Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/160406

Algorithms of Solar Energy Prediction Combined with Percentile Root Estimation of Three-Parameters Distributions

M. El Genidy[1,*], *W. Megahed*[2] *and K. Mahfouz*[1]

[1]Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt
[2]Department of Basic science, Obour Higher Institute for Management Informatics, Egypt

Received: 6 Mar. 2022, Revised: 5 Apr. 2022, Accepted: 27 May 2022
Published online: 1 Jul. 2022

Abstract: We propose three algorithms to the problem of solar energy prediction and Percentile Root Estimation (PRE) of three-parameters distributions. The first algorithm named Algorithm of Change Rate Matrix (ACRM), Our approach is based on creating a matrix of solar energy change rates for each month separately during successive years. ACRM is characterized by not relying on the transition matrix or Markov model. The second algorithm named Algorithm of Converting Dataset to Markov model (ACDM) depends on the transition states of the solar energy and Markov model for a month during successive years. The results were compared with the actual values to validate the algorithms ACRM and ACDM. We demonstrate the ability of the mentioned algorithms to perform on the other dataset in various applications. The third algorithm PRE applied on the distributions Lognormal, Fatigue lifetime, Erlang, Fréchet and Pert which it was validated using Goodness-fit-tests, Anderson-Darling test. We analyzed the influence of PRE algorithm, as a result it is more accurate and easier in coding than the maximum likelihood estimation method.

Keywords: Anderson-Darling test, Change rate matrix, Markov model, Percentile root

1 Introduction

Currently, the primary endeavour in the current age is to achieve the optimal use of available resources to manage both risk and opportunity and avoid the crisis. Analysis of energy crisis provides energy security and present solutions and renewable energy potential [1]. Eliminating potential risks also ensures a stable climate, providing enormous opportunities to support the civilization of the energy age effectively. Over the next few decades, global energy demand is expected to increase as shown [2,3,4,5,6,7]. Therefore, climate research such as wind and solar exposure have a great interest, especially in converting them to electric energy [8,9] Solar power has been the cornerstone for some groundbreaking research in the forecasting field. Since incorporating solar power into the national electricity grid proliferates, enhancing this highly dynamic renewable resource's forecast becomes indispensable.

Since there has been a great deal of research in recent years on improving solar panel power forecasts whether short-term forecasts last a few hours or longer, such as sub-seasonal climate predictions. Forecasts are primarily used to plan future deployments, maximize plant operations and effectiveness, and optimize load demand and supply [10]. The decision and the interest time scale, whether hourly, daily, or monthly, affect the forecasting methodology as shown by [11]. However, it is critical to create or apply simple methods to real-life model systems for estimating and predicting solar exposure reaching the earth's atmosphere using observed datasets of other weather conditions [12,13]. Although much research has focused on statistical approaches to improve solar energy forecasting, finite mixture distributions, mixture models, and techniques have received little attention [14,15,16].

Markov models have been used for forecasting and modelling the solar exposure for a long time. For several things such as demand dynamics, wind power, and solar

* Corresponding author e-mail: drmmg2016@yahoo.com

irradiates, Markov models have been used as a predictive analytic tool [17]. Markov model of solar exposure is proven to provide efficient modelling to obtain an accurate prediction. The Markov-chain forecasting models use various techniques to obtain accurate probability models, ranging from approximating time series for hidden Markov models combined with wavelet coefficients and the Markov-chain mixture distribution model [18,19].

Building a model reflecting global solar exposure is imperative since this needs to be simple and precise to predict the amount of solar exposure accurately. However, the more complicated the model is, the perfect modelling it gives with more accurate results. Most of the previous studies are complicated and difficult to apply and to understand. Therefore, in this study simpler methods were suggested to obtain the acceptable prediction algorithms. The study contributes to developing an interactive algorithm for obtaining the prediction of solar energy in next years with less standard error of the estimation. Researchers have faced severe difficulties in modelling, estimating, and obtaining the prediction equation of meteorological data like the solar energy amount and the corresponding maximum temperature. Various methods for estimating global solar exposure have been published using empirical correlations. The appropriate model must be depending on the statistical properties of global radiation with a variety of parameters to obtain accurate estimations of solar radiation and meet the growing demand for evaluation of solar energy system optimization and efficiency [20]. For regions where no real calculated values are, correlations estimate the values of meteorological data for a region of investigation from more widely accessible meteorological, climatologically, and geographical parameters using different meteorological and geographical variables [21].

Mathematical modelling and forecasting of energy remain a significant issue to detect and reinforce power management. Research papers keep track of potential risk factors, analyze and research all possible scenarios, especially for overcoming energy crises and managing power availability costs to achieve long-term growth. The existing probability forecasts have appropriate performance and reliability to provide beneficial guidance for energy decisions. Solar radiation can be accurately measured for areas with and without observed meteorological climate stations using various solar radiation prediction models [22]. Improving the economics of solar energy systems through the planned method yields solar forecasting recommendations and a broad perspective on improvements and their consequences. Almost all the approaches used applied forecasting parameters such as sunlight, position, temperature, and moisture. Therefore, methods must include more than two parameters to obtain reliable tests, carrying more complex and higher computational error risks. It has been found that the sampling approach transforms into statistical inference methods by relying on a broad capacity to make statistical decisions. Thus, it needs to select the appropriate statistical inference method for the significant studies terminal objective. The significance of estimation methods and selecting suitable models to match the data have been highlighted. The best practice modelling requires different assumptions to be verified and several factors to be taken into consideration [23,24,25,26].

The development of pattern similarity in solar radiation estimation in the clustering algorithm and its implementation have been presented in various studies. Bhardwaj et al. used the hidden Markov model to extract shape-based clusters from the input meteorological parameters, which is then processed by the generalized distribution to estimate solar radiation accurately [27]. The patterns of the data vectors are used as the similarity index for clustering instead of using distance function as an index of similarity, which overcomes a few of the disadvantages associated with distance-based clustering approaches [28].

In modelling the daily global solar exposure dataset, different techniques and procedures are used. Statistical model combined with generalized extreme values distribution is used by [29]. Non-linear regression and multiple nonlinear regression were used to get the closest probability distribution. Procedures such as the moment's method and the Kolmogorov test were performed to validate the generalized extreme value distribution parameters. (Quartiles-Moments) method was applied to estimate the solar exposure distribution. The parameters of the Exponentiated Gumbel Maximum Distribution (EGMD) were proposed to estimate the solar energy. The Australian Bureau of Meteorology's updated numerical climate prediction systems which developed a forecast solar exposure region that has been validated on multiple sites for 2012. ACCESS model, which was updated in August of 2010, became usable for the Australian. The transition matrix regulates the actions of Markov chains, which used to characterize the changes of a structure over time. The observed method is a strategy for determining the maximum estimate of the probability matrix unusual intervals. The methodology is with distinct advantages compared to other optimization approaches for that type of problem.

Transition states probabilities are usually obtained through Bayesian theory. Bayesian statistics is a method for analysing data and estimating parameters based on the Bayes theorem by creating a joint probability distribution that includes observed and unobserved parameters in a statistical model. The aggregation of available

information on a given parameter via a pre-determined distribution is inserted in a statistical model. The identification of the probability function using the information on parameters of observation data, and the combination of the previous distribution and likelihood function with Bayesian theory are the three key concepts in a typical Bayesian workflow. A solution to the Markovian state jump systems' estimation problem with a transition probability matrix has been proposed by [30]. Four algorithms for Minimum Mean-Square Error (MMSE) estimation of the transition probability matrix has been derived based on state recursion.

Markov chain models have been applied to analyze and predict data sequences in different time series. Welton and Ades have shown how to fill the gap between transition rates and probabilities in multi-states models using probability data. Standard uniform properties, Bayesian estimation, and transition rate uncertainty were used [31]. The initial and final states of completely observed and partially observed data after a fixed time interval were investigated. With an illustrative example for a 3-state model, the Markov chain method in WinBUGS was used to propose diagnostics for examining discrepancies between evidence from different starting states.

In Markov chain models, there is a variety of methods and techniques for estimating transformation probabilities. Transition probabilities Matrix are extracted to simulate and model the given data. It is often preferable to use a technique to convert the original dataset into a more compatible format that can be managed without significant issues. Techniques differ from one approach to another, depending on the nature of the dataset and the considered issue. The study redefines well-known modelling steps to predict solar energy amounts. Converting the rates of change in solar energy into a regression model to get the transition change rates. The technique reduces the standard error in estimating process by regenerate the regression equation for each new year adopting a specific structure to arrive at the best regression line. Also, the transition frequency with Markov modelling steps is merged with the regression polynomial equations to estimate the transition matrices in flexible methods. Solving the issue of state estimation matrices is a significant step in building an effective Markov model.

2 Materials and Methods

The best practice modelling necessitates the verification of various assumptions of some variables. Markov chains are used to model and solve several practical issues as they effectively represent the data sequences and data shift during different periods. The state transition probabilities can make optimal solutions to Markov decision problems susceptible. In many practical situations, the estimation of these probabilities is far from accuracy.

A set of mutually exclusive states, transition probabilities, and a valid cycle length are all required by the Markov model. The transition matrix, also known as the transition probabilities matrix, is the set of state probability transitions over a given period. Models of Markov chains explain how a system evolves. Transition probabilities are determined by a series of events that are more likely to occur at various periods. If the transition probability matrix of the Markov chain is known, predictions for each state can be produced.

The transfer probabilities between the various states can be calculated using the dataset history, much like the probabilities of each state. However, estimating the matrix is more complicated due to the dynamic relationship between transformation probabilities. A stochastic process is a big concept that uses a mathematical model that incorporates probabilities and matrices to analyze. The statistical properties of the Markov chain are investigated. Each month's values are checked to see if they are independent of one another. The assumptions of the independence test are checked before doing the test.

