Artificial neural network and neuro fuzzy inference modelling of global solar radiation data using bayesian algorithm for design of solar energy conversion system

S. Shanmuga Priya[1]*, Lisa Maria Ubbenjans[2], I. Thirunavukkarasu[3]

[1]*Department of Chemical Engineering, Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India.*
[2]*School of Engineering and Physical Sciences, Mechanical Engineering, Heriot Watt University, Edinburg, Scotland.*
[3]*Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal University,*
Manipal, Karnataka, India.
*Corresponding author E-mail: shanpriyanitt@yahoo.co.in

Abstract

Measurement of global solar radiation is particularly required for proper design of solar energy conversion systems. This study investigates the use of software tools like neural networks and fuzzy inference systems for modelling so as to predict global solar radiation using different input parameters based on available weather data. Advantages include simplicity, speed and efficiency, to make short term predictions of global solar radiation at different locations in India, Germany and United Kingdom. It helps in estimation of effectiveness of the applied model which matches solar radiation and other meteorological parameters which are in a non-linear relationship. Bayesian Inference algorithm is used for the current study in estimation of global solar radiation.

Keywords: *Global Solar Radiation, Artificial Neural Networks, Fuzzy Inference Modelling, Root Mean Square Error, Regression Coefficient.*

1. Introduction

The global temperature have risen by 1.4 degrees Celsius in 2016 since 1880 and the arctic minimum ice shrinks by 13.4% each decade. (Earth Science Communications Team 2016).Solar radiation is the primary and major source of energy on earth, is non-finite and clean to use. Renewable energy forms an important part in the industry to cover the increasingly expanding demand of energy while decreasing pollution and emissions. Global solar radiation (GSR) is useful data in several fields such as energy generation, agriculture, meteorology and hydrology. Measuring global solar radiation, the energy arriving on earth below the atmosphere, can be carried out directly by a sensor.(Teke et al. 2015) For instance, radiometers measure total radiation, solarimeters measure total incident radiation, pyranometers are used to determine diffuse solar irradiance and pyrheliometers measure direct normal irradiance. (Ahmet Teke et al.) However, this equipment might not always be available at the desired location due to high cost and time consumption, such as when considering domestic solar panels. For this case, a range of fairly reliable models have been developed to estimate solar radiation at different locations. Readily available meteorological data and geographical data is required as an input into the global solar radiation estimation model. (Teke et al. 2015)Climate change, which is associated with the burning of fossil fuels, impacts this data however, and therefore this effect should be considered when long term predictions are required.

Other techniques to conduct effective GSR predictions include support vector regression (SVR) using a polynomial basis function (-poly) or a radial basis function (-rbf) in which the last one outperforms ANFIS and SVR-poly. (Mohammadi et al. 2015) The ANN and fuzzy logic models can be further compared to previous models for the desired locations using the root means square error analysis (RMSE) as a short term performance indicator to compare estimated and measured data as well as a determination coefficient (R^2) to compare data to a statistical model. (Teke et al. 2015).

In this paper, two simple and effective, data-driven models with a high spatial resolution for the estimation of global solar radiation (GSR) at multiple locations were developed using soft computing models. These include neural network and Fuzzy Interference Systems (FIS). The advantage for these models over traditional methods is the close connection to real world measurements. (Araghinejad 2014).

2. Methodology

After having explored estimation techniques such as linear regression models and neural networking techniques, input data is collected from Indian weather portal, Meterological data from UK and Germany, NASA surface meterology and solar energy over a period of time for four locations each in India, UK and Germany which is relevant for the prediction of GSR. Then, the data is checked for faults and selected to reduce errors and minimise processing time as well as considering the limitation of five inputs into the ANFIS simulator in MATLAB. Matrices are prepared for the neural network and ANFIS simulator in order to input meteorological, geographic and GSR information. After model creation, training and checking takes places according to the splitting of data in the nntool and ANFIS simulator. Linear regression plots are used to compare the results of each model associated with certain input data.

