Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/130117

Forecasting the Climate Change through the Distributions of Solar Radiation and Maximum Temperature

M. M. El Genidy[1,], E. A. Hebeshy[1], B. S. El-Desouky[2] and R. S. Gomaa[2]*

[1]Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt
[2]Department of Mathematics, Faculty of Science, Mansoura University, 35516 Mansoura, Egypt

Received: 7 Feb. 2023, Revised: 4 May 2023, Accepted: 5 May 2023.
Published online: 1 Jan. 2024.

Abstract: The climate change crisis is negatively affecting the world and is the focus of many researchers' attention for its life-threatening economic and climate impact on Earth. Therefore, this study aims to estimate the joint distribution function (EF_{XY}) of both daily solar radiation (S) and daily maximum temperature (T) along with the Markov property. In this study, three-parameter distributions have been utilized with S and T, which are generalized extreme value (GEV) and Weibull (W-3P), respectively. Each of these parameters and the joint distribution function ($F(X,Y)$) have been estimated. Four real data of S and T in Queensland, Australia during two consecutive years are applied. The method of maximum likelihood estimation (MLE) is applied on the proposed distributions of S and T to estimate their parameters, which was validated using Goodness-of-Fit tests. In addition, the logarithmic (LF_{XY}) model and the multi-regression model (MF_{XY}) for $F(X,Y)$ are obtained. The results have been compared and the EF_{XY} and LF_{XY} are found to be non-equivalently, while the EF_{XY} and MF_{XY} are equivalent and homogeneous, confirming the validity of the joint distribution function estimate with the least error. Thus, the climate change probabilities are more accurately predictable by knowing both X and Y or by knowing both $F(X)$ and $F(Y)$ with minimal error.

Keywords: Generalized extreme value distribution, Maximum likelihood estimation, Multi-regression, Markov property, Weibull-3P distribution.

.

1 Introduction

Under the current conditions, knowledge of climate sciences has become a common practice since the world has been subject to climate changes that negatively affect the world and threaten life on earth. What confirms this is the greatest novelty on the planet in the last hundred years. This is the fossil fuel that loaded the atmosphere with huge amounts of emissions and pollutants [1].

Climate change has significant effects such as melting ice, extinction of many species of animals, and diseases. Therefore, alternative energy to fossil fuels should be used, and the course and term of the industrial revolution should be modified to sustainable development to preserve the environment and to overcome the climate change crisis [2].

Solar radiation, or irradiance, on the surface of the Earth has emerged in different fields of study due to its importance in understanding climate change. In the previous body of research, several authors acting of the solar radiation data by medium and long-term time series. Currently, the main objective is to construct statistical methods that are more accessible and accurate for predicting the global climate changes [3].

In statistical literature, some models of solar radiation models were developed by the analytical measuring data. This is done through employing some criteria in order to choose the probabilistic distributions shown in two models of artificial solar radiation generation in [4].

It should be added that the time series of solar radiation were discussed. They can be taken from data measured by determining periods of representative measurement data and calculating an average radiation year. More recently, configure Markov transition matrix model using the neural networks was studied by [5].

In [6, 7], authors explained climate change's influence on industry and used the artificial neural network for prediction of global incident solar radiation through the European Centre for Medium Range Weather Forecasting.

Further, several statistical models such as temperature model (TM), time series and Fourier series (TSFS) models were presented. Also, root mean square error (RMSE), mean absolute error (MAE) and mean bias error (MBE) were calculated. The random fluctuation model of the solar radiation time series by using beta distribution was proposed in [8, 9].

*Corresponding author e-mail: rashidau7745@gmail.com

On the other hand, threshold autoregressive (TAR) time series model was used on the logarithm of clear-sky index series to forecast multi-day solar radiation data and described its statistical properties in [10].

Further, in [11, 12], the probability of the daily average solar energy was computed using a multiple linear regression model. Moreover, numerous researches have centered on statistical approaches for improving solar energy forecasting which affected by the interest time scale whither by hour, day, or month [13, 14, 15, 16].

By using preprocessing and an array of synthetic intelligence approaches, a multiple nonlinear regression model could reliably forecast exceeding the daily maximum ozone threshold, as suggested by [17].

Predicting renewable energy consumption to eliminate the energy crisis and reduce emissions worldwide was discussed by [18, 19, 20]. Authors in [21] introduced an adaptive neural fuzzy inference system to predict the ground inflow using a sample of 110 data sets containing the most important characteristics influence on the ground inflow rate.

Multiple linear regression and artificial neural networks based on main components could forecast ozone concentrations. Additionally, [22, 23, 24] provided estimates of the global solar radiation over different areas.

Markov models were adopted in multiple studies to model solar radiation, where it has evidenced to give effective modeling to obtain accurately forecast. These models are blended with wavelet coefficients, the generalized Fuzzy model, and Markov-chain mixture distribution model as illustrated by [25, 26].

This paper aims to estimate the joint distribution function (EF_{XY}) of both daily solar radiation (S) and daily maximum temperatures (T) combined with Markov property. The parameters of both distributions GEV and W-3P are obtained by applying the method of maximum likelihood estimation (MLE). In addition, estimate the joint distribution function ($F(X,Y)$) for both S and T. Four real data analyses of S and T in Queensland, Australia during 2015-2016 are applied to assess the efficiency of the proposed distributions. Nonetheless, logarithmic (LF_{XY}) model and the multi-regression (MF_{XY}) model are constructed for $F(X,Y)$. Several metrics are employed to evaluate the effectiveness of the suggested distributions and compare the models EF_{XY}, LF_{XY}, and MF_{XY} such as Kolmogorov-Smirnov (K-S) test, P-value, R squared, Pearson correlation, Anderson Darling (A*), Chi-Squared, Mann-Whitney test, and homogeneity test.

The motivation behind this study is to predict the impact of climate change on life on Earth by estimating the co-distribution function of solar radiation and its corresponding temperature extremes. The Markov property of the joint distribution function was used to calculate the different probability values and to study the independence of the two variables solar radiation and extreme temperature. In addition, a method was proposed in this study to reduce the distributions of three parameters to the distributions of two parameters and compare them with the aim of obtaining the probabilities of the common distribution function at the maximum values of solar radiation and temperature in order to simplify the formulas of the probability distributions functions by converting them into regression models and thus the possibility of creating an algorithm for it and applying it in estimation of solar radiation and maximum temperature.

In addition, this study aims to find the probability of high maximum temperature with low solar radiation. This is an indication of an imbalance in the climate. Whereby there are reasons that led to this, such as environmental pollution arising from factory fumes, forest fires, and green spaces. Also, petroleum products and exhausts of various means of transportation are among the causes that negatively affect climate change.

Although the three-parameter distributions are reduced to two in the form of the logarithmic regression model, we find that there are some drawbacks from the result of this conversion. This is due to the fact that the logarithmic model does not include the range of small values for each of the solar radiation and extreme temperatures. The actual formula of the co-distribution function estimated by the distribution functions for both solar radiation and temperature extremes cover all the range of solar radiation values and temperature extremes. Such a matter is a great advantage. Although the actual formula covers the range of values of the data set for the two variables, we will need to use estimation methods for the three parameters of both distribution functions. That requires more time and effort, or the use of ready-made programs to estimate the parameters. If it is required to calculate the probabilities of extreme temperature values with medium or low values of solar radiation, the logarithmic regression model is easier to evaluate this probability.

This paper contains the following sections:

- Concepts of solar radiation and maximum temperature.
- Source of the datasets.
- Definitions of Generalized Extreme Value (GEV) and Weibull-three parameters (W-3P).
- Software.

- Theorems 2.1, 2.2 and 3.1.
- Remarks 2.1, 2.2, 2.3 and 3.1.
- Converting Three-Parameters $F(X, Y)$ into Two-Parameters Logarithmic Regression Model (LF_{XY}).
- Multiple Nonlinear Regression Model of $F(X, Y)$.
- Maximum Likelihood Estimation (MLE).
- Results and Discussion.
- Conclusions.
- Data Availability.
- Conflict of Interest.
- Funding Statement.
- Acknowledgments.
- Supplementary Materials.
- References.

2 Solar Radiation and Maximum Temperature

This section provided the following:

- Basic concepts of solar radiation and extreme temperature.
- Definitions of some probability distribution functions used in this study.
- Definition of Markov property.
- Theorems and remarks on the joint distribution function of solar radiation and temperature extremes.

Suppose that X and Y are distinct scale continuous random variables of the phenomena of daily solar radiation (S) and daily maximum temperature (T) which they have generalized extreme value (GEV) distribution and three-parameter Weibull (W-3P) distribution, respectively, and satisfy Markov property.

Definition 2.1 Solar radiation is a general term which refers to radiant energy produced by the sun. It can be used in numerous fields such as generating heat and electricity. It is measured by mega joules per square meter (MJ/m^2) [27].

Definition 2.2 Maximum temperature refers to the highest degree in a certain time and a certain place. This temperature has no limit in its degree. This means that it can reach to degrees that are cannot be measured. It is worth mentioning that the measure unit is degrees Celsius [28].