Consider X_t is a stochastic process $X_t, t = 1,2,3,...,$ where a X_t is a random variable with a discrete-time stochastic, and n is the month in a year through 30 years. Let $X_t = i$the process is in a state or position i in time t. The process undergoes transition matrix properties with probability $P_{i,j}$ since the process will be in position j at time t + 1. Transition probability states $P_{i,j}$ satisfies the Markov property (known as forgetfulness property), which was defined as follows:

$$\begin{aligned} &P(X_{t+1} = j | X_t = i_t, X_{t-1} = i_{t-1}, ..., X_0 = i_0) \\ &= P(X_{t+1} = j | X_0 = i_0) = P_{i,j} \end{aligned} \tag{1}$$

$$\begin{aligned} &P(X_{t+1} = j | X_t = i_t, X_{t-1} = i_{t-1}, ..., X_0 = i_0) = \\ &\frac{P(X_{t+1} \cap X_0 \cap X_1 \cap ... \cap X_{t-1} \cap X_t)}{P(X_0, X_1, ..., X_t)} = \\ &\frac{P(X_{t+1}|X_0)P(X_{t-1})P(X_{t-2})...P(X_1)P(X_0)}{P(X_0 P(X_1)...P(X_{t-1})} = \\ &P(X_{t+1}|X_t) = P(X_{t+1} = j | X_t = i_t = i) = P_{i,j} \end{aligned} \tag{2}$$

Eq. (2) holds the conditional distribution that any future state X_{t+1} depends only on the current state X_t, knowing all the history of states $X_0, X_1, ..., X_{t-1}$. Given the current

state of the process, the Markov chain reveals the future as independent of the past. The random process $X_t, t = 1,2,3,L$ is a continuous-time Markov chain, if for all $0 \leq s$, $0 \leq t$, and non-negative integers i,j. Equation.(3) simplifies Markov property where:

$$P[X(t+s) = j|X(s) = i, X(u) = u]; 0 \leq u \leq s, \\ P[X(t+s) = j|X(s) = i, X(u) = u] \tag{3}$$

The conditional probability satisfies the Markov process. The transition matrix p displays all various transition states. The probability of the Monthly Global Solar Exposure Average (MGSEA) will be predicted using the transition matrix p.

$$P = \begin{bmatrix} X_{(1,1)} & X_{(1,2)} & \dots & \dots & X_{(1,12)} \\ X_{(2,1)} & X_{(2,2)} & \dots & \dots & X_{(2,12)} \\ \vdots & \vdots & & \dots & \vdots \\ X_{(12,1)} & X_{(12,2)} & \dots & \dots & X_{(12,12)} \end{bmatrix} \tag{4}$$

where
$0 \leq P(X = x_{i,j}), \sum_{j=1}^{r} P_{ij} = 1 \; i, j = 1,2,...,r$

The probability of the states for each month will be estimated over a twenty six year to place it in the transition matrix p. The change rates of MGSEA between two successive years for the same month was also computed. Markov model of the MGSEA of January were applied to obtain the matrix p. The transition matrix is denoted by P_{ij} and their elements is represented by X(i,j).

Let $X_0, X_1, ..., X_{t-1}$ be a Markov chain with a random variable X_t. The probability distribution of X_t represented as a vector $N * 1$, where N is the total number of states and q_0 is the probability distribution of X_0

$$q_0 = \begin{bmatrix} (P(X_0 = 1) \\ P(X_0 =) \\ \vdots \\ P(X_0 = N) \end{bmatrix} = \begin{bmatrix} \pi^1 \\ \pi^2 \\ \vdots \\ \pi^N \end{bmatrix}$$

Since the probability vector π is every row of the matrix, with twelve entries, the sum of each row equals one.

$$\sum_{j=1}^{N} P_{ij} = \sum_{j=1}^{N} P(X_{i+1} = j|X_t = i) \sum_{j=1}^{N} P_{X_t=i}(X_{t+1} = j) = 1 \tag{5}$$

$$X_0 = \pi_t$$

where

$$\pi_t = \left[P(X_0 = 1) = X_1 \; P(X_0 = 1) = X_2 \; \dots \; \dots \; P(X_0 = 1) = X_{12}\right]$$

$$P(X_1 = j) = \sum_{i=1}^{N} P(X_1 = j|X_0 = i) p(X_0 = i) = \\ \sum_{j=1}^{N} P_{ij}\pi_i = \left(\pi^i P\right)_j \tag{6}$$

Similarly,

$$P(X_2 = j) = \sum_{i=1}^{N} P(X_2 = j|X_0 = i) p(X_0 = i) = \\ \sum_{j=1}^{N} P_{ij}^2 \pi_i = \left(\pi^t P^2\right)_j \tag{7}$$

Then,

$$\begin{aligned} X_0 &= \pi^t \\ X_1 &= X_0 p = \pi^t p \\ X_2 &= X_1 p^1 = \pi^t p^2 \\ &\vdots \\ X_T &= X_{T-1} p^{T-1} = \pi^t p^T \end{aligned} \tag{8}$$

Let $X_0, X_1, X_2, \cdots$ be a Markov chain with random variable x_t and transition matrix P $N * N$. If the probability distribution of $X_0$is given by $1 * N$ row vector π^t. The probability distribution of X_t will be given by $1 * N$ row vector$\pi^t P^T$.

$$\begin{aligned} X_T &= \pi^t p^T \\ X_{T+1} &= \pi^t p^{T+1} \end{aligned} \tag{9}$$

The probabilities of predicted values for the twelve months of the next year are determined by multiplying the probability vector of the year and the transition matrix (all subsequent single-step probabilities), where Y represents the transition probability for the following year.

$$Y = \begin{bmatrix} X_{(1,1)} & X_{(1,2)} & \dots & \dots & X_{(1,12)} \\ X_{(2,1)} & X_{(2,2)} & \dots & \dots & X_{(2,12)} \\ \vdots & \vdots & & \dots & \vdots \\ X_{(12,1)} & X_{(12,2)} & \dots & \dots & X_{(12,12)} \end{bmatrix} \begin{bmatrix} (P(x_0 = x_1) = X_1 \\ P(x_0 = x_2) = X_2 \\ \vdots \\ P(x_0 = x_{12}) = X_{12} \end{bmatrix} \tag{10}$$

The transition matrix and the initial vector (starting probabilities) provide all the necessary backgrounds to simulate the given data and obtain the desired information. According to Markovian assumptions, probabilities for X_{t+1} are solely dependent on the value of X_t for any given step.

Time series data on solar exposure can be collected from various sources, including IMD, NREL, Metronome, NASA, the World Radiation Data Centre (WRDC), etc. Some of these organizations' data that are for free, while others require payment. Solar exposure data could be denoted as $X_t, t = 1, 2, 3, \cdots$, where X_t is the solar exposure at discrete time t that could be the number of hours, days, months, or years. Solar exposure measurements are usually continuous values, but to fit for a Markov chain application, X_t must have a finite number of states.

The study area in Queensland, Australia (Terrey Hills) was chosen as the study site. The annual solar exposure map reveals that the southern coastal regions are more exposed to the sun that central and northern Australia. Inland Australian areas have a lower humidity in the air, resulting in less cloud cover. Yearly Global Solar Exposure Average (YGSEA) was one of the datasets used in this analysis. Climate data from Queensland, Australia, over the last 30 years were actual and reliable, as reported by the Bureau of Meteorology in the Australian government website: http://www.bom.gov.au. Figure 1, indicates the Monthly average of global solar exposure (GSEA) in 30 years successively.

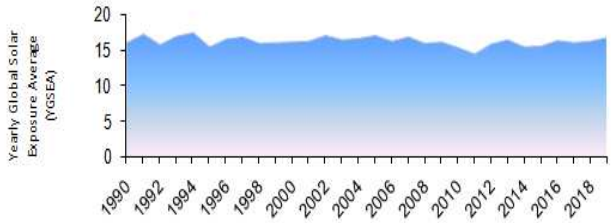

Fig. 1: Yearly Global Solar Exposure Average in Queensland-Australia

The values of daily global solar exposure are [1-35] MJ/m^2 (mega joules per square meter). Summer days with clear skies have the highest values, while winter days with heavy clouds have the lowest. As shown in Figure 2, for Monthly Global Solar Exposure Average over a twenty-six-years, while Figure 3, shown their percentages.

Processing missing data, there was a missing value in the month December in 2005. The regression equation for December values for the 30 years with the principle of Markov that future value depends only on the current value. Two points before and after the missing points are used to estimate the approximating curve to predict the missing point. Figure.4 depicts the best-fitting curve to the data of December.

Consequently, the nonlinear regression equation $y = 0.4583x^3 - 22.536x^2 + 367.43x - 1963.4$ was

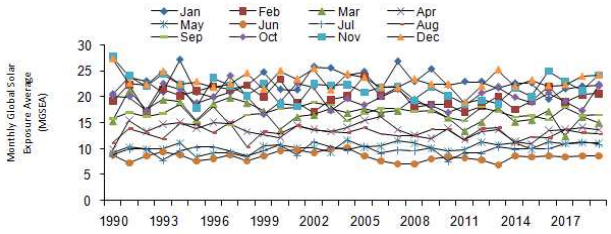

Fig. 2: Monthly Global Solar Exposure Average in Queensland-Australia

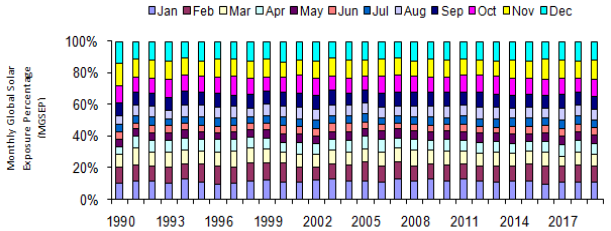

Fig. 3: Monthly Global Solar Exposure Percentage over 30 years in Queensland-Australia

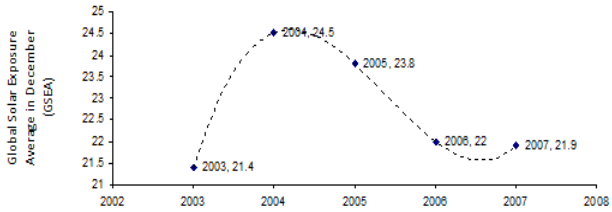

Fig. 4: Fitting regression curve of Global Solar Exposure Average in December, Queensland- Australia

obtained whereby R-Square almost 1, x is the rank of the year in twenty-six years from 1990 to 2019, and y is the global solar Exposure in Dec. Therefore, the missing estimated value of y at x = 16 is equal to 23.8 for 2005.

