Annals of Data Science (2022) 9(1):153–173
https://doi.org/10.1007/s40745-021-00351-y

Do Temperature and Humidity Affect the Transmission of SARS-CoV-2?-A Flexible Regression Analysis

Md. Rezaul Karim[1] · **Mst. Bithi Akter**[1] · **Sejuti Haque**[1] · **Nazmin Akter**[1]

Received: 27 March 2021 / Revised: 20 June 2021 / Accepted: 19 July 2021 /
Published online: 25 July 2021

Abstract
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmissible virus that causes Coronavirus disease 2019 (COVID-19). Temperature and humidity are two essential factors in the transmission of SARS-CoV-2 affect the respiratory system of human. This study aimed to investigate the effects of temperature and humidity on the transmission of SARS-CoV-2 and the Spread Covid-19. The daily number of SARS-CoV-2 infected new cases, and the number of death due to Covid-19 are considered the response variables. Data are collected from March 08, 2020 to January 31, 2021. A flexible regression model under the Generalized Additive Models for Location Scale and Shape framework is used to analyze data. The temperature and humidity have a significant impact on the transmission of SARS-CoV-2. The temperature is highly significant in the number of SARS-CoV-2 infected new cases and number of death due to COVID-19. In contrast, the humidity is significant on the number of SARS-CoV-2 infected new cases, but it is insignificant on the number of death due to COVID-19 at a 5% level of significance. The analysis revealed that both the temperature and humidity inversely affected the daily number of deaths and new cases of COVID-19.

Keywords SARS-CoV-2 · COVID-19 · GAMLSS · Temperature · Humidity

✉ Md. Rezaul Karim
rezaul@juniv.edu

Mst. Bithi Akter
bithiakter200stat@gmail.com

Sejuti Haque
sejutihaque68@gmail.com

Nazmin Akter
nazmin.swarna194@gmail.com

[1] Department of Statistics, Jahangirnagar University, Savar, Bangladesh

1 Introduction

Coronavirus disease 2019 (COVID-19), caused by the novel coronavirus (official name is SARS-CoV-2; formerly called 2019-nCoV), has become a major public health problem all over the world [1]. In light of the rising danger, World Health Organization (WHO) declared COVID-19 as an international public health emergency [2]. Although it is still unknown exactly where the outbreak first started, many early cases of COVID-19 have been attributed to people who have visited the Huanan Seafood Wholesale Market, located in Wuhan, Hubei, China [3]. Globally, as of March 21, 2021 there have been 123.55 million confirmed cases of COVID-19, including 2.72 million deaths and among confirmed cases 99.53 million are recovered [4].

Bangladesh is a well-known climate-vulnerable country due to its high population density and complex meteorological settings [5]. In Bangladesh, the first coronavirus cases were confirmed on March 08, 2020 by the country's epidemiology institute, the Institute of Epidemiology Disease Control and Research (IEDCR). It has been reported that the temperature, humidity, wind, and precipitation may favour either the spread or the inhibition of epidemic episodes [6, 7] reported that the transmission of viruses is influenced by weather conditions and the density of people. Although Bangladesh is an over-populated country (about 160 million), COVID-19 in Bangladesh seems less acute. As of March 21, 2021 there have been 570,878 confirmed cases including 8690 deaths and among confirmed cases 522,105 are recovered [4]. The reason for moderate transmission of COVID-19 might be an influence of tropical weather (consisting of high temperature, often excessive humidity).

Meteorological parameters are the important factors influencing infectious diseases such as severe acute respiratory syndrome (SARS) and influenza [8]. It is supposed that high temperature and humidity, together, have a combined effect on the inactivation of coronaviruses. In contrast, the opposite weather condition can support the prolonged survival time of the virus on surfaces and facilitate the transmission and susceptibility of the viral agent [9]. There is also some evidence that COVID-19 cases have particularly clustered around cooler, drier regions [10, 11]. Many articles have been published to examine the effects of temperature and humidity on the spread of COVID-19. A systematic review article has also been published in [12]. Most of the researches findings are that there is a significant effect of temperature and humidity on the spread of COVID-19. However, there is still a lack of evidence because some studies found no association between COVID-19 transmission with temperature (see for example, [13, 14]).

In addition, we know that the viruses continuously mutate, and SARS-CoV-2 also change similarly. Callaway [15] state that SARS-CoV-2 has been mutating at a rate of about 1–2 mutations per month. Mutations can have a negative or positive impact on the SARS-CoV-2 virus's capability to sustain and replicate, depending on where in the SARS-CoV-2 the genome misconstructions transpire. The researcher cautioned that these mutant genealogies of the SARS-CoV-2 strain would be continued uncontrolled transmission of SARS-CoV-2 in many parts of the world. Viral mutations and variants in the United States are regularly scanned through sequence-based surveillance, laboratory studies, and epidemiological investigations [16].

Recently, a novel SARS-CoV-2 mutated (known as lineage B.1.1.7) emerged in the United Kingdom (UK) in November 2020 and expanded quickly in other countries [17]. A total of 17 mutations have been recorded in the new strain found in the UK. Virologists in Bangladesh have announced that a new SARS-CoV-2 strain is a bit similar to the one discovered in the United Kingdom recently [18]. After the mutation, we do not know the effects of the temperature and humidity on the transmission of SARS-CoV-2 strain. Hence, it is crucial to understand the behaviour of the transmission of SARS-CoV-2 for the current data.

Therefore, the main objective of this research is to investigate the effects of temperature and humidity on the transmission of SARS-CoV-2 by using flexible regression models. We try to understand the seasonal behaviour of the transmission of SARS-CoV-2 and the spread of COVID-19. A detailed material and methods regarding data source and statistical models are explained in Sect. 2. Section 3 describe the data analysis and results. Finally, the discussions and conclusions are portrayed in Sect. 4.

2 Material and Methods

2.1 Data Source

Data of Covid-19 cases are collected from the daily reports of the Institute of Epidemiology Disease Control and Research (IEDCR), Dhaka, Bangladesh, during the period of March 08, 2020 to January 31, 2021. Data are available on the website with the link https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Bangladesh. The daily temperature (measured in °C) and humidity (%) of Bangladesh are collected from the website https://www.timeanddate.com/weather/bangladesh/dhaka.

2.2 Generalised Additive Models for Location Scale and Shape

Generalized Linear Models (GLM) and Generalized Additive Models (GAM) respectively introduced by [19, 20], are very popular in statistical data analysis. Rigby and Stasinopoulos [21] proposed a generalized additive model for location, scale and shape (GAMLSS) as a way of overcoming some of the limitations associated with GLM and GAM models for regression analysis. It is a general framework of (semi)parametric regression models where the distribution of response variable does not necessarily belong to the exponential family and includes highly skew and kurtotic continuous and discrete distribution. We consider the logarithmic transformation of the number of SARS-CoV-2 infected new cases and the number of death due to COVID-19 as response variables of the GAMLSS model. In the sequel we denote, for notational convenience, "number of SARS-CoV-2 infected new cases" as "number of new cases". To avoid the logarithmic transformation of zero and the computational complexity under the GAMLSS modelling framework, we add 1.1 to each response variable before the logarithmic transformation. For each response variable, we fit the GAMLSS model separately. The probability distribution of each response variable (Y) under the GAMLSS modelling framework is chosen based on the minimum Bayesian information criterion (BIC) and Akaike information criterion (AIC) values. The Normal Exponential-t distribution is selected for $Y = \log(\text{number of new cases})$, and the Gumbel distribution is selected for $Y = \log(\text{number of death})$. A detailed selecting procedure is described in Sect. 3.2.