Dataset: The datasets in Figure 1 were produced by the Bureau of Meteorology in Australian government during 2015-2016 in Queensland, Australia.

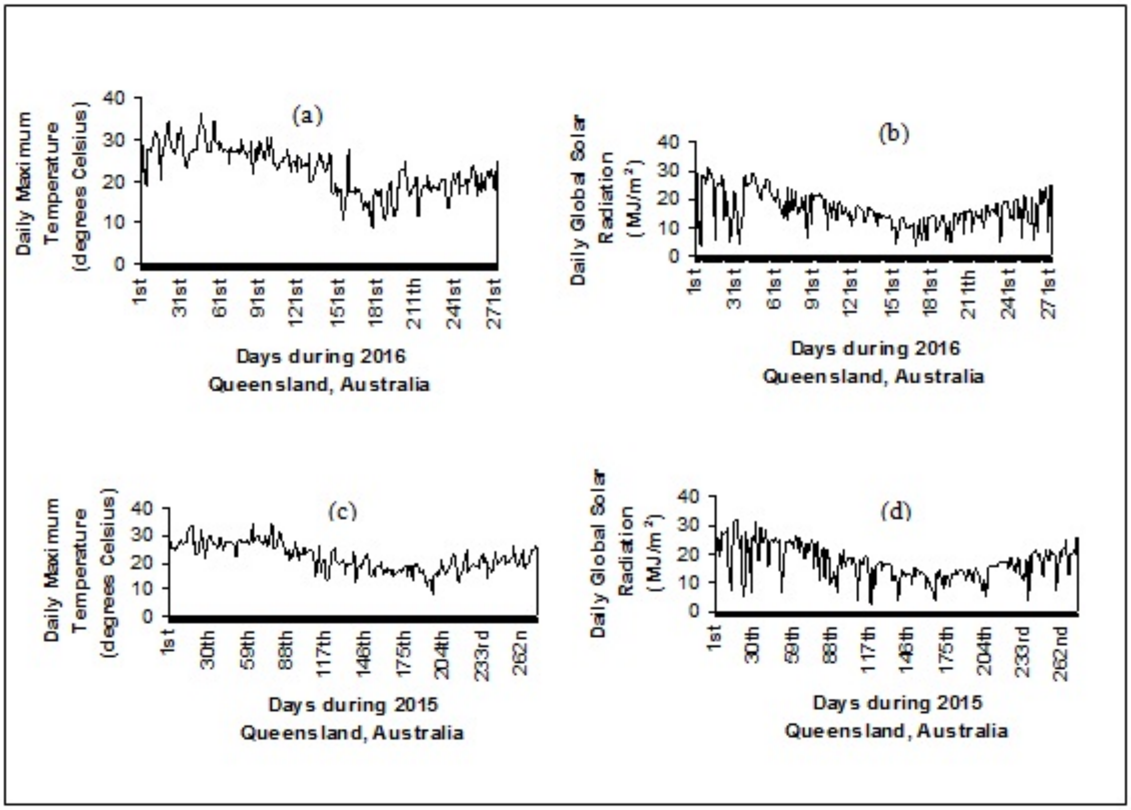


Fig. 1: Plots (a-d) illustrates datasets of S and T during 2015-2016 in Queensland, Australia.

Definition 2.3 Suppose that X is a continuous random variable representing the daily solar radiation (S) and it has (GEV) distribution, which includes three types of distributions: type I-Gumbel ($\alpha_S = 0$), type II-Fréchet ($\alpha_S > 0$), and the type III-Weibull ($\alpha_S < 0$). The (cdf) $F_{GEV}(x;\ \alpha_S, \theta_S, \lambda_S)$ is defined as follows:

$$F_{GEV}(x;\ \alpha_S, \theta_S, \lambda_S) = \begin{cases} e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} & -\infty < x \leq \lambda_S - \theta_S/\ \alpha_S \ \text{ for } \alpha_S < 0, \\ & \lambda_S - \theta_S/\ \alpha_S \leq x < \infty \ \text{ for } \alpha_S > 0, \\ e^{-e^{-\frac{(x-\lambda_S)}{\theta_S}}} & -\infty < x < \infty \ \text{ for } \alpha_S = 0. \end{cases} \tag{1}$$

Such that $\alpha_S \in R$, $\theta_S > 0$ and $\lambda_S \in R$ are the shape, scale and location parameters, respectively [29].

Definition 2.4 Suppose that Y is a continuous random variable representing the daily maximum temperature (T) and it has (W-3P) distribution. Then the (cdf) $F_W(y; \beta_T, \lambda_T, \alpha_T)$ is defined as follows:

$$F_W(y; \beta_T, \lambda_T, \alpha_T) = \begin{cases} 1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}} & y > \alpha_T, \\ 0 & y \leq \alpha_T. \end{cases} \tag{2}$$

Where $\beta_T \geq 0$, $\lambda_T \geq 0$ and $\alpha_T \geq 0$ are the shape, scale and location parameters, respectively [30].

The following Theorem 2.1 explores the relationship between the joint distribution function $F(X, Y)$ and the functions $F(X)$ and $F(Y)$ which is determined by the constant C which is computed from data sets for both solar radiation and maximum temperatures.

The study of the correlation relationship is based on the well-known rule that says that for any two events A and B they are independent if the value of the constant $C = 1$ in the relation $P(A \cap B) = C\ P(A)\ P(B)$.

Certainly, if the value of the constant is $C = 0$ or $C \approx 0$, then the correlation is non-existent, meaning that the two events A and B are not independent or that the phenomena of solar radiation and extreme temperatures are not independent.

This is what will be discovered in the application used in this study from the data set given about solar radiation and extreme temperatures.

In the case if the value of the constant is $C \approx 0$, then we will need another formula for the common distribution function using the OR relationship, as it is more comprehensive than the AND relationship.

The rule in probabilities about the OR relationship between any two events is known as $P(A \cup B) = \alpha P(A) + \beta P(B) - C P(A \cap B)$, where each of α, β and C are real constants and the formulas for these constants have been determined In Theorem 2.2 to calculate their values to estimate the joint distribution function $F(X, Y)$.

Theorem 2.1 If X and Y are continuous random variables representing two different phenomena S and T, which have GEV distribution and W-3P distribution, respectively, and satisfy the Markov property. Then, the joint distribution function $F(X, Y)$ is defined as

$$F(X, Y) = F(X < x \; and \; Y < y) = C \, F(x) F(y), \tag{3}$$

where

$$C = \frac{1}{n^2}\left(\mu\left(\gamma \, p_{\bar{D}_S}(x^+) + \eta \, p_{\bar{D}_S}(x^-)\right) + (\Theta + \varepsilon + \rho)\left(\delta \, p_{\bar{D}_T}(y^+) + \tau \, p_{\bar{D}_T}(y^-)\right) + S_{n-\gamma-\eta} \, T_{n-\delta-\tau}\right), \tag{4}$$

such that $x^+ = S^+ = \eta$, $x^- = S^- = \gamma$, $x^0 = S_{n-\gamma-\eta}$, $y^+ = T^+ = \tau$, $y^- = T^- = \delta$, $y^0 = T_{n-\delta-\tau}$, $\mu = T_{n-\delta-\tau} \, F(\bar{D}_S)$, $\Theta = \gamma \, F(\bar{D}_S) \, F(\bar{D}_T) \, p_{\bar{D}_S}(x^+)$, $\varepsilon = \eta \, F(\bar{D}_S) \, F(\bar{D}_T) \, p_{\bar{D}_S}(x^-)$, $\rho = S_{n-\gamma-\eta} \, F(\bar{D}_T)$, where $F(\bar{D}_S)$ and $F(\bar{D}_T)$ are cdfs of GEV and W-3P distributions, respectively.

Proof: Let the probabilities $p_{\bar{D}_S}(x^-)$, $p_{\bar{D}_T}(y^-)$, $p_{\bar{D}_S}(x^+)$, $p_{\bar{D}_T}(y^+)$, $p_{\bar{D}_S}(x^0)$, $p_{\bar{D}_T}(y^0)$ are defined as follows:

$p_{\bar{D}_S}(x^-)$ is the probability that x is decreased by $\bar{D}_S$,

$p_{\bar{D}_T}(y^-)$ is the probability that y is decreased by $\bar{D}_T$,

$p_{\bar{D}_S}(x^+)$ is the probability that x is increased by $\bar{D}_S$,

$p_{\bar{D}_T}(y^+)$ is the probability that y is increased by $\bar{D}_T$,

$p_{\bar{D}_S}(x^0)$ is the probability that reveals no change in the value of x,

$p_{\bar{D}_T}(y^0)$ is the probability that shows no change in the value of y.

Here, $\bar{D}_S$ and $\bar{D}_T$ are the average deviation of X and Y respectively.

On the other hand, the process is said to have the Markov property if

$$p(X(t+s) = j | X(t) = i, X(u) = x(u), 0 \le u < s) = p(X(t+s) = j | X(s) = i). \tag{5}$$

For all possible $x(u)$, $0 \le u < s$.