3. Geographic and meteorological data

Horizontal GSR (H) is the radiation coming into the earth's atmosphere, consisting of total incoming direct (H_{dir}) as well as diffuse short wave radiation from scattering due to the atmosphere (H_{dif}). Hence, $H = H_{dir} + H_{dif}$(Kamali et al. 2006). On a tilted surface, one additional component needs to be taken into account – the reflected radiation from the ground (H_{refl}).The solar constant is 1363 W/m^2, but solar activity varies from year to year due to the eleven-year-long cycle of the sun in which calm as well as stormy periods occur. This variation only makes a difference in solar radiation on earth of 0.1% or 1.3W/m^2. (Rani Gran 2008).The selected locations vary greatly in coordinates such as Chennai in India with latitude of 13.1 degrees towards the north and longitude of 80.3 degrees east as compared to Stornoway in the UK with latitude of 57.6 degrees north, and longitude of 3.9 degrees west. Due to the geographical spread of the countries to be investigated, coordinates such as latitude and longitude and altitude are also considered. (Teke et al. 2015)

Geographical data and meteorological data readily available online and collected over a period of time (between 18 and21 years) to forecast solar radiation on a monthly basis. The most common variables to be used, such as in the empirical Ångström-model, are air temperature, sunshine duration and cloud cover. Meteorological and geographical data is obtained from websites, including for India, for UK and for Germany. GSR data can be also referred from the Nasa Atmospheric Science Data Centre web site for all locations between the years 1983 and 2005. GSR differs from region to region due to different geographical locations, sun position in the sky as well as meteorological aspects. Similarly, the time of the year impacts the amount of GSR such as increased cloud cover during rainy season or less GSR in certain seasons due to the orientation of earth towards the sun. Therefore, average, maximum and minimum GSR data can be essential, to predict potential energy generation and reduce risk by considering the minimum. (Gulin et al. 2013)

Based on the latitude of the location to be investigated, the optimum solar radiation changes according to the time of the year due to the earth's seasonal weather changes associated with the orientation towards the sun. Therefore, tilting the solar panel towards the sun can significantly increase the solar radiation on that surface. For instance, even during the day, there will be differences in angles due to the movement of the sun in the sky from morning to evening. (Landau 2015). From data charts retrievable from the NASA webpage, during rainy seasons in southwest India, clearness index and direct solar radiation are lower than in sunny seasons and diffuse radiation increases due to scattering of the light in the atmosphere. Additionally, it can be observed that at the optimum angle, solar radiation is significantly higher compared to radiation onto horizontal surface. In this study, only the horizontal GSR will be considered for modeling, and GSR can be converted accordingly at certain angles.

India

Global solar radiation, extra-terrestrial radiation, minimum temperature, average temperature, maximum temperature, cloud cover, precipitation, vapour pressure and relative humidity data was collected for the locations Chennai, Delhi, Kolkata and Pune from NASA and the Indian Weather Portal over the years 1984 to 2002, resulting in a matrix size of 228 elements for each variable and location in MATLAB. Additionally, the geographical data for latitude, longitude and altitude are included as well as the month of year of the associated data.

UK

Similarly, GSR data, extra-terrestrial radiation, relative humidity, minimum temperature, maximum temperature, precipitation, data for sunshine hours, month of the year, latitude, longitude and altitude data is used for the UK locations Stornoway Airport, Eskdalemuir, Sheffield and Heathrow from NASA and the UK Met Office over the years 1984 to 2005, having 264 elements for each variable and selected locations in UK.

Germany

For the locations Berlin Tempelhof, Bremen, Frankfurt and Munich in Germany GSR data, extra-terrestrial radiation, relative humidity, average temperature, maximum and minimum temperature, cloud cover, precipitation, wind speed, sunshine hours, month of the year, latitude, longitude and altitude data from NASA and the German weather portal over the years 1985 to 2005 is used. Hence, there are 252 elements for each variable and selected locations in Germany.

4. Soft computing for GSR prediction

There are different soft computing techniques for the model creation for GSR prediction. Depending on the selected location, weather data available and computing time and software, different methods are more effective than others as demonstrated by a broad range of studies. Highly effective models, using neural networks, include the radial basis function type, the Bayesian type and the multilayer-perceptron type depending on the number of hidden layers and neurons using input data such as maximum temperature, sunshine hours, relative humidity and wind speed. (Hejase et al. 2013)Adaptive neuro-fuzzy inference systems (ANFIS) are another commonly used method using either the Sugeno and Mamdani type. Artificial neural networks (ANN) are expected to produce quicker output with multiple input compared to fuzzy inference system while FIS might incur some advantages over ANN such as higher accuracy.