2.2.1 Normal Exponential-*t* Distribution

The Normal Exponential-t Distribution (NET) distribution was first introduced by [22] as a robust method of fitting the mean and scale parameters of symmetric distribution as functions of explanatory variables (X). The probability density function (pdf) of NET distribution, which is denoted as $\text{NET}(\mu, \sigma, \nu, \tau)$, is given by [22] and defined by

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{c}{\sigma}\begin{cases} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}, & \text{when } |\frac{y-\mu}{\sigma}| \le \nu \\ \exp\left\{-\nu|\frac{y-\mu}{\sigma}| + \frac{\nu^2}{2}\right\}, & \text{when } \nu < |\frac{y-\mu}{\sigma}| \le \tau \\ \exp\left\{-\nu\tau\log\left(\frac{|y-\mu|}{\tau\sigma}\right) - \nu\tau + \frac{\nu^2}{2}\right\}, & \text{when } |\frac{y-\mu}{\sigma}| > \tau \end{cases}$$

for $-\infty < y < \infty$, where $-\infty < \mu < \infty, \sigma > 0, \nu > 1, \tau > \nu$, and $c = (c_1 + c_2 + c_3)^{-1}$, where $c_1 = \sqrt{2\pi}[1 - 2\Phi(-\nu)]$, $c_2 = \frac{2}{\nu}\exp\left\{-\frac{\nu^2}{2}\right\}$ and $c_3 = \frac{2}{(\nu\tau-1)\nu}\exp\left\{-\nu\tau + \frac{\nu^2}{2}\right\}$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variate. Note that the location parameter μ is the mean of Y, for detailed density can be found in [23]. We are interested in estimating the mean function in the regression settings.

2.2.2 Gumbel Distribution

The pdf of the Gumbel distribution (also called an extreme value or Gompertz distribution), denoted by $\text{GU}(\mu, \sigma)$, is defined by:

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma}\exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right]$$

for $-\infty < y < \infty$, where $-\infty < \mu < \infty$ and $\sigma > 0$, with $E(Y) = \mu - \gamma\sigma$, where $\gamma \approx 0.577$ is Euler- Mascheroni constant and $\text{Var}(Y) = \pi^2\sigma^2/6$, for detailed density can be found in [23].

The covariates for both response variables are time (in days), temperature, and humidity are considered for this article. The beauty of the GAMLSS model is that the systematic part of it can be elaborated to endorse modelling not only the location (usually, mean) but other parameters of the distribution such as scale, shape. These parameters could be linear parametric and/or additive non-parametric functions of covariates and/or random effects. In this research, we choose flexible predictor models via fractional polynomial and B-spline functions for finding the smoothing function of the predictor time. To estimate the conditional mean of the response variable Y given covariate $X = (\text{time}, \text{temperature}, \text{humidity})$, we have to estimate the parameters (as a function of X) of the conditional distribution of Y given X. Therefore, the flexible regression models for the location function $\mu(X)$ and the scale function $\sigma(X)$ under the flexible GAMLSS modeling framework can be written as

$$\mu(X, \boldsymbol{\beta}) = \beta_0 + f(\text{time};\boldsymbol{\beta}_1) + \beta_2 \times \text{temperature} + \beta_3 \times \text{humidity}, \tag{1}$$

and

$$\log\left(\sigma(X;\boldsymbol{\gamma})\right) = \gamma_0 + f(\text{time};\boldsymbol{\gamma}_1) + \gamma_2 \times \text{temperature} + \gamma_3 \times \text{humidity}. \tag{2}$$

The (penalized) maximum likelihood estimation is used to estimate the parameters of the model (1) and (2).

2.2.3 Flexible Regression with Fractional Polynomial Function

The fractional polynomial in flexible predictor models is a generalization of the polynomial function. The general form of a fractional polynomial in x of degree m can be written as

$$f_p(x; \boldsymbol{\theta}, p_1, p_2, \ldots, p_m) = \sum_{l=0}^{m} \theta_l H_l(x), \tag{3}$$

where m is an integer and

$$H_l(x) = \begin{cases} x^{p_l} & \text{if} \quad p_l \neq p_{l-1} \\ H_{l-1}(x) \times \log(x) & \text{if} \quad p_l = p_{l-1}, \end{cases}$$

with $p_0 = 0$ and $H_0(x) = 1$, for a sequence of powers $p_1 \le p_2 \le \cdots \le p_m$ from the grid

$$\{-2, -1, -0.5, 0, 0.5, 1, 2, \max(3, m)\}.$$

The optimal combination of powers will be selected by using the smallest value of BIC.

We select $p_1 = 0$, $p_2 = 0$, $p_3 = 0.5$ and $m = 3$ for the response variable of log(number of new cases) and hence the model (3) can be written as

$$f_p(\text{time};\boldsymbol{\theta}_1;0, 0, 0.5) = \theta_{10} + \theta_{11}\log(\text{time}) + \theta_{12}[\log(\text{time})]^2 + \theta_{13}(\text{time})^{0.5}. \tag{4}$$

For the response variable of log(number of death), we select $p_1 = 1$, $p_2 = 2$ and $p_3 = 2$, and the fractional polynomial in time (in days) variable of degree $m = 3$ for the model (3) can be written as

$$f_p(\text{time}; \boldsymbol{\theta}_1; 1, 2, 2) = \theta_{10} + \theta_{11} \times \text{time} + \theta_{12} \times (\text{time})^2 + \theta_{13} \times (\text{time})^2 \times \log(\text{time}). \tag{5}$$

2.2.4 Flexible Smoothing Regression with B-Splines model

Flexible smoothing function with basis spline (B-spline) models were also fitted in order to get a more flexible approximation to the data. A general form of B-spline predictor model of x for the degree D can be written as

$$f_b(x; \theta_0, D, K) = \sum_{j=0}^{D} \theta_{0j} x^j + \sum_{k=D+1}^{D+K} \theta_{0k}(x - b_k)^D H(x > b_k), \tag{6}$$

where K is the number of knot values, b_k is the knot value at kth interval or piece and $H(x > b_k)$ is the Heaviside function taking value 1 if $x > b_k$, otherwise 0. The combination of D, K, and the number knots values will be chosen based on the lowest value of BIC.

3 Data Analysis and Results

3.1 Exploratory Data Analysis

To explore the raw data and find an indication for selecting the more sophisticated statistical model, we provide descriptive statistics and some graphical presentation of the variables in this section. Table 1 summarizes the descriptive statistics of the daily number of death due to COVID-19, SARS-CoV-2 infected new cases, and meteorological variables such as temperature and humidity for $n = 324$ days.

This study included 8033 total death, and 535,139 confirmed cases during that period. The average of the daily number of death due to COVID-19 and number of SARS-CoV-2 infected new cases are 24.79 and 1626.2, respectively. Besides, other factors showed that the lowest temperature of 20 °C with the highest temperature of 37 °C, and the lowest humidity of 21% with the highest humidity of 100%.

