University of Tehran Press

Journal of the Earth and Space Physics

Homepage: https://jesphys.ut.ac.ir

Vol. 49, No. 4, Winter 2024, P. 143- 159

(Research Article)

Regression Models and Hybrid Intelligent Systems for Estimating Clear-Sky Downward Longwave Radiation in Equatorial Africa

Obot, N. I.[1]

1. Department of Physics, Faculty of Science, University of Lagos, Akoka, Lagos, Nigeria.

Corresponding Author E-mail: nobot@unilag.edu.ng

(Received: 25 Dec 2022, Revised: 11 Jan 2023, Accepted: 2 May 2023, Published online: 20 Feb 2024)

Abstract

Modelling downward longwave radiation (DLR) in Equatorial Africa is challenging due to dense cloud cover and data scarcity. In this twofold study, daily cloudless DLR in Ilorin (8° 32′ N, 4° 34′ E), Nigeria, was modelled using two atmospheric factors, namely water vapour pressure and air temperature. Firstly, four cloudless DLR models were reformed and tested with others. Secondly, both particle swarm optimization (PSO) and genetic algorithm (GA) were deployed to optimize the adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANN). The statistical measures used to evaluate the performance of the models were the coefficient of determination (r^2), the mean bias error, and the mean square error (*MSE*). While restructuring clear skies DLR models typically reduces the estimation errors, it may not necessarily impact r^2 positively. The regression models have r^2 values ranging from approximately 0.82 to 0.87, while *MSE* lies between 56.6 W/m^2 and 767.5 W/m^2. There are instances where *MSE* drastically reduces from 692.6 to 72.3 (W/m^2) and from 767.5 to 66.2 (W/m^2) after restructuring two different models. A recently developed expression for the region remains the best, possibly because of its format. During the training phase of the computationally intelligent systems, r^2 approximately ranges between 88% and 92% but lies between 55% and 76% during testing. Although reproducibility inclusion in the code can meaningfully improve ANN systems at training, GA optimizes better than PSO. Furthermore, hybrid intelligent systems had higher r^2 values than standalone computationally intelligent modes at the testing phase. Due to the efficient generalization based on r^2 during the testing phase, ANN-GA is viable for modelling cloudless DLR at this site, though ANFIS has the lowest *MSE* at this same stage.

Keywords: Reformed modes, Particle swarm optimisation, Genetic algorithm, Adaptive neuro-fuzzy inference system, Artificial neural networks.

1. Introduction

The wavelength of downward longwave radiation (DLR) is approximately 4–100 μm, and it is received continuously at the Earth's surface with or without sunlight. Though shortwave radiation is the Earth's primary energy source, DLR is a vital energy flux responsible for a reasonably large amount of atmospheric heat energy in the Equatorial region and the nocturnal atmospheric heat at any location. Factors like water vapour pressure, relative humidity, cloudiness, air temperature, land cover, latitude, rainfall, wind direction and topography impact DLR. Beside water vapour, which is the major contributor, the radiation also depends on other atmospheric gases like O_3, CO_2 and aerosols. DLR can be obtained via three means: it comes directly from the sun; the atmosphere attenuates the more energetic shortwave radiation from the sun to the less active longwave radiation, and the atmosphere acts like a blackbody by absorbing and reradiating in the longwave band. Thus, transmission, absorption, scattering and reflection activities influence DLR in the atmosphere (Jimenez et al., 1987; Pluss & Ohmura, 1997).

Beside the downward radiations, we also have respective upward parts of solar radiations reflected from the ground. The two segments of both solar shortwave and longwave radiations control the energy balance of the Earth's surface, and the differences between the upward and downward radiative components give the respective net radiations. Unlike the energetic shortwave radiation balance that occurs while the sun shines, longwave radiation balance remains evaluable during daytime and night. DLR can be used for heating and cooling

Cite this article: Obot, N. I. (2024). Regression Models and Hybrid Intelligent Systems for Estimating Clear-Sky Downward Longwave Radiation in Equatorial Africa. *Journal of the Earth and Space Physics*, 49(4), 143- 159. DOI: http//doi.org/10.22059/jesphys.2023.352877.1007484

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/jesphys.2023.352877.1007484

Print ISSN: 2538-371X
Online ISSN: 2538-3906

applications since it has a thermal nature in addition to its uses for solar collectors, soil evapotranspiration, architecture engineering and agricultural applications (Jones et al., 2003; Scherba et al., 2011). Furthermore, DLR also affects atmospheric processes like the El Nino Southern Oscillation (ENSO), hydrology cycle and sea surface temperature (SST) (Jacobs et al., 2004; Pavlakis et al., 2007). The atmospheric layer above 500 m has a meagre contribution of approximately 16–20% to the Earth's surface DLR, in contrast to the lowest 10 m above the ground, which accounts for about 32–36%. However, the rest is from the space between 10 and 500 m. Under clear skies (i.e., when cloud impact is negligible), water vapour and ambient conditions are suitable predictors of DLR (Schmetz, 1989).

DLR can be measured directly or indirectly: pyrgeometers are used for direct measurements, and the indirect method involves subtracting shortwave solar radiation from global solar radiation. Radiometers for measuring shortwave and global solar radiations are the pyrheliometers and pyranometers, respectively. The primary deficiency of indirect measurement is the absence of data during the night periods because both shortwave and global solar radiations can be measured only when there is sunshine. Another problem with the indirect method is the compounding of errors from two instruments rather than one. A common source of error for radiometers is the daytime heating of their basements by shortwave radiation, which affects sensor sensitivity. Unlike pyrheliometers and pyranometers that work periodically, pyrgeometers are relatively fragile because they function all day (Kano et al., 1975; Udo, 2000).

Ground measurements of DLR are scarce because pyrgeometers are relatively expensive, require regular calibration, are delicate, and are mostly mounted and managed by experts. DLR can be modelled to alleviate data paucity using predictors such as cloud index, relative humidity, water vapour pressure, and air and dew temperatures. Since DLR variation depends on the prevailing atmospheric conditions of a given location, most regression models are specific to the sites used to derive them. Consequently, DLR models have to be localized or developed entirely for other places. Even though pyrgeometers are still being enhanced, it is also essential to improve the models for estimating the radiation (Alvarez & Garcia, 1990; Lhomme et al., 2007; Obot et al., 2019a).

There are several cloudless DLR expressions, though none seems superior to the others; therefore, more studies still have to be done for a comprehensive yet straightforward model. Restructuring elsewhere, DLR expressions may be unjustified if they do not fit suitably. Since only a handful of studies have been conducted in Africa and her Equatorial regions, techniques for deriving DLR expressions somewhere else may not apply to the peculiarities of tropical Africa. The behaviour of clouds around the equator differs from other areas; for instance, Equatorial clouds exhibit short longitudinal wave propagation and are very thick. There is a unique dual diurnal cycle associated with the Equatorial clouds due to the intertropical convergence zone (ITCZ). Furthermore, due to regional high cloud reflectance and proximity to the sun, direct incidence rays from the sun enormously energize the Equatorial atmosphere. Still, to sustain global energy equilibrium, atmospheric heat energy dissipates from this region to the cold places of the globe (Ramanathan et al., 1989). Equatorial Africa is also exceptional due to the frequently occurring wood-burning and dusty atmosphere (Anuforom et al., 2007; Bede-Ojimadu & Orisakwe, 2020).

There are two notable regression divisions for modelling DLR, namely theoretical and analytical (Brutsaert, 1975). While the theoretical techniques mainly involve using model development concepts, analytical methods use the data distribution nature. The accuracy of equations or any other procedure for data gaining is highly desirable. Moreover, the coefficients of DLR models need recalibration for better fit due to their inability to accurately interpret conditions of other climates (Alvarez & Garcia, 1990; Juszak & Pellicciotti, 2013). For example, Li et al. (2017) localized Brunt (1932) and recommended it for the USA due to its simplicity and accuracy. Meanwhile, Konzelmann et al. (1994) restructured Brutsaert (1975) for the Greenland Ice Sheet.