Generally, the higher the complexity of a model (e.g. higher order instead of linear), the more challenging the developing and running time associated with more issues. In this case, due to the geographical spread and different types of data available for each location, the complexity of the model needs to be balanced with the quality, quantity and type of input and output variables available.

Neural network model

The neural networking tool is used to determine the most adequate variables for the ANFIS input. The nf-tool is used first, due to the restriction of five inputs into the ANFIS solver, while the nf-tool allows for a higher input number.70 % of the data will be used for training, 15% for validation and 15% for testing. The number of neurons is left at 10 and the network is trained using the Bayesian regularization, which has the advantage of better results with small or noisy datasets but being slightly more time consuming than the Levenberg-Marquardt method.The regression plot will be used to evaluate the accuracy of the model. Values close to one indicate a good correlation between original input and predicted data when using the model. Starting with all variables included in the first matrix, variables will be removed individually to check the influence of each onto the model.

ANFIS model

The neuro-fuzzy toolbox in MATLAB allows for the creation, training and testing of a fuzzy inference system of Sugeno-type. This is a relatively fast and efficient way to create FIS models, reducing trial and error processing, provided that input and output data is available for training and testing the system. In this study, the supervised training method with a hybrid approach, a combination of the least square method (forward pass) and back propagation, will be used to train and tune membership functions using input and output data. The weight and functions between

input and output will affect the system behaviour, such as the type and grade of membership functions for each input.

In the GRID partition section, the number of membership functions (MF) can be specified for each input variable. Generally, increasing the number of membership functions for the input variables, lowers the training error but the running time to train the model increases. A higher number indicates a higher belongingness to the system for that variable. Additionally, the type of MF can be specified as well as whether the output is constant or linear. There are eight different types of MFs. The triangular (trimf) and trapezoidal (trapmf) MFs are the simplest ones, having a triangular and trapezoidal shape, respectively. Two MFs (gaussmf and gauss2mf) are based on the Gaussian distribution and the generalized bell-shaped MF (gbellmf) is similar to these, but with one more parameter. These three MFs are widely used due to their smoothness.(Zadeh 1995) The last three MFs are based on sigmoidal functions, in which one is asymmetric and open (sigmf) and two are asymmetric and closed in which one is the product of two sigmoidal MFs (psigmf) and one is the difference between two MFs (dsigmf).For the ANFIS model three locations are used for training the network and one location is used for checking. The hybrid training method will be used with three to five bell-shaped, linear or constant membership functions, depending on the processing time and computer memory available.

5. Results and discussion

Bayesian neural network

India

From the simulation results it can be determined that the most influential parameter for predicting GSR is the cloud cover closely followed by extra-terrestrial radiation. Latitude and precipitation similarly have a positive effect on GSR prediction and relative humidity and average temperature can be also included. Parameters to neglect include minimum and maximum temperature and longitude. Even though altitude and vapour pressure may have a positive effect, there is no significant difference in the overall result to predict GSR and with this dataset it might even worsen the model. Increasing or decreasing the number of hidden neurons did not enhance the linear regression model, however the Bayesian regulation method was most suitable as compared to the Levenberg-Marquardt training algorithm or the scaled conjugate training method. Fig 1 shows the ANN results for training and testing for all inputs for India.

UK

The most influential parameter for the prediction of GSR in the UK with the data collected is sunshine hours followed by maximum temperature and relative humidity. Altitude and minimum temperature only had a slight effect to predict GSR as well as longitude and latitude which might be due to the small geographical size of the UK with a small variation. Precipitation and ETR neither contributed significantly to benefit the model compared to the other parameters. Fig 2 shows ANN simulation results including all inputs for UK.