There are five different cases of the daily solar radiation X and the daily maximum temperature Y:

$$p(A_I) = p(X < x \pm \bar{D}_S | Y < y \pm \bar{D}_T) \cdot p(Y < y \pm \bar{D}_T) \cdot p_{\bar{D}_S}(x^\mp) \cdot p_{\bar{D}_T}(y^\mp). \tag{6}$$

$$p(A_J) = p(X < x \pm \bar{D}_S | Y < y \pm \bar{D}_T) \cdot p(Y < y \mp \bar{D}_T) \cdot p_{\bar{D}_S}(x^\mp) \cdot p_{\bar{D}_T}(y^\pm). \tag{7}$$

$$p(A_K) = p(X < x | Y < y \pm \bar{D}_T) \cdot p(Y < y \pm \bar{D}_T) \cdot p_{\bar{D}_S}(x^0) \cdot p_{\bar{D}_T}(y^\mp). \tag{8}$$

$$p(A_L) = p(X < x \pm \bar{D}_S | Y < y) \cdot p(Y < y) \cdot p_{\bar{D}_S}(x^\mp) \cdot p_{\bar{D}_T}(y^0). \tag{9}$$

$$p(A_M) = p(X < x | Y < y) \cdot p(Y < y) \cdot p_{\bar{D}_S}(x^0) \cdot p_{\bar{D}_T}(y^0). \tag{10}$$

Where $F(x \pm \bar{D}_S) = F(x) \, F(\bar{D}_S) \, p_{\bar{D}_S}(x^\pm)$, $F(y \pm \bar{D}_T) = F(y) \, F(\bar{D}_T) \, p_{\bar{D}_T}(y^\pm)$.

The probabilities equations from (6) to (10) represent the three different states of increase, decrease, and non-change of the values of solar radiation and maximum temperatures. Whereby the joint distribution function $F(X, Y)$ is in the case of the correlation of $F(X)$, $F(Y)$ to AND. The values of solar radiation x and maximum temperatures y has three cases: (i) That they were increased in the current year by $\bar{D}_S$ and $\bar{D}_T$, then they decreased by the same amounts in next year, thus they returned to their actual values x and y. (ii) Likewise, if they were decreased in the current year by $\bar{D}_S$ and $\bar{D}_T$, then they increased in next year by the same amounts and return to the actual values. (ii) Last case that they are not change during two consecutive years. In order to prevent writing many equations, the positive and negative signs were merged together into one equation.

Then,

$$p(A_1)=\frac{\gamma\,\delta}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{11}$$

$$p(A_2)=\frac{\eta\,\tau}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{12}$$

$$p(A_3)=\frac{\gamma\,\tau}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{13}$$

$$p(A_4)=\frac{\eta\,\delta}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{14}$$

$$p(A_5)=\frac{\delta\,S_{n-\gamma-\eta}}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{15}$$

$$p(A_6)=\frac{\tau\,S_{n-\gamma-\eta}}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{16}$$

$$p(A_7)=\frac{\gamma\;T_{n-\delta-\tau}}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{17}$$

$$p(A_8)=\frac{\eta\;T_{n-\delta-\tau}}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{18}$$

$$p(A_9)=\frac{S_{n-\gamma-\eta}\;T_{n-\delta-\tau}}{n^2}e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1-e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{19}$$

Where the equations (11) to (19) represent the substitution of the cumulative distribution functions of GEV and W-3P in the equations (6) to (10).

Substituting from Eqs. (11-19) in Eq. (3), such that

$$F(x,y)=F(X<x\ and\ Y<y)=C\,F(x)F(y)=\sum_{i=1}^{9}p(A_i)$$

$$=\frac{1}{n^2}\Big(\big(\gamma F(x+\bar{D}_S)+S_{n-\gamma-\eta}\,F(x)+\eta F(x-\bar{D}_S)\big)\big(\delta F(y+\bar{D}_T)+\ T_{n-\delta-\tau}F(y)+\tau F(y-\bar{D}_T)\big)\Big)$$

$$=\left(\frac{1}{n^2}\Big(\big(\gamma F(\bar{D}_S)\,p_{\bar{D}_S}(x^+)+S_{n-\gamma-\eta}\ +\eta F(\bar{D}_S)\,p_{\bar{D}_S}(x^-)\big)\big(\delta F(\bar{D}_T)\,p_{\bar{D}_T}(y^+)+\ T_{n-\delta-\tau}+\tau F(\bar{D}_T)\,p_{\bar{D}_T}(y^-)\big)\Big)\right)F(x)F(y).$$

Then, we obtain the formula of the constant C in Eq. (4).

Remark 2.1 If $C\to1$, then the joint distribution function $F(X,Y)$ has a complete positive correlation with the multiplying $F(x)$ in $F(y)$ which means that X and Y are independent variables.

Remark 2.2 If $C\to0$, then the joint distribution function $F(X,Y)$ has almost no correlation with the multiplying $F(x)$ in $F(y)$ which means that X and Y are not independent variables. For this reason, the following theorem is considered.

Theorem 2.2 If X and Y are continuous random variables representing two different phenomena S and T, which have GEV distribution and W-3P distribution, respectively, and satisfy Markov property. Then, the joint distribution function $F(X+Y)$ (the estimation of the theoretical model (EFxy)) is depending on Theorem 2.1 and it is defined as

$$\text{EFxy} = F(X+Y) = F(X < x \text{ or } Y < y) = A_{GEV}\, F(x) + B_{Weibull} F(y) - C\, F(x)F(y). \quad (20)$$

Where

$$A_{GEV} = \frac{3}{n}\left(\gamma\, F(\overline{D}_S)\, p_{\overline{D}_S}(x^+) + \eta\, F(\overline{D}_S)\, p_{\overline{D}_S}(x^-) + S_{n-\gamma-\eta} \right).$$

The constant C will be evaluate from Theorem 2.1 and $B_{Weibull}$ from Eq. (20), where $F_{max.}(X+Y) = 1$, $F(x_{max.}) = 1$ and $F(y_{max.}) = 1$. Such that $x^+ = S^+ = \eta$, $x^- = S^- = \gamma$, $x^0 = S_{n-\gamma-\eta}$, $y^+ = T^+ = \tau$, $y^- = T^- = \delta$, $y^0 = T_{n-\delta-\tau}$. $F(\overline{D}_S)$ and $F(\overline{D}_T)$ are the cdfs of GEV and W-3P distributions, respectively.

Proof: There are five different cases of the daily solar radiation X and the daily maximum temperature Y:

$$p(A_I) = p(X < x \pm \overline{D}_S | Y < y \pm \overline{D}_T) \cdot p_{\overline{D}_S}(x^{\mp}) + p(Y < y \pm \overline{D}_T) \cdot p_{\overline{D}_T}(y^{\mp}). \quad (21)$$

$$p(A_J) = p(X < x \pm \overline{D}_S | Y < y \pm \overline{D}_T) \cdot p_{\overline{D}_S}(x^{\mp}) + p(Y < y \mp \overline{D}_T) \cdot p_{\overline{D}_T}(y^{\pm}). \quad (22)$$

$$p(A_K) = p(X < x | Y < y \pm \overline{D}_T) \cdot p_{\overline{D}_S}(x^0) + p(Y < y \pm \overline{D}_T) \cdot p_{\overline{D}_T}(y^{\mp}). \quad (23)$$

$$p(A_L) = p(X < x \pm \overline{D}_S | Y < y) \cdot p_{\overline{D}_S}(x^{\mp}) + p(Y < y) \cdot p_{\overline{D}_T}(y^0). \quad (24)$$

$$p(A_M) = p(X < x | Y < y) \cdot p_{\overline{D}_S}(x^0) + p(Y < y) \cdot p_{\overline{D}_T}(y^0). \quad (25)$$

Where $F(x \pm \overline{D}_S) = F(x)\, F(\overline{D}_S)\, p_{\overline{D}_S}(x^{\pm})$, $F(y \pm \overline{D}_T) = F(y)\, F(\overline{D}_T)\, p_{\overline{D}_T}(y^{\pm})$.

The probabilities equations from (21) to (25) represent the three different states of increase, decrease, and non-change of the values of solar radiation and maximum temperatures. Whereby the joint distribution function $F(X,Y)$ is in the case of the correlation of $F(X)$, $F(Y)$ to OR.

Consequently,

$$p(A_1) = \frac{\gamma}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\delta}{n}\left(1 - e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (26)$$

$$p(A_2) = \frac{\eta}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\tau}{n}\left(1 - e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (27)$$

$$p(A_3) = \frac{\gamma}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\tau}{n}\left(1 - e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (28)$$

$$p(A_4) = \frac{\eta}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\delta}{n}\left(1 - e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (29)$$

$$p(A_5) = \frac{S_{n-\gamma-\eta}}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\delta}{n}\left(1 - e^{-\left(\frac{(y+\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (30)$$

$$p(A_6) = \frac{S_{n-\gamma-\eta}}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{\tau}{n}\left(1 - e^{-\left(\frac{(y-\bar{D}_T)-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (31)$$

$$p(A_7) = \frac{\gamma}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x+\bar{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{T_{n-\delta-\tau}}{n}\left(1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \quad (32)$$

$$p(A_8) = \frac{\eta}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}((x-\overline{D}_S)-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{T_{n-\delta-\tau}}{n}\left(1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{33}$$

$$p(A_9) = \frac{S_{n-\gamma-\eta}}{n} e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + \frac{T_{n-\delta-\tau}}{n}\left(1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{34}$$

Such that the equations (26) to (34) represent the substitution of the cumulative distribution functions of GEV and W-3P in the equations (21) to (25).