Apart from regression methods, machine learning techniques are occasionally used to

model solar radiation. Machine learning methods include neural networks (NN), support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS), differentiate evolution (DE), case reasoning, swamp intelligence (e.g., particle swamp optimization (PSO), artificial bee colony (ABC), spider optimization, and ant colony optimization for continue domain), imperialist competitive algorithm (ICA), genetic algorithm (GA), hybrids, etc. (Zhang et al., 2015; Azad et al., 2019). Though there are several studies on the regression modelling of DLR, except as retrieval methods from satellite or other airborne systems and for conditioning pyrgeometers, the use of machine learning methods for modelling DLR is rare (Wang et al., 2012; Obot et al., 2019a). Furthermore, hybrid intelligent systems like ANN-PSO, ANN-GA, ANFIS-GA, ANFIS-PSO and others should also be considered for modelling DLR given the growing interest in their utilization for diverse kinds of applications (Zhang et al., 2015; Azad et al., 2019; Shariati et al., 2019; Singh & Thongam, 2020). Though some of the same classes of hybrids have been compared, it is relatively rare across different intelligent systems. For instance, in the case of neural networks, likely because Koopialipoor et al. (2019) showed that ANN-PSO performed better than ANN-GA, ANN-ICA and ANN-ABC, Shariati et al. (2019) considered only the ANN-PSO system for the prediction of a mechanical engineering process. However, in another comparison of the aforementioned hybrid systems in an engineering study, ANN-GA was better than the others (Le et al., 2019). They did not compare, for example, ANN-GA with ANFIS-GA. Nonetheless, it is worth noting that there have been some comparisons of the hybrids of ANFIS, like ANFIS-PSO, ANFIS-GA and others in a handful of meteorological studies (Kamp & Savenije, 2006; Halabi et al., 2018).

Efficiencies of solar energy appliances are better guaranteed under clear skies than in a cloudy atmosphere, and daily mean values are less prone to errors than hourly (solar irradiance) values (Gueymard, 2003). Furthermore, verifications of atmospheric models and the correlation between climatic factors are easily understood and are more accurate under clear-sky than cloudy conditions (Marty & Philipona, 2000). Subsequently, the actual behaviour of atmospheric parameters is suitably studied under clear skies with average values.

So far, a handful of studies on DLR at Ilorin, Nigeria, have been made. For example, regression techniques have been used to model all skies and cloudless DLR using relative humidity, air temperature and water vapour pressure. Machine learning modes were also used to model the radiation, but no hybrid intelligent system has yet been deployed, and the regression models developed elsewhere have also not been localized (Udo, 2003; Obot et al., 2019a, 2019b). Given the global rarity of the ground-measured DLR, it is necessary to investigate the appropriate approach for alleviating the scarcity. This work aims to ascertain the need for calibrating cloudless DLR models beside using hybrid intelligent systems to model the radiation. Hence, the study comprises two sections: (1) localization of regression constants, and (2) comparison of ANFIS and ANN with hybrid intelligent systems such as ANFIS-PSO and ANN-GA.

2. Study Area

The study area of this work is Ilorin (8° 32′ N, 4° 34′ E), Kwara State, Nigeria. Information on the site and regional climate, instrument types, calibration factors, and measurement and data treatment of DLR from September 1992 to August 1994 have been previously reported (Miskolczi et al., 1997; Udo & Aro, 1999; Udo, 2003). Though DLR was also measured at other periods, the inclusion of additional data seemingly leads to low models' performance (Obot et al., 2019a; Obot, 2019); consequently, this study uses only the mentioned two-year data. Nevertheless, Ilorin is situated in the hinterland of the central region of Nigeria and has two major seasons, namely rainy and dry. However, the typical Harmattan season of the West African region, a subset of the dry season, also occurs in this vicinity.

As formerly reported (Obot et al., 2019a; Obot, 2019), the data were obtained from two different organizations. DLR was sourced from Baseline Surface Radiation Network (BSRN) at https://www.pangaea.de/PHP/BSRN_Status.php, while water vapour pressure and air

temperature records were obtained from the Nigerian Meteorological Agency (NIMET). The data from BRSN were recorded in intervals of three minutes, but they were averaged to daily values. A pyrgeometer used to measure DLR at the University of Ilorin, Nigeria, was an Eppley PIR 2046813 made in 1980 and recalibrated in 1988 at Newport Eppley Lab, USA, with a new constant of 3.92 $\mu VW^{-1}m^2$ from an initial value of 4.21 $\mu VW^{-1}m^2$. However, the BSRN station in Ilorin had no air temperature and water vapour pressure measurements, so they were sourced from NIMET. Data obtained from NIMET were taken on a three-hourly basis at the airport situated about 12 km from the campus (Figure 1). The air temperature was measured using a dry bulb thermometer, while the barometer was used to measure water vapour pressure at mean sea level. Accordingly, the average data for air temperature and water vapour pressure on clear days were mined from the archive of NIMET.

Only 82 out of 730 days had clear skies. Table 1 has the descriptive statistics of the data used in this work. Mean, median, min, max and mode values for DLR (W/m^2) are approximately 386, 395, 337, 415 and 398, respectively. The min and max values for water vapour pressure and air temperature are 8.4 millibar and 26.4 millibar, 297.3 K and 303.9 K, respectively. Like the range, the standard deviation is highest for DLR and least for temperature. Unlike the consistent performance of regression models on any given data, intelligent systems may produce different outcomes on segmented data depending on the computing technique. Conventionally, data for modelling with machine learning modes require partitioning into training and testing sets. It is also important to avoid overfitting associated with machine learning systems. Consequently, in this study, while there was no data partitioning for regression models, the approximate ratio of 7:3 was used for segmenting the data into the respective training and testing groups for the intelligent systems.

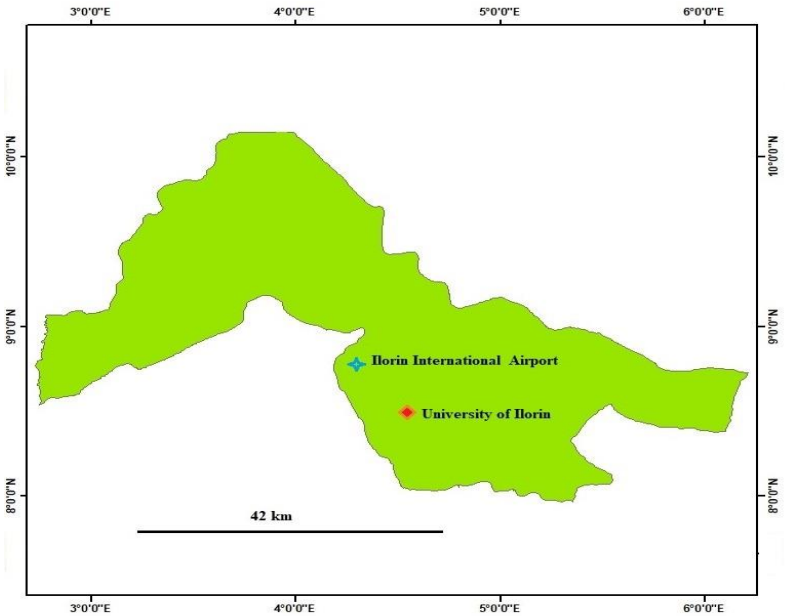

Figure 1. The study area of the university and airport in Ilorin, Kwara State, Nigeria.

Table 1. Descriptive statistics of the data used in this study.

Statistics	Water vapour pressure (mb)	Air temperature (K)	DLR (W/m^2)
Mean	22.1902	300.3220	386.9695
Median	25.02	300.3	394.7
Standard deviation	5.8696	1.4817	20.7882
Minimum	8.4	297.3	337.0
Maximum	26.4	303.9	414.7
Mode	25.8	300.0	397.9

3. Regression modelling

The model format for longwave radiation coming down to the Earth's surface under cloudless conditions can be given as;

$$L = \varepsilon\,(\mathrm{T},\mathrm{e})\sigma T^4 \quad (1)$$

where, ε is the sky emissivity, σ is the Stefan Boltzmann constant, and T and e are the air temperature (Kelvin) and water vapour pressure (millibar), respectively measured at screen level. In most cases, the defining factor for DLR expressions is the sky emissivity, which can be given as:

$$\varepsilon\,(\mathrm{T},\mathrm{e}) = \frac{L}{\sigma T^4} \quad (2)$$

While some models were tested in their original forms, others were localized to determine the impact of transformed coefficients. Moreover, the theoretical background for models was ignored while the experimental method was used for coefficients restructuring. For instance, the sky emissivity for Brutsaert (1975) is given as:

$$\varepsilon\,(\mathrm{T},\mathrm{e}) = A\left(\frac{e}{T}\right)^{1/7} \quad (3)$$

For an approximate standard atmosphere at sea level, the coefficient A represents the exponential temperature and pressure profiles' attenuation factor. The above expression has a theoretical background. However, in this study, using the experimental technique, the constant for the Brutsaert (1975) model was obtained from:

$$A = \frac{\overline{\varepsilon\,(\mathrm{T},\mathrm{e})}}{\overline{\left(\frac{e}{T}\right)^{1/7}}} \quad (4)$$

The bars in the above expression indicate the least square method.

