S. Faleri, E. Trebbi, et al., Mental health outcomes of the COVID-19 pandemic, *Riv. Psichiatr.*, **55** (2020), 137–144. <https://doi.org/10.1708/3382.33569>
7. E. A. Wastnedge, R. M. Reynolds, S. R. Van Boeckel, S. J. Stock, F. C. Denison, J. A. Maybin, et al., Pregnancy and COVID-19, *Physiol. Rev.*, **101** (2021), 303–318. <https://doi.org/10.1152/physrev.00024.2020>
8. W. Bo, Z. Ahmad, A. R. Alanzi, A. I. Al-Omari, E. H. Hafez, S. F. Abdelwahab, The current COVID-19 pandemic in China: an overview and corona data analysis, *Alexandria Eng. J.*, **61** (2021), 1369–1381. <https://doi.org/10.1016/j.aej.2021.06.025>
9. V. H. Moreau, Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths, *Braz. J. Microbiol.*, **51** (2020), 1109–1115. <https://doi.org/10.1007/s42770-020-00331-z>
10. S. Tuli, S. Tuli, R. Tuli, S. S. Gill, Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing, *Internet Things*, **11** (2020), 100222. <https://doi.org/10.1016/j.iot.2020.100222>
11. S. M. Rahman, J. Kim, B. Laratte, Disruption in Circularity? Impact analysis of COVID-19 on ship recycling using Weibull tonnage estimation and scenario analysis method, *Resour. Conserv. Recycl.*, **164** (2021), 105139. <https://doi.org/10.1016/j.resconrec.2020.105139>
12. E. M. Almetwally, R. Alharbi, D. Alnagar, E. H. Hafez, A new inverted topp-leone distribution: applications to the COVID-19 mortality rate in two different countries, *Axioms*, **10** (2021), 25. <https://doi.org/10.3390/axioms10010025>
13. M. Alizadeh, G. M. Cordeiro, A. D. Nascimento, M. D. C. S. Lima, E. M. Ortega, Odd-Burr generalized family of distributions with some applications, *J. Stat. Comput. Simul.*, **87** (2017), 367–389. <https://doi.org/10.1080/00949655.2016.1209200>

14. F. Chipepa, B. Oluyede, B. Makubate, A new generalized family of odd Lindley-G distributions with application, *Int. J. Stat. Probab.*, **8** (2019), 1–22. <https://doi.org/10.5539/ijsp.v8n6p1>
15. L. Handique, S. Chakraborty, T. A. de Andrade, The exponentiated generalized Marshall–Olkin family of distribution: its properties and applications, *Ann. Data Sci.*, **6** (2019), 391–411. <https://doi.org/10.1007/s40745-018-0166-z>
16. M. H. Tahir, M. A. Hussain, G. M. Cordeiro, M. El-Morshedy, M. S. Eliwa, A new Kumaraswamy generalized family of distributions with properties, applications, and bivariate extension, *Mathematics*, **8** (2020), 1989. <https://doi.org/10.3390/math8111989>
17. S. M. Zaidi, M. M. A. Sobhi, M. El-Morshedy, A. Z. Afify, A new generalized family of distributions: properties and applications, *AIMS Math.*, **6** (2021), 456–476. <https://doi.org/10.3934/math.2021028>
18. F. H. Riad, E. Hussam, A. M. Gemeay, R. A. Aldallal, A. Z. Afify, Classical and Bayesian inference of the weighted-exponential distribution with an application to insurance data, *Math. Biosci. Eng.*, **19** (2022), 6551–6581. <https://doi.org/10.3934/mbe.2022309>
19. M. E. Bakr, A. A. Al-Babtain, Z. Mahmood, R. A. Aldallal, S. K. Khosa, M. M. Abd El-Raouf, et al., Statistical modelling for a new family of generalized distributions with real data applications, *Math. Biosci. Eng.*, **19** (2022), 8705–8740. <https://doi.org/10.3934/mbe.2022404>
20. A. Xu, S. Zhou, Y. Tang, A unified model for system reliability evaluation under dynamic operating conditions, *IEEE Trans. Reliab.*, **70** (2019), 65–72. <https://doi.org/10.1109/TR.2019.2948173>
21. C. Luo, L. Shen, A. Xu, Modelling and estimation of system reliability under dynamic operating environments and lifetime ordering constraints, *Reliab. Eng. Syst. Saf.*, **218** (2022), 108136. <https://doi.org/10.1016/j.ress.2021.108136>
22. A. Alzaatreh, C. Lee, F. Famoye, A new method for generating families of continuous distributions, *Metron*, **71** (2013), 63–79. <https://doi.org/10.1007/s40300-013-0007-y>
23. H. M. Almongy, E. M. Almetwally, H. M. Aljohani, A. S. Alghamdi, E. H. Hafez, A new extended rayleigh distribution with applications of COVID-19 data, *Results Phys.*, **23** (2021), 104012. <https://doi.org/10.1016/j.rinp.2021.104012>
24. M. Qi, G. P. Zhang, An investigation of model selection criteria for neural network time series forecasting, *Eur. J. Oper. Res.*, **132** (2001), 666–680. [https://doi.org/10.1016/S0377-2217\(00\)00171-5](https://doi.org/10.1016/S0377-2217(00)00171-5)
25. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.*, **20** (1995), 273–297. <https://doi.org/10.1007/BF00994018>
26. M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, L. dos Santos Coelho, Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil, *Chaos, Solitons Fractals*, **135** (2020), 109853. <https://doi.org/10.1016/j.chaos.2020.109853>
27. N. Bibi, I. Shah, A. Alsubie, S. Ali, S. A. Lone, Electricity spot prices forecasting based on ensemble learning, *IEEE Access*, **9** (2021), 150984–150992. <https://doi.org/10.1109/ACCESS.2021.3126545>

28. C. J. Lu, T. S. Lee, C. C. Chiu, Financial time series forecasting using independent component analysis and support vector regression, *Decis. Support Syst.*, **47** (2009), 115–125. <https://doi.org/10.1016/j.dss.2009.02.001>
29. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
30. T. G. Dietterich, Ensemble methods in machine learning, in *International Workshop on Multiple Classifier Systems*, Springer, Berlin, Heidelberg, **1857** (2000), 1–15. [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)
31. Z. Peng, F. U. Khan, F. Khan, P. A. Shaikh, Y. H. Dai, I. Ullah, et al., An application of hybrid models for weekly stock market index prediction: empirical evidence from SAARC countries, *Complexity*, **2021** (2021). <https://doi.org/10.1155/2021/5663302>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)