Converting dataset to change rates using, Markov transition matrix was developed to model and forecast the percentages of monthly solar exposure averages ACRM is used to determine the potential divisions of solar exposure quantities for each month in 30 years. The change rates matrix of solar exposure were calculated, then ACRM was performed. Matrix of change rates reserve the main idea of change rate principles. Calculate the change rates of monthly solar exposure for two successive years in matrix A during 1990 $\rightarrow$ 1991 which represented by

$$\left[G_{(1,1)}\ G_{(1,2)}\ \cdots\ \cdots\ G_{(1,11)}\ G_{(1,12)}\right]$$

, the change rate matrix of 1991 $\rightarrow$ 1992 can be

$$\left[G_{(2,1)}\ G_{(2,2)}\ \ldots\ \ldots\ G_{(2,11)}\ G_{(2,12)}\right]$$

. Similarly, we obtained the change rates values over the twenty-six years. Create the change rates matrix named C_1 , with dimension $25*12$, where the rows represent the yearly change rates and columns are the monthly change rates in solar exposure.

$$C_1 = \begin{bmatrix} G_{(1,1)} & G_{(1,2)} & \ldots\ \ldots & G_{(1,12)} \\ G_{(2,1)} & G_{(2,2)} & \ldots\ \ldots & G_{(2,12)} \\ \vdots & \vdots & \ldots\ \ldots & \vdots \\ G_{(24,1)} & G_{(24,2)} & \ldots\ \ldots & G_{(24,12)} \\ G_{(25,1)} & G_{(25,2)} & \ldots\ \ldots & G_{(25,12)} \end{bmatrix}$$

Determine a non-linear regression polynomial equation for each month for the solar exposure: $\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \cdots$. The prediction of the change rate values $\hat{G}$ have order $(m+1, j)$, where $j : 1,2,\cdots,12, m : 1,2,\cdots,25$

$$\begin{aligned} \hat{G}_{m+1,i} &= \frac{G_{m+1,i} - G_{m,i}}{G_{m,i}} \\ G_{m+1,i} &= \hat{G}_{m+1,i} G_{m,i} + G_{m,i} \end{aligned} \tag{11}$$

Add the change rates as a new row in the change rates matrix C_N , where N is the total number of change rates. Then the prediction of the solar radiation will be obtain at $(m+n, j)$.

$$C_n = \begin{bmatrix} G_{(1,1)} & G_{(1,2)} & \ldots\ \ldots & G_{(1,12)} \\ G_{(2,1)} & G_{(2,2)} & \ldots\ \ldots & G_{(2,12)} \\ \vdots & \vdots & \ldots\ \ldots & \vdots \\ G_{(24,1)} & G_{(24,2)} & \ldots\ \ldots & G_{(24,12)} \\ G_{(25,1)} & G_{(25,2)} & \ldots\ \ldots & G_{(25,12)} \\ \vdots & \vdots & \ldots\ \ldots & \vdots \\ G_{(m,1)} & G_{(m,2)} & \ldots\ \ldots & G_{(m,12)} \\ G_{(m+n,1)} & G_{(m+n,2)} & \ldots\ \ldots & G_{(m+n,12)} \end{bmatrix}$$

Then the steps of the Algorithm 1 (ACRM) will be as follows:

Algorithm 1 (ACRM)

Rem Create Change Rate Matrix CRM (m x j)

Step 1. Read m "No. of Change rates=m where No. of years is m+1".
Step 2. For j=1 to 12 "Month No."
Step 3. For x=1 to m
Step 4. Compute change rates $G(x, j)$ which represents the elements of CRM, where Y represents Solar Energy (SE) value:
$G(x, j) = Y(x+1, j) - Y(x, j)/Y(x, j)$.
Step 5. Next x
Step 6. Next j **Rem** Compute the Prediction of CRM $(m+1xj)$ **Rem** Forecasting SE values for Y(m+r+1,j) where r=1,. . . ,n
Step 7. Read n
Step 8. For j=1 to 12
Step 9. For r=1 to n
Step 10. Estimate the non-linear regression $\hat{G}(m+r, j) = a(m+r, j)x^2 + b(m+r, j)x + c(m+r, j)$ by the given points (x,G(x,j)) where $x = 1,2,\cdots,m+r-1$ and $j = 1,2,\cdots,12$
Step 11. Compute the estimated SE for the month j at the year No. m+r+1: (m+r+1,j) using the equation: $\hat{G}(m+r, j) = Y(m+r+1, j) - Y(m+r, j)/Y(m+r, j)$
Step 11. Next r
Step 11. Next j

Converting dataset to Markov model using ACDM, the algorithm 2 converts the dataset into three – states of Markov model which classifies the change rates of solar radiation to three states Low (L), Medium (M), and High (H). The three – states of Markov model represent the values of the solar exposure over successive years. Then, the transition matrix was constructed for the given dataset and predict the probability of the states L, M, and H for the same month in next year.

Algorithm 2 (ACDM)

Step 1. Compute the frequencies of the transition states L, M and H for the month:

X_1: L $\rightarrow$ L $\rightarrow$ n.(LL)	Y_1: M $\rightarrow$ M $\rightarrow$ n.(MM)	Z_1: H $\rightarrow$ H $\rightarrow$ n.(HH)
X_2: L $\rightarrow$ M $\rightarrow$ n.(LM)	Y_2: M $\rightarrow$ L $\rightarrow$ n.(ML)	Z_2: H $\rightarrow$ L $\rightarrow$ n.(HL)
X_3: L $\rightarrow$ H $\rightarrow$ n.(LH)	Y_3: M $\rightarrow$ H $\rightarrow$ n.(MH)	Z_3: H $\rightarrow$ M $\rightarrow$ n.(HM)

$$F_{ij} = \frac{\sum_{i=1}^{n} X_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}} \tag{12}$$

Step 2. Compute the transition state probability:

$$P(X_1)=\frac{n.(LL)}{n.(LL)+n.(LM)+n.(LH)} \quad P(Y_1):\frac{n.(MM)}{n.(MM)+n.(ML)+n.(MH)}$$

$$P(X_2):\frac{n.(LM)}{n.(LL)+n.(LM)+n.(LH)} \quad P(Y_2):\frac{n.(ML)}{n.(MM)+n.(ML)+n.(MH)}$$

$$P(X_3):\frac{n.(LH)}{n.(LL)+n.(LM)+n.(LH)} \quad P(Y_3):\frac{n.(MH)}{n.(MM)+n.(ML)+n.(MH)}$$

$$P(Z_1):\frac{n.(HH)}{n.(HH)+n.(HL)+n.(HM)} \quad P(Z_2):\frac{n.(HL)}{n.(HH)+n.(HL)+n.(HM)}$$

$$P(Z_3):\frac{n.(HM)}{n.(HH)+n.(HL)+n.(HM)}$$

The corresponding Markov transition state probability diagram over each state showing all possible transitions among all states with the possible transition probabilities using equation 12, seems as follow in Figure 5.

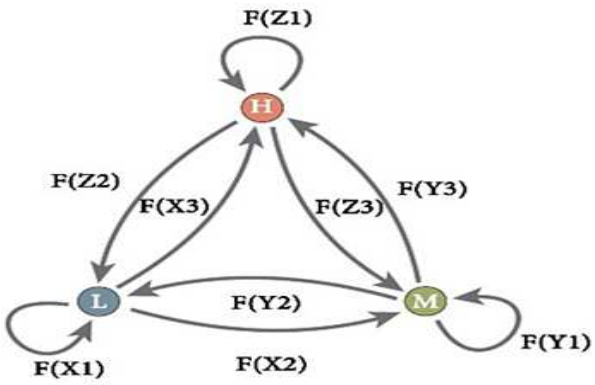

Fig. 5: Markov-Transition diagram.

Step 3. The state space of model depends on the number of existing states used in predicting the next state. The possible states used in the previous model are L, M, H. The states can be classified into more states to insure more reliability. The Markov model predicts the next state using the last state known as the first order Markov model is used in the study. More complicated model computes the prediction using the last two actions known as the second order Markov model, and its states correspond to all possible pairs that can be performed in sequence. Transition matrix is formed easily with the possible probabilities of states.

$$T=\begin{bmatrix} P(X_1) & P(Y_3) & P(Z_2) \\ P(X_2) & P(Y_1) & P(Z_3) \\ P(X_3) & P(Y_2) & P(Z_1) \end{bmatrix} = \begin{bmatrix} P(X_{LL}) & P(Y_{ML}) & P(Z_{HL}) \\ P(X_{LM}) & P(Y_{MM}) & P(Z_{HM}) \\ P(X_{LH}) & P(Y_{MH}) & P(Z_{HH}) \end{bmatrix} \tag{13}$$

where

$K = 0,1,\cdots,8$ and $X_{LL} = 0, X_{LM} = 1, X_{LH} = 2$, $Y_{ML} = 3, Y_{MM} = 4, Y_{MH} = 5$, $Z_{HL} = 6, Z_{HM} = 7, Z_{HH} = 9$

Step 4. Construct the regression polynomial models.