Models whose sky emissivity is unrelated to water vapour pressure perform poorly in this region (Obot et al., 2019a). As such, Swinbank (1963) and Guest (1999) were overlooked, and models like Satterlund (1979) that accommodate water vapour pressure but do not yield high correlation coefficient values in this vicinity, possibly due to their formats, were also not considered. The examined models were:

$$Brunt\ (1932) = \left(0.605 + 0.048e^{1/2}\right)\sigma T^4 \quad (5)$$

$$Efimova\ (1961) = (0.746 + 0.0066e)\sigma T^4 \quad (6)$$

$$\text{Brutsaert}\ (1975) = 1.24(\mathrm{e/T})^{1/7}\ \sigma T^4 \quad (7)$$

$$\text{Idso}\ (1981) = [0.70 + 5.95 * 10^{-5}\mathrm{e}\ \exp(1500/\mathrm{T}^4)]\sigma T^4 \quad (8)$$

$$\text{Prata}\ (1996) = [1 - (((1 + 46.5(\mathrm{e/T})) * \exp(-(1.2 + 3 * 46.5 * \mathrm{e/T})^{1/2})]\,\sigma T^4 \quad (9)$$

$$\text{Iziomon et al.}\ (2003) = \left[1 - 0.35 * \exp\left(-10.0\,x\,\frac{e}{T}\right)\right]\sigma T^4 \quad (10)$$

$$\text{Obot et al.}\ (2019a) = (0.699 + \frac{(1.014\,x\,e\,x\,E30}{T^{13}})\sigma T^4 \quad (11)$$

Note: Other equations obtained after calibrations are given in the result section.

4. Computational intelligence models

4-1. Adaptive neuro-fuzzy inference system

The adaptive neuro-fuzzy inference system (ANFIS) amalgamates two modes, namely artificial neural network and fuzzy set theory. Despite the precision of fuzzy set theory, its convergence rate is comparatively slow, especially when variables are many. In contrast, the fast-converging neural network scheme is associated with undesirable imprecision. Therefore, Jang (1993) used the multilayer neural networks algorithm to implement the if-then rules in fuzzy set theory. ANFIS uses the well-known Sugeno–Takagi fuzzy procedures and local parameters, which can handle linguistic and numerical data, to map nonlinear input to output.

Fuzzy logic and artificial neural networks can be fused into a unit while retaining their fundamental properties. Sometimes, human reasoning can be vague, but rather than being only crisp (yes or no), its degree of certainty can be varied in-between truth and false or maximum and minimum. Without ambiguity, reasoning can be reduced to an appropriate interpretation using the procedures of Boolean algebra and probability found in mathematics. The solution takes cognisance of the consequence of the initial statement(s). In this case, a set can be defined as a space of real numbers or/and linguistic values. In set theory, the membership of an element in a given set is either yes or no, whereas fuzziness can be descriptive, comparative, perceptive, conceptual, transitional and characterizing in nature. The combination of set theory and fuzziness makes the fuzzy set theory (Zimmermann, 2001). A fuzzy set denotes a class of objects with information

on their respective degrees of membership (Zadeh, 1965).

Numerous methods can be used in characterizing a fuzzy set. For instance, Zadeh (1965) used addictive, associative and distributive operations to describe unions, containments, relations, interceptions and other fuzzy set properties. Furthermore, the inclusion of fuzzy or/and statements (logic operators) adds other descriptions such as axiomatic strength, adaptability, empirical fit and monotonicity (Zimmermann, 2001). Additionally, the logic operators give rise to rules used in conditioning the fuzzy set. The rules can be expressed in terms of premises and implications like; if x is A_1 and y is B_1, then z is C_1, as the first conditions, while the other conditions can also be expressed as: if x is A_2 and y is B_2, then z is C_2. A_1, A_2, B_1 and B_2 are the labels (otherwise regarded as premises or antecedents) of the inputs x and y, while C_1 and C_2 are the consequences or conclusions of their respective labels.

The uniqueness of a fuzzy system lies in its ability to handle uncertainties. Numerous connective techniques that produce shapes like trapeziums, triangles and other trajectories can solve the fuzzy system. Applying neural networks to the fuzzy system, to form the neural fuzzy inference system, improves the accuracy and speed of the solution. Data can be categorically labelled depending on their qualities, and when fuzzified, they can belong to two or more membership levels. A fuzzy set can be defined as:

$$A = \{(x, \mu_i(x)/\ x\ \in X)\} \tag{12}$$

$\mu_i(x)$ is the membership function; where 0 mapping means the element is not a member of the set, 1 indicates full membership, and values between 0 and 1 indicate the strength of the fuzzy membership. Furthermore, X is the universal collection consisting of x objects.

ANFIS operates under five stages: the fuzzy layer, product layer, normalisation layer, defuzzification layer and the output layer (Figure 2).

Layer 1: This is the fuzzification section where the adaptive sum of the inputs is taken to produce the membership grade of the fuzzy sets, which is afterwards forwarded to the next layer. The number of inputs to the system ranges from i = 1, 2, to n, and it is otherwise regarded as the nodes. Each node i is associated with an output of:

$$O_{1,i} = \mu_A(x) \text{ for } i = 1, 2, \ldots, n$$
$$O_{1,i} = \mu_B(y) \text{ for } i = 1, 2, \ldots, n \tag{13}$$

where, $\mu_A(x)$ and $\mu_B(y)$ are the membership functions for the fuzzy sets A and B, while x and y can be regarded as the linguistic values of their respective sets.

The fuzzification in the layer is achieved by taking the mean of the outputs of the nodes using the Gaussian function;

$$\mu_{Ai}(x) = 1/(1 + \left[\frac{x-ci}{ai}\right]^{bi}) \tag{14}$$

ai, bi, and ci are regarded as the precise parameters, and the changes in these parameters affect the linguistic values of Ai. During fuzzification, the maximum and minimum values of these parameters are 1 and 0, respectively.

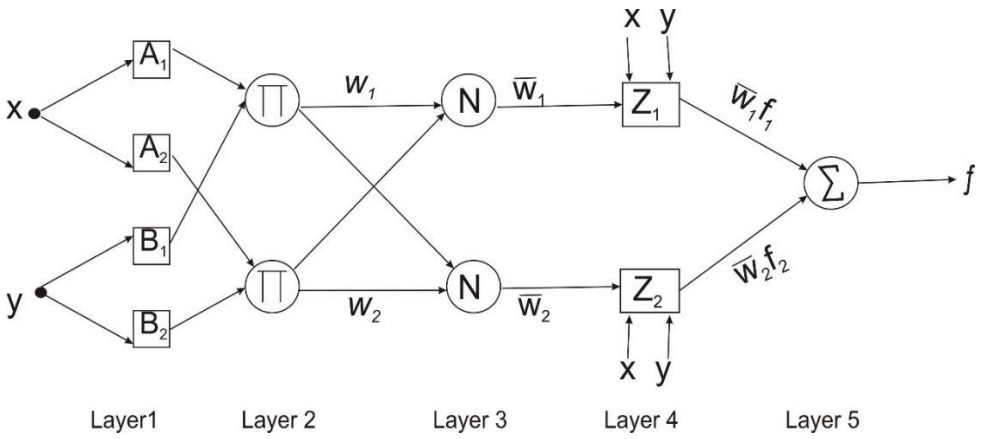

Figure 2. The five sections of ANFIS.