Substituting from Eqs. (26-34) in Eq. (20), then

$$\text{EFxy} = F(X+Y) = F(X<x \text{ } or \text{ } Y<y) = \sum_{i=1}^{9} p(A_i) - C\,F(x)F(y)$$

$$= \frac{3}{n}\Big(\gamma\,F\big(x+\overline{D}_S\big) + \eta\,F\big(x-\overline{D}_S\big) + S_{n-\gamma-\eta}F(x) + \delta\,F\big(y+\overline{D}_T\big) + \tau\,F\big(y-\overline{D}_T\big) + T_{n-\delta-\tau}F(y)\Big) - C\,F(x)F(y),$$

$$\text{EFxy} = \frac{3}{n}\Big(\big(\gamma\,F\big(\overline{D}_S\big)\,p_{\overline{D}_S}(x^+) + \eta\,F\big(\overline{D}_S\big)\,p_{\overline{D}_S}(x^-) + S_{n-\gamma-\eta}\big)F(x) +$$

$$\big(\delta\,F\big(\overline{D}_T\big)\,p_{\overline{D}_T}(y^+) + \tau\,F\big(\overline{D}_T\big)\,p_{\overline{D}_T}(y^-) + T_{n-\delta-\tau}\big)F(y)\Big) - C\,F(x)F(y),$$

$$\text{EFxy} = \left(\frac{3}{n}\big(\gamma\,F\big(\overline{D}_S\big)\,p_{\overline{D}_S}(x^+) + \eta\,F\big(\overline{D}_S\big)\,p_{\overline{D}_S}(x^-) + S_{n-\gamma-\eta}\big)\right)F(x) + B_{Weibull}F(y) - C\,F(x)F(y). \tag{35}$$

From the previous equation, we obtain A_{GEV}.

Remark 2.3 If $A_{GEV} \neq 0$, $B_{Weibull} \neq 0$ and $C \approx 0$, then $F(X+Y)$ has a linear correlation with $F(x)$ and $F(y)$.

The aim of getting the $F(X+Y)$ is to predict probabilities at the different values of both X, Y and especially non corresponding values. If X has a small value and represents solar radiation and Y has a large value and represents the maximum temperature and the probability of both is high, this means that there is a dysfunction or climate change caused by the pollution or the ozone layer damaged.

3 Converting Three-Parameters $F(X,Y)$ into Two-Parameters Logarithmic Regression Model (LF_{XY})

This section introduced:

- Theorem on the logarithmic regression model of the joint distribution function of solar radiation and maximum temperature.
- Algorithm of solving nested percentile of backward and forward equations to estimate the distribution functions used in this study.

Theorem 3.1 If the logarithmic regression model (LF_{XY}) is defined as:

$$F(x) = \varphi_1 \ln(x) + \varphi_2, \tag{36}$$

$$F(y) = \omega_1 \ln(y) + \omega_2. \tag{37}$$

Then,

$$\varphi_1 = \frac{F(x_{min.}) - F(x_{max.})}{\ln(x_{min.}) - \ln(x_{max.})}, \tag{38}$$

$$\varphi_2 = \left(\left(\frac{F(x_{min.})}{\ln(x_{min.})} - \frac{F(x_{max.})}{\ln(x_{max.})}\right) \Bigg| \left(\frac{1}{\ln(x_{min.})} - \frac{1}{\ln(x_{max.})}\right)\right), \tag{39}$$

$$\omega_1 = \frac{F(y_{min.}) - F(y_{max.})}{\ln(y_{min.}) - \ln(y_{max.})}, \tag{40}$$

$$\omega_2 = \left(\left(\frac{F(y_{min.})}{\ln(y_{min.})} - \frac{F(y_{max.})}{\ln(y_{max.})}\right) \middle| \left(\frac{1}{\ln(y_{min.})} - \frac{1}{\ln(y_{max.})}\right)\right). \tag{41}$$

Proof: From Eq. (20) in Theorem 2.2

$$EF_{XY} = F(X+Y) = A_{GEV}\, e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} + B_{Weibull}\left(1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right) - C\, e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\left(1 - e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right). \tag{42}$$

Then, the logarithmic regression model of $F(X,Y)$ is defined as follows:

$$LF_{XY} = \hat{F}(X+Y) = A_{GEV}(\varphi_1 \ln(x) + \varphi_2) + \hat{B}_{Weibull}(\omega_1 \ln(y) + \omega_2) - C(\varphi_1 \ln(x) + \varphi_2)(\omega_1 \ln(y) + \omega_2). \tag{43}$$

Where $\hat{F}(X+Y)$ is the estimation of $F(X+Y)$ and $\hat{B}_{Weibull}$ will be evaluate from Eq. (43) such that $F_{max.}(X+Y) = 1$, $F(x_{max.}) = 1$ and $F(y_{max.}) = 1$. Comparing Eq. (42) and Eq. (43), then

$$e^{-\left(1+\frac{\alpha_S}{\theta_S}(x_{min.}-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} = \varphi_1 \ln(x_{min.}) + \varphi_2, \tag{44}$$

$$e^{-\left(1+\frac{\alpha_S}{\theta_S}(x_{max.}-\lambda_S)\right)^{\frac{-1}{\alpha_S}}} = \varphi_1 \ln(x_{max.}) + \varphi_2, \tag{45}$$

by subtracting Eqs. (44) and (45), we get Eq. (38). Also, by dividing the Eq. (44) on $\ln(x_{min.})$ and dividing the Eq. (45) on $\ln(x_{max.})$, we get the following Eqs. (44) and (45), respectively

$$\frac{F(x_{min.})}{\ln(x_{min.})} = \varphi_1 + \frac{\varphi_2}{\ln(x_{min.})}, \tag{46}$$

$$\frac{F(x_{max.})}{\ln(x_{max.})} = \varphi_1 + \frac{\varphi_2}{\ln(x_{max.})}, \tag{47}$$

by subtracting Eqs. (46) and (47), we get Eq. (39). In addition to, comparing transactions of the $B_{Weibull}$ we obtain the following

$$1 - e^{-\left(\frac{y_{min.}-\alpha_T}{\lambda_T}\right)^{\beta_T}} = \omega_1 \ln(y_{min.}) + \omega_2, \tag{48}$$

$$1 - e^{-\left(\frac{y_{max.}-\alpha_T}{\lambda_T}\right)^{\beta_T}} = \omega_1 \ln(y_{max.}) + \omega_2, \tag{49}$$

subtracting Eqs. (48) and (49), we obtain Eq. (40). Also, by dividing the Eq. (48) on $\ln(y_{min.})$ and dividing the Eq. (49) on $\ln(y_{max.})$, we get the following Eqs. (50) and (51), respectively

$$\frac{F(y_{min.})}{\ln(y_{min.})} = \omega_1 + \frac{\omega_2}{\ln(y_{min.})}, \tag{50}$$

$$\frac{F(y_{max.})}{\ln(y_{max.})} = \omega_1 + \frac{\omega_2}{\ln(y_{max.})}, \tag{51}$$

by subtracting Eqs. (50) and (51), we obtain Eq. (41).

On the other hand, Eqs. (36) and (37) can be solved to obtain the solutions of the parameters of distributions using Nested Percentiles Algorithm (NPA) [31]. The percentiles equations were applied to obtain the values of x_i. Then 98 equations in terms of φ_1 and φ_2 were obtained. Thus, every two equations were solved together for all $i = 1$ to 49 to obtain 49 estimation values of $\varphi_1, \varphi_2, \omega_1$ and ω_2.

Let cdfs at any two points of S dataset x_{fi} and x_{li} are:

$$P_{fi} = \varphi_1 \ln\left(x_{fi}\right) + \varphi_2,$$

and

$$P_{li} = \varphi_1 \ln(x_{li}) + \varphi_2,$$

such that $P_{li} = 1 - P_{fi}, i = 1, 2, 3, \cdots, N/2$, and N is an even number. Then, we obtain Eqs. (38) and (39).