$$P(X) = \sum_{i=0}^{2} a_i x^i$$

$$P(Y) = \sum_{i=3}^{5} a_i y^i$$

$$P(Z) = \sum_{i=6}^{8} a_i z^i$$

$$T=\begin{bmatrix} a_2X_{LL}^2+a_1X_{LL}^1+a_0 & a_5Y_{LM}^2+a_4X_{LM}^1+a_3 & a_8Z_{HL}^2+a_7Z_{HL}^1+a_6 \\ a_2X_{LM}^2+a_1X_{LM}^1+a_0 & a_5Y_{MM}^2+a_4X_{MM}^1+a_3 & a_8Z_{HM}^2+a_7Z_{HM}^1+a_6 \\ a_2X_{LH}^2+a_1X_{LH}^1+a_0 & a_5Y_{MH}^2+a_4X_{MH}^1+a_3 & a_8Z_{HH}^2+a_7Z_{HH}^1+a_6 \end{bmatrix} \tag{14}$$

where

$$P(X_{LH}) = 1-[P(X_{LL})+P(X_{LM})];$$

$$\sum_{i=1}^{3} P_{i1}(x) = 1$$

$$P(Y_{MH}) = 1-[P(Y_{ML})+P(Y_{MM})];$$

$$\sum_{i=1}^{3} P_{i2}(Y) = 1$$

$$P(Z_{HH}) = 1-[P(Z_{HL})+P(Z_{HM})];$$

$$\sum_{i=1}^{3} P_{i3}(Z) = 1$$

Step 5. Then we obtain the coefficients by solving the following equations:

$$\begin{bmatrix} X_{LL}^2 & X_{LL}^1 & 1 \\ X_{LM}^2 & X_{LM}^1 & 1 \\ X_{LH}^2 & X_{LH}^1 & 1 \end{bmatrix} \begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} P(X_{LL}) \\ P(X_{LM}) \\ P(X_{LH}) \end{bmatrix}$$

$$\begin{bmatrix} Y_{ML}^2 & Y_{ML}^1 & 1 \\ Y_{MM}^2 & Y_{MM}^1 & 1 \\ Y_{MH}^2 & Y_{MH}^1 & 1 \end{bmatrix} \begin{bmatrix} a_5 \\ a_4 \\ a_3 \end{bmatrix} = \begin{bmatrix} P(Y_{ML}) \\ P(Y_{MM}) \\ P(Y_{MH}) \end{bmatrix}$$

$$\begin{bmatrix} Z_{HL}^2 & Z_{HL}^1 & 1 \\ Z_{HM}^2 & Z_{HM}^1 & 1 \\ Z_{HH}^2 & Z_{HH}^1 & 1 \end{bmatrix} \begin{bmatrix} a_8 \\ a_7 \\ a_6 \end{bmatrix} = \begin{bmatrix} P(Z_{HL}) \\ P(Z_{HM}) \\ P(Z_{HH}) \end{bmatrix}$$

Step 6. Solve the following equation to predict the probability of the states L, M, and H for the same

month in next year.

$$S_1 = T * \begin{bmatrix} P_{CU}(L) \\ P_{CU}(M) \\ P_{CU}(H) \end{bmatrix} = \begin{bmatrix} P_{next}(L) \\ P_{next}(M) \\ P_{next}(H) \end{bmatrix}$$

3 Results and Discussion

Since Markov chains can be modelled at the state level, random walks are a great example of their mathematical utility. Modelling the method of solar exposure and forecasting the potential amount of solar exposure over the year is done using the Markov Transition Matrix. The previous procedures are used to estimate the predicted quantities of MGSE in 2018 and 2019 by working on the data to get the optimum structure to obtain the transition matrix probabilities with correct prediction results for next year.

Results of Algorithm 1: The procedures were applied over the twelve months to obtain the average change rates of the solar exposure by ACRM. The change rates of the solar exposure were computed as shown in Table 1, then the regression polynomial models for each month were determined. Consequently, the estimated change rates for each month in the next years can be obtained using the previous regression polynomial models.

Table 1: Change Rates Matrix for each month over the twenty-six years

	Jan	Feb	⋯	Nov	Dec
1	0.160	0.160	⋯	-0.139	-0.103
2	-0.037	-0.236	⋯	-0.089	-0.012
3	-0.019	0.125	⋯	0.102	0.171
4	0.029	-0.051	⋯	-0.017	-0.096
5	-0.024	0.033	⋯	-0.209	0.019
⋮	⋮	⋮	⋮	⋮	⋮
24	0.032	0.121	⋯	0.241	0.133
25	0.021	0.136	⋯	0.2431	-0.041

The non-linear regression curve equation for January change rates over the data set is shown in Figure 6, as $y = 0.0005x^4 - 0.0088x^3 + 0.0726x^2 - 0.2732x + 0.3444$.

Similarly, the regression equation for each month is obtained with considering R^2 approximately equal one, to ensure the best curve-fitting to the data. As shown in Figure 7, there are two regressions curves representing the data. The first regression curve equation is $y = -0.06x^3 + 0.205x^2 - 0.0005x + 0.0018$ with R^2 equal 0.4348. The second regression curve equation is $y = -0.09x^5 - 0.8507x^4 + 0.2205x^3 - 0.0002x^2 + 0.0008x + 0.0205$ with R^2 equal 0.8495. However, the selected curve is the one with R^2 equal to one.

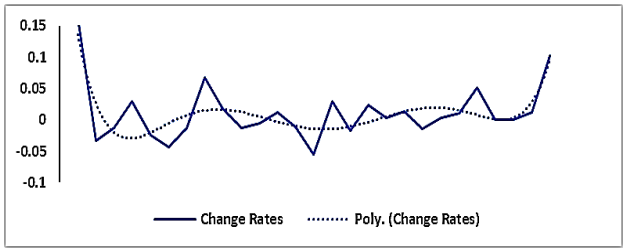

Fig. 6: Regression polynomial curve of the change rates for the solar exposure in January.

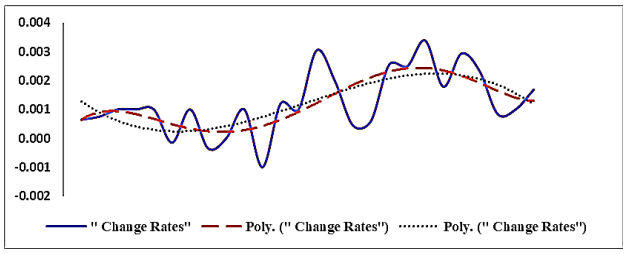

Fig. 7: Regression polynomial curve of the change rates for the solar exposure in February.

Using the regression equations, we estimate the change rates among current months and future months. The estimation of the change rates make it possible to predict the amount of solar exposure for next month. For $C_{m+1} = C_{26}$, we obtained January change rate $\hat{C}_{m=26} = 0.01$. Substituting in Eq. 11, to obtain the predicted value of January $G_{m+1,i} = \hat{G}_{m+1,i} G_{m,i} + G_{m,i} = 0.01 * 22.7 + 22.7 = 22.927$.

Similarly, the estimated change rates of the twelve months are added as new row in change rate matrix $C_{m+1} = 26$. The prediction values of Solar Exposure of the twelve months over the years from 2018 – 2020 are

Table 2: Solar exposure for each month in 2018, 2019 and 2020

	2018	2019	2020
Jan	23.589	21.168	23.328
Feb	21.8283	24.882	21.6021
Mar	17.16	19.03	13.75
Apr	13.31	16.456	16.698
May	11.88	13.68	13.2
Jun	8.585	8.686	8.585
Jul	11	12	12.21
Aug	12.423	12.322	13.938
Sep	15.908	15.132	17.46
Oct	20.9	22.88	21.12
Nov	24.321	30.25	27.83
Dec	23.0437	21.9558	22.1536

shown in Table 2, using the algorithm of the change rates.

$$C_{26} = \begin{bmatrix} 0.16 & 0.01 & \dots & -0.139 & -0.103 \\ -0.37 & -0.26 & \dots & 0.08 & -0.012 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.021 & 0.136 & \dots & 0.23 & -0.041 \\ 0.017 & 0.131 & \dots & 0.21 & -0.11 \end{bmatrix}$$

The procedures of ACRM are pursued the track, until the change rates matrix of the order $m+n$,C_{m+N}, is obtained, in addition to the years of the order $m+n+1$.

Results of Algorithm 2: Classify and convert the dataset of each month to three states. The classification depends on the length of daily solar exposure where the states will be as: low (L) in [1 – 12], medium (M) in [12 – 24], and high (H) in [24 – 35]. The probability transition matrix, T, is obtained through the history of the twenty-six years of monthly Solar Exposure. Current state for each month is calculated as the initial vector S_0 represent the current state of the month. Using ACDM algorithm, the transition matrix T was obtained by substituting K values in the Eq.14, to the coefficients a_i. Then, Markov property is used to predict the next state for each month. That means (The next state, S_1) = (Transition matrix, T) * (current state, S_0); $S_1 = P * S_0$

$$S_1 = \begin{bmatrix} 0.75 & 0.151 & 0 \\ 0.25 & 0.822 & 0.25 \\ 0 & 0.027 & 0.75 \end{bmatrix} \begin{bmatrix} P_{CU}(L) \\ P_{CU}(M) \\ P_{CU}(H) \end{bmatrix} = \begin{bmatrix} P_{next}(L) \\ P_{next}(M) \\ P_{next}(H) \end{bmatrix}$$

Where cu is symbol for current state. The possible states used in the used model are L, M, H as shown. The probabilities of each state are evaluated. The transition probabilities of each state are computed, and the transition probability matrix is shown in Figure 6. T matrix indicates that no transition is possible between H and L. changes in the amounts of solar exposure must be gradually over time. Therefore, if it is high, to get low amounts of SE, it must get medium amounts first, but the opposite is correct; this means there are no sudden changes of SE amounts.

Nevertheless, in real life, things are not constant, especially in climate conditions. That means there may be sudden changes among SE amounts in January. To obtain a specific transition probability matrix of a specific month, like January, the daily amount of SE of that month through the twenty-six year are classified into the three-state Markov chain. Markov transition diagram of SE in January and February shown in Figure 8.

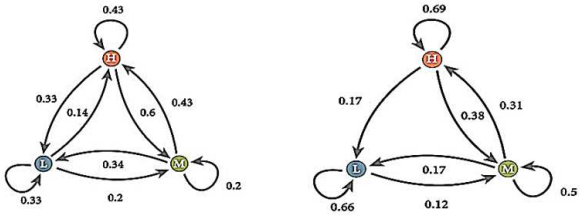

Fig. 8: Markov- Transition diagram of Solar Exposure of January and February.

As a result of the prediction, the percentage of MGSEA was acceptable with a low error of average and dispersion. The estimated transition probability for solar exposure data shown in Table 3. While, the results of the two algorithms ACRM and ACDM are shown in Table 4.