Layer 2: The product layer takes the multiplication of each node's incoming signals from the first layer to produce rules firing strength. It achieves that by applying the *min* (AND) operator to the linguistic value of each node. The multiplication operation is expressed as:

$$w_i = \mu_{Ai}(x) \times \mu_{Bi}(y) \text{ for } i = 1, 2, \ldots, n \quad (15)$$

Layer 3: This layer produces the normalized firing strength from the specific rules of the nodes. Normalization is achieved by determining the ratio of the firing strength of each node to the sum of the firing strengths of all rules. The normalization is given by:

$$\overline{w}_i = \frac{w_i}{w_1 + w_2} \text{ for } i = 1, 2, \ldots, n \quad (16)$$

Layer 4: Rules defuzzification or its inference in this layer is by paring the normalized firing strength from the former layer with the linear function of the system. This layer sums up all the results from each node of the previous layer and produces an output of:

$$\overline{w}_i f = \overline{w}_i(p_1 x + q_1 y + r_1) \quad (17)$$

p_1, q_1 and r_1 are known as the consequent parameters, and their relationship with the inputs can be expressed as: $p_1 x + q_1 y + r_1 = f$. The defuzzification can otherwise be represented by Z_1 and Z_2 (Figure 2).

Layer 5: The last section, which is the output layer, takes the algebraic sum of the inferences from all nodes as:

$$y = \overline{w}_1 f_1 + \overline{w}_2 f_2 + \ldots + \overline{w}_n f_n \quad (18)$$

The 'anfis' syntax, which represents the Sugeno-type fuzzy inference system, was used to model the adaptive neuro-fuzzy inference system in MATLAB. Its training and validation procedures combine the backpropagation gradient descent with the least square method. While fuzzy centres-means clustering of 2 was deployed, the maximum iteration was set to 100, and the membership function was tied to the number of input-output pairings during training. On the other hand, the 'evalfis' code was adopted to estimate the output data. The same ANFIS structure was used in hybridization instances with particle swarm optimization and genetic algorithm.

4-2. Artificial neural networks

Artificial neural networks or neural networks (ANN or NN) are algorithms for information processing in electronic devices like computers and robots that mimic the human nervous system. Like the synaptic modifications of electrical signals in biological neurons, excitation in ANN is via the adjustments of connecting weights. Initially, connecting weights are empty, but they acquire and store information during training and use such to perform future tasks. Instead of the linear connectivity of the Neumann computer, ANN is a black box unit that uses parallel connectivity between its components and functions by learning the patterns of any given data using a learning modality. Thus, ANN is an experience-based learner, flexible, error-tolerant, nonlinear and nonparametric. It is, therefore, uniquely suitable for modelling problems without well-defined mathematical procedures. The primary unit of neural networks is the node, which is similar to the biological neuron, though the brain consists of billions of neurons. Furthermore, the crude approximations of several brain activities like learning, recalling and decision-making have produced different neural networks like the perceptron, multiple-layer feedforward neural networks or multi-perceptron (which is arguably the most popular), Hopfield neural networks, and self-organization maps (Zhang, 2007; Aggarwal, 2018).

The first attempt to solve arithmetic problems using mathematical expressions for biological neurons was around 1943 (McCulloch & Pitts, 1943). Immense contributions from numerous scientists followed, but progress ceased when the inability of the perceptron to solve nonlinear separable problems emerged (Hebb, 1949; Rosenblatt, 1958; Widrow & Hoff, 1960). Although multiple-perceptron, as a solution, was suggested around 1969, it was not realisable then (Minsky & Papert, 1969). Moreover, the backpropagation method was also proposed in the 1970s. However, not much was achieved until the 1980s when statistical mechanics, self-organization mapping, and backpropagation techniques were separately used to overcome the issue (Werbos, 1974; Hopfield, 1982; Kohonen, 1982; Rumelhart et al., 1986). Afterwards, the field witnessed notable developments (Zhang, 2007; Aggarwal, 2018).

In contrast to the perceptron, which has a

single computational layer consisting of input-to-output mapping, the multilayer feedforward neural networks (MFNN) comprise three computational sections, namely the input layer, hidden layer and output layer. The system is known as the feedforward neural network because it processes information unidirectionally like the perceptron (Figure 3). However, the hidden layer may consist of more than one section. Connecting weights between layers in MFNN are arranged in matrix formats contrary to the linear arrangements in a perceptron.

The perceptron inputs are usually mapped using binary notations to link the weights to the output, and they may be associated with a bias. For example, the output $y \in \{-1, +1\}$ associated with inputs such as $x_1, x_2, x_3, \ldots, x_n, x \in X$ in a d-dimensional space during training, can be transformed to:

$$\dot{y} = f\left\{\sum_{i=1}^{d} w_i x_i + b\right\} = f\left(\overline{W} \cdot \bar{X} + b\right) \quad (19)$$

where f is the activation function, w_i is the weight connection of each i^{th} node, x_i is the input of the i^{th} node, b is the bias, which is equal to zero in some instances, and $\overline{W} \cdot \bar{X}$ is the linear multiplier.

The error, e existing between the predicted values, $\dot{y}$ and the observed values, y can be given as; $e = y - \dot{y}$, though any loss function like the goodness of fit or gradient descent can be used for system performance assessment and to update connecting weights. Furthermore, several activation functions such as linear, sigmoid (or logistic), sine and cosine, and tan can be applied to the system. ANN with a hidden layer combines the outputs of input and hidden layers for its final result. While the learning process for updating connecting weights uses several well-known algorithms, the popular backpropagation algorithm uses an error function known as quadratic error or gradient loss to improve node performance.

On the one hand, Hopefield neural networks (HNN) may fall under recurrent networks. They have no hidden layers but consist of a section with both forward and backward interconnectivity between the nodes. Recurrent networks like the recurrent temporal restricted Boltzmann machines and layer recurrent neural networks have at least one feedback loop between nodes/layers or an output back to an input section. Nodes in HNN commonly have a triangular arrangement, and the system is unique because it is akin to human memory. Accordingly, HNN is suitable for classification, reasoning and identification problems.

On the other hand, the self-organising map (SOM) uses unsupervised training (where no output information is provided to the system) and is rather suitable for cluttering and dimension problems. SOM has no hidden layer but consists of input and output layers. It uses competitive learning and adaptive resonance theory techniques that require only input data to update connecting weights. Altogether, the algorithms for artificial neural networks regularize and avoid overfitting to an extent. Irrespective of its structure or mechanism, ANN can be used for numerous applications, ranging from machines (e.g., robotics) and data mining to financial and educational purposes. In this work, 10 nodes were in the hidden layer, two nodes were in the input layer, and a node was in the output layer of the feedforward neural network. The ANN structure had fixed connecting weights and was also used in the hybridisations with particle swarm optimization and genetic algorithm. The systems were implemented with codes in MATLAB.

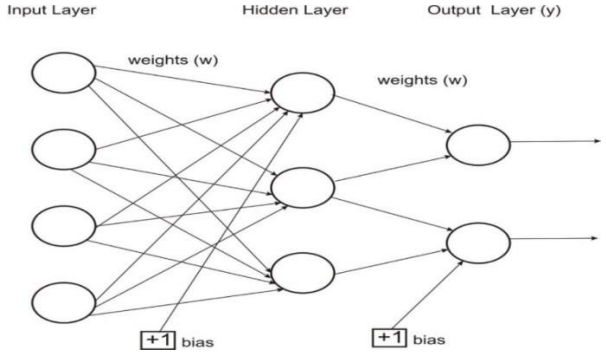

Figure 3. Schematic diagram of the feedforward neural network with one hidden layer.

4-3. Particle swarm optimization

Particle swarm optimization (PSO) comprises swarm intelligence algorithms for computers and robots tailored after the biological behaviours in decentralized or self-organized systems. PSO is conditioned after food search, nest building, unit protection, or any other conduct of swarm and flocking animals like birds, fishes, microorganisms (e.g., bacteria) and insects (e.g., bees and ants). Over the years, several evolutionary intelligence systems of animals-in-group, namely ant colony optimization, bee algorithm, bat algorithm, bacterial foraging optimization, ant system, monkey search, grey wolf optimizer, firefly optimization, etc., have been developed. The essence of swarming is to achieve a specific goal through group interaction, where members share and process information with collective responsibilities while sacrificing any contradicting self-desire. Generally, the goal is unattainable but beneficial to an individual (Krause et al., 2010). Even though swarm intelligence had been in use in applications like robotics, it was made famous by Kennedy and Eberhart (1995) from the inspiration of flocking birds, schooling fishes and human social lives.