The histogram with kernel density plot of the number of death due to COVID-19 and the number of SARS-CoV-2 new cases are presented in Fig. 1. Figure (a) shows that the distributional shape of the number of death due to COVID-19 seems symmetric, indicating that the bell-shape distribution would be one of the best probability models for this variable. In contrast, Figure (b) reveals that the distributional shape of the number of SARS-CoV-2 infected new cases looks similar to a skewed pattern, indicating a skewed distribution would be more suitable for predicting this variable's values.

Table 1 Descriptive statistics of daily number of death due to COVID-19, number of SARS-CoV-2 infected new cases, temperature and humidity for March 08, 2020–January 31, 2021

Variables	Mean (SD)	Lowest	Highest
Number of death	24.79 (14.078)	0	64
Number of new cases	1626.2 (1032.006)	0	4019
Temperature	30.87 (3.738)	20	37
Humidity	60.71 (17.026)	21	100

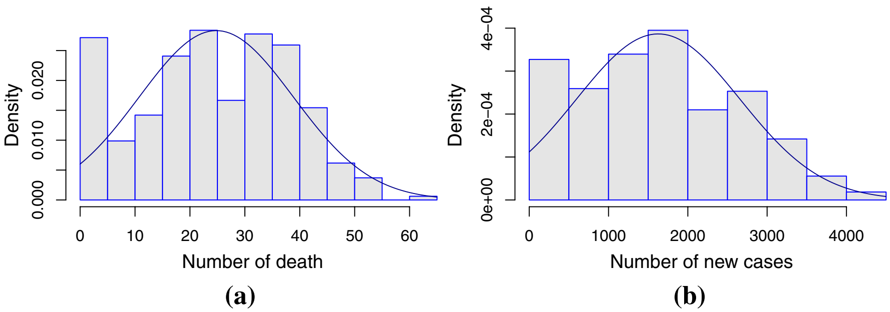

Fig. 1 Histrogram **a** Number of death due to the Covid-19; **b** number of SARS-CoV-2 infected new cases for Covid-19

The scatter plot of the number of death due to COVID-19 and the number of SARS-CoV-2 infected new cases against time index for the period from August 03, 2020 to January 31, 2021 are drawn in Fig. 2. We clearly see a nonlinear relationship between the response variables and the time index. We depict the scatter plot of the number of death due to COVID-19, and the number of SARS-CoV-2 infected new cases against humidity in Fig. 3. The relationship between the number of death due to COVID-19 and the number of SARS-CoV-2 infected new cases against temperature are shown in Fig. 4. It is observed from these figures that there is a connection between both response variables and temperature and humidity covariates.

Without adjusting time effect in the model, we consider the following regression model to explore only the conditional relationship between two response variables $Y = \log(\text{number of new cases})$ and $Y = \log(\text{number of death})$ and two covariates named temperature and humidity. The mean regression model is, for $i = 1, 2, \dots, n$

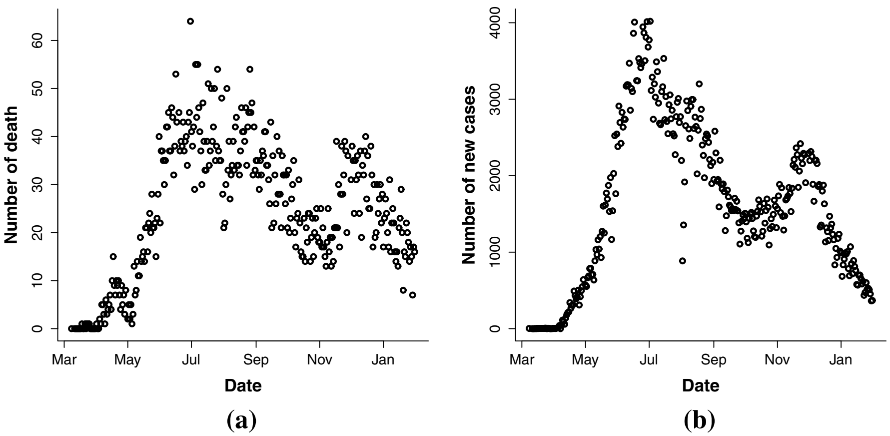

Fig. 2 Scatter diagram **a** number of death due to Covid-19; **b** number of SARS-CoV-2 infected new cases for Covid-19 against time index during the period August 03, 2020–January 31, 2021

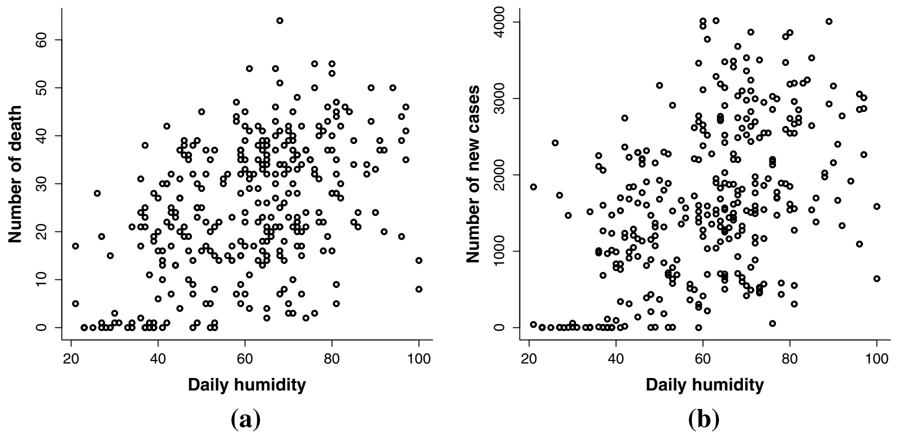

Fig. 3 Scatter diagram **a** number of death due to the Covid-19 versus humidity during the period August 03, 2020–January 31, 2021; **b** number of SARS-CoV-2 infected new cases for Covid-19 versus humidity during the period August 03, 2020–January 31, 2021

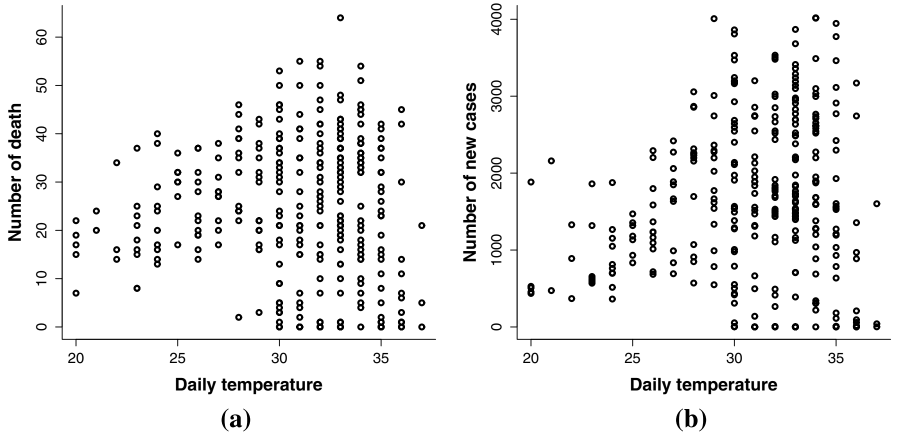

Fig. 4 Scatter diagram **a** number of death due to the Covid-19 versus daily temperature during the period August 03, 2020–January 31, 2021; **b** number of new cases for Covid-19 versus daily temperature during the period August 03, 2020–January 31, 2021

$$y_i = \beta_0 + \beta_1 \times \text{temperature}_i + \beta_2 \times \text{humidity}_i + \epsilon_i, \qquad (7)$$

where $y_i = \log(\text{number of new cases})$ (and log(number of death)) and ϵ_i is the disturbance term for ith individual. Under the classical regression model assumptions (see for example, [24]), the summary statistics of the model (7) are tabulated in Table 2. The exploratory results show that the temperature is not significant on the number of new cases and on the number of death. In contrast, the humidity is highly significant on both response variables.