Germany

The results for the collected weather data in Germany reveal that the most influencing variables for positive GSR prediction include sunshine hours followed by cloud cover and wind speed as well as maximum temperature and ETR, minimum temperature as well as as relative humidity. Geographical data such as altitude, longitude and latitude can be neglected as well average temperature.
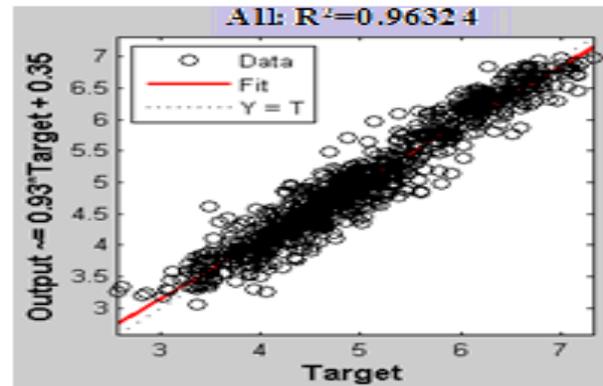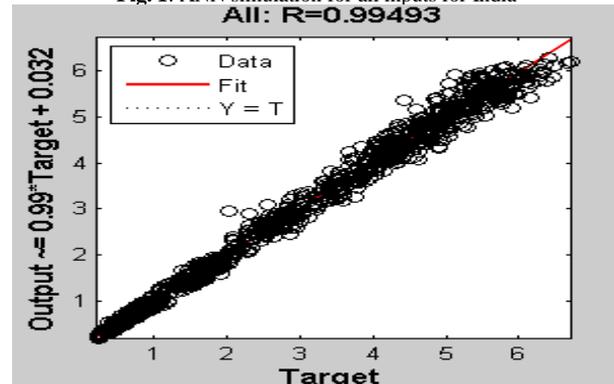

Fig. 1: ANN simulation for all inputs for India
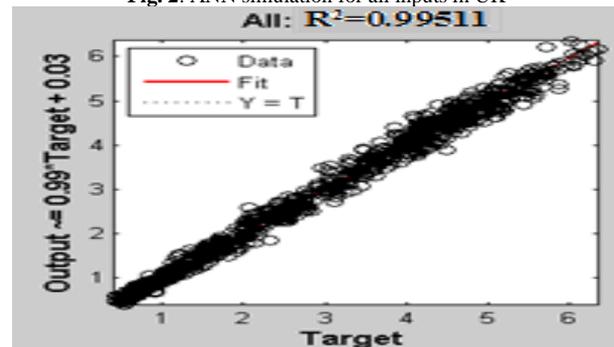

Fig. 2: ANN simulation for all inputs in UK

Fig. 3: ANN simulation for all inputs in Germany

Overall, the Bayesian regularization method worked best in the nf-tool with data from UK and Germany with R-values exceeding 0.994 and 0.995, respectively. Also for the data acquired for India, R-values were high with values above 0.96. Hence, neural networks are a suitable option to model solar radiation and to make predictions with the data collected. Considering similar variables for the locations, especially cloud cover, sunshine hours, ETR, latitude, precipitation, relative humidity, average and maximum temperature were significant for the GSR model while neglecting minimum temperature, altitude and longitude.

ANFIS

In order to enable data comparison to the GSR prediction model when using the Bayesian neural networking method, a linear regression plot will be created for the ANFIS output and target data. (Zadeh 1995)

ANFIS india

The nf-tool simulation revealed that the most influential parameters for predicting GSR are cloud cover and ETR followed by latitude, precipitation, relative humidity and average temperature. Hence matrices with these variables will be created, including GSR at the end as the output variable. Other than in the neural networking tool, in the Neuro fuzzy designer, the data needs to be split before. For India, the locations Chennai and Pune

will be used for training, Delhi will be used for testing and Kolkata for checking to avoid over-fitting.

The Table 1 shows the matrices to be trained and tested for the GSR model. The letter "x" needs to be replaces by Chennai, Pune, Delhi and Kolkata for training and checking, respectively.