In principle, PSO works by randomly triggering some population of particles (otherwise known as agents or candidates) to search around a given region while being governed by a pre-set mathematical function. Optimization in mathematics implies finding the maxima or minima of a given function subject to some constraints, or seeking and using the easiest route to solve problems. In a d-dimensional space, each particle adjusts to a new position, preferably the best using its experience and those of other particles. All particles follow basic rules like flocking together, avoiding collision, matching the group velocity, and knowing the current position, beside remembering the local best spot and the global best place while searching their domain.

For n-particles in the region, the position and velocity of each particle can be expressed by: $Y_i = Y_{i1}, Y_{i2}, \ldots, Y_{id}$ and $V_i = V_{i1}, V_{i2}, \ldots, V_{id}$, for $i = 1, 2, 3, \ldots, n$. The respectively updated new velocity and position for any particle during each iteration is given as:

$$V_{id}(t+1) = wV_{id}(t) + C_1 \cdot rand_1 \cdot \left(pbest_{id} - Y_{id}(t)\right) + C_2 \cdot rand_2 \cdot \left(gbest_{id} - Y_{id}(t)\right) \quad (20)$$

and

$$Y_{id}(t+1) = Y_{id}(t) + V_{id}(t+1) \quad (21)$$

where w is the inertia weight, C_1 and C_2 are the acceleration coefficients that control personal and group properties, $pbest_{id}$ is the best experience of the i^{th} particle, $gbest_{id}$ is the best experience of the group, and $rand_1$ and $rand_2$ are uniform random variables that change with each iteration.

Furthermore, the inertia weight can be expressed as follows:

$$w = w_{max} - \left(\frac{iter}{iter_{max}}\right)(w_{max} - w_{min}) \quad (22)$$

where w_{max} is the maximum inertia weight, $iter$ is the current iteration number, $iter_{max}$ is the maximum iteration number, and w_{min} is the minimum iteration.

ANFIS and ANN optimization fitness function was the mean square error, defined in Equation (26). Furthermore, boundary limits between -1.5 and 1.5 were used in the respective optimization processes with ANFIS and ANN (Heris, 2015; Alam, 2016). While optimizing ANFIS, the initial weight for PSO was at unity, the weight damping ratio was 0.99, and the personal learning and global learning coefficients were 1 and 2, respectively. Velocity update for the particles was linked with a population size of 25 and varied between the pre-set minimum and maximum limits. However, a population of 10 and a tolerance interval between 1 and 10^{-8} were used for the ANN-PSO system, though both optimization instances used a maximum iteration of 1000.

4-4. Genetic algorithm

Genetic algorithm (GA) is another popular metaheuristic method like PSO that can be used to optimize intelligent systems like ANFIS and ANN. GA is an evolutionary algorithm that mimics the development stages of living organisms, involving the manoeuvring of DNA (deoxyribonucleic acid) from the formation stage to the fully grown phase. It is therefore based on natural selection and genetic procedures, where living organizms have to overcome hostile forces of nature to survive and reproduce (Fraser, 1957; Holland, 1975). Instead of deploying sophisticated mathematical techniques to optimize, GA simplifies the problem by approximating the solutions

using encoding and domain mapping of chromosome-type data.

For many solutions such as $x_1, x_2, \ldots, x_n$ to a problem, where $x_i \in X$ for $i = 1, 2, \ldots, n$, and $F: X \to R$, the maximal (or minimal) solution can be sought in the form of: $F(x_i) = \max_{x_o \in X} f(x)$. In this case, F is the assigned fitness function and the variables $x_1, x_2, \ldots, x_n$ are encoded in binary or any other style like ternary, alphabet, integer or quaternary. Precisely, the evolutionary solution is given as:

$$F(x_i) = \frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \quad (23)$$

$F(x_i)$ becomes the individual fitness, n is the population size, x_i represents the phenotypic value of an i^{th} individual, while $f(x_i)$ is the raw performance of the specified individual.

GA has three main mechanisms, namely selection, crossover and mutation. Selection occurs when, against all the odds, the best chromosomes favourably mate and reproduce while overlooking others. Survived healthy genes, regarded as the fittest, fuse with their unique attributes such that their offspring inherit similar traits. The most crucial aspect of GA is the selection, which involves picking the best parents, and it uses most of the computation powers. Methods like the Roulette Wheel and Stochastic Universal Sampling techniques can be used for coding selection and related features like representation and repairing infeasible individuals from sum operation, single-point, minimum spread and zero bias procedures. However, fitness function coding is a prerequisite for selection. Probability sampling through the fitness ranking of individuals produces the desired population for the next phase. Randomly picked individuals continue to increase their fitness values to the maximum limits.

Crossover or recombination is an evolutional mechanism that produces new chromosomes from randomly selected parents that breed. It is achievable through several crossover methods, such as one-point, two-point, multi-point, ordered, uniform, shift, partially matched, precedence preservative and reduced surrogate. Primarily, the new chromosomes consist of genes or bits exchanged from spitted phenotypes of any distinctive individuals. Recombination rules initiation acquire specific genes for any given offspring, although a new gene contains variations of the initial genes.

Moreover, mutation sustains parent genes in their progeny variants, which helps GA avoid local minimum by ensuring previous behaviours are unlost in offerings due to crossover procedures over time. Whereas crossover combines non-repetitive patterns from two parents to constitute a new population with entirely new chromosomes, mutation reintroduces an ancient phenotype of a single parent to offspring for diversity in the new generation. Mutation techniques like the inversion operator, flipping operator, reshuffling operator, etc., are available in the GA computer package. Usually, the stopping criteria for ending these procedures, which evaluate the new generation with desired limits, are defined in the programmes. Otherwise, if the end is unreached, the candidates return to an initial stage for reconsideration (Hınçal et al., 2011; Mirjalili, 2019).

The Roulette Wheel and tournament selection algorithms in GA were handy in the respective optimizations of ANFIS and ANN. In the case of ANFIS-GA, the coding comprised a population size of 25, 70% mutation occurrence at a 15% mutation rate, and a gamma of 0.7. However, the optimization of ANN with GA involved a population size of 200, generation of 50, and stall limit of 100. Additionally, the feasible mutation function with tolerance function and fitness limit of 10^{-10} and 10^{-5} were adopted in coding ANN-GA.

The statistical measures used for assessing the performance of the models were the coefficient of determination (r^2), mean bias error (MBE), and mean square error (MSE). In this case, the coefficient of determination indicates the extent to which the cloudless DLR, as estimated by the individual models, deviate from the ground-measured values. It also points to the linear relationship between the estimated and the measured radiation. The coefficient of determination typically ranges from a minimum of 0.0 to a maximum of 1.0, but values of 70% and above are desirable. Furthermore, the error terms indicate the degree of dispersion between the measured and the calculated values, though the mean square error also served as the fitness function. The lower the estimation errors, the better the model. The expressions

for the statistical indexes are:

$$r^2 = \frac{n(\sum R_m R_c) - (\sum R_m)(\sum R_c)}{\sqrt{[n\sum {R_m}^2 - (\sum R_m)^2][n\sum {R_c}^2 - (\sum R_c)^2]}} \quad (24)$$

where R_m is the measured daily cloudless downward longwave radiation, R_c is calculated daily cloudless downward longwave radiation, and *n* is the total amount of data.

$$MBE = \frac{\sum(R_m - R_c)}{n} \quad (25)$$

and

$$MSE = \frac{\sum(R_m - R_c)^2}{n} \quad (26)$$

5. Results

5-1. Model restructuring impact

The calibration of cloudless DLR models such as Brunt (1932), Efimova (1961), Brutsaert (1975) and Idso (1981), respectively, using the data of Ilorin, gives the following expressions:

$$\text{Brunt (1932) } (2nd) = \left(0.593 + 0.053e^{1/2}\right)\sigma T^4 \quad (27)$$

$$\text{Efimova (1961) } (2nd) = (0.697 + 0.0064e)\sigma T^4 \quad (28)$$

$$\text{Brutsaert (1975) } (2nd) = 1.22(\text{e/T})^{1/7}\ \sigma T^4 \quad (29)$$

$$\text{Brutsaert (1975) } (3rd) = \left(0.0477 + 1.155(\text{e/T})^{1/7}\right)\sigma T^4 \quad (30)$$

$$\text{Idso (1981) (2nd)} = [0.69 + 4.41 * 10^{-5}\text{e exp}(1500/\text{T}^4)]\sigma T^4 \quad (31)$$

As expected, the constants of Equations (27) to (31) are different from those of the original expressions due to differences between the climate conditions of Ilorin and the other locations used to derive the previous models. Seemingly, both theoretical and analytical techniques for obtaining regression coefficients produce the same outcome. Regarding Equation (7), there is arguably a good fit between the theoretical approach and the experimental method in Ilorin. The first factor of Equation (7) given by Brutsaert (1975) is 1.24. In contrast, Lhomme et al. (2007) also used the theoretical technique and got 1.18, while in the study, the analytical method was used, and 1.22 was obtained (Equation 29).