Table 2 Summary statistics of the estimated model given in (7)

Estimate	Log(Number of new cases)			Log(number of death)		
	$\hat{\beta}\left(se(\hat{\beta})\right)$	*t* value	*P* value	$\hat{\beta}\left(se(\hat{\beta})\right)$	*t* value	*P* value
Intercept	3.885 (0.866)	4.489	< 0.001	1.840 (0.462)	3.983	< 0.001
Temperature	– 0.018 (0.024)	– 0.734	0.464	– 0.196 (0.013)	– 1.509	0.132
Humidity	0.056 (0.005)	10.518	< 0.001	0.028 (0.003)	9.915	< 0.001

Next, we consider the time (in days) variable as a covariate in the model. Since the exploratory data analysis shows a nonlinear relationship between time and response variables, we need advanced computer-intensive statistical models for further research.

3.2 Generalised Additive Models for Location Scale and Shape (GAMLSS) family

For selecting the best probability model for the response variable $Y = \log(\text{number of new cases})$, the summary including AIC and BIC values with their degrees of freedom of all selected candidate distributions coming from the GAMLSS family, are provided in Table 7 in "Appendix". Above all of the distributions, we selected five possible candidate distributions based on the minimum BIC provided in Table 9 in "Appendix". It is noticed that the smallest BIC and AIC are observed for the NET model. In contrast, the highest value of BIC (and also AIC) is observed for the Skew-*t* type-3 model. Based on the minimum BIC, we select the NET model to explain the transmission of SARS-CoV-2 for further investigation.

Similarly, for the response variable $Y = \log(\text{number of death})$, the summary results including AIC and BIC values with degrees of freedom of all selected candidate distributions coming from the GAMLSS family, are provided in Table 8 in "Appendix". Above all of the distributions, we selected five possible candidate distributions based on minimum BIC presented in Table 10 in "Appendix". It is noted that the smallest values of BIC and AIC are observed for the Gumbel model. In contrast, the highest value of BIC (and also AIC) is observed for the Skew-*t* type-4 model. Therefore, the Gumbel model is chosen as the best model to describe the number of death due to COVID-19 for further analysis.

3.2.1 Flexible Regression with Fractional Polynomial Function

A fractional polynomials flexible models for log(number of new cases) given in (4) and for $Y = \log(\text{number of death})$ given in (5) are estimated within the GAMLSS modelling framework via the best chosen probability distribution of each response variable. The fitted flexible predictor model for the $\mu(X)$ for log(number of new cases) is

$$\hat{\mu}(X_i;\hat{\boldsymbol{\beta}}) = 10.524 + f_p(\text{time}_i, \hat{\boldsymbol{\beta}}_1) - 0.089 \times \text{temperature}_i - 0.015 \times \text{humidity}_i, \tag{8}$$

and the estimated flexible predictor model (2) is $\sigma(X;\hat{\boldsymbol{\gamma}}) = \exp(-1.322) = 0.267$. We, here, leave out the insignificant effects of the estimated model. The corresponding estimated fractional polynomial model for the $\mu(X)$ in time (in days) of degree 3 is

$$f_p(\text{time}_i, \hat{\boldsymbol{\beta}}_1) = 35.067 + 18.203 \times \log(\text{time}_i) + 2.204 \times [\log(\text{time}_i)]^2 - 33.766 \times (\text{time}_i)^{0.5}.$$

The summary statistics of this estimated flexible predictor model (8) is tabulated in Table 3. Hence, the estimated flexible regression model for mean function $E(Y|X) = \mu(X)$ of the conditional NET distribution under the GAMLSS modeling framework is

$$\hat{\mu}(X_i;\hat{\boldsymbol{\beta}}) = 45.591 + 18.203 \times \log(\text{time}_i) + 2.204 \times [\log(\text{time}_i)]^2 - 33.766 \times (\text{time}_i)^{0.5} - 0.089 \times \text{temperature}_i - 0.015 \times \text{humidity}_i. \quad (9)$$

Note that the values of two fixed parameters ν is 1.5 and τ is 2 in the GAMLSS modelling framework. We found the Global Deviance is 288.373, AIC is 308.373, and SBC is 346.181 for the final fitted model. Table 3 shows that the temperature and humidity are highly significant on the number of SARS-CoV-2 infected new cases. In addition, the regression coefficients for both temperature and humidity are negative which indicates that there is a negative relationship between these variables and the number of SARS-CoV-2 infected new cases.

Similarly, for the response variable of log(number of death), the estimated flexible regression model under the GAMLSS modelling framework of the location function $\mu(X)$ of Gumbel distribution is:

$$\hat{\mu}(X_i;\hat{\boldsymbol{\beta}}) = 5.495 + f_p(\text{time}_i, \hat{\boldsymbol{\beta}}_1) - 0.062 \times \text{temperature}_i - 0.008 \times \text{humidity}_i, \quad (10)$$

and the estimated flexible regression model of (2) is $\hat{\sigma}(X;\hat{\boldsymbol{\gamma}}) = \exp(-1.264) = 0.283$. The estimated fractional polynomial model for the $\mu(X)$ in time (in days) of degree 3 is: for $i = 1, 2, \dots, n$

$$f_p(\text{time}_i, \hat{\boldsymbol{\beta}}_1) = -3.584 + 9.832 \times \text{time}_i - 5.549 \times \text{time}_i^2 + 2.398 \times \text{time}_i^2 \times \log(\text{time}_i).$$

The summary statistics of the estimated model are provided in Table 4. Hence, the estimated flexible regression model of (10) can be written as

Table 3 Summary statistics of the estimated flexible predictor model given in (9)

Estimate	Coefficients	SE	t value	P value
Intercept	10.524	0.138	75.75	< 0.0001
Temperature	– 0.089	0.005	– 21.40	< 0.0001
Humidity	– 0.015	0.001	– 14.57	< 0.0001

Table 4 Summary statistics of the estimated flexible regression model via fractional polynomial function

Estimate	Coefficients	SE	t value	P value
Intercept	5.495	0.155	35.380	< 0.0001
Temperature	− 0.062	0.005	− 13.579	< 0.0001
Humidity	− 0.008	0.001	− 8.073	< 0.0001

$$\hat{\mu}(X_i;\hat{\boldsymbol{\beta}}) = 1.911 + 9.832 \times \text{time}_i - 5.549 \times \text{time}_i^2 + 2.398 \times \text{time}_i^2 \times \log(\text{time}_i) - 0.062 \times \text{temperature}_i - 0.008 \times \text{humidity}_i. \quad (11)$$

Finally, we obtain the estimated flexible regression model for mean function $E(Y|X) = \mu(X) - \gamma\sigma(X)$ of conditional Gumbel distribution under the GAMLSS modeling framework is

$$\begin{aligned}\widehat{E(Y_i|X_i)} &= \hat{\mu}(X_i;\hat{\boldsymbol{\beta}}) - \gamma\hat{\sigma}(X_i;\hat{\boldsymbol{\gamma}}) \\ &= 1.748 + 9.832 \times (\text{time}_i) - 5.549 \times (\text{time}_i)^2 + 2.398 \times (\text{time}_i)^2 \times \log(\text{time}_i) \\ &\quad - 0.062 \times \text{temperature}_i - 0.008 \times \text{humidity}_i.\end{aligned}$$

For this model, the global deviance is 199.555, AIC is 219.555 and SBC is 257.363. Table 4 shows that the temperature and humidity are highly significant on the number of death due to COVID-19.