Table 1: Input Data Matrices for the ANFIS Toolbox in MATLAB for India

Matrix name	Characteristics	Variables
xANFISbest	Best inputs	month of year, cloud cover, ETR and GSR
xANFIS1	Best inputs plus latitude	month of year, cloud cover, ETR, latitude, GSR
xANFIS2	Best inputs plus precipitation	month of year, cloud cover, ETR, precipitation, GSR
xANFIS3	Best inputs plus relative humidity	month of year, cloud cover, ETR, relative humidity, GSR
xANFIS4	Best inputs plus average temperature	month of year, cloud cover, ETR, average temperature, GSR
xANFIS5	Best inputs plus latitude, precipitation, average temperature	month of year, cloud cover, ETR, latitude, precipitation, average temperature and GSR

Membership functions (3 or 4) are adjusted using hybrid back-propagation and least square method. Hence the ANFIS can be trained according to the input and output data. The bell shaped function was found to work best with data from India.

Trying different input methods for the simplest input data, only training and checking is necessary to obtain reasonably accurate results as compared to the Bayesian model. Training is done by using Chennai, Pune and Delhi weather data while Kolkata is used for checking.

From the results it can be seen that with 10 membership functions for each input variable, the ANFIS model performed better than the Bayesian regulation neural networking model. However, considering time, efficiency and simplicity for the creation of a neural network model for estimating GSR when using Matlab, the Bayesian regularization networking method achieves highly accurate outputs while being simple to use and having a low processing time as compared to ANFIS which takes a longer processing time for achieving similar results. In addition, when more than three input variables are used for the ANFIS model, the number of membership functions is limited due to the processing time (depending of the processor speed of the computer). In this case with 4 inputs, a maximum of 5 mfs (and generally 4 mfs) was matrices for ANFIS toolbox for UK.

used and with 5 inputs a maximum of 4mfs (and generally 3mfs) was used.

More flexibility in tuning makes ANFIS an interesting option as well as when considering the different types of input data specific to a certain problem and not suitable for all soft computing methods. Fig 4 shows the ANFIS best results for India.

ANFIS UK

The most influential parameter for the prediction of GSR in the UK with the data collected is sunshine hours followed by maximum temperature and relative humidity.

Stornoway Airport, Sheffield and Heathrow are used for training and Eskdalemuir for checking the ANFIS model as described previously, with UK weather data, in which various techniques to train the network were tested. Similarly, a bell-shaped function will be used due to the small training error incurred with this function and 4 or 5 membership functions for each input variables, depending on the computer's capacity.

Stornoway, Sheffield and Heathrow (x = Storno Sheffield Heathrow) are used for training and Eskdalemuir (x = Eskdalemuir) for testing. Table 2 shows the input data

Table 2: Input Data Matrices for the ANFIS Toolbox in MATLAB for UK

Matrix name	Characteristics	Variables
xANFISbest	Best inputs	Month, sunshine hours, maximum temperature and GSR
xANFIS1	Best inputs plus relative humidity	Month, sunshine hours, maximum temperature, relative humidity and GSR
xANFIS2	xANFIS1 plus altitude	Month, sunshine hours, maximum temperature, relative humidity, altitude and GSR
xANFIS3	xANFIS1 plus minimum temperature	Month, sunshine hours, maximum temperature, relative humidity, minimum temperature and GSR
xANFIS4	xANFIS1 plus ETR	Month, sunshine hours, maximum temperature, relative humidity, ETR and GSR
xANFIS5	xANFIS1 plus precipitation	Month, sunshine hours, maximum temperature, relative humidity, precipitation and GSR

Due to long processing time, the membership function was changed to 3 with five input variables in the second set of results. Furthermore, instead of a linear membership function a constant function was used, as well as a linear and trapezoidal type for comparison and accuracy check. Fig 5 shows ANFIS best result for UK.

ANFIS germany

The results for the collected weather data in Germany reveal that the most influencing variables for positive GSR prediction include sunshine hours followed by maximum temperature and cloud cover as well as Evapo-Transpiration Rate (ETR), precipitation, as relative humidity and wind speed.Geographical data such as altitude, longitude and latitude can be neglected as well average and minimum temperature. Fig 6 shows the ANFIS best result for Germany.
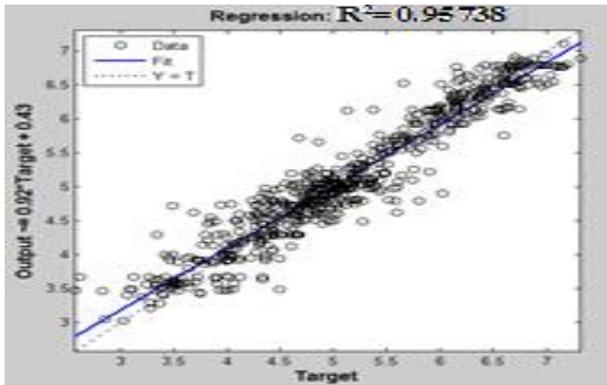
Table 3: Input data matrices for the ANFIS toolbox in MATLAB for Germany