Iterate the above steps for each pair of equations that satisfy $P_{li} = 1 - P_{fi}$. Thus, $N/2$ different values for each φ_1 and φ_2 were obtained. Then, the averages are calculated to get the final value of the two parameters.

$$\varphi_1 = \frac{\sum_{i=1}^{N/2} \varphi_{1i}}{N/2}. \tag{52}$$

$$\varphi_2 = \frac{\sum_{i=1}^{N/2} \varphi_{2i}}{N/2}. \tag{53}$$

Similarly for T and we obtain ω_1 and ω_2 in Eqs. (40) and (41).

$$\omega_1 = \frac{\sum_{i=1}^{N/2} \omega_{1i}}{N/2}. \tag{54}$$

$$\omega_2 = \frac{\sum_{i=1}^{N/2} \omega_{2i}}{N/2}. \tag{55}$$

Figure 2 shows the method of solving the nested percentiles of backward and forward equations. The Algorithm was applied on the datasets of $z = \{x, y\}$, such that $F(x)$ and $F(y)$ are the cdfs of GEV and W-3P distributions, respectively.

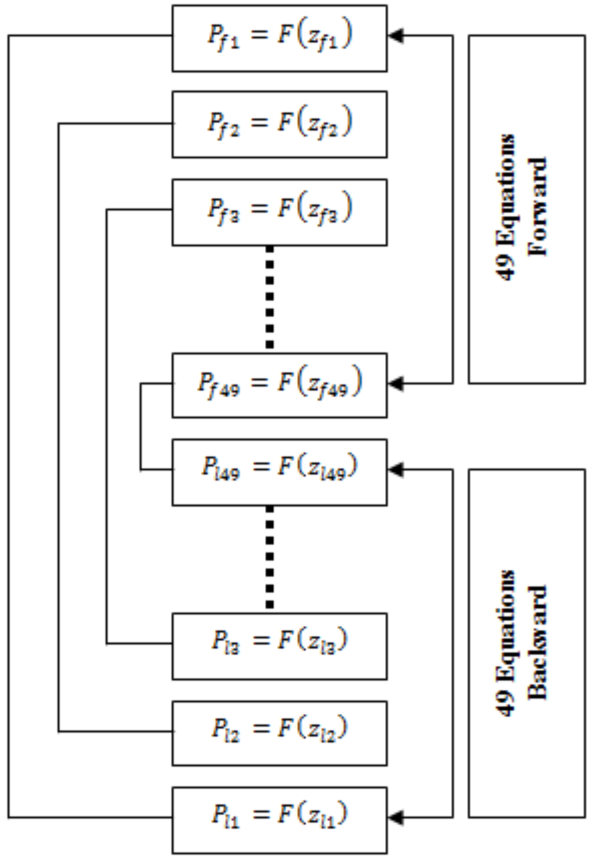


Fig. 2: Algorithm of solving nested percentile of backward and forward equations.

The logarithmic regression equation for $F(z)$ over $\ln(z)$ can be defined on the following formula

$$\begin{vmatrix} F(z) & \ln(z) & 1 \\ \sum_{i=1}^{n} F(z_i) & \sum_{i=1}^{n} \ln(z_i) & N \\ \sum_{i=1}^{n} \ln(z_i) F(z_i) & \sum_{i=1}^{n} (\ln(z_i))^2 & \sum_{i=1}^{n} \ln(z_i) \end{vmatrix} = 0. \tag{56}$$

Remark 3.1 The Adjusted R^2 for $F(x)$ and $F(y)$ must be close to 1 such that

$$R_{adj.}^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-k-1}\right), \tag{57}$$

where n is the size of dataset, k is the number of variables in the model and R^2 is the coefficient of determination:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{Z}_i - \bar{Z})^2}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}, 0 \leq R^2 \leq 1. \tag{58}$$

Such that $\hat{Z}_i = F(Z) = A\ln(Z) + B$, where A and B are constants. Also, the Adjusted R^2 will always be less than or equal to R^2.

4 Multiple Nonlinear Regression Model of $F(X, Y)$

In this section, we introduced the estimation model, multi-regression model and logarithmic model for $F(X, Y)$.

To validate the theoretical model estimate in Eq. (20), we utilize the multi-regression model (MF_{xy}) by knowing both $F(X)$ and $F(Y)$ and comparing results with the EF_{xy}. The estimated theoretical model can thus be obtained from Eq. (20) in Theorem 2.2 as follows:

$$EF_{xy} = F(X,Y) = A\left(e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\right) + B\left(1-e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right) - C\left(e^{-\left(1+\frac{\alpha_S}{\theta_S}(x-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}\right)\left(1-e^{-\left(\frac{y-\alpha_T}{\lambda_T}\right)^{\beta_T}}\right), \tag{59}$$

and the multi-regression model is given by

$$MF_{xy} = F^*(X,Y) = A^*F(X) + B^*F(Y) - C^*F(X)F(Y), \tag{60}$$

where $F^*(X,Y)$, A^*, B^* and C^* are the estimated values of $F(X,Y)$, A, B and C, respectively. The constants A^*, B^* and C^* are estimated from the following determinant

$$\begin{vmatrix} F^*(X,Y) & F(X) & F(Y) & -F(X)F(Y) \\ \sum_{i=1}^{n} F^*(X_i,Y_i) & \sum_{i=1}^{n} F(X_i) & \sum_{i=1}^{n} F(Y_i) & -\sum_{i=1}^{n} F(X_i)\,F(Y_i) \\ \sum_{i=1}^{n} F(X_i)F^*(X_i,Y_i) & \sum_{i=1}^{n}\left(F(X_i)\right)^2 & \sum_{i=1}^{n} F(X_i)\,F(Y_i) & -\sum_{i=1}^{n}\left(F(X_i)\right)^2 F(Y_i) \\ \sum_{i=1}^{n} F(Y_i)F^*(X_i,Y_i) & \sum_{i=1}^{n} F(X_i)\,F(Y_i) & \sum_{i=1}^{n}\left(F(Y_i)\right)^2 & -\sum_{i=1}^{n} F(X_i)\left(F(Y_i)\right)^2 \end{vmatrix} = 0. \tag{61}$$

Finally, Mann-Whitney test and homogeneity test of both logarithmic model and multi-regression model of the $F(X, Y)$ were applied.

5 Maximum Likelihood Estimation (MLE)

This section presented the estimation method (MLE) for the distributions GEV and W-3P.

The parameters estimation of GEV and Weibull models are derived by using the method of MLE. Let $X_1, X_2, X_3, \cdots, X_n$ and $Y_1, Y_2, Y_3, \cdots, Y_n$ be a random samples from GEV$(x;\ \alpha_S, \theta_S, \lambda_S)$ and W$(y; \beta_T, \lambda_T, \alpha_T)$, respectively. The likelihood functions are defined respectively as

$$L(x;\ \alpha_S,\theta_S,\lambda_S) = \prod_{i=1}^{n} f_{GEV}(x_i;\ \alpha_S,\theta_S,\lambda_S)$$
$$= \prod_{i=1}^{n} \frac{1}{\theta_S}\left(1+\frac{\alpha_S}{\theta_S}(x_i-\lambda_S)\right)^{-\left(1+\frac{1}{\alpha_S}\right)} e^{-\left(1+\frac{\alpha_S}{\theta_S}(x_i-\lambda_S)\right)^{\frac{-1}{\alpha_S}}}.$$

Thus, the log-likelihood function is

$$\ell = -n\log\theta_S - \sum_{i=1}^{n}\left(\left(1+\frac{1}{\alpha_S}\right)\log(z_i) + (z_i)^{\frac{-1}{\alpha_S}}\right), \tag{62}$$

where $\varepsilon = (\alpha_S, \theta_S, \lambda_S)$, $\alpha_S \neq 0$ and $z_i = \left(1+\frac{\alpha_S}{\theta_S}(x_i-\lambda_S)\right)$.

The following equations are formed by taking the derivatives of the previous equation with respect to the three parameters and equating it to zero

$$\frac{\partial \ell}{\partial \alpha_S} = \frac{1}{(\alpha_S)^2} \sum_{i=1}^{n} \left(\log(z_i) \left(1 - (z_i)^{\frac{-1}{\alpha_S}}\right) - \frac{\left(1+\alpha_S - (z_i)^{\frac{-1}{\alpha_S}}\right)}{z_i} \left(\frac{\alpha_S}{\theta_S}(x_i - \lambda_S)\right) \right) = 0 \tag{63}$$

$$\frac{\partial \ell}{\partial \theta_S} = -\frac{n}{\theta_S} + \frac{1}{\theta_S} \sum_{i=1}^{n} \left(\frac{\left(1+\alpha_S - (z_i)^{\frac{-1}{\alpha_S}}\right)}{z_i} \left(\frac{x_i - \lambda_S}{\theta_S}\right) \right) = 0 \tag{64}$$

$$\frac{\partial \ell}{\partial \lambda_S} = \frac{1}{\theta_S} \sum_{i=1}^{n} \left(\frac{\left(1+\alpha_S - (z_i)^{\frac{-1}{\alpha_S}}\right)}{z_i} \right) = 0 \tag{65}$$

$$L(y; \beta_T, \lambda_T, \alpha_T) = \prod_{i=1}^{n} f_W(y_i; \beta_T, \lambda_T, \alpha_T)$$

$$= \prod_{i=1}^{n} \frac{\beta_T}{\lambda_T} \left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T - 1} e^{-\left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T}}.$$