The efficiency of the algorithms may be validated using the Sum of Squared Residuals (SSR). The difference between predicted and actual data is measured to insure precision. MSE, Mean Square Error and RRMSE, Relative Root Mean Square Error was defined to be used to assess this methodology performance. The accuracy of the SSR significantly depends on the models' capacity to adapt to the relevant data conditions of the various states.

Table 3: States probability for each month in 2018

	P(L)	P(M)	P(H)
Jan	0.23	0.33	0.44
Feb	0.24	0.33	0.43
Mar	0.40	0.41	0.19
Apr	0.38	0.43	0.18
May	0.54	0.35	0.10
Jun	0.75	0.25	0.00
Jul	0.59	0.33	0.08
Aug	0.40	0.43	0.18
Sep	0.30	0.48	0.32
Oct	0.17	0.40	0.44
Nov	0.13	0.33	0.53
Dec	0.12	0.37	0.51

Table 4: Estimated values of ACRM and ACDM in 2018 – 2019

	2018			2019		
	Actual	ACRM	AcDM	Actual	ACRM	AcDM
Jan	22.1	23.0859	20.77	22.3	21.168	19.45
Feb	20.5	21.8283	20.25	20.7	24.882	19.051
Mar	16.2	17.16	12.82	15.1	19.03	12.63
Apr	14.2	13.31	13.51	13.7	16.456	12.34
May	11.2	11.88	5.57	11.1	13.68	5.10
Jun	8.7	8.585	7.74	8.7	8.686	6.45
Jul	11.4	11	6.07	10.9	12	5.41
Aug	13.1	12.423	13.32	12.9	12.322	12.57
Sep	16.4	15.908	14.81	16.5	15.132	11.49
Oct	17.4	20.9	13.57	22.1	22.88	13.40
Nov	21.2	24.321	24.97	24.2	30.25	22.63
Dec	24	23.0437	23.79	24.3	21.956	22.19

Some of estimation errors such as R^2, MSE, MPE, RMSE, erMAX, and RRMSE were computed in Table 5 to validate the results.

The model is very accurate as there is a strong correlation between the model's predictions and the actual results Figure 10. The remarkable aspect of the system is that the prediction error exact value was established before any prediction was made. The forecast is the best possible outcome from global solar exposure data. The technique of ACRM is better than ACDM, see the Tables (4,5) and Figures (9,10). On the other hand, ACDM algorithm can be improved by shortening the length of the

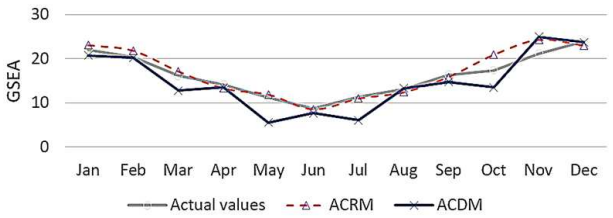

Fig. 9: Comparison between ACRM and ACDM with the actual data of GSEA in 2018.

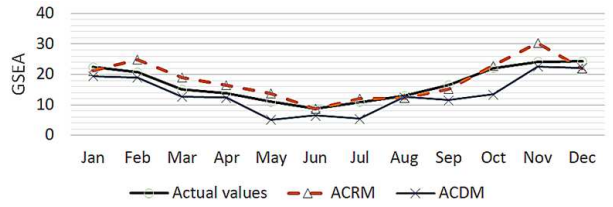

Fig. 10: Comparison between ACRM and ACDM with the actual data of GSEA in 2019.

Table 5: Estimation errors of solar radition

	ACRM		ACDM	
	2018	2019	2018	2019
R^2	0.83	0.93	0.83	0.74
MSE	0.86	0.64	0.86	1.17
MPE	-0.53	0.50	-0.53	-1.27
RMSE	0.80	0.68	0.93	1.08
erMAX	0.18	0.15	1.01	1.18
RRMSE	0.01	0.01	0.02	0.02

interval. In other words, the results of ACDM will be more accurate by classifying the dataset into six-states.

Percentile Root Estimation (PRE), the algorithm 3 based on solving equations to get numerical approximated values for the unknown parameters. Theoretically, the practical steps to apply the Algorithm 3, on the three-parameters distribution are indicated below.

Algorithm 3 (PRE)

Step 1. Determine the required probability distributions for the dataset that will be checked.

Step 2. Evaluate the empirical properties as the mean and standard deviation for the dataset.

Step 3. Write the parameters in terms of one parameter, $R:\alpha\rightarrow f(\theta),\beta\rightarrow g(\theta)$
Suppose that the scale parameter α shape parameter β and location parameter θ with CDF $F(X;\alpha,\beta,\theta)$. A relation R, mathematical elicitation, is figured out among the parameters using theoretical properties of the under-study distribution and the empirical values of the data to obtain the scale parameter α as a function F of location parameter θ, and to obtain the shape parameter β as function g of location parameter θ. Where $R:\alpha\rightarrow f(\theta),\beta\rightarrow g(\theta)$

Step 4. Substitute parameters in CDF equation to obtain $F[X;f(\theta),g(\theta),\theta]=F(X;\theta)=U_i$.

Step 5. Solve the CDF equation to get the value of the parameter $\theta=F^{-1}(X;\theta)$.

Step 6. Obtain the value of other parameters.
Step 7. Check the validation of parameters through theoretical mean, variance, and
$$0<F(X_i;\alpha,\beta,\theta)<1;$$
$$F(X_1)<F(X_2)<\cdots<F(X_N)$$
Step 8. If step 7 is valid; repeat the steps 5,6, and 7 for N times to obtain α,β,θ.
Step 9. If step 7 is not valid, repeat step 3 with different relations:$R_2:\alpha\rightarrow l(\theta),\beta\rightarrow w(\theta)$, then continue the steps 4-7, and check again.
Step 10. The estimated parameter value is the median of all obtained values.

The percentiles equations of the Cumulative Distribution Function (CDF) will be solved for all values of the variable X. Moreover, the algorithm PRE will check the percentile roots with the actual statistics measures of the given dataset. To demonstrate the impact of PRE, the method was applied on five different three-parameter distributions: Lognormal distribution, Fatigue lifetime distribution, Fréchet distribution, Erlang distribution, and Pert distribution.

Suppose the random variable $X_i:x_1,x_2,\cdots,x_N$ follows the exponential distribution with probability density function $f(x;\lambda)=\lambda e^{-\lambda x}$, and cumulative function $F(x;\lambda)=1-e^{-\lambda x}$. There is only one parameter, shape parameter λ. Then, the steps of the algorithm PRE will be as follows: determine the dataset and no need to evaluate empirical properties from the data to exploit. Therefore, compute the numerical estimation for the parameter as follows: $\theta=F^{-1}(X;\theta)$
$U_i=\frac{i}{N},i=1,2,\cdots,N$
$\lambda=\frac{-ln(1-U_i)}{x_i}$

All possible values for the parameter λ are obtained through substituting the values of variable X_i and percentile U_i in the above equation. Then, checking the validation of estimated parameters through theoretical properties of the distribution, such as theoretical mean and variance is a priority. Ensuring whether the estimated values imply the following basics:
$0<F(X_i;\lambda)<1;$
$F(X_1)<F(X_2)<\cdots<F(X_N)$

All the possible values for the rate parameter λ were obtained. There are extra steps for two parameters and more, including finding relations among parameters to obtain parameters in terms of one parameter.

Suppose a random variable $X=x_1,x_2,\cdots,x_{364},x_{365}$ of the daily global solar exposure in Queensland, Australia over 2018. Whereby, $X=X_i:ln(x_i-\gamma),i=1,\cdots,365$ follows the three-parameters Lognormal distribution. The three parameters were defined as location parameter (γ), scale parameter (μ), and shape parameter (σ). Then

$$f(x:\mu,\sigma,\gamma)=\frac{1}{(x-\gamma)\sigma\sqrt{2\pi}}exp\left[\frac{-(ln(x-\gamma)-\mu)^2}{2\sigma^2}\right] \tag{15}$$

$-\infty<\mu<\infty,\gamma<x<\infty$

$$F(x:\mu,\sigma,\gamma)=\Phi\left[\frac{(ln(x-\gamma)-\mu)}{\sigma}\right] \tag{16}$$

Mean:

$$E(x)=\gamma+exp\left[\mu+\frac{\sigma^2}{2}\right] \tag{17}$$

Median:

$$Med=\gamma+exp\left[\mu\right] \tag{18}$$

$$\mu=ln(Med-\gamma) \tag{19}$$

$$\sigma=\sqrt{\frac{1}{2}ln\left[\frac{E(x)-\gamma}{Med-\gamma}\right]} \tag{20}$$

Finally, the location parameter γ could be calculated by applying PRE with percentile equation CDF for the given dataset. Then, the values of the location parameter will be substituted in Eq.19 and Eq.20 to obtain the parameters μ and σ.

$$F(x:\gamma)=\Phi\left[\frac{(ln(x-\gamma)-[ln(Med-\gamma)])}{\sqrt{\frac{1}{2}ln\left[\frac{E(x)-\gamma}{Med-\gamma}\right]}}\right] \tag{21}$$

Equation.21 contains one unknown parameter, the location parameter (γ) is now easy to solve knowing the values of variable X_i and the percentile values.

Substituting the value of the estimated location parameter to obtain the values of parameters μ and σ. The process is repeated several times $N;i = 1,2,3,\cdots,N$. Since N represents the size of the dataset. The mathematical calculations can be done manually or with Software programs. Mathematica program version 11.1.1 is used to solve the equations using the code FindRoot.