An extension of the Brutsaert (1975) model from one to two constants, as suggested by Konzelmann et al. (1994), does not fit better than the one-constant model regarding the coefficient of determination, though the estimation errors reduce. The reformed one-constant Brutsaert (1975) model (Equation 29) had both mean bias error and mean square error values of 1.2155 W/m^2 and 80.3524 W/m^2, whereas the same error terms for the restructured two-constants model (Equation 30) were -0.2365 W/m^2 and 76.3620 W/m^2. Despite the reduced errors, the coefficient of determination value of 0.8281 for Equation (29) was higher than that for Equation (30), which was 0.8222. Nonetheless, it should be noted that Equation (11), with the highest coefficient of determination value and corresponding lowest mean square error, remains the best regression model for the region, notwithstanding the localization of several other expressions (Table 2). Apart from Equation (11), both the original and localized Idso (1981) expressions (Equations 8 and 31) were better than the others vis-à-vis the coefficient of determination. However, the localized Efimova (1961) (i.e., Equation 28) had the lowest mean bias error.

Except in rare instances, the localized equations had remarkable enhancements over the original cloudless DLR models. For example, the respective values of r^2, MBE and MSE for the original Efimova (1961) expression (Equation 6) were 0.8332, -24.8185 W/m^2 and 692.6361 W/m^2, but for the restructured model (Equation 28), those values were 0.8554, -0.1624 W/m^2 and 72.3364 W/m^2, respectively. Noticeably, there were meaningful error depreciations in the restructured Efimova (1961) model. However, the comparison of a reformed Brutsaert (1975) model with the original version (i.e., Equation 30 and Equation 7) shows that even though the MBE and MSE values improved from -5.1084 W/m^2 and 106.7184 W/m^2 to 0.2365 W/m^2 and 76.3620 W/m^2, respectively, r^2 depreciated slightly from 0.8281 to 0.8222. Similarly, while the coefficient of determination reduced from 0.8665 to 0.8562 after Idso (1981) was reformed, the mean bias error and mean square error improved respectively from -25.8249 W/m^2 to 2.0504 W/m^2, and from 767.4839 W/m^2 to 66.1975 W/m^2. Aside from the occasional depreciation in the coefficient of determination due to restructuring, it could also remain unchanged: the r^2 value of 0.8281 was the same for Equations (7) and (30) of the one-

constant Brutsaert models (Table 2). Thus, estimation errors generally reduce after a regression model is localized.

5-2. Intelligent systems performance

The performances of the soft computing systems during the training and the testing phases slightly differ. Probably due to the small quantity of data, lower coefficients of determination and higher estimation errors were often associated with the testing phase. Otherwise, the results reflect the capability of intelligent systems to generalize the input-to-output data pattern. Table 3 indicates that the performances of computationally intelligent modes are comparably close. Nevertheless, the hybrids are somewhat better than the standalone modes when based solely on the coefficient of determination. For example, during testing, the r^2 values for ANFIS, ANFIS-GA and ANFIS-PSO were 54.91%, 56.87% and 56.09%, respectively, while they were 45.90%, 75.83% and 55.23% for ANN, ANN-GA and ANN-PSO, respectively. The outstanding result of the hybrid intelligent systems could be due to their long implementation time occasioned by the procedures (i.e., the respective reproduction and search activities) of GA and PSO. Because ANN-GA did generalize the most, it is considered better than the other three hybrid intelligent systems (Table 3).

Table 2. The result of estimating cloudless DLR with regression models at Ilorin, Nigeria, where n = 82.

Model	Expression	r^2	MBE (W/m^2)	MSE (W/m^2)
Brunt (1932)	Equation (5)	0.8168	4.6308	99.6423
Localized Brunt (1932)	Equation (27)	0.8269	-0.5882	76.3678
Efimova (1961)	Equation (6)	0.8332	-24.8185	692.6361
Localized Efimova (1961)	Equation (28)	0.8354	-0.1642	72.3364
Brutsaert (1975)	Equation (7)	0.8281	-5.1084	106.7184
Localized Brutsaert (1975)	Equation (29)	0.8281	1.2155	80.3524
Extended and localized Brutsaert (1975)	Equation (30)	0.8222	-0.2365	76.3620
Idso (1981)	Equation (8)	0.8665	-25.8249	767.4839
Localized Idso (1981)	Equation (31)	0.8562	2.0504	66.1975
Prata (1996)	Equation (9)	0.8179	-3.9193	93.2345
Iziomon et al. (2003)	Equation (10)	0.8318	4.2859	91.9234
Obot et al. (2019a)	Equation (11)	0.8691	0.4443	56.6126

Table 3. The result of estimating cloudless DLR with soft computing systems at Ilorin, Nigeria.

Training stage (n = 58)				Testing stage (n = 24)		
System	r^2	MBE (W/m^2)	MSE (W/m^2)	r^2	MBE (W/m^2)	MSE (W/m^2)
ANFIS	0.8854	$1.41 * 10^{-13}$	56.9933	0.5491	-2.4372	57.2881
ANN	0.9160	0.1471	42.0165	0.4590	-6.9822	134.8013
ANFIS-GA	0.9132	0.5828	43.7016	0.5687	-3.3594	59.9439
ANFIS-PSO	0.9033	0.2158	48.2379	0.5609	-3.2097	64.6735
ANN-GA	0.9103	0.6365	45.9011	0.7583	-4.1418	66.3895
ANN-PSO	0.9099	0.9919	46.1288	0.5523	-2.3512	58.0546

Still, ANN and ANFIS are viable due to their short execution time and competitive performances. Whereas restructuring regression models always leads to estimation error reductions, such is uncertain in the optimizations of ANFIS and ANN with GA and PSO. During the training phase, the MBE and MSE values for ANFIS were lower than those for ANFIS-GA and ANFIS-PSO. The same scenario of higher estimation errors played out between ANN and its hybridization with GA and PSO during the same phase. In addition, there were other instances where ANN and ANFIS performed better than the hybrids. For example, the coefficient of determination for ANN was highest, while the mean bias error of ANFIS was lowest during the training phase. Instead of using the backpropagation method to update connecting weights, learning in ANN is enhanced using the genetic algorithm or particle swarm optimization in either ANN-GA or ANN-PSO, respectively.

ANN-GA is better than ANN-PSO in all instances during the training phase, but the opposite is the situation regarding the error terms at the testing stage. Such mixed outcomes also occurred between ANFIS-GA and ANFIS-PSO. For illustration, in the training section, the respective values of r^2, MBE, and MSE for ANN-GA were 0.9103, 0.6365 W/m^2 and 45.9011 W/m^2, but 0.9099, 0.9919 W/m^2 and 46.1288 W/m^2 for ANN-PSO, in which case ANN-GA is always better than ANN-PSO. However, during the testing phase, the r^2, MBE and MSE values for ANN-GA were 0.7583, -4.1418 W/m^2 and 66.3895 W/m^2, respectively, while for ANN-PSO, they were 0.5523, -2.3512 W/m^2 and 58.0546 W/m^2, respectively. Thus, the coefficient of determination for ANN-GA outrightly surpasses that for ANN-PSO at the training phase, but the estimation errors for ANN-PSO are better than those for ANN-GA.