Thus, the log-likelihood function is

$$\ell = \sum_{i=1}^{n} \left(\log(\beta_T) - \beta_T \log(\lambda_T) + (\beta_T - 1)\log(y_i - \alpha_T) - \left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T} \right). \tag{66}$$

The following equations are formed by taking the derivatives of the previous equation with respect to the three parameters and equating it to zero

$$\frac{\partial \ell}{\partial \beta_T} = \sum_{i=1}^{n} \left(\frac{1}{\beta_T} + \log(y_i - \alpha_T) - \log(\lambda_T) - \left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T} \log\left(\frac{y_i - \alpha_T}{\lambda_T}\right) \right) = 0. \tag{67}$$

$$\frac{\partial \ell}{\partial \lambda_T} = \sum_{i=1}^{n} \left(-\frac{\beta_T}{\lambda_T} + \left(\frac{\beta_T}{\lambda_T}\right) \left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T} \right) = 0. \tag{68}$$

$$\frac{\partial \ell}{\partial \alpha_T} = \sum_{i=1}^{n} \left(-\frac{(\beta_T - 1)}{(y_i - \alpha_T)} + \left(\frac{\beta_T}{\lambda_T}\right) \left(\frac{y_i - \alpha_T}{\lambda_T}\right)^{\beta_T - 1} \right) = 0. \tag{69}$$

Multiple nonlinear equations were solved numerically to obtain the estimation of the parameters $\alpha_S, \theta_S, \lambda_S, \beta_T, \lambda_T$ and α_T. The value A^2 of Anderson-Darling (A-D) test was applied to test the fitting of the datasets with GEV and W-3P distributions.

$$A^2 = -\sum_{i=1}^{n} \left((2i-1)\left(\ln F_X(x_i) + \ln\left(1 - F_X(x_{n+1-i})\right) \right) / n \right) - n. \tag{70}$$

The adjusted AD test statistic of GEV and W-3P distributions is given by

$$A^* = A^2 \left(1 + \frac{0.3}{n}\right). \tag{71}$$

6 Results and Discussion

In this section, we presented all numerical results and their discussion.

Various methods and methodologies are applied for modeling the daily global solar radiation dataset. El Genidy [32] employed a statistical model along with the GEV distribution. The closest probability distribution was determined using nonlinear regression and multiple nonlinear regression. The GEV distribution parameters were validated using methods like the moment's technique and the K-S test. The distribution of solar radiation was estimated using the (Quartiles-Moments) approach. EGMD was presented for estimating solar energy. Also, forecasting the monthly solar energy average in Queensland-Australia utilizing percentile root estimation and Markov transition probability matrices were studied by [33, 34].

This study deals with estimate the joint distribution function (EF_{XY}) of S and T combined with Markov property. MLE has utilized to estimate the parameters of both GEV distribution and W-3P distribution with S and T, respectively. Logarithmic (LF_{XY}) model and the multi-regression (MF_{XY}) model have created for the $F(X,Y)$ and comparing them with EF_{XY}. Evaluate the performance of the proposed distributions and models using the Goodness-of-Fit tests. This is conducted with the least error rate.

Tables (1-4) are listed the observed outcomes and illustrate the descriptive statistics for the datasets of S and T during 2015-2016 in Queensland – Australia. The results demonstrate the precision of estimation methodologies and indicate the efficiency of MLEs in estimating parameters.

Table 1: Descriptive statistics for the datasets of S and T during 2015(2016) in Queensland, Australia.

Statistics	S	T	Statistics	S	T
Average	18.3(16.337)	22.99(22.7)	Skewness	-0.004(0.257)	-0.115(-0.116)
Variance	43.399(34.876)	25.818(28.991)	Kurtosis	-0.57(-0.344)	-0.727(-0.604)
StDev[a]	6.588(5.906)	5.08(5.384)	Q1	13.2(12.4)	18.7(18.4)
Min	2.4(3.1)	8(8.9)	Q2	18.1(15.6)	23.3(23.2)
Max	32.1(31.3)	34(36)	Q3	23.6(20.1)	27.3(26.9)
Mode	17(14.6)	18.1(26.7)	Q4	32.1(31.3)	34(36)
Median	18.1(15.6)	23.3(23.2)	AveDev[b]	5.43(4.714)	4.29(4.57)

a. StDev refers to Standard Deviation.
b. AveDev stands for Average Deviation.

Table 2: Frequencies and probabilities of difference cases of S and T during 2015-2016.

S	f	P_S	T	f	P_T
Decreasing (X^-)	147	$P_S(X^-) = 147/274$	Decreasing (Y^-)	119	$P_T(Y^-) = 119/274$
Unchangeable (X^0)	3	$P_S(X^0) = 3/274$	Unchangeable (Y^0)	2	$P_T(Y^0) = 2/274$
Increasing (X^+)	124	$P_S(X^+) = 124/274$	Increasing (Y^+)	153	$P_T(Y^+) = 153/274$

Table 3: Some statistics of the absolute difference-values (AD) of S and T in 2015-2016.

Statistics	ADS	ADT
Average	$\overline{D}_S = 4.515$	$\overline{D}_T = 3.211$
Variance	$Var(D_S) = 18.73$	$Var(D_T) = 6.737$

Table 4: MLEs of parameters and Goodness-of-Fit measures of S and T.

Model	MLEs	K-S	A*	Chi-Squared	Pearson Correlation
S (GEV)	α_S= -0.26493 θ_S= 6.5895 λ_S= 15.899	0.03743	0.57116	15.113	0.967
T (W(3P))	β_T= 4.072 λ_T = 20.036 α_T= 4.8369	0.05756	1.594	21.864	

According to Tables (5-8), the functions of both GEV distribution and the logarithm of S data, functions of both W-3P distribution and the logarithm of T data, comparison between the estimated theoretical model and logarithmic regression model of S and T datasets, and comparison between EFxy and LFxy are performed. Furthermore, the logarithmic regression equation LF_X for $F(x)$ over $\ln(x)$ is presented in Figure 3. Also, the logarithmic regression equation LF_Y for $F(y)$ over $\ln(y)$ is displayed in Figure 4.

Table 5: Functions of both GEV distribution and logarithm of S data.

N	x	F(x) Actual	ln x	$\ln^2 x$	$\ln x * F(x)$
1	2.4	0.005875	0.875469	0.766446	0.005144
2	2.9	0.00753	1.064711	1.133609	0.008017
3	3.0	0.007904	1.098612	1.206949	0.008684
4	3.6	0.010499	1.280934	1.640792	0.013449
5	3.9	0.012044	1.360977	1.852257	0.016392
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
359	31.6	0.977115	3.453157	11.92429	3.374132
360	31.8	0.978923	3.459466	11.96791	3.386552
361	31.9	0.979787	3.462606	11.98964	3.392618
362	32.0	0.980626	3.465736	12.01133	3.39859
363	32.1	0.981439	3.468856	12.03296	3.40447
Sum		182.5516	1031.188	2984.177	558.8398

The logarithmic regression equation for $F(x)$ over $\ln(x)$ can be obtained from Eq. (56) as follow

$$\begin{vmatrix} F(x) & \ln(x) & 1 \\ 182.5516 & 1031.188 & 363 \\ 558.8398 & 2984.177 & 1031.188 \end{vmatrix} = 0$$

then, we get that $F(x) = 0.608\ \ln(x) - 1.217$. Thus, $\varphi_1 = 0.608$ and $\varphi_2 = 1.217$. Also, from Eqs. (57) and (58) we obtain that $R^2 = 0.999894$, $R^2_{adj.} = 0.99979$ and the coefficient of variation (CV) = 35.9988%.

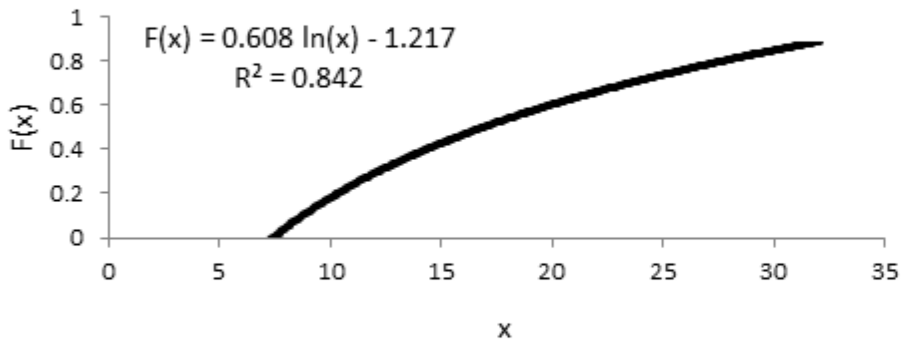


Fig. 3: Illustrate the plot of the logarithmic regression equation LF_X for $F(x)$ over $\ln(x)$.

Table 6: Functions of both W-3P distribution and logarithm of T data.