3.2.2 Flexible Smoothing Regression with B-Splines Function

For the response variable log(number of new cases), we also use B-spline function given in (6) for estimating $\mu(X;\boldsymbol{\beta})$ and $\sigma(X;\boldsymbol{\gamma})$ of the NET distribution. With $D = 3$ and $K = 4$ in the model (6), the B-spline predictor function for estimating $\mu(X;\boldsymbol{\beta})$, the estimated B-spline smoothing function of $f_b(\text{time}_i;\boldsymbol{\beta}_0, 3, 4)$; $\forall i = 1, 2, \dots, n$ is

$$\begin{aligned}f_b(\text{ time}_i;\hat{\boldsymbol{\beta}}_0, 3, 4) = &-0.635 + 5.703 \times \text{ time}_i - 9.591 \times \text{time}_i^2 + 9.919 \times \text{time}_i^3 \\ &+ H(\text{time}_i > b_4)[8.534 \times (\text{time}_i - 65.6)^3 + 9.285 \times (\text{time}_i - 130.2)^3 \\ &+ 8.219 \times (\text{time}_i - 194.8)^3 + 7.295 \times (\text{time}_i - 259.4)^3].\end{aligned} \quad (12)$$

With $D = 3$ and $K = 1$ in the model (6), the estimated function of $f_b(\text{time}_i;\boldsymbol{\gamma}_0, 3, 1)$ for $i = 1, 2, \dots, n$, is

$$\begin{aligned}f_b(\text{time}_i;\hat{\boldsymbol{\gamma}}_0, 3, 1) = 1.529 - 3.627 \times \text{time}_i - 2.365 \times \text{time}_i^2 - 2.726 \times \text{time}_i^3 \\ - 2.992(\text{time}_i - 162.5)^3 H(\text{time}_i > b_1).\end{aligned} \quad (13)$$

Using the estimated B-spline function for estimating $\mu(X;\boldsymbol{\beta})$ given in (12), we find the estimated flexible regression function of $E(Y|X) = \mu(X;\boldsymbol{\beta})$ which is

$$
\begin{aligned}
\mu(X_i;\hat{\boldsymbol{\beta}}) = & -0.635 + 5.703 \times \text{ time}_i - 9.591 \times \text{time}_i^2 + 9.919 \times \text{time}_i^3 \\
& + H(\text{time}_i > b_4)[8.534 \times (\text{time}_i - 65.6)^3 + 9.285 \times (\text{time}_i - 130.2)^3 \\
& + 8.219 \times (\text{time}_i - 194.8)^3 + 7.295 \times (\text{time}_i - 259.4)^3] \\
& - 0.022 \times \text{temperature}_i - 0.003 \times \text{humidity}_i.
\end{aligned}
$$

The summary statistics of the estimated function $\hat{\mu}(X,\hat{\boldsymbol{\beta}})$ and $\log\left(\hat{\sigma}(X,\hat{\boldsymbol{\gamma}})\right)$ are presented in Table 5. For this estimated model, the Global Deviance, AIC and SBC are −46.572, −12.572 and 51.700, respectively. In the estimated mean function $\widehat{E(Y|X)} = \hat{\mu}(X_i;\hat{\boldsymbol{\beta}})$, we see the slope co-efficient of temperature (β_1) and humidity (β_2) are negative which indicates that there is a negative relationship between these variables. In addition, the regression co-efficients for both temperature and humidity are highly significant on the number of SARS CoV-2 infected new cases. Similarly for estimated $\log\left(\hat{\sigma}(X,\hat{\boldsymbol{\gamma}})\right)$, we see the slope co-efficient of temperature (γ_1) and humidity (γ_2) are also negative which indicates that there is a negative relationship between these variables where both regression coefficients are not significant on the number of SARS-CoV-2 infected new cases at 5 % level of significance.

For the response variable log(number of death), we use the B-spline function of time predictor to estimate $\mu(X)$ and $\sigma(X)$ of the Gumbel distribution. For estimating $\mu(X;\boldsymbol{\beta})$, we select $D = 3$ and $K = 4$ in the (6) and the estimated flexible function of $f_b(\text{time}_i;\boldsymbol{\beta}_0, 3, 4)$ for $i = 1, 2, \ldots, n$, is

$$
\begin{aligned}
f_b(\text{time}_i;\hat{\boldsymbol{\beta}}_0, 3, 4) = & \ 0.978 + 1.409 \times \text{time}_i + 3.809 \times \text{time}_i^2 + 4.361 \times \text{time}_i^3 \\
& + H(\text{time}_i > b_4)[3.733 \times (\text{time}_i - 65.6)^3 + 3.289 \times (\text{time}_i - 130.2)^3 \\
& + 3.667 \times (\text{time}_i - 194.8)^3 + 2.600 \times (\text{time}_i - 259.4)^3].
\end{aligned}
\tag{14}
$$

Table 5 The summary statistics of flexible regression models of $\mu(X;\boldsymbol{\beta})$ and $\log(\sigma(X;\boldsymbol{\gamma}))$ via B-spline smoothing function for the response variable log(number of new cases)

$\mu(X,\boldsymbol{\beta})$				$\log(\sigma(X,\boldsymbol{\gamma}))$			
$\boldsymbol{\beta}$	$\hat{\boldsymbol{\beta}}\left(se(\hat{\boldsymbol{\beta}})\right)$	t value	P value	$\boldsymbol{\gamma}$	$\hat{\boldsymbol{\gamma}}\left(se(\hat{\boldsymbol{\gamma}})\right)$	t value	P value
β_{00}	– 0.635 (0.938)	– 0.677	0.499	γ_{00}	1.529 (1.398)	1.093	0.275
β_{01}	5.703 (0.896)	6.366	< 0.005	γ_{01}	– 3.627 (0.832)	– 4.361	< 0.005
β_{02}	9.591 (0.848)	11.311	< 0.005	γ_{02}	– 2.365 (0.652)	– 3.630	0.0003
β_{03}	9.919 (0.851)	11.655	< 0.005	γ_{03}	– 2.726 (0.726)	– 3.753	0.0002
β_{04}	8.534 (0.844)	10.114	< 0.005	γ_{04}	– 2.992 (0.639)	– 4.685	< 0.005
β_{05}	9.285 (0.869)	10.675	< 0.005				
β_{06}	8.219 (0.861)	9.547	< 0.005				
β_{07}	7.295 (0.865)	8.437	< 0.005				
β_1	– 0.022 (0.007)	– 2.985	0.003	γ_1	– 0.013 (0.033)	– 0.394	0.694
β_2	– 0.003 (0.001)	– 2.178	0.031	γ_2	– 0.005 (0.006)	– 0.786	0.433