During training, the coefficient of determination for ANFIS-GA (0.9133) was better than that for ANFIS-PSO (0.9033); however, the reserve was the situation regarding the respective mean bias error values of 0.5828 W/m^2 and 0.2158 W/m^2. The same scenario of mixed results between ANFIS-GA and ANFIS-PSO continued in the testing phase. From the above analysis, GA performs better than PSO in terms of the overall optimization of ANN and ANFIS. Furthermore, a slight overestimation of the measured data occurred during the training phase, though the opposite arose during testing. Although not shown, the inclusion of the default reproductive code improved the performance of stochastic systems like ANN and its hybrids. Stochastic implies having different results each time the programme runs due to diverse connecting weights or candidates' usage at one point or the other. In contrast, such systems transform when their synaptic triggering or individuals remain fixed. Reproducibility inclusion enabled ANN to have the highest coefficient of determination during the training phase.

6. Discussion

As earlier clarified, owing to data quality and scarcity, only the duration from September 1992 to August 1994 was considered in this work. Due to similarities in atmospheric conditions, the localization of an elsewhere model could be suitable for another region, else such parametrization may fail altogether. Since the coefficients of regression models change according to atmospheric and topological peculiarities from one site to another, several constants can exist for a particular cloudless DRL model (Iziomon et al., 2003). Though some techniques for acquiring the regression models with their coefficients are somewhat complicated, it is needless to adopt relatively complex and lengthy regression techniques where the easy methods have similar or even better accuracies. Most DLR models were developed in temperate regions, so they may not suit the Equatorial provinces, yet Equation (11), developed for Ilorin, Nigeria, is ideal for the tropics (Obot et al., 2019a).

Though water vapour is a minor atmospheric gas compared to others like nitrogen and oxygen, it plays a vital role in the weather system aside from being the major DLR gas. The gas impacts reasonably on rain, snow and ice, in addition to having a direct correlation with air temperature, though it exhibits a low mixing ratio and high spatiotemporal variability. Unlike the polar sections, tropical regions have distinguishable dependence on water vapour due to hot regional temperatures arising from their proximity to the sun. The proper relationship between atmospheric factors like

ambient water vapour and temperature should certainly occur for accurate parameterisation of DLR. Though not deemed the best cloudless DLR model, the reformed Efimova (1961) model, which has the lowest mean bias error, to a level depicts the high and right dependence of water vapour on DLR around the tropics (Obot et al., 2019b). Thus, there is room for more research to find a possibly superior cloudless DLR model in the future due to the failure of Obot et al. (2019a) not being an all-around best model.

Recalibrating clear skies and cloudy DLR models reduces estimation errors, while an existing model may apply to other locations having similar climatic conditions under cloudless conditions (Bilbao & De Miguel, 2007). Despite some variations in the atmospheric compositions and mechanisms, such as different cloud behaviours from one area to another, some similarities exist between global locations. A comprehensive model should therefore require varying the meteorological factors instead of the coefficients. Consequently, it could be essential to account for latitudinal differences in an all-inclusive cloudless DLR regression model.

On the contrary, it is crucial to depend on more effective intelligent systems for modelling, but the major disadvantage is the relative difficulty in their usage. Since most of them are being recently developed, they operate more with codes, which everyone except the experts cannot readily deploy. Additionally, GA and PSO need a fitness function to optimise another system. For example, the mean square error or any other error statistic is used as a fitness function to optimise modes like ANFIS and ANN. However, the choice of fitness function affects the performance of the hybrids in addition to the type of codes. Subsequently, the systems depend, to some extent, on coding competence. In this study, the difference between soft computing systems was minute, and when time is not an issue, the hybrids are likely preferable to native or standalone modes. Halabi et al. (2018) found that ANFIS-PSO was slightly better than ANFIS-GA, ANFIS-DE and ANFIS in all cases of estimating global solar irradiance in Malaysia. However, there were mixed results in this study, although ANN-GA was the best. ANN-GA had the best results primarily because it generalized the most during testing based on the coefficient of determination. Besides, contrasting with Koopialipoor et al. (2019), this study agrees with Le et al. (2019) that ANN-GA is superior to ANN-PSO.

7. Conclusion

Apart from assessing the impacts of restructuring cloudless DLR regression models at Ilorin (8° 32′ N, 4° 34′ E), Kwara State, Nigeria, computationally intelligent systems like ANFIS, ANN, ANFIS-GA, ANFIS-PSO, ANN-GA and ANN-PSO implemented with MATLAB were also used to estimate the radiation. Estimation errors generally reduce after restructuring, and GA is a better optimizer than PSO. After the reformation of Efimova (1961), the MSE of the model reduced from approximately 693 W/m^2 to 72 W/m^2, while r^2 increased slightly from 83% to 84%. However, whereas MSE improved from around 767 to 66 (W/m^2), r^2 depreciated from 87 to 86 (%) after reforming Idso (1980). Though restructuring somewhat improves performance, a model developed with data from the site remains the best for this region, possibly due to its format (Obot et al, 2019a). Ensuring computationally intelligent systems' reproducibility during coding improves their performance, but optimizing ANFIS and ANN with PSO and GA does not always guarantee better results, particularly in the training segment. While it usually takes a relatively long time to execute hybrid intelligent systems due to complicated computing procedures, ANN-GA had a high correlation between the measured and estimated radiation during the testing phase and is supposedly the best. Except for ANN-GA, no other computationally intelligent mode had r^2 of up to 70% and above during testing.

Acknowledgments

The author is really grateful to the Department of Physics, University of Ilorin, Baseline Surface Radiation Network, and Nigerian Meteorological Agency for the data used in this research.

References

Aggarwal, C. C. (2018). Neural networks and deep learning. *Springer*, 10(978), 3.

Alam, M. N. (2016). Codes in MATLAB for training artificial neural networks using the particle swarm optimization. *Research Gate*, 1-14.

Alvarez, C. C., & Garcia, C. G. (1990). Calculation and measurements of infrared radiation at the surface. *Atmosfera*, 3, 291–304.

Anuforom, A. C., Akeh, L. E., Okeke, P. N., & Opara, F. E. (2007). Inter-annual variability and long-term trend of UV-absorbing aerosols during Harmattan season in sub-Saharan West Africa. *Atmos. Environ.*, 41, 1550–1559.

Azad, A., Manoochehri, M., Kashi, H., Farzin, S., Karami, H., Nourani, V., & Shiri, J. (2019). Comparative evaluation of intelligence algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modeling. *J. Hydrol.*, 571, 214–224.

Bede-Ojimadu, O., & Orisakwe, O. E. (2020). Exposure to wood smoke and associated health effects in sub-Saharan Africa: a systematic review. *Ann. Glob. Health*, 86, 1–27.

Bilbao, J., & De Miguel, A. H. (2007). Estimation of daylight downward longwave atmospheric irradiance under clear sky and all-sky conditions. *J. Appl. Meteorol. Climatol.*, 46, 878–889.

Brunt, D. (1932). Notes on radiations in the atmosphere. *Q. J. R. Meteorol. Soc.*, 58, 389–420.

Brutsaert, W. (1975). On a derivable formula for long-wave radiation from clear skies. *Water Resour. Res.*, 11, 742–744.

Efimova, N. A. (1961). On methods of calculating monthly values of net longwave radiation. *Meteorol. Gidrol.*, 10, 28–33.

Fraser, A. S. (1957). Simulation of genetic systems by automatic digital computers. II: Effects of linkage on rates under selection. *Austral. J. Biol. Sci.*, 10, 492–499.

Guest, P. B. (1999). Surface longwave radiation conditions in the eastern Weddell Sea during winter. *J. Geophys. Res.*, 103, 30761–30771.

Gueymard, C. A. (2003). Direct solar transmittance and irradiance predictions with broadband models. Part 1: detailed theoretical performance assessment. *Sol. Energy*, 74, 355–379.

Halabi, L. M., Mekhilef, S., & Hossain, M. (2018). Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Appl. Energy*, 213, 247–261.

Hebb, D. O. (1949). The organization of behavior, Wiley, New York.

Heris, S. M. K. (2015). Evolutionary ANFIS training in MATLAB, www.yarpiz.com.

Hınçal, O., Altan-Sakarya, A. B., & Metin Ger, A. (2011). Optimization of multireservoir systems by genetic algorithm. *Water Resour. Manag.*, 25, 1465–1487.

Holland, J. H. (1975). Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, MI.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci.*, 79, 2554–2558.