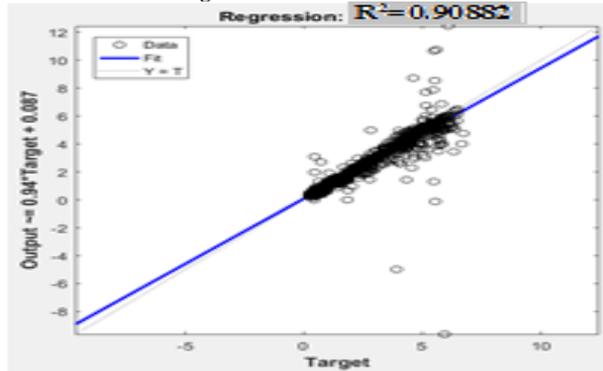| Matrix name	Characteristics	Variables
xANFISbest	Best inputs	Month, sunshine hours, cloud cover and GSR
xANFIS1	ANFISbest plus wind speed	Month, sunshine hours, cloud cover, wind speed and GSR
xANFIS2	ANFISbest plus maximum temperature	Month, sunshine hours, cloud cover, maximum temperature and GSR
xANFIS3	ANFISbest plus ETR	Month, sunshine hours, cloud cover, ETR and GSR
xANFIS4	ANFISbest plus minimum temperature	Month, sunshine hours, cloud cover, minimum temperature and GSR
xANFIS5	ANFISbest plus relative humidity	Month, sunshine hours, cloud cover, relative humidity and GSR
xANFIS6	ANFISbest plus precipitation	Month, sunshine hours, cloud cover, precipitation and GSR

Fig. 4: ANFIS result for India
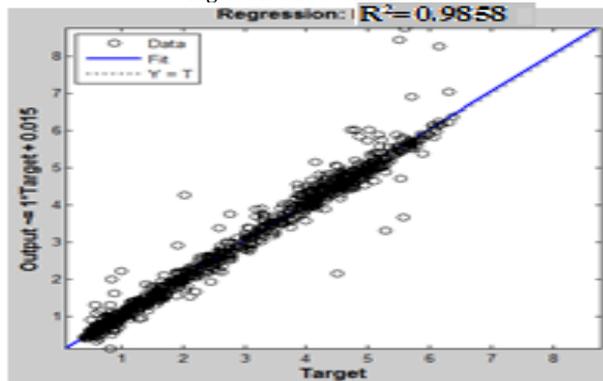

Fig. 5: ANFIS result for UK

Fig. 6: ANFIS result for Germany

From the results it can be observed that according to each type of variable a different combination of number of mfs, type of mf and linearity needs to be chosen as well as varied between hybrid or back propagation. It is highly important that each ANFIS

simulation is complete, which means that the training error does not change anymore with each epoch passed or only changes by a small amount. Otherwise, the output obtained for GSR values will have a high likelihood to be wrong. Additionally, enough time and epochs should be allowed to pass, until the training error becomes insignificantly small. Depending on the type and number of membership function selected as well as whether it was selected to be constant or linear or the back propagation or hybrid technique was selected, the final results of the ANFIS model would vary for each data set and thus perform worse or better at predicting GSR. Generally, the smaller the training error in the simulations, which required more time than in those simulations with larger training errors, the better the linear regression and ability to predict GSR.