N	y	F(y) Actual	ln y	$\ln^2 y$	$\ln y * F(y)$
1	8.0	0.000544	2.079442	4.324077	0.001131
2	11.2	0.009323	2.415914	5.836639	0.022522
3	12.3	0.017769	2.509599	6.298088	0.044593
4	12.4	0.018749	2.517696	6.338796	0.047205
5	12.6	0.02083	2.533697	6.41962	0.052776
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
359	33.3	0.984655	3.505557	12.28893	3.451766
360	33.5	0.986401	3.511545	12.33095	3.463794
361	33.7	0.98798	3.517498	12.37279	3.475219
362	34.0	0.990062	3.526361	12.43522	3.491314
363	34.0	0.990062	3.526361	12.43522	3.491314
Sum		181.5427	1128.374	3527.736	589.2964

Further, Pearson correlation for the daily solar radiation (x) and the daily maximum temperature (y) is 0.995. Similarly, the logarithmic regression equation for $F(y)$ over $\ln(y)$ can be obtained from Eq. (56) as follow

$$\begin{vmatrix} F(y) & \ln(y) & 1 \\ 181.5427 & 1128.374 & 363 \\ 589.2964 & 3527.736 & 1128.374 \end{vmatrix} = 0$$

then, we get that $F(y) = 1.235\ \ln(y) - 3.338$. Thus, $\omega_1 = 1.235$ and $\omega_2 = 3.338$. Also, from Eqs. (57) and (58) we obtain that $R^2 = R^2_{adj.}$=0.999914 and the coefficient of variation (CV) = 22.104%.

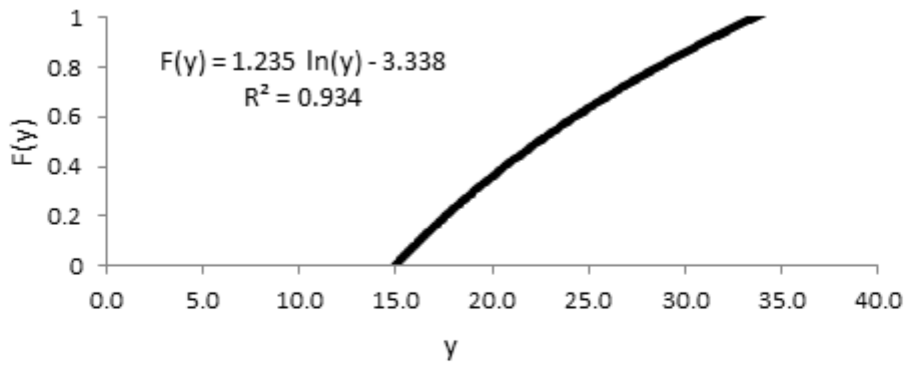


Fig. 4: Illustrate the plot of the logarithmic regression equation LF_Y for $F(y)$ over $\ln(y)$.

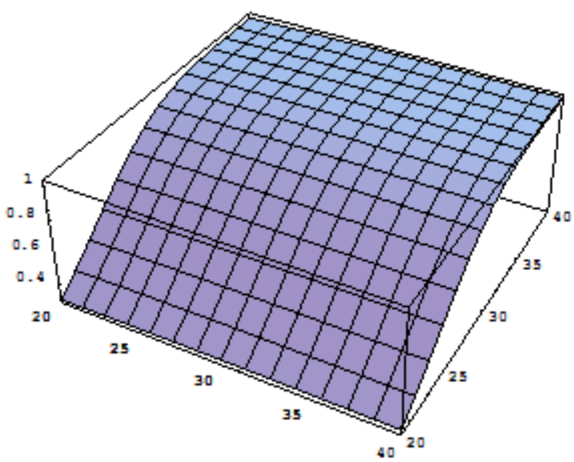
Table 7: Comparison between the estimated theoretical model and the logarithmic regression model of S and T.

EF_X	LF_X	EF_Y	LF_Y
0.602562	0.604405	0.61914	0.622402
0.608097	0.607438	0.634145	0.632362
0.608097	0.607438	0.634145	0.632362
0.613608	0.610455	0.634145	0.632362
0.619095	0.613457	0.634145	0.632362
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
0.967575	0.864952	0.979476	0.972679
0.971022	0.870864	0.980611	0.976439
0.97211	0.872822	0.982729	0.983924
0.973168	0.874774	0.984655	0.991363
0.976171	0.880592	0.986401	0.998759

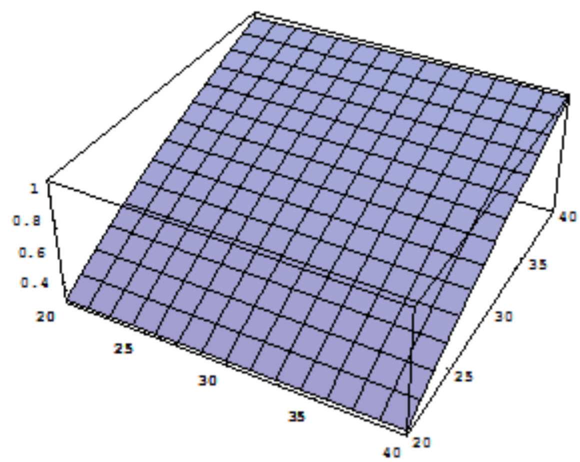
Table 8: Comparison between EFxy and LFxy.

EFxy	LFxy
0.623383	0.551929
0.638147	0.560509
0.638147	0.560509
0.638356	0.560623
0.638564	0.560737
⋮	⋮
⋮	⋮
0.986706	0.859509
0.987938	0.862928
0.990034	0.869364
0.991942	0.875761
0.993749	0.882267

Figure 5 illustrates the estimation of $F(X,Y)$ for each of the estimated theoretical model and logarithmic regression model. In addition, there is a positive relation between solar radiation and maximum temperature. Moreover, there is a convergence between the two models in the shape and location of the diagram for a specific values only but do not cover all data values.



(a)



(b)

Fig. 5: Comparison $F(X,Y)$ between (a) EFxy and (b) LFxy.

By using Mann-Whitney test and homogeneity test of EFxy and (LFxy, MFxy), respectively. We show that the P-value of the comparison between EFxy and LFxy of the logarithmic regression model is equal to zero; consequently EFxy and LFxy are not equivalent. On the other hand, the P-value of the comparison between EFxy and MFxy of the multi-regression model is equal to $0.911 \geq 0.05$, then EFxy and MFxy are equivalent and homogeneous as shown in Tables (9-11). Therefore, the multi-regression model is fitting to the datasets of S and T in Queensland during 2015-2016.

Table 9: Comparison between EFxy, LFxy and MFxy.

Estimation model (EFxy)	LFxy		MFxy	
	Estimates	P-value	Estimates	P-value
A = 0.0379103 B = 0.97 C = 0.00007	$\hat{A}$=0.0379103 $\hat{B}$=0.85 $\hat{C}$=0.00007	0.0	A^*= 0.038 B^*= 0.970 C^*= 7.000E-5	0.911

Such that EFxy refers to the estimate of the joint distribution function $F(X,Y)$ of the theoretical model and LFxy refers to the estimate of $F(X,Y)$ in logarithmic regression model and MFxy refers to the estimate of $F(X,Y)$ in the multi-regression model.

From Eqs. (59) and (60), then the estimation model is EF_{xy} = (0.0379103) $F(X)$ + (0.97) $F(Y)$ – (0.00007) $F(X)F(Y)$ and the multi-regression model is MF_{xy}= $F^*(X,Y)$= (0.38) $F(X)$ + (0.970) $F(Y)$– (0.00007) $F(X)F(Y)$, respectively.

Table 10: ANOVA of the estimated theoretical model.

Source	Sum of Squares	df	Mean Squares
Regression	125.612	3	41.871
Residual	0.000	360	0.000
Uncorrected Total	125.612	363	
Corrected Total	33.423	362	

Where R^2 for the estimated theoretical model = 1.

Table 11: Homogeneity Test of EFxy and MF_{xy}.

Var of EF_{xy}	Var of MF_{xy}	F-Max Test
0.092327485	0.0923433	0.999828734

Computed F-Max = Var of EFxy / Var of MFxy = 0.999828734≈ 1, then EF_{xy} and MF_{xy} are Homogeneity.

Finally, estimation of the theoretical model allows us to estimate the value of the joint distribution function (EF_{XY}) by knowing both X and Y while (MF_{XY}) enables us to estimate the value of the joint distribution function with high accuracy by knowing the values of both $F(X)$and $F(Y)$ since EF_{XY} and MF_{XY} are equivalent and homogeneous if $F(X)$and $F(Y)$ are known. On the other hand, the logarithmic regression model (LF_{XY}) was found to be non-equivalently with both EF_{XY} and MF_{XY} and inaccurate and did not cover all the range of the variables X and Y, so exclude its use in predicting of the joint distribution function.

7 Conclusions

Many researchers need easier-to-use and more accurate statistical methods to predict the climate changes occurring in many regions of the world. Therefore, we have estimated $\left(F(X,Y)\right)$ alongside the Markov property (EF_{XY}). The method of MLE has utilized for determining the unknown parameters of the GEV distribution and W-3P distribution for both S and T, respectively. In addition, datasets of S and T in Queensland, Australia during 2015-2016 were analyzed to obtain the fitting distributions the climate change. Logarithmic (LF_{XY}) model and the multi-regression (MF_{XY}) model are constructed for the $F(X,Y)$ to validate the theoretical model estimate. Comparing the results of EF_{XY} with both LF_{XY} and MF_{XY}, the MF_{XY} model offers a better modeling to the data in which EF_{XY} and MF_{XY} are equivalent, homogeneous and including all the data.