To estimate $\sigma(X;\boldsymbol{\gamma})$ of the Gumbel distribution, we select $D = 3$ and $K = 0$ in the model given in (6). The estimated flexible function of $f_b(\text{time}_i; \boldsymbol{\gamma}_0, 3, 0)$ for ith individual is

$$f_b(\text{time}_i;\widehat{\boldsymbol{\gamma}}_0, 3, 0) = 0.086 - 1.894 \times \text{time}_i - 0.929 \times \text{time}_i^2 - 1.118 \times \text{time}_i^3. \tag{15}$$

Using the estimated B-spline function given in (14), the estimated function of $\mu(X; \boldsymbol{\beta})$ can be written as

$$\begin{aligned}\widehat{\mu}(X_i; \widehat{\boldsymbol{\beta}}) = 0.978 &+ 1.409 \times \text{time}_i + 3.809 \times \text{time}_i^2 + 4.361 \times \text{time}_i^3\\ &+ H(\text{time}_i > b_4)[3.733 \times (\text{time}_i - 65.6)^3 + 3.289 \times (\text{time}_i - 130.2)^3\\ &+ 3.667 \times (\text{time}_i - 194.8)^3 + 2.600 \times (\text{time}_i - 259.4)^3]\\ &- 0.032 \times \text{temperature}_i - 0.003 \times \text{humidity}_i.\end{aligned}$$

By using the estimated B-spline model given in (15), the estimated scale function $\widehat{\sigma}(X_i; \widehat{\boldsymbol{\gamma}})$ for $i = 1, 2, \ldots, n$ is

$$\begin{aligned}\widehat{\sigma}(X_i; \widehat{\boldsymbol{\gamma}}) = \exp(&0.086 - 1.894 \times \text{time}_i - 0.929 \times \text{time}_i^2 - 1.118 \times \text{time}_i^3\\ &- 0.008 \times \text{temperature}_i - 0.004 \times \text{humidity}_i).\end{aligned}$$

The summary statistics of the estimated models are tabulated in Table 6. In the estimated mean function $\widehat{\mu}(X_i; \widehat{\boldsymbol{\beta}})$, we see the slope co-efficient of temperature (β_1) and humidity (β_2) are positive which indicates that there is a positive relationship between these variables. Table 6 shows that, the temperature is highly significant but the humidity is not significant on the number of death due to COVID-19 at 5% level of significance. Similarly for estimated $\log\left(\widehat{\sigma}(X, \widehat{\boldsymbol{\gamma}})\right)$, we see the slope co-efficient of temperature (γ_1) and humidity (γ_2) are negative which indicates that there is a negative relationship between these variables and the number of death due to

Table 6 The summary statistics of flexible regression models of $\mu(X; \boldsymbol{\beta})$ and $\log(\sigma(X; \boldsymbol{\gamma})$ via B-spline smoothing function for the response variable log(number of death)

$\mu(X;\boldsymbol{\beta})$				$\log(\sigma(X;\boldsymbol{\gamma}))$			
$\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}}\left(se(\widehat{\boldsymbol{\beta}})\right)$	t value	P value	$\boldsymbol{\gamma}$	$\widehat{\boldsymbol{\gamma}}\left(se(\widehat{\boldsymbol{\gamma}})\right)$	t value	P value
β_{00}	0.978 (0.485)	2.018	0.045	γ_{00}	0.086 (1.073)	0.080	0.936
β_{01}	1.409 (0.332)	4.242	< 0.005	γ_{01}	– 1.894 (0.794)	– 2.387	0.018
β_{02}	3.809 (0.248)	15.334	< 0.005	γ_{02}	– 0.929 (0.371)	– 2.506	0.013
β_{03}	4.361 (0.268)	16.245	< 0.005	γ_{03}	– 1.118 (0.444)	– 2.521	0.012
β_{04}	3.733 (0.254)	14.678	< 0.005				
β_{05}	3.289 (0.256)	12.867	< 0.005				
β_{06}	3.667 (0.264)	13.886	< 0.005				
β_{07}	2.600 (0.256)	10.142	< 0.005				
β_1	– 0.032 (0.012)	– 2.688	0.008	γ_1	– 0.008 (0.028)	– 0.291	0.771
β_2	– 0.003 (0.002)	– 1.703	0.089	γ_2	– 0.004 (0.005)	– 0.777	0.438

COVID-19. In addition, both regression co-efficients are not significant on the number of death due to COVID-19 at 5% level of significance. Based on these results, we obtain the estimated flexible regression model via B-spline smoothing function for $E(Y|X) = \mu(X) - \gamma\sigma(X)$, where $\gamma \approx 0.577$ is Euler- Mascheroni constant of conditional Gumbel distribution. Hence, the estimated mean function for ith individual ($\forall i = 1, 2, \dots, n$) can be written as

$$\begin{aligned}\widehat{E(Y_i|X_i)} &= \hat{\mu}(X_i; \hat{\boldsymbol{\beta}}) - \gamma\hat{\sigma}(X_i; \hat{\boldsymbol{\gamma}}) \\ &= 0.978 + 1.409 \times \text{time}_i + 3.809 \times \text{time}_i^2 + 4.361 \times \text{time}_i^3 \\ &\quad + H(\text{time}_i > b_4)[3.733 \times (\text{time}_i - 65.6)^3 + 3.289 \times (\text{time}_i - 130.2)^3 \\ &\quad + 3.667 \times (\text{time}_i - 194.8)^3 + 2.600 \times (\text{time}_i - 259.4)^3] \\ &\quad - 0.032 \times \text{temperature}_i - 0.003 \times \text{humidity}_i \\ &\quad - 0.629 \exp(-1.894 \times \text{time}_i - 0.929 \times \text{time}_i^2 - 1.118 \times \text{time}_i^3 \\ &\quad - 0.008 \times \text{temperature}_i - 0.004 \times \text{humidity}_i).\end{aligned}$$

We also calculate the predicted values of response variable via fractional polynomial and B-spline models. The graphical presentation of actual values and predicted values are depicted in Fig. 5.

We see the estimated curve via B-spline function is a smooth curve which is expected. On the other hand, the estimated curve via fractional polynomial function is not smooth. However, estimated both curves are very close.