Fatigue lifetime distribution, often known as the distribution of Birnbaum–Saunders, is a widely used probability distribution to estimate breakdown times in reliability applications. Let X be a random variable with a shape parameter $\alpha > 0$, a scale parameter$\beta > 0$, and a location parameter $\gamma > 0$. The probability density function of X follows Fatigue distribution is

$$f(x:\alpha,\beta,\gamma) = \phi\left[\frac{1}{\alpha}\sqrt{\frac{x-\gamma}{\beta}}+\sqrt{\frac{\beta}{x-\gamma}}\right]* \frac{\sqrt{\frac{x-\gamma}{\beta}}+\sqrt{\frac{\beta}{x-\gamma}}}{2\alpha(x-\gamma)} \tag{22}$$

$$F(x:\alpha,\beta,\gamma) = \Phi\left[\frac{1}{\alpha}\sqrt{\frac{x-\gamma}{\beta}}+\sqrt{\frac{\beta}{x-\gamma}}\right] \tag{23}$$

Mean:

$$E(x) = \gamma+\beta\left[1+\frac{\alpha^2}{2}\right] \tag{24}$$

Skewness:

$$Sk = \frac{4\alpha(11\alpha^2+6)}{(5\alpha^2+4)^{\frac{3}{2}}} \tag{25}$$

Variance:

$$var(x) = (\alpha\beta)^2\left[1+\frac{5\alpha^2}{4}\right] \tag{26}$$

The scale parameter (β), and location parameter (γ) were described in terms of the shape parameter (α) in Eq. 27 and 28. Otherwise, the skewness in Eq. 25 can be used to obtain the shape parameter (α), then Eq. 26 is used to get the scale parameter (β). The percentiles values with variable Xi, location parameter, and scale parameter values are substituted in Eq. 23 to obtain the values of shape parameter (α).

$$\beta = \sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}} \tag{27}$$

$$\gamma = E(x)-\left[1+\frac{\alpha^2}{2}\right]\sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}} \tag{28}$$

$$F(x:\alpha) = \Phi\left[\frac{1}{\alpha}\sqrt{\frac{x-\left[E(x)-\left[1+\frac{\alpha^2}{2}\right]\sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}}\right]}{\sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}}}}\right] + \sqrt{\frac{\sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}}}{x-\left[E(x)-\left[1+\frac{\alpha^2}{2}\right]\sqrt{\frac{var(x)}{(\alpha)^2\left[1+\frac{5\alpha^2}{4}\right]}}\right]}} \tag{29}$$

The Erlang distribution, known as the Erlang-k distribution is a specific case of the Gamma distribution, where the shape parameter is an integer number. Erlang distribution was developed for application in the queuing theory, stochastic processes, and mathematical biology. In the study of queuing systems, Erlang distribution represents the time between the arrivals and time for services of a unit. As a result of multi-steps models, the Erlang distribution was suggested as a good approach to the distribution of cell cycle times, and inter-purchase intervals in business economics. In the case of the two-parameter Erlang distribution, the location parameter equals zero.

$$f(x:m,\beta,\gamma) = e^{\left[\frac{x-\gamma}{\beta}\right]}\frac{(x-\gamma)^{m-1}}{\beta^m\Gamma(m)} \tag{30}$$

where m is a positive integer, $\beta > 0$, and $\gamma < x < \infty$

$$F(x:m,\beta,\gamma) = \frac{\Gamma_{\frac{x-\gamma}{\beta}}(m)}{\Gamma(m)} \tag{31}$$

where $\Gamma(.)$ is the gamma function and $\Gamma - z(.)$ is the incomplete gamma function.

Mean:

$$E(x) = \gamma+m\beta \tag{32}$$

Variance:

$$var(x) = m\beta^2 \tag{33}$$

$$\gamma = E(x)-\frac{var(x)}{\beta} \tag{34}$$

$$m = \frac{var(x)}{\beta^2} \tag{35}$$

$$F(x:\beta) = \frac{\Gamma_{\frac{x-\left[E(x)-\frac{var(x)}{\beta}\right]}{\beta}}\left(\left[\frac{var(x)}{\beta^2}\right]\right)}{\Gamma\left(\frac{var(x)}{\beta^2}\right)} \tag{36}$$

In the case of the two-parameter Erlang distribution, we used the mean in Eq. 34 to obtain the shape parameter in terms of the scale parameter as in Eq. 37.

$$m = \frac{E(x)}{\beta} \tag{37}$$

$$F(x:\beta) = \frac{\Gamma_{\frac{x}{\beta}}(\left[\frac{E(x)}{\beta}\right])}{\Gamma(\frac{E(x)}{\beta})} \tag{38}$$

Fréchet distribution, known as the Extreme Value Distribution (EVD) Type II, is used to represent the maximum values for a given data. The distribution is used to simulate a large range of phenomena such as flood evaluation, horse racing, human life cycles, extreme weather, and hydrological discharges.

$$f(x:\alpha,\beta,\gamma) = \frac{\alpha}{\beta}\left[\frac{\beta}{x-\gamma}\right]^{\alpha+1} \exp\left[-\left[\frac{\beta}{x-\gamma}\right]^{\alpha}\right] \tag{39}$$

where $\alpha > 0, \beta > 0$, and $\gamma < x < \infty$

$$F(x:\alpha,\beta,\gamma) = \exp\left[-\left[\frac{\beta}{x-\gamma}\right]^{\alpha}\right] \tag{40}$$

Median:

$$Med = \gamma + \frac{\beta}{\sqrt[\alpha]{log(2)}} \tag{41}$$

Mode:

$$M = \gamma + \beta\left[\frac{\alpha}{1+\alpha}\right]^{\frac{1}{\alpha}} \tag{42}$$

$$\beta = \frac{Med - m}{[log(2)]^{\frac{-1}{\alpha}} - \left[\frac{\alpha}{1+\alpha}\right]^{\frac{1}{\alpha}}} \tag{43}$$

$$\gamma = Med - \left[\frac{Med - m}{[log(2)]^{\frac{-1}{\alpha}} - \left[\frac{\alpha}{1+\alpha}\right]^{\frac{1}{\alpha}}}\right][log(2)]^{\frac{-1}{\alpha}} \tag{44}$$

Pert distribution is the most common form of risk distribution performed at the risk-based estimation method. It is defined by three extremely well-defined points; (a) the minimum parameter, (b) the maximum parameter, and (m) the most likely parameter. Pert distribution is a special case of the Beta distribution specified by the parameters α_1 and α_2. $\alpha_1 = \frac{4m+b-5a}{b-a}$, $\alpha_2 = \frac{5b-a4m}{b-a}$

$$f(x:a,m,b) = \frac{1}{B(\alpha_1,\alpha_2)}\frac{[x-a]^{\alpha_1-1}[b-x]^{\alpha_2-1}}{[b-a]^{\alpha_1+\alpha_2-1}} \tag{45}$$

where $a < x < b$

$$f(x:a,m,b) = I_z(\alpha_1,\alpha_2) \tag{46}$$

where $z = \frac{x-a}{b-a}$, $B(\alpha_1,\alpha_2)$ is the Beta function, $I_z(\alpha_1,\alpha_2)$ is the Regularized Incomplete Beta function. Where $I_z(\alpha_1,\alpha_2) = \frac{B_x(\alpha_1,\alpha_2)}{B(\alpha_1,\alpha_2)}$.

$$Mode : mode = m \tag{47}$$

$$Variance : var(x) = \frac{(b-a)^2}{36} \tag{48}$$

$$b = 6\sqrt{var(x)} + a \tag{49}$$

The most likely parameter is easily obtained by the mode of the given dataset, leaving the boundary parameters of the distribution to be found. Using the variance Eq. 48, the maximum parameter is described in terms of the minimum parameter as in Eq.49. The estimated parameters by the proposed algorithm PRE were obtained through substituting the pre-defined parameters in the percentiles of CDF equations. The pre-defined parameters are mathematical elicitation from the theoretical properties of the probability distribution in the form of relation between the parameters and the actual properties of the dataset. Through solving the illustrated percentile equations, the parameters are estimated in sequence as shown for each probability distribution. The estimated parameters of the distributions: Lognormal distribution, Fatigue life, two-parameter Erlang, three-parameter Erlang, Fréchet, and Pert were obtained in Tables 6 – 10.

The results of PRE algorithm was compared with MLE method. The goodness of fit tests determines whether a sample is compatible with a theoretical distribution or not. They are the essential measures of accuracy to any estimator. In other words, the tests illustrate how well the chosen distribution matches the given data. A popular test to evaluate the compatibility of observed data set to a probability distribution function is the Anderson-Darling test. The validity of PRE results is measured using Anderson-Darling test (AD). AD is considered effective fitting statistical test.

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[ln(F(X_i)) + ln(1 - F(X_{n-i+1}))\right] \tag{50}$$

AD test statistic value of PRE is obtained by applying Eq.50 and compared to the AD test value of MLE for the chosen distributions for the given dataset of global Solar Exposure. The estimated parameters and AD test statistic value of three-parameter lognormal distribution obtained by PRE and MLE are shown in Table 11. AD test statistic value of PRE is less than the AD test statistic value of MLE. The histogram-pdf plot for the probability distribution was used to indicate how well the distribution fits the data. The fitting of the three-parameters Lognormal distribution were estimated by PRE and MLE to the dataset of global solar exposure whereby are shown in Figure 11. The results of PRE were compared with the results of MLE and found that the estimated parameters are more accurate in PRE than MLE.