Idso, S. B. (1981). A set of equations for full spectrum and 8- to 14 um and 10.5- to 12.5 um thermal radiation from cloudless skies. *Water Resour. Res.*, 17, 295–304.

Iziomon, M. G., Mayerb, H., & Matzarakisb, A. (2003). Downward atmospheric longwave irradiance under clear and cloudy skies: measurement and parameterization. *J. Atmos. Sol. Terr. Phys.*, 65, 1107–1116.

Jacobs, J. M., Anderson, M. C., Friess, L. C., & Diak, G. R. (2004). Solar radiation, longwave radiation and emergent wetland evapotranspiration estimates from satellite data in Florida, USA. *Hydrolog. Sci. J.*, 49, 461–476.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Mon. Cybern.*, 23, 665–685.

Jimenez, J. I., Alados-Arboledas, L., Castro-Diezy, Y., & Ballester, G. (1987). On the estimation of longwave radiation flux from clear skies. *Theor. Appl. Climatol.*, 38, 37–42.

Jones, H.G., Archer, N., Rotenberg, E., & Casa, R. (2003). Radiative measurement for plant ecophysiology. *J. Exp. Bot.*, 54, 879–889.

Juszak, I., & Pellicciotti, F. A. (2013). Comparison of parameterizations of incoming longwave radiation over glaciers: model robustness and seasonal variability. *J. Geophys. Res.*, 118, 3066–3084.

Kamp, R. G., & Savenije, H. H. G. (2006). Optimising training data for ANNs with genetic algorithms. *Hydrol. Earth Syst. Sci.*, 10, 603–608.

Kano, M., Suzuki, M., & Miyanchi, M. (1975). On the measurement of hemispherical longwave radiation flux in the daytime. *Papers Meteorol. Geophys.*, 26, 111–119.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95 – International Conference on Neural Networks, 4, 1942–1948.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43, 59–69.

Konzelmann, T., Van de Wal, R. S. W., Greuell, W., Brntanja, R., Henneken, E. A. C., & Abe-Ouchi, A. (1994). Parametrization of global and longwave incoming radiation from the Greenland Ice Sheet. *Glob. Planet Change*, 9,

143–164.

Koopialipoor, M., Armaghani, D. J., Hedayat, A., Marto, A., & Gordan, B., 2019, Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. *Soft Comput.*, 23, 5913–5929.

Krause, J., Ruxton, G. D., & Krause, S. (2010). Swarm intelligence in animals and humans. *Trends Ecol. Evol.*, 25, 28–34.

Le, T. L., Nguyen, H., Dou, J., & Zhou, J. (2019). A comparative study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings' energy efficiency for smart city planning. *Appl. Sci.*, 9, 2630.

Lhomme, J. P., Vacher, J. J., & Rocheteau, A. (2007). Estimating downward long-wave radiation on the Andean Altiplano. *Agri. For. Meteorol.*, 145, 139–148.

Li, M., Jiang, Y., & Coimbra, C. F. M. (2017). On the determination of atmospheric longwave irradiance under all-sky conditions. *Sol. Energy*, 144, 40–48.

Marty, C., & Philipona, R. (2000). The clear-sky index to separate clear-sky from cloud-sky situations in climate research. *Geophys. Res. Lett.*, 27, 2649–2652.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5, 115–133.

Minsky, M. L., & Papert, S. A. (1969). Perceptrons, MIT Press, MA.

Mirjalili, S. (2019). Genetic algorithm, in: Evolutionary algorithms and neural networks. *Stud. Comput. Intell.*, 780, 43–55.

Miskolczi, F., Aro, T. O., Iziomon, M., & Pinker, R. T. (1997). Surface radiative fluxes in Sub-Sahel Africa. *J. Appl. Meteorol.*, 36, 521–530.

Obot, N. I. (2019). Comparison between ground measured and satellite estimates of downward longwave radiation. *MAUSAM*, 70, 533–540.

Obot, N. I., Humphrey, I., Chendo, M. A. C., & Udo, S. O. (2019a). Deep learning and regression modelling of cloudless downward longwave radiation. *Beni-Suef. Uni. J. Basic Appl. Sci.*, 8, 23.

Obot, N. I., Humphrey, I., Chendo, M. A. C., Oyeyemi, E. O., & Udo, S. O. (2019b). Periodicity of downward longwave radiation at an equatorial location. *J. Earth Space Phys.*, 44, 165–177.

Pavlakis, K. G., Hatzidimitrion, D., Drakakis, E., Matsoukas, C., Fotiadi, A., Hatzianastassiou, N., & Vardavas I. (2007). ENSO surface longwave radiation forcing over the tropical Pacific. *Atmos. Chem. Phys.*, 7, 2013–2026.

Pluss, C., & Ohmura, A. (1997). Longwave radiation on snow-covered mountainous surfaces. *J. Appl. Meteorol.*, 36, 818–824.

Prata, A. J. (1996). A new long-wave formula for estimating downward clear-sky radiations at the surface. *Q. J. R. Meteorol. Soc.*, 122, 1127–1151.

Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., & Hartmann, D. (1989). Cloud-radiative forcing and climate: results from the Earth Radiation Budget Experiment. *Science*, 243, 57–63.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 65, 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representation by error propagating: in Rumelhart, D. E., McCleland, J. L., and the PDP Research Group (Eds). Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1, Foundations, MIT Press, Cambridge, MA, pp. 318–362.

Satterlund, D. R. (1979). An improved equation for estimating long-wave radiation from the atmosphere. *Water Resour. Res.*, 15, 1649–1650.

Scherba, A., Sailor, D. J., Rosenstiel, T. N., & Wamser, C. C. (2011). Modeling impacts of roof reflectivity, integrated photovoltaic panels and green roof systems on sensible heat flux into the urban environment. *Build. Environ.*, 46, 2542–255.

Schmetz, J. (1989). Towards a surface radiation climatology: retrieval of downward irradiances from satellites. *Atmos. Res.*, 23, 287–321.

Shariati, M., Mafipour, M. S., Mehrabi, P., Bahadori, A., Zandi, Y., Salih, M. N. A., Nguyen, H., Dou, J., Song, X., & Poi-Ngian, S. (2019). Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete. *Appl. Sci.*, 9, 5534.

Singh, N. H., & Thongam, K. (2020). Mobile robot navigation in cluttered environment using spider monkey optimization algorithm. *Iran J. Sci. Technol. Trans. Electr. Eng.*, 44, 1673–1685.

Swinbank, W. C. (1963). Long-wave radiation from clear skies. *Q. J. R. Meteorol. Soc.*, 89, 339–348.

Udo, S. O., & Aro, T. O. (1999). Technical note measurement of global solar global photosynthetically active and downward infrared radiations at Ilorin, Nigeria. *Renew. Energy*, 17, 113–122.

Udo, S. O. (2000). Quantification of solar heating of the dome of a pyrgeometer for a tropical location: Ilorin, Nigeria. *J. Atmos. Ocean. Tech.*, 17, 995–1000.

Udo, S. O. (2003). Modeling of infrared radiation for all skies from easy-to-measure meteorological parameters at tropical location: Ilorin, Nigeria. *Turk. J. Phys.*, 27, 61–68.

Wang, T., Yan, G., & Chen, L. (2012). Consistent retrieval methods to estimate land surface shortwave and longwave radiative flux components under clear-sky conditions. *Remote Sens. Environ.*, 124. 61–71.

Werbos, P. J. (1974). Beyond regression: new tools for prediction and analysis in the behavioral sciences, Ph.D. Thesis, Harvard Uni.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. 1960 IRE WESCON Convention Record, New York, IRE Part 4, 96–104.

Zadeh, L. A. (1965). Fuzzy sets, Info. Control, 8, 338–355.

Zhang, G. P. (2007). Neural networks for data mining, in: Maimon O and Rokack L (Eds). Soft computing and knowledge discovery for data mining, Springer, 17–44.

Zhang, S., Lee, C. K. M., Chan, H. K., Choy, K. L., & Wu, Z. (2015). Swam intelligence applied in green logistics: a literature review. *Eng. Appl. Artif. Intel.*, 37, 154–169.

Zimmermann, H. J. (2001). Fuzzy set theory – and its applications, Kluwer Academic Publishers, Boston.