Data Availability

(1) Daily global solar exposure in Queensland (Terrey Hills), Australia, in 2015, data utilized to support this study's results are listed in the complementary information file named [IDCJAC0016_040395_2015_Solar Radiation_Queensland_Australia].
(2) Daily global solar exposure in Queensland (Terrey Hills), Australia, in 2016, data utilized to support this study's results are listed in the complementary information file named [IDCJAC0016_040395_2016_Solar Radiation_Queensland_Australia].
(3) Daily maximum temperature in Queensland (Terrey Hills), Australia, in 2015, data utilized to support this study's results are listed in the complementary information file named [IDCJAC0010_041529_2015_Maximum Temperature_Queensland_Australia].
(4) Daily maximum temperature in Queensland (Terrey Hills), Australia, in 2016, data utilized to support this study's results are listed in the complementary information file named [IDCJAC0010_041529_2016_Maximum Temperature_Queensland_Australia].

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

There are no sources of funding for this work.

Acknowledgments

I should give many thanks to Australian Bureau of Meteorology for providing online climate data of Queensland, Australia. I also extend my appreciation to the authors who have contributed to the field of this study. Moreover, the authors express their gratitude towards the reviewers for their valuable feedback and suggestions on the original manuscript. Without their efforts, this study would have not been in its final version.

Supplementary Materials

(1) The file named [Solar Radiation 2015] shows the daily global solar exposure (MJ/m2) in Queensland (Terrey Hills), Australia, in 2015.
(2) The file named [Solar Radiation 2016] represents the daily global solar exposure (MJ/m2) in Queensland (Terrey Hills), Australia, in 2016.
(3) The file named [Maximum Temperature 2015] shows the daily maximum temperature (degrees Celsius) in Queensland (Terrey Hills), Australia, in 2015.
(4) The file named [Maximum Temperature 2016] shows the daily maximum temperature (degrees Celsius) in Queensland (Terrey Hills), Australia, in 2016; (*Supplementary Materials*).

References

[1] H. V. Asselt and F. Green, COP26 and the dynamics of anti-fossil fuel norms, *WIREs Climate Change*, 1-12 (2022).

[2] A. Rempel and J. Gupta, Equitable, effective, and feasible approaches for a prospective fossil fuel transition, *WIREs Climate Change*, **13(2)**, 1-32 (2022).

[3] J. W. Harder, S. Béland, S. Penton and T. N. Woods, Long-Term Trend Analysis in the Solar Radiation and Climate Experiment (SORCE)/Spectral Irradiance Monitor (SIM), *Solar Physics*, **297(69)**, 1-28 (2022).

[4] L. Linguet, Y. Pousset and C. Olivier, Identifying statistical properties of solar radiation models by using information criteria, *Solar Energy*, **132(4356)**, 236–246 (2016).

[5] A. Linares-Rodríguez, J. A. Ruiz-Arias, D.Pozo-Vazquez and J. Tovar-Pescador, Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks, *Energy*, **36(8)**, 5356–5365 (2011).

[6] S. Ghimire, R. C. Deo, N. J. Downs and N. Raj, Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia, *Cleaner Production*, **216**, 288–310 (2019).

[7] M. K. Linnenluecke, C. Zhou, T. Smith, N. Thompson and N. Nucifora, The impact of climate change on the Australian sugarcane industry, *Journal of Cleaner Production*, **246**, 118974 (2020).

[8] J. Polo, L. F. Zarzalejo, R. Marchante and A. A. Navarroa, A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution, *Solar Energy*, **85(5)**, 1164–1170 (2011).

[9] M. Larrañeta, S. Moreno-Tejera, M. A. Silva-Pérez and I. Lillo-Bravo, An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series, *Solar Energy*, **122**, 517–528 (2015).

[10] C. Euán, Y. Sun and B. J. Reich, Statistical analysis of multi-day solar irradiance using athreshold time series model, *Environmetrics: John Wiley & Sons Ltd*, 1–21 (2022).

[11] M. M. El Genidy, Multiple linear regression formula for the probability of the average daily solar energy using the queue system, *Asian Journal of Mathematics and Statistics*, **5(3)**, 93–98 (2012).

[12] M. El Genidy, Multiple nonlinear regression of the Markovian arrival process for estimating the daily global solar radiation, *Communications in Statistics—Theory and Methods*, **48(22)**, 5427–5444 (2019).

[13] Ö. Ayvazoğluyüksel and Ü. B. Filik, Estimation methods of global solar radiation, cell temperature and solar power forecasting: A review and case study in Eskişehir, *Renewable and Sustainable Energy Reviews*, **91**, 639-653 (2018).

[14] M. AlKandari and I. Ahmad, Solar power generation forecasting using ensemble approach based on deep learning and statistical methods, *Applied Computing and Informatics*, (2020).

[15] H. X. Li, D. J. Edwards, M. R. Hosseini and G. P. Costin, A review on renewable energy transition in Australia: An updated depiction, *Cleaner Production*, **242**, 118475 (2020).

[16] A. Manzano, M. L. Martín, F. Valero and C. Armenta, A single method to estimate the daily global solar radiation from monthly data, *Atmospheric Research*, **166**, 70-82 (2015).

[17] B. Gong and J. Ordieres-Mere, Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong, *Environmental Modelling and Software*, **84**, 290–303 (2016).

[18] Y. Chen, S. Li and S. Guo, A Novel fractional hausdorff discrete grey model for forecasting the renewable energy consumption, *Journal of Mathematics*, **11**, 1-23 (2022).

[19] P. S. Balakrishnan, M. Shabbir, F. A. Siddiqi and X. Wang, Current status and future prospects of renewable energy: A case study, *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, **42(21)**, 2698-2703 (2020).

[20] P. A. Qstergaard, N. Duic, Y. Noorollahi, H. Mikulcic and S. Kalogirou, Sustainable development using renewable energy technology, *Renewable Energy*, **146**, 2430-2437 (2020).

[21] M. S. Mecibah, T. E. Boukelia, R. Tahtah, and K. Gairaa, Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (case study: Algeria), *Renewable and Sustainable Energy Reviews*, **36**, 194–202 (2014).

[22] K. Doost and M. Akhlaghi, Estimation and comparison of solar radiation intensity by some models in a region of Iran, *Journal of Power and Energy Engineering*, **2(4)**, 345-351 (2014).

[23] A. Das, J. K. Park and J. H. Park, Estimation of available global solar radiation using sunshine duration over South Korea, *Journal of Atmospheric and Solar-Terrestrial Physics*, **134**, 22-29 (2015).

[24] E. Quansah, L. K. Amekudzi, K. Preko et al., Empirical models for estimating global solar radiation over the Ashanti region of Ghana, *Journal of Solar Energy*, **3**, 1-6 (2014).

[25] T. Liu, Application of Markov chains to analyze and predict the time series, *Modern Applied Science*, **4(5)**, 162-166 (2010).

[26] S. Bhardwaj, V. Sharma, S. Srivastava et al., Estimation of solar radiation using a combination of hidden Markov model and generalized Fuzzy model, *Solar Energy*, **93**, 43-54 (2013).

[27] E. Rodríguez, J. M. Cardemil, A. R. Starke and R. Escobar, Modelling the Exergy of Solar Radiation: A Review, *Energies*, **15(4)**, 1-26 (2022).

[28] X. Meng and J. W. Taylor, Comparing Probabilistic Forecasts of the Daily Minimum and Maximum Temperature, *International Journal of Forecasting*, **38(1)**, 267-281 (2022).

[29] S. Nadarajah, The exponentiated Gumbel distribution with climate application, *Environmetrics*, **17**, 13–23 (2006).

[30] F. Yang, H. Ren and Z. Hu, Maximum Likelihood Estimation for Three-Parameter Weibull Distribution Using Evolutionary Strategy, *Mathematical Problems in Engineering*, 1-8 (2019).

[31] M. M. El Genidy and D. A. E. Abd El-Rahman, A New high accurate estimation method for evaluating the daily solar energy by nested percentiles algorithm, *Asian Journal of Scientific Research*, **12(4)**, 480-487 (2019).

[32] M. M. El Genidy, Statistical modelling of the daily global solar radiation in Queensland, Australia, *Songklanakarin Journal of Science Technology*, **41(6)**, 1381-1389 (2019).

[33] M. El Genidy, W. Megahed and K. Mahfouz, Algorithms of solar energy prediction combined with percentile root estimation of three-parameters distributions, *Applied Mathematics & Information Sciences*, **16 (4)**, 1-19 (2022).

[34] M. M. El Genidy and M. S. Beheary, Forecasting methods in various applications using algorithm of estimation regression models and converting data sets into Markov model, *Complexity*, 1-20 (2022).