Table 6: Estimated parameters of Lognormal distribution using PRE

N	x	F(x)	γ	μ	σ
1	1.2	0.00274	-1.8204	2.80064	0.41945
2	1.5	0.005479	-1.7525	2.79886	0.41972
3	2.1	0.008219	-1.29935	2.78681	0.42158
4	3.6	0.010959	7-1.324	2.787473	0.421482
5	3.6	0.013699	-1.7942	2.799956	0.419551
6	3.8	0.016438	-1.9946	2.80517	0.418752
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
360	31.9	0.986301	-1.50611	2.79235	0.420724
361	31.9	0.989041	-1.53734	2.801084	0.419378
362	32	0.991781	-1.8076	2.800307	0.419497
363	32	0.994521	-1.82413	2.800739	0.419431
364	32.1	0.99726	1.93254	2.689493	0.437491
365	32.1	0.99999	-1.81076	2.80039	0.419485
Median			-0.046977	2.791381	0.420874

Table 7: Estimated parameters of Fatigue life distribution using PRE

N	x	F(x)	α	β	γ
1	1.2	0.00274	0.238002	27.04502	-13.8693
2	1.5	0.005479	0.247711	27.83011	-13.4229
3	2.1	0.008219	0.23603	27.30368	-13.5397
4	3.6	0.010959	0.23602	27.305	-13.2564
5	3.6	0.013699	0.237303	27.13622	-13.4267
6	3.8	0.016438	0.249899	27.56902	-13.8099
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
360	31.9	0.986301	0.24803	27.79176	-13.5797
361	31.9	0.989041	0.2392	27.88991	-13.5608
362	32	0.9917817	0.24989	27.57009	-13.9989
363	32	0.994521	0.24797	27.79897	-13.0203
364	32.1	0.99726	0.24787	27.81098	-13.6213
365	32.1	0.99999	0.23031	27.07839	-13.6697
Median			0.24764	27.79969	-13.5397

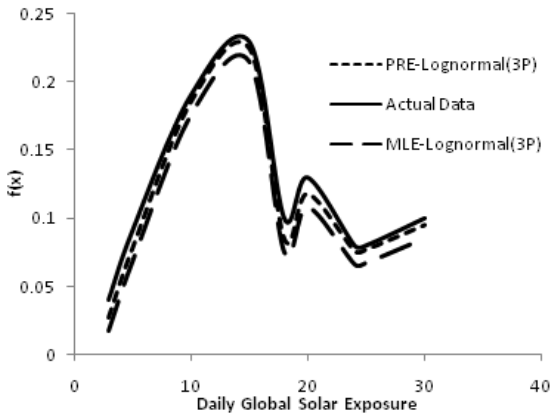

Fig. 11: Estimated three-parameter Lognormal distribution

Table 8: Estimated parameters of Erlang distribution using PRE

N	x	F(x)	β	γ	m	β	m
1	1.2	0.00274	3.153719	0.319938	5.144813	3.099318	5.33762
2	1.5	0.005479	3.183447	0.383226	5.049174	3.05469	5.40159
3	2.1	0.008219	3.17813	0.308738	5.06608	3.05469	5.417495
4	3.6	0.010959	3.187107	0.318689	5.037581	3.067985	5.34213
5	3.6	0.013699	3.306602	0.309882	4.680063	3.21901	5.13914
6	3.8	0.016438	3.183126	0.30961	5.050189	3.057331	5.41092
⋮	⋮	⋮	⋮	⋮	⋮		
⋮	⋮	⋮	⋮	⋮	⋮		
360	31.9	0.986301	3.167075	0.300813	5.101511	3.089141	5.35535
361	31.9	0.989041	3.176646	0.306882	5.070814	3.058131	5.40805
362	32	0.991781	3.186465	0.309845	5.039614	3.081029	5.35183
363	32	0.994521	3.179449	0.316301	5.06188	3.092131	5.34821
364	32.1	0.99726	3.184379	0.308793	5.046216	3.076118	5.37711
365	32.1	0.99999	3.17827	0.305705	5.065634	3.067708	5.40485
Median			3.187537	0.30934	5.04502	3.088082	5.357047

The estimated parameters and AD test statistic value of Fatigue lifetime distribution were obtained by PRE and MLE in Table 12.

The Fatigue distribution fitting data was estimated by PRE and MLE as shown in Figure 12. Although the histogram-fit line shown the similarity of the estimated values between the PRE and MLE, we found that the PRE covers the boundaries of the dataset better than MLE.

The estimated parameters of the three-parameters and two-parameters for Erlang distribution were obtained by PRE and MLE in Table 13. Moreover, the AD test statistic value and the standard errors of the estimated parameters were obtained.

The Erlang distribution three-parameters and two-parameters were estimated by PRE and MLE as shown in Figure 13. The histogram data-fit illustrates the

Table 9: Estimated parameters α, β, γ of Fréchet distribution using PRE

N	x	F(x)	α	β	γ
1	1.2	0.00274	1.8661632	11.77603	1.623546
2	1.5	0.005479	1.8625325	11.76241	1.6603329
3	2.1	0.008219	1.8615641	11.75703	1.603355
4	3.6	0.010959	1.8615641	11.76513	1.588823
5	3.6	0.013699	1.8629139	11.76042	1.656818
6	3.8	0.016438	1.8661632	11.76311	1.623546
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
360	31.9	0.986301	1.8629139	11.76042	1.668179
361	31.9	0.989041	1.8685159	11.7847	1.647330
362	32	0.9917817	1.8691487	11.7579	1.627343
363	32	0.994521	1.8698151	11.7794	1.608132
364	32.1	0.99726	1.8632417	11.67512	1.663724
365	32.1	0.99999	1.8625325	11.76701	1.670331
Median			1.86769	11.766	1.66996

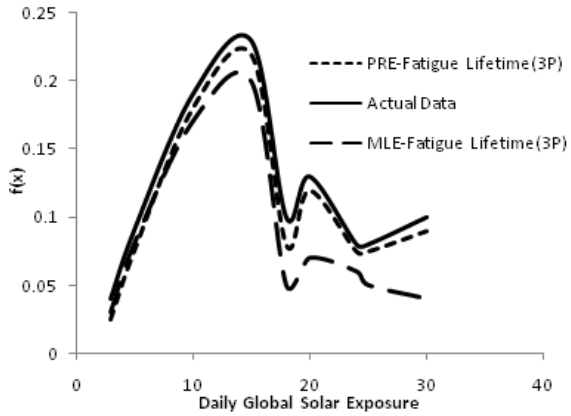

Fig. 12: Estimated Three-parameter Fatigue lifetime distribution

efficient of PRE method to cover the boundaries of the data. The plot gives the comparison between PRE and MLE method with the best-fit parameters for the given dataset.

The values of the estimated parameters of three-parameter Fréchet distribution by PRE and MLE, with the AD test statistic value are shown in Table 14.

Table 10: Estimated parameters (a) and (b) of Pert distribution using PRE

N	x	F(x)	a	b
1	1.2	0.00274	0.76317	36.353
2	1.5	0.005479	0.76038	36.3528
3	2.1	0.008219	0.76647	36.3537
4	3.6	0.010959	0.76088	36.3578
5	3.6	0.013699	0.7621	36.352
6	3.8	0.016438	0.7681	36.358
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
360	31.9	0.986301	0.76657	36.35647
361	31.9	0.989041	0.761056	36.3556
362	32	0.9917817	0.76307	36.35607
363	32	0.994521	0.765892	36.3592
364	32.1	0.99726	0.762784	36.3584
365	32.1	0.99999	0.76186	36.3085
Median			0.76631	36.3601

Table 11: Estimated parameters and AD test value of lognormal distribution using PRE and MLE.

Distribution	Parameter	PRE	Std.Error	MLE	Std.Error
LogNormal	μ	2.7955	0. 003464	3.3395	0.19198
	σ	0.4183	0.021548	0.2451	0.0479
	γ	-1.46977	0. 744878	-12.705	5.24189
	AD	1.5127		1.6	

The standard error of the estimated parameter for each method PRE and MLE were derived.

The fitting of Fréchet distribution was estimated by PRE and MLE are shown in Figure 14 The estimated distribution by PRE method seems more fitting to the dataset that estimated by MLE method. The estimated distribution by PRE covers the boundaries of the data set better.

Table 12: Results of estimated parameters and AD test of Fatigue distribution using PRE and MLE.

Distribution	Parameter	PRE	Std.Error	MLE	Std.Error
Fatigue	α	0.248	0.012962	0.24808	0.012985
	β	27.7888	1.455102	28.08	1.469774
	γ	-13.54	0.57087	-12.587	0.658834
	AD	2.328		4.4119	

Table 13: Estimated parameters of Erlang distribution by PRE and MLE with AD test

Distribution	Parameter	PRE	Std.Error	MLE	Std.Error
Erlang (3p)	m	5	1.475409	8	2.51038
	β	3.18	0.496094	2.5189	0.44963
	γ	0.3093	2.058859	-4.2532	2.81509
	AD	1.867		2.2176	
Erlang(2P)	m	5	0.52758	5	0.33916
	β	3.088	0.71205	3.2895	0.25981
	AD	2.939		3.5062	

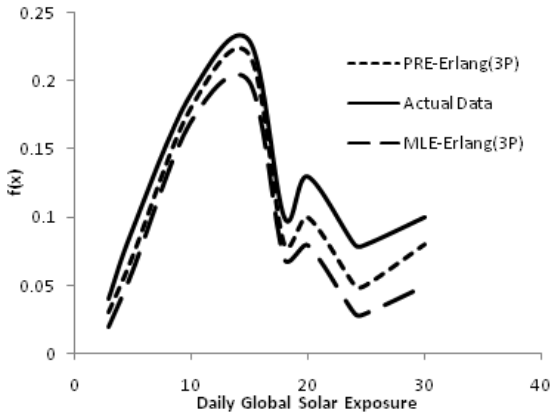

Fig. 13: Estimated Three-parameter Erlang distribution

The estimation of the parameters of Pert distribution shown in Table 15. AD test statistic values for PRE and MLE methods with the standard errors were also obtained and compared. The estimated results appear much similar from both PRE and MLE methods. The data are fitting plots of Pert distribution of estimated parameters by PRE and MLE method shown in Figure 15. Thus the estimation of Pert distribution by PRE is more fitting than MLE method.

Table 14: Results of estimated parameters and AD test of Fatigue distribution using PRE and MLE.

Distribution	Parameter	PRE	Std.Error	MLE	Std.Error
Fréchet	α	1.868	0.025725	2.0868	0.709228
	β	11.766	0.34963	9.7662	0.511186
	γ	1.6794	0.30103	1.0794	0.356498
	AD	5.38		27.892	

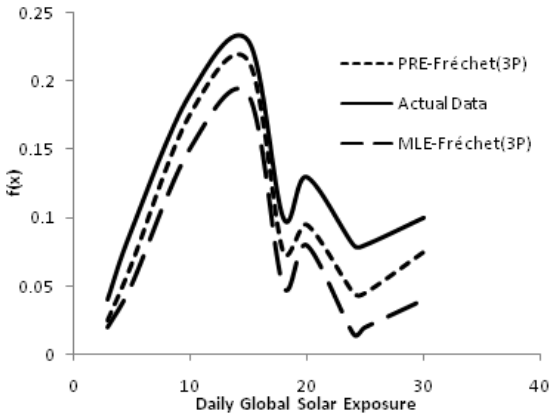

Fig. 14: Estimated Three-parameter the Fréchet distribution

Table 15: Estimated parameters and AD test values of the Pert distribution by PRE and MLE.