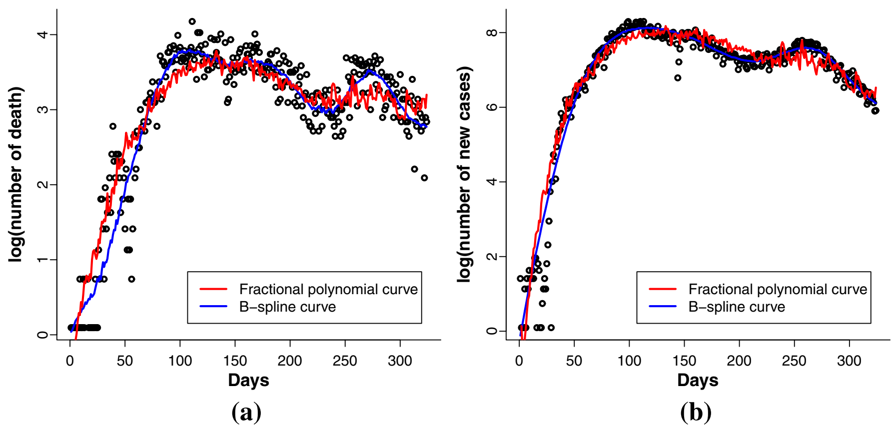

Fig. 5 Fractional polynomial curve versus basis spline curve **a** number of death due to the Covid-19 versus days during the period August 03, 2020–January 31, 2021; **b** number of SARA-CoD-2 infected new cases for Covid-19 versus days during the period August 03, 2020–January 31, 2021

4 Discussion and Conclusions

This study examined whether the temperature and humidity in the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) affect humans' respiratory system that causes Coronavirus disease (COVID-19). We relied on the daily count of the number of confirmed SARS-CoV-2 infected new cases and the total number of death due to COVID-19 per day from Institute of Epidemiology Disease Control and Research (IEDCR), Dhaka, Bangladesh. A generalised additive model location scale and shape (GAMLSS) model is used to examine the effect of temperature and humidity on the number of confirmed SARS-CoV-2 infected daily new cases and the total number of death due to COVID-19 separately. Without adjusting the time effect in exploratory data analysis, we did not find the significant impact of temperature on both response variables.

To investigate the significant effects of temperature and humidity after adjusting the time variable, we used the flexible GAMLSS model. The best response distribution is chosen based on the minimum BIC under the GAMLSS modeling framework. The Normal Exponential-*t* distribution for log(number of new cases) and Gumbel distribution for log(number of death) are selected. To estimate the systematic part of the GAMLSS model, we have employed two flexible predictor models such as (i) fractional polynomial model and (ii) B-spline smoothing model. Both models suggested that high temperature and high humidity significantly reduce the transmission of SARS-CoV-2. A fractional polynomial model indicates that high temperature and high humidity significantly reduce the number of deaths due to COVID-19. Many researches support these results (see, for example in [12]) but these are opposite of the findings of [25]. According to the fitted fractional polynomial model, for every 1 °C increase in temperature, the number of deaths due to COVID-19 reduced by 8.9% (95% CI: 7.3%, 10.5%) and daily new cases reduced by 6.2% (95% CI: 4.6%, 7.8%); for every 1% increase in humidity, the number of deaths due to COVID-19 reduced by 1.5% (95% CI: 1.2%, 1.8%) and daily new cases reduced by 0.8% (95% CI: 0.48%, 1.1%), holding all the other factors constant.

On the other hand, the B-spline model suggested that high temperature and high humidity minimise the number of death due to COVID-19, where the temperature significantly affects. However, the humidity significantly affects the number of deaths at a 10% level of significance but not significantly affects at a 5% level of significant. Note that there are a number of reasons for getting the insignificant effect of the humidity in the B-Spline model. It might happen that the sample size ($n = 324$) is not enough to find the significant humidity effect in the B-spline model. Moreover, the temperature and humidity are correlated. As the response variable is already well explained by the temperature and B-spline function of the time variable, it is possible to get a high p value of the regression coefficient of humidity. According to the fitted B-spline model, for every 1°C increase in temperature, the daily number of deaths due to COVID-19 reduced by 0.8% (95% CI: 7.9%, 9.5%) and the daily new cases reduced by 2.2% (95% CI: 0.03%, 4.4%); for every 1% increase in humidity the number of deaths due to COVID-19 reduced by 0.4% (95% CI: 1.2%, 1.9%) and daily new cases reduced by 0.3% (95% CI: 0.02%, 0.62%), holding all the other factors constant.

Although our analysis shows that the temperature and humidity will be affected by the transmission of SARS-CoV-2, we notice that the temperature and humidity alone do not explain most of the variability of the transmission of SARS-CoV-2 infection. To find the actual behaviour and variability of transmission of SARS-CoV-2 infection, we have to consider the temperature and humidity with other confounding factors such as population density, public health policies, public health intervention, social isolation campaigns, actual diagnosis, transportation system, people lifestyle, etc. in the computer-intensive statistical model.

Appendix

See the Tables 7, 8, 9 and 10.

Table 7 Distribution of the response variable log(number of new cases) under the GAMLSS modelling framework

Distribution	AIC	BIC	*df*
Beta	475.5216	520.8905	12
Box–Cox transformation	369.5741	411.1623	11
Box–Cox cole and green	648.6555	686.4629	10
Exponential	1845.012	1871.478	7
Exponential generalized beta type 2	346.9286	392.2975	12
Gamma	781.051	818.8591	10
Generalized beta Type 2	761.1340	806.5029	12
Generalized inverse Gaussian	1316.182	1353.990	10
Generalized gamma	731.2409	772.8290	11
Generalized **t**	304.0473	**349.4163**	12
Gumble	363.6858	401.4932	10
Inverse Gaussian	2103.315	2133.561	8
Johnson's SU	318.6647	364.0336	12
Logarithmic	318.6647	364.0336	12
Lognormal	847.3616	885.1690	10
Log-normal (Box–Cox)	1485.072	1522.879	10
Normal	529.8718	567.6792	10
Normal family	351.9066	393.4947	11
Normal linear quadratic	1218.531	1256.338	10
Normal exponential **t**	305.7230	**343.5305**	10
Pareto type 2	1877.099	1907.345	8
Pareto type 2 original	4383.501	4406.186	6
Power exponential (type 1)	368.8753	410.4635	11
Power exponential (type 2)	369.5741	411.1623	11
Reverse Gumble	741.2917	779.0992	10
Skew power exponential type 1	349.2828	394.6517	12
Skew power exponential type 2	354.3330	399.7020	12
Skew power exponential type 3	350.9642	396.3331	12
Skew power exponential type 4	337.4964	382.8654	12
Shash	325.7598	371.1287	12
Shash original	342.5825	387.9514	12
Skew t type 1	310.3102	**355.6792**	12
Skew t type 2	314.6741	461.9935	10
Skew t type 3	316.4191	**361.788**	12
Skew t type 4	302.2429	**347.6118**	12
Skew t type 5	309.8542	458.0314	12
Weibull	960.9114	998.7188	10
Weibull (PH parameterization)	1225.135	1262.943	10
Weibull (mu as mean)	963.949	1001.756	10
Zero adjusted IG	2105.315	2139.342	9

Bold values represent importance of related component and compare with other results

Table 8 Distribution of the response variable log(number of death) under the GAMLSS modelling framework