6. Conclusion

Simple empirical regression models can be used to predict GSR with statistical methods. This study investigated the effectiveness of the applied model matching solar radiation and other meteorological parameters which are in a non-linear relationship. Therefore, an improvement to the empirical models are ANN and ANFIS models, capable of determining the relationship between input and output data depending on the selected input and output variables available as well as the fine tuning such as adjusting the number of neurons within the ANN or selecting the optimum number and type of membership functions within the ANFIS. Generally, for the three locations, apart from cloud cover, the most significant variables to influence the GSR prediction positively, include relative humidity, precipitation, and temperature (and wind speed for Germany) for all locations. From the weather data collected for each location and in combination with the models created using neural networks, the accuracy of GSR prediction models depends on the accuracy of data collected and the combination of data, e.g. month of year and cloud cover. For the NN and ANFIS simulation two different versions of MATLAB were used, including R201a and R2016a running on two different computers which might have had an effect on the overall outcome of each neural and fuzzy network model.

Acknowledgement

Appendix A

Table 4: Selected locations

Country	City	Coordinates	Altitude
India	Pune	19.0760° N, 72.8777° E	14 m
	Delhi	28.6139° N, 77.2090° E	216 m
	Kolkata	22.5726° N, 88.3639° E	9.1 m
	Chennai	13.0827° N, 80.2707° E	6.7 m
	(Manipal)	13.3605° N, 74.7864° E	73 m
UK	Stornoway Airport	57.5864° N, 3.8685° W	13 m
	Eskdalemuir	55.2678° N, 3.1757° W	242 m
	Sheffield	53.3811° N, 1.4701° W	90 m
	Heathrow	51.4700° N, 0.4543° W	25 m
	(Edinburgh)	55.9533° N, 3.1883° W	205 m
Germany	Bremen	53.0793° N, 8.8017° E	11 m
	Berlin-Dahlem (FU)	52.5200° N, 13.4050° E	34 m
	Frankfurt	50.1109° N, 8.6821° E	103 m
	Munich	48.1351° N, 11.5820° E	520 m
	(Diepholz)	52.6056° N, 8.3708° E	39 m

Nomenclature

H	Horizontal radiation

H_{dir}	Total incoming direct radiation
H_{dif}	Diffuse short wave radiation
GSR	Global Solar Radiation
SVR	Support Vector Regression

FIS   Fuzzy Interference Systems
ANFIS   Adaptive Neuro-Fuzzy Inference Systems
RMSE   Root Mean Square Error
R^2 Regression coefficient
ANN   Artificial Neural Networks
MF   Membership Functions
RBF   Radial Basis Function
ETR   Evapo-Transpiration Rate

References

[1]   Araghinejad S, *Data-driven modeling: using MATLAB® in water resources and environmental engineering*, Springer Science & Business Media, (2013).

[2]   Gulin M, Vašak M & Baotic M, "Estimation of the global solar irradiance on tilted surfaces", *17th International Conference on Electrical Drives and Power Electronics,* Vol.6, No.4, (2013), pp. 347-353.

[3]   Hejase HA, Assi AH & Al Shamisi MH, "Use of Empirical Regression and Artificial Neural Network Models for Estimation of Global Solar Radiation in Dubai, UAE", *Causes, Impacts and Solutions to Global Warming*, (2013), pp.61-86.

[4]   Kamali Gh.A, Morad I & Khalili A, "Estimating solar radiation on tilted surfaces with various orientations. A study case in Karaj (Iran)", *Theor. Appl. Climatol,* Vol.84, No.4, (2006), pp.235–241.

[5]   Landau CR, "Optimum tilt of solar panels", *Optimum Tilt of Solar Panels,* (2012).

[6]   Mohammadi K, Shamshirband S, Danesh AS, Abdullah MS & Zamani M, "Temperature-based estimation of global solar radiation using soft computing methodologies", *Theoretical and applied climatology*, Vol.125, No.1-2, (2016), pp.101-112.

[7]   Gran R, "Solar Variability: Striking a Balance with Climate Change", *Hg. v. NASA. NASA's Goddard Space Flight Center.*, (2008).

[8]   Teke A, Yıldırım HB & Çelik Ö, "Evaluation and performance comparison of different models for the estimation of solar radiation", *Renewable and Sustainable Energy Reviews*, Vol.50, (2015), pp.1097-1107.

[9]   Zadeh LA, *Fuzzy Logic Toolbox*, Membership Functions. World sci books, (1995), pp.19-34.