Distribution	Parameter	PRE	Std.Error	MLE	Std.Error
Pert	a	0.766	0.040047	0.55186	0.027145
	m	14.8	0.774667	14.863	0.777965
	b	36.36	1.902907	38.69	2.025127
	AD	3.689		3.856	

Consequently, PRE algorithm is more accurate than MLE method as shown in Table 16 whereby AD is lower in the case of PRE algorithm than MLE method. Thus PRE algorithm given more accurate estimations with the

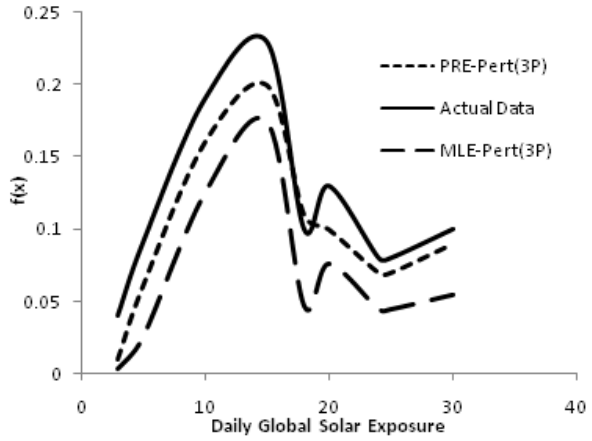

Fig. 15: Estimated Three-parameter Pert distribution

least estimation error for the three-parameters distributions as in the above-mentioned results. On the other hand, many researchers always found a complexity to obtain an accurate estimated three-parameters distributions using the traditional estimation methods like as: MLE, moments and Bayesian inference.

Significant statement: In this paper, three algorithms were suggested providing accurate results of the prediction of the daily global solar radiation. The point that is highly beneficial to the applications of environment and energy sciences. It also helps uncover the critical areas of the prediction and estimation methods. Now, researchers can work out a new statistical method of sustainable energy sources and possibly of other new distributions. Moreover, the results of this study enable us to better understand the causes of high temperatures in the atmosphere.

4 Conclusion

In this study, the forecasting approaches were implemented to predict the states of solar energy with the least estimation error using three proposed algorithms ACRM, ACDM and PRE. The first algorithm ACRM, basically recasts to control the prediction error by computing the change rates of solar energy during successive years. The ACRM algorithm converts the change rates to the regression polynomial models to predict the change rates of solar energy in the next years. The second algorithm ACDM reduces the standard error in estimating process by regenerate the regression polynomial equations for each month. The performance of the prediction algorithm could achieve interesting improvement in diagnostic and decision making processes. Converting the dataset to six-states, Markov model will be provided more accuracy results. Moreover, this study concluded that PRE algorithm is less

Table 16: Comparison AD values of estimated distributions using PRE and MLE.

Distribution	Parameter	PRE	MLE
LogNormal	μ	2.7955	0. 3.3395
	σ	0.4183	0.2451
	γ	-1.46977	-12.705
	AD	1.5127	1.6
Fatigue	α	0.248	0.24808
	β	27.7888	28.08
	γ	-13.54	-12.587
	AD	2.328	4.4119
Erlang (3p)	m	5	8
	β	3.18	2.5189
	γ	0.3093	-4.2532
	AD	1.867	2.2176
Erlang(2P)	m	5	5
	β	3.088	3.2895
	AD	2.939	3.5062
Fréchet	α	1.868	2.0868
	β	11.766	9.7662
	γ	1.6794	1.0794
	AD	5.38	27.892
Pert	a	0.766	0.55186
	m	14.8	14.863
	b	36.36	38.69
	AD	3.689	3.856

complicated than MLE method in coding and it also gives more accurate estimations. Availability of data and materials The data used to support the finding of this study are from Bureau of Meteorology, http://www.bom.gov.au/climate/data.

Competing Interests

The authors declare that they have no competing interests.

Funding

There are no sources of funding for this work.

Acknowledgement

It is a pleasure to thank Australian Bureau of Meteorology for providing online climate data of Queensland, Australia. In addition, I thank the contributing authors in this field of study. Furthermore, the authors gratefully acknowledge the reviewers for their valuable comments and recommendations on the original version of this paper, which resulted in this improved version.

References

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

[2] P. S. Balakrishnan, M. Shabbir, F. A. Siddiqi, 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.

[3] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun , K. H. Kim, Solar energy: Potential and future prospects, *Renewable and Sustainable Energy Reviews*, **82**, 894-900, (2018).

[4] A. Qazi, F. Hussain, N. A. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban and K. Haruna, Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions. *IEEE Access*, **7**, 63837-63851,2019.

[5] L.Bonkaney, I. SeidouSanda , A. A. Balogun, Wavelet analysis of daily energy demand and weather variables, *Journal of Energy*, 2019.

[6] K. Latif, M. Y. Raza, G. M. Chaudhary , A. Arshad, Analysis of Energy Crisis, Energy Security and Potential of Renewable Energy: Evidence from Pakistan, *Journal of Accounting and Finance in Emerging Economies*, **6(1)**, 167-182,2020.

[7] B. Awan , Z. A. Khan, Recent progress in renewable energy–Remedy of energy crisis in Pakistan, *Renewable and Sustainable Energy Reviews*, **33**, 236-253, (2014).

[8] W. Tang, K. Yang, J. He , J. Qin, Quality control and estimation of global solar radiation in China, *Solar Energy*, **84(3)**, 466-475, 2010.

[9] R. R. Kumar, P. J. Stauvermann, N. Loganathan, R. D. Kumar, Exploring the role of energy, trade and financial development in explaining economic growth in South Africa: A revisit, *Renewable and Sustainable Energy Reviews*, **52**, 1300-1311, 2015.

[10] M. Shahbaz, H. Mallick, M. K. Mahalik, P. Sadorsky, The role of globalization on the recent evolution of energy demand in India: Implications for sustainable development, *Energy Economics*, **55**, 52-68. doi:10.1016/j.eneco, 2016.

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

[12] L. Dewangan, S. N. Singh , S. Chakrabarti, Combining forecasts of day-ahead solar power, *Energy*, **202**, 117743, 2020.

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

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

[15] Ö. Ayvazoğluyüksel 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.

[16] N. Tosun, E. Sert, E. Ayaz, E. Yılmaz , M. Göl, *Solar Power Generation Analysis and Forecasting Real-World Data Using LSTM and Autoregressive CNN*, In 2020 International Conference on Smart Energy Systems and Technologies (SEST) (pp. 1-6), IEEE, 2020.

[17] P. Nystrup, E. Lindström , H. Madsen , Learning hidden Markov models with persistent states by penalizing jumps, *Expert Systems with Applications*, **150**, 113307, 2020.

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

[19] R. Levy, The rise of Markov chain Monte Carlo estimation for psychometric modelling, *Journal of Probability and Statistics*, 2009.

[20] H. Ahmad , N. Hayat , Modelling and prediction of primary energy supply and electricity generation structures based on Markov chain: an insight with focus on the role of natural gas in Pakistan, *Journal of the Chinese Institute of Engineers*, **44(2)**, 177-191, 2021.

[21] P. A. Gregory , L. J. Rikus , Validation of the Bureau of Meteorology's global, diffuse, and direct solar exposure forecasts using the ACCESS numerical weather prediction systems , *Journal of Applied Meteorology and Climatology*, **55(3)**, 595-619, 2016.

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

[23] F. Besharat, A. A. Dehghan , A. R. Faghih , Empirical models for estimating global solar radiation: A review and case study, *Renewable and Sustainable Energy Reviews*, **21**, 798-821.doi:10.1016/j.rser., 2013.

[24] E. Quansah, L. K. Amekudzi, K. Preko, J. Aryee, O. R. Boakye, D. Boli , M. R. Salifu , Empirical models for estimating global solar radiation over the Ashanti region of Ghana, *Journal of Solar Energy*, **6**, 897970, 2014.

[25] A. Das, J. K. Park , 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.

[26] K. Doost , 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.

[27] S. Bhardwaj, V. Sharma, S. Srivastava, O. S. Sastry, B. Bandyopadhyay, S. S. Chandel , J. R. P. Gupta , Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model , *Solar Energy*, **93**, 43-54, 2013.

[28] L. Zou, L. Wang, L. Xia, A. Lin, B. Hu , H. Zhu , Prediction and comparison of solar radiation using improved empirical models and Adaptive Neuro-Fuzzy Inference Systems, *Renewable energy*, **106**, 343-353, 2017.

[29] M. M. El Genidy , Statistical modelling of the daily global solar radiation in Queensland, Australia, *Songklanakarin Journal of Science Technology*, **41(6)**,doi:10.1080/03610926.2016.1167909, 2019.

[30] V. P. Jilkov , X. R. Li, Online Bayesian estimation of transition probabilities for Markovian jump systems, *IEEE Transactions on signal processing*, **52(6)**, 1620-1630, 2004.

[31] N. J. Welton , A. E. Ades, Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration, *Medical Decision Making*, **25(6)**, 633-645, 2005.

Mohammed El Genidy received the Ph.D. degree in statistics and computer science from the Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt, in 2001. He is currently an Assistant Professor of statistics with the Faculty of Science, Mathematics and Computer Science Department, Port Said University, Port Said, Egypt. His research interests include statistics, probability, order statistics, queues theory, regression, estimation, probability distributions, algorithms, mathematical programming, statistical tests, SPSS, Markov models, and hypotheses tests.

Wesal Megahed received the B.Sc. degree in statistics from the Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt in 2019. Her research interests include statistics, probability, queues theory, estimation, algorithms, Markov models, and probability distributions.

Khaled Mahfouz received the Ph.D. degree in mathematical statistics from the Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt in 2015. He is currently a lecturer of statistics with the Faculty of Science, Mathematics and Computer Science Department, Port Said University, Port Said, Egypt. His research interests include probability, statistics, statistical analysis and operations research.