Distribution	AIC	BIC	*df*
Box–Cox transformation	391.2810	436.6499	12
Box–Cox cole and green	445.7253	487.1573	11
Exponential	1278.677	1305.142	7
Exponential generalized beta type 2	224.2077	**269.5766**	12
Gamma	781.0517	818.8591	10
Generalized beta type 2	461.8899	507.2588	12
Generalized gamma	484.7071	526.2953	11
Generalized *t*	244.7373	290.1062	12
Gumbel	221.9905	**259.7980**	10
Inverse Gaussian	1425.243	1463.051	10
Johnson's SU	222.2492	**267.6181**	12
Log-normal	847.3616	885.1690	10
Logit-normal	847.3616	885.1690	10
Logistic	246.0621	283.8695	10
Normal exponential *t*	244.4399	282.2473	10
Normal	269.5845	307.3920	10
Normal family	250.9953	292.5835	11
Normal linear quadratic	910.6652	942.8346	10
Pareto type 2	1310.657	1340.903	8
Pareto type 2 original	1311.364	1344.314	6
Power exponential	247.6591	289.2472	11
Power exponential (type 2)	247.6752	289.2634	11
Reverse Gumbel	377.6024	415.4098	10
Reverse generalized extreme	223.9409	**265.5291**	11
Skew power exponential type 1	229.1898	274.5587	12
Skew power exponential type 2	249.6290	294.9979	12
Skew power exponential type 3	228.4612	273.8301	12
Skew power exponential type 4	223.3892	268.7581	12
Shash	224.6456	270.0145	12
Skew t type 1	225.9048	271.2737	12
Skew t type 2	230.5579	275.9268	12
Skew t type 3	225.1310	270.4999	12
Skew t type 4	223.8488	**269.2177**	12
Skew t type 5	224.6678	270.0367	12
Weibull	607.6390	645.4465	10
Weibull (PH parameterization)	648.1237	685.9312	10
Weibull (mu as mean)	607.6391	645.4465	10
Zero adjusted inverse Gaussian	1427.243	1468.831	11

Bold values represent importance of related component and compare with other results

Table 9 Goodness-of-fit statistics for selecting the best response distribution of log(number of new cases)

Distribution	AIC	BIC	*m*	*df*
Normal exponential *t*	308.373	**346.181**	3	10
	430.145	460.391	2	8
	839.973	862.658	1	6
Skew-*t* type 4	304.744	**350.113**	3	12
	422.128	459.935	2	10
	802.695	832.941	1	8
Generalized *t*	306.172	**351.541**	3	12
	426.326	464.134	2	10
	817.814	848.060	1	8
Skew-*t* type 1	312.875	**358.244**	3	12
	422.449	460.257	2	10
	817.572	847.818	1	8
Skew-*t* type 3	318.883	**364.252**	3	12
	425.063	462.870	2	10
	813.698	843.944	1	8

Bold values represent importance of related component and compare with other results

Table 10 Goodness-of-fit statistics for selecting the best response distribution of log(number of death)

Distribution	AIC	BIC	*m*	*df*
Gumbel	219.555	**257.363**	3	10
	306.370	336.616	2	8
	473.084	495.768	1	6
Reverse generalized extreme	221.531	**263.119**	3	11
	305.942	339.968	2	9
	471.106	497.572	1	7
Johnson’s SU	220.237	**265.606**	3	12
	311.545	349.353	2	10
	473.223	503.468	1	8
Exponential generalized beta type 2	222.236	**267.605**	3	12
	309.918	347.725	2	10
	480.896	511.142	1	8
Skew-*t* type 4	222.241	**267.610**	3	12
	320.571	358.378	2	10
	482.300	512.546	1	8

Bold values represent importance of related component and compare with other results

Authors' contributions Md. Rezaul Karim designed and directed the research. Mst. Bithi Akter, Sejuti Haque and Nazmin Akter collected data and carried out the implementation. Md. Rezaul Karim and Mst. Bithi Akter analyzed the data and shared the R code with other authors. Bithi Akter, Sejuti Haque and Nazmin Akter finalized the data analysis and tabulated the results. They performed other calculations and wrote a draft copy of the manuscript. Md. Rezaul Karim finalized the manuscript with input from all authors.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data and Code Availability The data sources are provided the Section 2. We will provide data and R code if anyone needs these.

Declarations

Ethics approval We have conducted ourselves with integrity, fidelity, and honesty. We have not intentionally engaged in or participated in any form of malicious harm to another person or animal.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 76:71–76
3. Sun J, He WT, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J et al (2020) Covid-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 26(5):483–495
4. Worldometers (2021) COVID-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/. Accessed 25 February 2021
5. Siddiqur Rahman Md, Towfiqul Islam Abu Reza Md (2019) Are precipitation concentration and intensity changing in Bangladesh overtimes? Analysis of the possible causes of changes in precipitation systems. Sci Total Environ 690:370–387
6. Cheval S, Mihai Adamescu C, Georgiadis T, Herrnegger M, Piticar A, Legates DR (2020) Observed and potential impacts of the COVID-19 pandemic on the environment. Int J Environ Res Public Health 17(11):4140
7. Bashir MF, Ma B, Komal B, Bashir MA, Tan D, Bashir M et al (2020) Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci Total Environ 728:138835
8. Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, Yan J, Niu J, Zhou J, Luo B (2020) Effects of temperature variation and humidity on the death of COVID-19 in Wuhan. China. Sci Total Environ 724:138226
9. Chan K-H, Peiris JSM, Lam S, Poon LL, Yuen KY, Seto WH (2011) The effects of temperature and relative humidity on the viability of the SARS coronavirus. Adv Virol 2011
10. Wang J, Tang K, Feng K, Lv W (2020) High temperature and high humidity reduce the transmission of COVID-19 (March 9, 2020). SSRN 3551767
11. Tan J, Mu L, Huang J, Yu S, Chen B, Yin J (2005) An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation. J Epidemiol Commun Health 59(3):186–192
12. Mecenas P, Bastos RTRM, Vallinoto ACR, Normando D (2020) Effects of temperature and humidity on the spread of COVID-19: a systematic review. PLOS ONE 15(9):e0238339
13. Yao Y, Pan J, Liu Z, Meng X, Wang W, Kan H, Wang W (2020) No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. Eur Respir J 55(5)

14. Xie J, Zhu Y (2020) Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ 724:138201
15. Callaway E (2020) The coronavirus is mutating-does it matter? Nature 585(7824):174–177
16. Fontanet A, Autran B, Lina B, Kieny MP, Karim SSA, Sridhar D (2021) SARS-CoV-2 variants and ending the COVID-19 pandemic. The Lancet
17. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD et al (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. Science
18. The Daily Star (2020) No need to p]anic over new strain-experts say; local virologists found the variant in early November, almost similar to the UK one. https://www.thedailystar.net/frontpage/news/no-need-panic-over-new-strain-2016793. Accessed 17 March 2021
19. Nelder JA, Wedderburn RWM (1972) Generalized linear models. J R Stat Soc Ser A (General) 135(3):370–384
20. Hastie TJ, Tibshirani RJ (1990) Generalized additive models, vol 43. CRC Press, Boca Raton
21. Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. J R Stat Soc Ser C (Appl Stat) 54(3):507–554
22. Rigby RA, Stasinopoulos DM (1994) Robust fitting of an additive model for variance heterogeneity. In: COMPSTAT: proceedings in computational statistics. Physica, Heidelberg, pp 263–268
23. Stasinopoulos M, Rigby B, Akantziliotou C (2008) Instructions on how to use the gamlss package in R second edition. https://www.researchgate.net/publication/228429663_Instructions_on_how_to_use_the_gamlss_package_in_R_Second_Edition
24. Gujarati DN, Porter DC (2010) Essentials of econometrics
25. Gupta D (2020) Effect of ambient temperature on COVID-19 infection rate. SSRN 3558470